# Appunti per le lezioni di Ingegneria dei Sistemi Software

**bozza**

Andrea Domenici

A.A. 2008–2009

# Indice

# Capitolo 1

# Introduzione

*L'ingegneria del software non mi piace. Mi piacciono le cose pratiche, mi piace programmare in assembler.*
*– confessione autentica di uno studente*

L'ingegneria del software è l'insieme delle teorie, dei metodi e delle tecniche che si usano nello sviluppo industriale del software. Possiamo iniziarne lo studio considerando le seguenti definizioni:

**software**

> *"i programmi, le procedure, le regole e l'eventuale documentazione associata, e i dati relativi all'operatività di un sistema di elaborazione"*
>
> *– IEEE*

**ingegneria del software**

> *"approccio sistematico allo sviluppo, all'operatività, alla manutenzione ed al ritiro del software"*
>
> *– IEEE*

> *"disciplina tecnologica e manageriale che riguarda la produzione sistematica e la manutenzione dei prodotti software, ... sviluppati e modificati entro i tempi e i costi preventivati"*
>
> *– D. Fairley [2]*

È particolarmente importante, nelle due definizioni di "ingegneria del software", il concetto di sistematicità. La necessità di sottolineare l'importanza

di questo concetto, che è data per scontata nelle altre discipline ingegneristiche, può essere meglio compresa se consideriamo la storia dell'informatica, schematizzata nelle tre fasi di *arte*, *artigianato*, e *industria*.

La prima fase, quella delle origini, è caratterizzata dal fatto che i produttori e gli utenti del software condividevano una formazione scientifica, per cui i programmatori avevano una certa familiarità con i problemi degli utenti, e questi ultimi erano in grado di capire il funzionamento dei calcolatori e quindi le esigenze dei programmatori. Anzi, accadeva spesso che utente e programmatore fossero la stessa persona. Un altro aspetto caratteristico di quel periodo è il fatto che spesso un'applicazione venisse prodotta *ad hoc* per un particolare problema e venisse abbandonata una volta soddisfatta quella particolare necessità. La produzione del software in quel periodo era quindi paragonabile ad un'arte, nel senso che era dominata dall'inventività individuale e da un'organizzazione del lavoro molto lasca (tralasciamo qui altri aspetti, pur importanti, legati alle tecniche di programmazione ed agli strumenti di sviluppo, in quel periodo assai limitati).

In una fase successiva, in seguito allo sviluppo delle applicazioni informatiche, estese dal campo scientifico a quello commerciale e amministrativo, i nuovi utenti del software hanno formazioni culturali diverse da quella dei programmatori. Questi ultimi devono quindi imparare ad affrontare problemi di tipo nuovo, ed a comunicare efficacemente con persone che non solo non hanno le conoscenze necessarie a capire gli aspetti tecnici della programmazione, ma non hanno neppure alcuna motivazione pratica per interessarsene. Inoltre, le applicazioni richieste hanno un peso economico sempre maggiore ed un ruolo sempre piú critico nelle organizzazioni che ne fanno uso. I programmatori assumono quindi una figura professionale distinta, ed il lavoro viene organizzato su gruppi di programmatori, che possono essere imprese autonome (*software houses*) o parti di organizzazioni (servizi di elaborazione dati o di sviluppo del software). In questa fase c'è una maggiore strutturazione dell'attività di produzione, però rimangono l'approccio individualistico alla programmazione e la mancanza di strumenti teorici e di pratiche standardizzate. In questo senso si può parlare di artigianato, cioè di un'attività "pre-industriale", motivata economicamente ma sempre basata sulle capacità individuali e poco standardizzata.

L'inadeguatezza degli strumenti teorici e metodologici nella fase artigianale della produzione di software fece sentire i propri effetti nella famosa *crisi del software*, fra gli anni '60 e '70: la produzione del software non riusciva a tenere il passo con le richieste degli utenti e con i progressi dello hardware, aumentavano i costi di sviluppo, i ritardi nelle consegne, ed i malfunziona-

menti, con un conseguente aumento dei costi di manutenzione. Si capí quindi che la via d'uscita era il passaggio da artigianato a industria: era cioè necessario (e lo è tuttora) usare nella produzione di software lo stesso approccio che si usa nelle industrie mature, come l'industria meccanica o l'industria elettronica. Nella conferenza di Garmisch [11] del 1968 (anno cruciale anche per altri aspetti) venne coniato il termine di "ingegneria del software".

La transizione da artigianato a industria è tuttora in corso: la disciplina dell'ingegneria del software è ancora in fase di crescita e lontana, in molti dei suoi settori, dal consolidamento di teorie e metodi, e molte imprese sono ancora legate alla vecchia impostazione artigianale. Però la maggior parte dei produttori di software ha adottato processi di sviluppo strutturati e metodologie di carattere ingegneristico.

## 1.1 Il software è diverso

Per comprendere meglio la storia ed i problemi della produzione del software, è bene aver presenti alcuni fatti che rendono il software un prodotto abbastanza diverso dagli altri manufatti:

- Prima di tutto, il software è "soft": non occupa praticamente spazio fisico, è invisibile, e gli aspetti materiali della sua produzione incidono pochissimo sul costo globale del prodotto; il software è informazione pura, e la sua produzione è progetto puro.
- Il software è estremamente *malleabile*, cioè modificabile con pochissimo sforzo. Questa proprietà è un vantaggio ma anche un pericolo, poiché incoraggia un modo di lavorare non pianificato.
- L'immaterialità del software fa sí che sia difficile visualizzare i concetti che gli sono propri, come algoritmi, strutture dati, e flussi di controllo. In effetti, una buona parte dell'ingegneria del software è rivolta allo studio e all'applicazione di notazioni espressive e sintetiche per rappresentare tali concetti.
- All'immaterialità del software è legata la difficoltà di quantificare e definire rigorosamente le caratteristiche del prodotto che ne determinano la qualità: mentre, per esempio, la qualità di una parte meccanica è largamente riconducibile a proprietà fisiche e quindi misurabili, questo generalmente non avviene per il software.
- Il software è *complesso*: un programma può essere composto da un numero grandissimo di istruzioni, e ciascuna di esse potenzialmente

condiziona il comportamento di tutte le altre. Inoltre, il software è "*non lineare*", nel senso che un piccolo cambiamento nel codice può portare a grandi cambiamenti (spesso indesiderati!) nel funzionamento di un'applicazione.

A proposito della complessità del software, osserviamo che questa è tanto maggiore quanto meno il software è *strutturato*. Diciamo che un sistema software è strutturato se è suddiviso in componenti (*moduli*), ciascuno dei quali interagisce con pochi altri componenti. Le istruzioni contenute in ciascun componente hanno un campo d'azione limitato al componente stesso, eccettuate quelle istruzioni che devono espressamente interagire con altri componenti. Questo argomento verrà ripreso nel capitolo relativo al progetto del software.

Oltre alle caratteristiche del software sú esposte, sulla produzione del software pesa il fatto che si tratta di un'attività ancora giovane, sviluppatasi per di piú ad un ritmo velocissimo. Questo comporta varie conseguenze, fra cui la principale è forse una certa difficoltà di comprensione fra progettista e committente/utente: spesso un'organizzazione (o un individuo) sente il bisogno di uno strumento informatico per risolvere un problema, ma non ha un'idea chiara di cosa si possa chiedere allo strumento. Dall'altra parte, il progettista può non avere un'idea chiara delle esigenze dell'utente. Inoltre, in certi casi può anche mancare un linguaggio comune e consolidato sia fra progettista e utente che fra progettisti provenienti da ambienti diversi.

Si può anche osservare che, almeno fino a tempi recenti, l'impiego di strumenti automatici nello sviluppo del software è stato relativamente limitato, cosa abbastanza paradossale. Questo problema è collegato anche ad altre proprietà del software, come la sua complessità ed il carattere intellettuale (quindi difficilmente automatizzabile) delle attività legate alla sua produzione.

Sebbene tutti questi problemi siano ancora sentiti, è però incoraggiante notare una generale maturazione nel mondo dell'industria informatica, per cui le conoscenze relative all'ingegneria del software sono sempre piú richieste ed apprezzate.

## 1.2 Il software non è diverso

Pur con le sue particolarità, il software è un prodotto industriale che come tutti i prodotti industriali deve soddisfare le esigenze dell'utente, quelle del

produttore[1], e quelle della società. Quindi il prodotto deve svolgere le funzioni richieste, e deve farlo secondo i criteri di qualità richiesti. Deve anche essere realizzato e consegnato nei limiti di tempo richiesti (dettati sia dalle esigenze dell'utente che da fattori economici piú generali), e nei limiti di costo richiesti.

La possibilità di rispettare i vincoli sú esposti, e in generale di produrre il software in modo economicamente vantaggioso, dipende naturalmente dal *processo* di produzione, cioè dall'insieme delle attività volte alla produzione. Sebbene buona parte dell'ingegneria del software sia rivolta al prodotto, cioè alle teorie e alle tecniche necessarie per descriverne i requisiti, per realizzarlo e per valutarne la qualità, il suo obiettivo principale è lo studio del processo di produzione. Per questo si parla, come della definizione di Fairley, di una disciplina tecnologica e manageriale, che attinge da campi diversi, come la matematica, i vari settori dell'informatica, l'organizzazione aziendale, la psicologia, l'economia ed altri ancora.

L'insieme degli strumenti a disposizione dell'ingegnere del software è del tutto analogo a quello di cui si servono le altre discipline ingegneristiche:

**Teorie,** le basi formali e rigorose del lavoro. Per esempio, l'elettromagnetismo per l'ingegneria elettrica, la logica nell'ingegneria del software.

**Linguaggi,** notazioni (testuali o grafiche) e formalismi per esprimere i concetti usati e le loro realizzazioni: disegno tecnico per l'ingegneria meccanica, UML per il software.

**Standard,** documenti che stabiliscono requisiti uniformi che devono essere soddisfatti da processi, procedure, eccetera: norme UNI per l'ingegneria industriale, standard ANSI per l'informatica.

**Strumenti,** ambienti di sviluppo e programmi che assistono l'ingegnere nella progettazione, anche automatizzandone alcune fasi: programmi di calcolo delle strutture per l'ingegneria civile, strumenti CASE per il software.

Naturalmente, questo armamentario non è completo senza la capacità, da parte del progettista, di supplire con l'esperienza e l'inventiva ai limiti delle teorie e delle procedure formalizzate: è fondamentale conoscere le regole del mestiere, ma queste da sole non bastano per trovare tutte le soluzioni.

[1]Non quelle del progettista, che comunque, nel caso dell'ingegnere informatico, si limitano alla disponibilità di un calcolatore, di lunghe notti al terminale, e di un po' di bevande e generi di conforto durante il lavoro `:)`.

### 1.2.1 Deontologia

Si è accennato al fatto che ogni prodotto deve rispondere alle esigenze della società, oltre che a quelle del produttore e del cliente. Le esigenze della società sono in primo luogo quelle espresse da leggi e normative di vario genere. Al di là degli adempimenti legali, ogni progettista (come ogni altro individuo) è responsabile delle conseguenze che le proprie scelte ed i propri comportamenti possono avere sulla società: possiamo pensare, per esempio, all'impatto di un prodotto sull'ambiente, sui rapporti sociali, o sull'occupazione.

È particolarmente importante tener conto dei rischi economici e umani connessi all'uso del software. Il software è estremamente pervasivo, essendo un componente di sistemi disparati, come, per esempio, sistemi di comunicazione, elettrodomestici, veicoli, impianti industriali. Inoltre, alcuni prodotti software possono essere riusati in applicazioni diverse da quelle per cui sono stati concepiti originariamente, per cui il progettista di software può non essere in grado di anticipare la destinazione finale del proprio lavoro. Se sottovalutiamo la pervasività e l'adattabilità del software rischiamo di non valutare i rischi in modo adeguato. Le situazioni catastrofiche portate spesso ad esempio dei rischi del software (distruzione di veicoli spaziali, guasti in apparati militari o in centrali nucleari) possono non mettere in evidenza il fatto che del software difettoso si può trovare in qualsiasi officina o in qualsiasi automobile, mettendo a rischio la salute e la vita umana anche nelle situazioni piú comuni e quotidiane. Un ingegnere deve sentirsi sempre responsabile per il proprio lavoro, ed eseguirlo col massimo scrupolo, qualunque sia il suo campo di applicazione.

## 1.3 Concetti generali

In questo corso ci limiteremo agli aspetti piú tradizionalmente ingegneristici della materia, e purtroppo (o forse per fortuna) sarà possibile trattare solo pochi argomenti. Questi dovrebbero però dare una base sufficiente ad impostare il lavoro di progettazione nella vita professionale. Il corso si propone di fornire le nozioni fondamentali sia su argomenti di immediata utilità pratica, sia su temi rilevanti per la formazione culturale dell'ingegnere del software.

In questa sezione vogliamo introdurre alcuni motivi conduttori che ricorreranno negli argomenti trattati in queste dispense.

### 1.3.1 Le parti in causa

Lo sviluppo e l'uso di un sistema software coinvolge molte persone, ed è necessario tener conto dei loro ruoli, delle esigenze e dei rapporti reciproci, per ottenere un prodotto che risponda pienamente alle aspettative. Per questo l'ingegneria del software studia anche gli aspetti sociali ed organizzativi sia dell'ambiente in cui viene sviluppato il software, sia di quello in cui il software viene applicato.

Questo aspetto dell'ingegneria del software non verrà trattato nel nostro corso, ma qui vogliamo chiarire alcuni termini che saranno usati per designare alcuni ruoli particolarmente importanti:

**Sviluppatore:** sono sviluppatori coloro che partecipano direttamente allo sviluppo del software, come *analisti*, *progettisti*, *programmatori*, o *collaudatori*. Due o piú ruoli possono essere ricoperti da una stessa persona. In alcuni casi il termine "sviluppatore" verrà contrapposto a "collaudatore", anche se i collaudatori partecipano al processo di sviluppo e sono quindi sviluppatori anch'essi.

**Produttore:** per "produttore" del software si intende un'organizzazione che produce software, o una persona che la rappresenta. Uno sviluppatore generalmente è un dipendente del produttore.

**Committente:** un'organizzazione o persona che chiede al produttore di fornire del software.

**Utente:** una persona che usa il software. Generalmente il committente è distinto daagli utenti, ma spesso useremo il termine "utenti" per riferirsi sia agli utenti propriamente detti che al committente, ove non sia necessario fare questa distinzione.

Osserviamo che l'utente di un software può essere a sua volta uno sviluppatore, nel caso che il software in questione sia un ambiente di sviluppo, o una libreria, o qualsiasi altro strumento di programmazione.

Osserviamo anche che spesso il software non viene sviluppato per un committente particolare, ma viene messo in vendita (o anche distribuito liberamente) come un prodotto di consumo.

### 1.3.2 Specifica e implementazione

Una *specifica* è una descrizione precisa dei *requisiti* (proprietà o comportamenti richiesti) di un sistema o di una sua parte. Una specifica descrive una certa entità "dall'esterno", cioè dice quali servizi devono essere forniti o quali proprietà devono essere esibite da tale entità. Per esempio, la specifica di un palazzo di case popolari potrebbe descrivere la capienza dell'edificio dicendo che deve poter ospitare cinquanta famiglie. Inoltre la specifica ne può indicarne il massimo costo ammissibile.

Il grado di precisione richiesto per una specifica dipende in generale dallo *scopo* della specifica. Dire che uno stabile deve ospitare cinquanta famiglie può essere sufficiente in una fase iniziale di pianificazione urbanistica, in cui il numero di famiglie da alloggiare è effettivamente il requisito piú significativo dal punto di vista del committente (il Comune). Questo dato, però, non è abbastanza preciso per chi dovrà calcolare il costo e per chi dovrà progettare lo stabile. Allora, usando delle tabelle standard, si può esprimere questo requisito come volume abitabile. Questo dato è correlato direttamente alle dimensioni fisiche dello stabile, e quindi diviene un parametro di progetto. Possiamo considerare il numero di famiglie come un *requisito dell'utente* (nel senso che questa proprietà è richiesta *dall'*utente, o piú precisamente, in questo caso, dal committente) e la cubatura come un *requisito del sistema* (cioè una proprietà richiesta *al* sistema). Anche nell'industria del software è necessario, in generale, produrre delle specifiche con diversi livelli di dettaglio e di formalità, a seconda dell'uso e delle persone a cui sono destinate.

Una specifica descrive *che cosa* deve fare un sistema, o quali proprietà deve avere, ma non *come* deve essere costruito per esibire quel comportamento o quelle proprietà. La specifica di un palazzo non dice come è fatta la sua struttura, di quanti piloni e travi è composto, di quali dimensioni, e via dicendo, non dice cioè qual è la *realizzazione* di un palazzo che soddisfi i requisiti specificati (nel campo del software la realizzazione di solito si chiama *implementazione*).

Un palazzo è la realizzazione di una specifica, ma naturalmente la costruzione del palazzo deve essere preceduta da un progetto. Un *progetto* è un insieme di documenti che descrivono *come* deve essere realizzato un sistema. Per esempio, il progetto di un palazzo dirà che in un certo punto ci deve essere una trave di una certa forma con certe dimensioni. Questa descrizione della trave è, a sua volta, una specifica dei requisiti (proprietà fisiche richieste) della trave stessa. Infine, la trave "vera", fatta di cemento, è la realizzazione di tale specifica. Possiamo quindi vedere il "processo di svilup-

po" di un palazzo come una serie di passaggi: si produce una prima specifica orientata alle esigenze del committente, poi una specifica orientata ai requisiti del sistema, poi un progetto che da una parte è un'implementazione delle specifiche, e dall'altra è esso stesso una specifica per il costruttore.

Come si è già detto, una specifica dice cosa fa un sistema, elencandone i requisiti, e l'implementazione dice come. Dal punto di vista del progettista, un requisito è un obbligo imposto dall'esterno (e quindi fa parte della specifica) mentre l'implementazione è il risultato di una serie di scelte. Certe caratteristiche del sistema che potrebbero essere scelte di progetto, in determinati casi possono diventare dei requisiti: per esempio, i regolamenti urbanistici di un comune potrebbero porre un limite all'altezza degli edifici, oppure le norme di sicurezza possono imporre determinate soluzioni tecniche. Queste caratteristiche, quindi, non sono piú scelte dal progettista ma sono *vincoli* esterni. Piú in generale, un vincolo è una condizione che il sistema deve soddisfare, imposta da esigenze ambientali di varia natura (fisica, economica, legale...) o da limiti della tecnologia, che rappresenta un limite per le esigenze dell'utente o per le scelte del progettista. Un vincolo è quindi un requisito indipendente dalla volontà dell'utente.

Data la potenziale ambiguità nel ruolo (requisito/vincolo o scelta di progetto) di certi aspetti di un sistema, è importante che la documentazione prodotta durante il suo sviluppo identifichi tali ruoli chiaramente. Se ciò non avviene, si potrebbe verificare una situazione di questo tipo: un sistema viene realizzato rispettando vincoli e requisiti, e dopo qualche tempo se ne fa una nuova versione. Nello sviluppare questa versione, una soluzione implementativa imposta dai requisiti o dai vincoli viene scambiata per una scelta di progetto, e viene sostituita da una versione alternativa, che può aggiungere qualcosa al sistema originale, però non rispetta i requisiti e i vincoli, risultando quindi insufficiente. Supponiamo, per esempio, che i documenti di specifica chiedano che i record di un database possano essere elencati in ordine alfabetico, e che durante lo sviluppo dell'applicazione venga comunicato, *senza aggiornare la documentazione*, deve essere rispettato un limite sull'uso di memoria centrale. Conseguentemente, sceglieremo un algoritmo di ordinamento efficiente in termini di memoria, anche se piú lento di altri. Se una versione successiva dell'applicazione viene sviluppata da persone ignare del vincolo sulla memoria, queste potrebbero sostituire l'algoritmo con uno piú veloce, ma in questo modo verrebbe violato il vincolo, con la possibilità di malfunzionamenti.

### 1.3.3 Modelli e linguaggi

Per quanto esposto nella sezione precedente, il processo di sviluppo del software si può vedere come la costruzione di una serie di *modelli*. Un modello è una descrizione astratta di un sistema, che serve a studiarlo prendendone in considerazione soltanto quelle caratteristiche che sono necessarie al conseguimento di un certo scopo. Ogni sistema, quindi, dovrà essere rappresentato per mezzo di piú modelli, ciascuno dei quali ne mostra solo alcuni aspetti, spesso a diversi livelli di dettaglio.

Ovviamente un modello deve essere espresso in qualche linguaggio. Una parte considerevole di questo corso verrà dedicata a linguaggi concepiti per descrivere sistemi, sistemi software in particolare.

Un linguaggio ci offre un *vocabolario* di simboli (parole o segni grafici) che rappresentano i concetti necessari a descrivere certi aspetti di un sistema, una *sintassi* che stabilisce in quali modi si possono costruire delle espressioni (anche espressioni grafiche, cioè diagrammi), e una *semantica* che definisce il significato delle espressioni.

Nel seguito si userà spesso il termine "*formalismo*", per riferirsi ad una famiglia di linguaggi che esprimono concetti simili (avendo quindi un modello teorico comune) e hanno sintassi simili (e quindi uno stile di rappresentazione comune). Questi linguaggi peraltro possono presentare varie differenze concettuali o sintattiche.

## Letture

**Obbligatorie:** Cap. 1 Ghezzi, Jazayeri, Mandrioli, oppure Sez. 1.1–1.2 Ghezzi, Fuggetta et al., oppure Cap. 1 Pressman.

# Capitolo 2

# Ciclo di vita e modelli di processo

Ogni prodotto industriale ha un *ciclo di vita* che, a grandi linee, inizia quando si manifesta la necessità o l'utilità di un nuovo prodotto e prosegue con l'identificazione dei suoi requisiti, il progetto, la produzione, la verifica, e la consegna al cliente. Dopo la consegna, il prodotto viene usato ed è quindi oggetto di manutenzione e assistenza tecnica, e infine termina il ciclo di vita col suo ritiro. A queste *attività* se ne aggiungono altre, che spesso vi si sovrappongono, come la pianificazione e la gestione del processo di sviluppo, e la documentazione. Ciascuna attività ne comprende altre, ovviamente in modo diverso per ciascun tipo di prodotto e per ciascuna organizzazione produttrice.

Un *processo di sviluppo* è un particolare modo di organizzare le attività costituenti il ciclo di vita, cioè di assegnare risorse alle varie attività e fissarne le scadenze. Una *fase* è un intervallo di tempo in cui si svolgono certe attività, e ciascuna attività può essere ripartita fra piú fasi. I diversi processi possono essere classificati secondo alcuni *modelli di processo*; un modello di processo è quindi una descrizione generica di una famiglia di processi simili, che realizzano il modello in modi diversi.

Il ciclo di vita del software (Fig. 2.1) segue lo schema generale appena esposto, ma con alcune importanti differenze, particolarmente nella fase di produzione. Come abbiamo visto, nel software la riproduzione fisica del prodotto ha un peso economico ed organizzativo molto inferiore a quello che si trova nei prodotti tradizionali, per cui nel ciclo di vita del software il segmento corrispondente alla produzione è costituito dall'attività di programmazione,

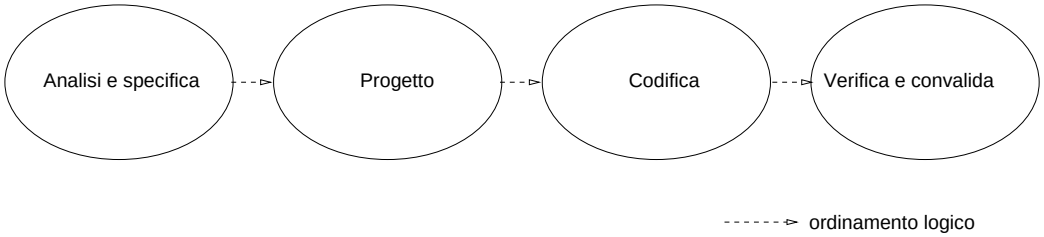


Figura 2.1: Il ciclo di vita del software.

che, al pari delle fasi precedenti di analisi e di progetto, è un'attività di carattere intellettuale piuttosto che materiale. Un'altra importante differenza sta nella fase di manutenzione, che nel software ha un significato completamente diverso da quello tradizionale, come vedremo piú oltre.

Il ciclo di vita del software verrà studiato prendendo come esempio un particolare modello di processo, il modello a cascata, in cui ciascuna attività del ciclo di vita corrisponde ad una fase del processo (Fig. 2.2). Successivamente si studieranno altri modelli di processo, in cui l'associazione fra attività e fasi del processo avviene in altri modi.

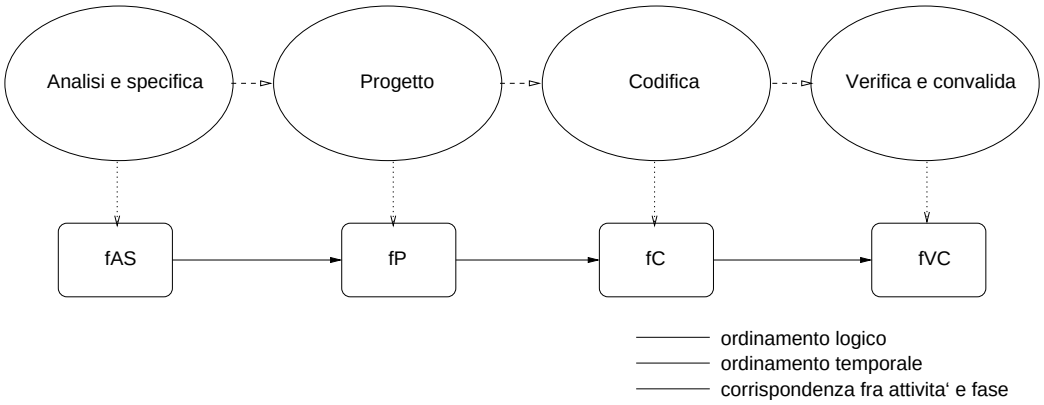


Figura 2.2: Il modello a cascata.

## 2.1 Il modello a cascata

Il *modello a cascata* (*waterfall*) prevede una successione di fasi consecutive. Ciascuna fase produce dei *semilavorati* (*deliverables*), cioè documenti relativi al processo, oppure documentazione del prodotto e codice sorgente o compilato, che vengono ulteriormente elaborati dalle fasi successive.

Il modello presuppone che ciascuna fase sia conclusa prima dell'inizio della fase successiva e che il flusso dei semilavorati sia, pertanto, rigidamente unidirezionale (come in una catena di montaggio): i risultati di una fase sono il punto di partenza della fase successiva, mentre non possono influenzare una fase precedente. Questo implica che in ogni fase si operi sul prodotto nella sua interezza: la fase di analisi produce le specifiche di tutto il sistema, quella di progetto produce il progetto di tutto il sistema. Quindi questo modello è adatto a progetti in cui i requisiti iniziali sono chiari fin dall'inizio e lo sviluppo del prodotto è prevedibile. Infatti, se in una certa fase si verificassero degli imprevisti, come la scoperta di errori od omissioni nel progetto o nelle specifiche, oppure l'introduzione di nuovi requisiti, allora si renderebbero necessarie la ripetizione delle fasi precedenti e la rielaborazione dei semilavorati prodotti fino a quel punto. Ma questa rielaborazione può essere molto costosa se la pianificazione del processo non la prevede fin dall'inizio: per esempio, il gruppo responsabile della fase di progetto potrebbe essere stato assegnato ad un altro incarico, o addirittura sciolto, all'inizio della fase di codifica. Inoltre, le modifiche sono rese costose dalle grandi dimensioni dei semilavorati.

Il numero, il contenuto e la denominazione delle fasi può variare da un'organizzazione all'altra, e da un progetto all'altro. Nel seguito ci riferiremo ad un processo articolato nelle seguenti fasi:

- *studio di fattibilità*;
- *analisi e specifica dei requisiti*, suddivisa in
  - *analisi (definizione e specifica) dei requisiti dell'utente*, e
  - *specifica dei requisiti del software*;
- *progetto*, suddiviso in
  - progetto *architetturale*, e
  - progetto *in dettaglio*;
- *programmazione e test di unità*;
- *integrazione e test di sistema*;
- *manutenzione*.

Contemporaneamente a queste fasi, e nel corso di tutto il processo, si svolgono anche queste attività di supporto:

- *gestione*;
- *controllo di qualità*;
- *documentazione*.

Lo studio del modello a cascata è importante sia perché è il modello piú noto, sia perché l'analisi delle varie fasi permette di descrivere delle attività che fanno parte di tutti i modelli di processo, anche se raggruppate e pianificate in modi diversi.

### 2.1.1 Studio di fattibilità

Lo *studio di fattibilità* serve a stabilire se un dato prodotto può essere realizzato e se è conveniente realizzarlo, ad accertare quali sono le possibili strategie alternative per la sua realizzazione, a proporne un numero ristretto, a valutare la quantità di risorse necessarie, e quindi i costi relativi.

I metodi ed i criteri di questa fase dipendono sensibilmente dal rapporto fra committente e produttore: il prodotto di cui si vàluta la fattibilità può essere destinato alla stessa organizzazione di cui fa parte il produttore (p. es., il produttore vuole realizzare uno strumento CASE per uso interno), oppure ad un particolare committente esterno (il produttore realizza un database per un'azienda), oppure, genericamente, al mercato (il produttore realizza un database generico). Nei primi due casi è importante il dialogo fra produttore e committente (o fra sviluppatore e utente), che permette di chiarire i requisiti. Nel terzo caso il ruolo del committente viene assunto da quel settore dell'azienda che decide le caratteristiche dei nuovi prodotti.

In base al risultato dello studio di fattibilità, il committente decide se firmare o no il contratto per la fornitura del software. La rilevanza economica di questa fase può influenzare negativamente la qualità dello studio di fattibilità poiché il produttore di software, per non perdere il contratto, può sottovalutare i costi e le difficoltà della proposta, ovvero sopravvalutare le proprie capacità e risorse. Questi errori di valutazione rischiano poi di causare ritardi e inadempienze contrattuali, con perdite economiche per il fornitore o per il committente.

A volte il committente paga lo studio di fattibilità, che in questo caso si può considerare un "prodotto finito" anche se il progetto iniziale cade. In questo caso lo studio di fattibilità è servito ad evitare il danno economico derivante dalla decisione di sviluppare un prodotto eccessivamente costoso o addirittura irrealizzabile. Inoltre, le conoscenze acquisite e rese accessibili nel corso dello studio di fattibilità contribuiscono ad arricchire il patrimonio di competenze del committente.

Il semilavorato prodotto dallo studio di fattibilità è un documento che dovrebbe contenere queste informazioni:

- una descrizione del problema che deve essere risolto dall'applicazione, in termini di obiettivi e vincoli;
- un insieme di scenari possibili per la soluzione, sulla base di un'analisi dello stato dell'arte, cioè delle conoscenze e delle tecnologie disponibili;
- le modalità di sviluppo per le alternative proposte, insieme a una stima dei costi e dei tempi richiesti.

### 2.1.2 Analisi e specifica dei requisiti

Questa fase serve a capire e descrivere nel modo piú completo e preciso possibile *che cosa vuole* il committente dal prodotto software. La fase di analisi e specifica può essere suddivisa nelle sottofasi di *analisi dei requisiti dell'utente* e *specifica dei requisiti del software*. La prima di queste sottofasi è rivolta alla comprensione del problema dell'utente e del contesto in cui dovrà operare il sistema software da sviluppare, richiede cioè una *analisi del dominio*, che porta all'acquisizione di conoscenze relative all'attività dell'utente stesso e all'ambiente in cui opera.

L'analisi dei requisiti dell'utente può essere ulteriormente suddivisa [13] in *definizione dei requisiti* e *specifica dei requisiti*: la definizione dei requisiti descrive ad alto livello servizi e vincoli mentre la specifica è piú dettagliata. Anche se la specifica dei requisiti contiene tutte le informazioni già fornite dalla definizione dei requisiti, a cui ne aggiunge altre, sono necessari tutti e due i livelli di astrazione, in quanto la definizione dei requisiti, essendo meno dettagliata, permette una migliore comprensione generale del problema. Inoltre, le descrizioni dei requisiti vengono usate da persone che hanno diversi ruoli e diverse competenze (committenti, amministratori, progettisti...), che richiedono diversi livelli di dettaglio.

Un livello di dettaglio ancora piú fine si ha nella specifica dei requisiti del software, che descrive le caratteristiche (*non* l'implementazione) del software che deve essere prodotto per soddisfare le esigenze dell'utente. Nella sottofase di specifica dei requisiti del software è importante evitare l'introduzione di scelte implementative, che in questa fase sono premature.

Come esempio di questi tre livelli di dettaglio, consideriamo il caso di uno strumento che permette di operare su file prodotti da altri strumenti, usando un'interfaccia grafica, come avviene, per esempio, col *desktop* di un PC (esempio adattato da [13]). Uno dei requisiti potrebbe essere espresso nei seguenti modi:

**Analisi dei requisiti dell'utente**

**Definizione dei requisiti dell'utente**

**1** L'applicazione deve permettere la rappresentazione e l'elaborazione di file creati da altre applicazioni (detti *file esterni*).

**Specifica dei requisiti dell'utente**

**1.1** L'applicazione deve permettere all'utente di definire i tipi dei file esterni.

**1.2** Ad ogni tipo di file esterno corrisponde un programma esterno ed opzionalmente un'icona che viene usata per rappresentare il file. Se al tipo di un file non è associata alcuna icona, viene usata un'icona default non associata ad alcun tipo.

**1.3** L'applicazione deve permettere all'utente di definire l'icona associata ai tipi di file esterni.

**1.4** La selezione di un'icona rappresentante un file esterno causa l'elaborazione del file rappresentato, per mezzo del programma associato al tipo del file stesso.

**Specifica dei requisiti del software**

**1.1.1** L'utente può definire i tipi dei file esterni sia per mezzo di menú che di finestre di dialogo. È opzionale la possibilità di definire i tipi dei file esterni per mezzo di file di configurazione modificabili dall'utente.

**1.2.1** L'utente può associare un programma esterno ad un tipo di file esterno sia per mezzo di finestre di dialogo che di file di configurazione.

**1.2.2** L'utente può associare un'icona ad un tipo di file esterno per mezzo di una finestra di selezione grafica (*chooser*).

**1.3.1** L'applicazione deve comprendere una libreria di icone già pronte ed uno strumento grafico che permetta all'utente di crearne di nuove.

**1.4.1** La selezione di un'icona rappresentante un file esterno può avvenire sia per mezzo del mouse che della tastiera.

### Requisiti funzionali e non funzionali

I requisiti possono essere *funzionali* o *non funzionali*. I requisiti funzionali descrivono cosa deve fare il prodotto, generalmente in termini di relazioni fra dati di ingresso e dati di uscita, mentre i requisiti non funzionali sono caratteristiche di qualità come, per esempio, l'affidabilità o l'usabilità, oppure vincoli di varia natura. Di questi requisiti si parlerà piú diffusamente in seguito.

Altri requisiti possono riguardare il processo di sviluppo anziché il prodotto. Per esempio, il committente può richiedere che vengano applicate

determinate procedure di controllo di qualità o vengano seguiti determinati standard.

### Documenti di specifica

Il prodotto della fase di analisi e specifica dei requisiti generalmente è costituito da questi documenti:

**Documento di Specifica dei Requisiti (DSR).** È il fondamento di tutto il lavoro successivo, e, se il prodotto è sviluppato per un committente esterno, ha anche un valore legale poiché viene incluso nel contratto.

**Manuale Utente.** Descrive il comportamento del sistema dal punto di vista dell'utente ("se tu fai questo, succede quest'altro").

**Piano di Test di Sistema.** Definisce come verranno eseguiti i test finali per convalidare il prodotto rispetto ai requisiti. Anche questo documento può avere valore legale, se firmato dal committente, che cosí accetta l'esecuzione del piano di test come collaudo per l'accettazione del sistema.

È di fondamentale importanza che il DSR sia *completo* e *consistente*. "Completo" significa che contiene esplicitamente tutte le informazioni necessarie, e "consistente" significa che non contiene requisiti reciprocamente contraddittori. Idealmente, dovrebbe essere scritto in un linguaggio formale, tale da permettere un'interpretazione non ambigua ed una verifica rigorosa, ma di solito è in linguaggio naturale, al piú strutturato secondo qualche standard, e accompagnato da notazioni semiformali (diagrammi etc.).

A proposito dell'uso del linguaggio naturale nei documenti di specifica, osserviamo che il significato dei termini usati può essere definito dai *glossari*. Per ogni progetto è opportuno preparare un glossario dei termini ad esso specifici; inoltre, esistono numerosi glossari standard, come, per esempio, la norma ISO-8402 relativa alla terminologia della gestione e assicurazione della qualità. Un documento può anche contenere istruzioni relative ad usi particolari (generalmente piú ristretti) di termini appartenenti al linguaggio ordinario. Per esempio, è uso comune che il verbo (inglese) "shall" denoti comportamenti o caratteristiche obbligatori, il verbo "should" denoti comportamenti o caratteristiche desiderabili (ma non obbligatori), ed il verbo "may" denoti comportamenti o caratteristiche permessi o possibili.

### 2.1.3 Progetto

In questa fase si stabilisce *come* deve essere fatto il sistema definito dai documenti di specifica (DSR e manuale utente). Poiché, in generale, esistono diversi modi di realizzare un sistema che soddisfi un insieme di requisiti, l'attività del progettista consiste essenzialmente in una serie di *scelte* fra le soluzioni possibili, guidate da alcuni principi e criteri che verrano illustrati nei capitoli successivi.

Il risultato del progetto è una *architettura software*, cioè una scomposizione del sistema in elementi strutturali, detti *moduli*, dei quali vengono specificate le funzionalità e le relazioni reciproche. La fase di progetto può essere suddivisa nelle sottofasi di *progetto architetturale* e di *progetto in dettaglio*. Nella prima fase viene definita la struttura generale del sistema, mentre nella seconda si definiscono i singoli moduli. La distinzione fra queste due sottofasi spesso non è netta, e nelle metodologie di progetto piú moderne tende a sfumare.

Il principale semilavorato prodotto da questa fase è il *Documento delle Specifiche di Progetto* (DSP).

Anche il DSP dovrebbe poter essere scritto in modo rigoroso ed univoco, possibilmente usando notazioni formali (*linguaggi di progetto*). In pratica ci si affida prevalentemente al linguaggio naturale integrato con notazioni grafiche.

In questa fase può essere prodotto anche il *Piano di Test di Integrazione*, che prescrive come collaudare l'interfacciamento fra i moduli nel corso della costruzione del sistema (Sez. 6.5.1).

### 2.1.4 Programmazione (codifica) e test di unità

In questa fase i singoli moduli definiti nella fase di progetto vengono implementati e testati singolarmente. Vengono scelte le strutture dati e gli algoritmi, che di solito non vengono specificati dal DSP.

Evidentemente questa fase è cruciale in quanto consiste nella realizzazione pratica di tutte le specifiche e le scelte delle fasi precedenti. Il codice prodotto in questa fase, oltre ad essere conforme al progetto, deve avere delle qualità che, pur essendo invisibili all'utente, concorrono in modo determinante sia alla bontà del prodotto che all'efficacia del processo di produzione. Fra queste qualità citiamo la *modificabilità* e la *leggibilità*. Per conseguire tali

qualità è necessario adottare degli standard di codifica, che stabiliscono, ad esempio, il formato (nel senso tipografico) dei file sorgente, le informazioni che devono contenere oltre al codice (autore, identificazione del modulo e della versione...), ed altre convenzioni.

L'attività di codifica è strettamente collegata a quella di testing e di debugging. Tradizionalmente queste tre attività sono affidate alla stessa persona, che le esegue secondo i propri criteri. Tuttavia l'attività di testing di unità, per la sua influenza critica sulla qualità del prodotto, deve essere svolta in modo metodico e pianificato, e si deve avvalere di metodologie specifiche.

In questo corso non si parlerà della programmazione, di cui si suppongono noti i principî e le tecniche, ma si indicheranno le caratteristiche di alcuni linguaggi che permettono di applicare alcuni concetti relativi alla progetto del software. Inoltre vogliamo accennare in questa sezione ad alcuni aspetti del lavoro di programmazione e ad alcuni strumenti relativi:

**Gestione delle versioni.** Durante la programmazione, vengono prodotte numerose versioni dei componenti software, ed è importante conservare e gestire tali versioni. Uno strumento molto diffuso è il *Concurrent Versioning System* (*CVS*), che permette di gestire un archivio (*repository*[1]) del codice sorgente a cui gli sviluppatori possono accedere in modo concorrente, anche da locazioni remote[2]. Un'altro strumento di questo tipo è *Subversion* (*SVN*)[3].

**Configurazione e compilazione automatica.** Esistono strumenti che permettono di automatizzare il processo di costruzione (compilazione e collegamento) del software, e di configurare questo processo, cioè di adattarlo a diverse piattaforme software. Alcuni di questi strumenti sono i programmi *Make*, *Automake*, *Autoconf*, e *Libtool*, noti collettivamente come *Autotools* [4].

**Notifica e archiviazione di malfunzionamenti.** Lo strumento *Bugzilla*, basato su web, permette di segnalare agli sviluppatori i guasti rilevati nell'uso del software, di archiviare tali notifiche, e di tenere utenti e sviluppatori al corrente sui progressi nell'attività di *debugging*[5].

**Test di unità.** Il test di unità può essere parzialmente automatizzato grazie

[1] Si pronuncia con l'accento tonico sulla seconda sillaba.
[2] v. `www.cvshome.org`
[3] v. `subversion.tigris.org`
[4] v. `www.gnu.org/manual`
[5] v. `www.mozilla.org/projects/bugzilla`

a strumenti come *DejaGNU*[6], *CppUnit*[7], *mockpp*[8].

Il prodotto della fase di programmazione e test di unità è costituito dal codice dei programmi con la relativa documentazione e dalla documentazione relativa ai test.

### 2.1.5 Integrazione e test di sistema

In questa fase viene assemblato e collaudato il prodotto completo. Il costo di questa fase ovviamente è tanto maggiore quanto maggiori sono le dimensioni e la complessità dell'applicazione. Specialmente se l'applicazione viene sviluppata da gruppi di lavoro diversi, questa fase richiede un'accurata pianificazione. Il lavoro svolto in questa fase viene tanto piú facilitato quanto piú l'applicazione è modulare.

Nel corso dell'integrazione vengono assemblati i vari sottosistemi a partire dai moduli componenti, effettuando parallelamente il *test di integrazione*, che verifica la corretta interazione fra i moduli. Dopo che il sistema è stato assemblato completamente, viene eseguito il *test di sistema*.

Il test di sistema può essere seguito, quando l'applicazione è indirizzata al mercato, da questi test:

**alfa test:** l'applicazione viene usata all'interno all'azienda produttrice;
**beta test:** l'applicazione viene usata da pochi utenti esterni selezionati (*beta tester*).

### 2.1.6 Manutenzione

Il termine "manutenzione" applicato al software è improprio, poiché il software non soffre di usura o invecchiamento e non ha bisogno di rifornimenti. La cosiddetta manutenzione del software è in realtà la riparazione di difetti presenti nel prodotto consegnato al cliente, oppure l'aggiornamento del codice allo scopo di fornire nuove versioni. Questa riparazione consiste nel modificare e ricostruire il software. Si dovrebbe quindi parlare di *riprogettazione* piuttosto che di manutenzione. La scoperta di difetti dovrebbe portare ad un riesame critico del progetto e delle specifiche, ma nella prassi comune

[6] v. `www.gnu.org/software/dejagnu`
[7] v. `cppunit.sourceforge.net`
[8] v. `mockpp.sourceforge.net`

questo non avviene, particolarmente quando si adotta il modello a cascata. La manutenzione avviene piuttosto attraverso un "rattoppo" (*patch*) del codice sorgente. Questo fa sí che il codice non corrisponda piú al progetto, per cui aumenta la difficoltà di correggere ulteriori errori. Dopo una serie di operazioni di manutenzione, il codice sorgente può essere talmente degradato da perdere la sua struttura originaria e qualsiasi relazione con la documentazione. In certi casi diventa necessario ricostruire il codice con interventi di *reingegnerizzazione* (*reengineering* o *reverse engineering*).

Per contenere i costi e gli effetti avversi della manutenzione è necessario tener conto fin dalla fase di progetto che sarà necessario modificare il software prodotto. Questo principio è chiamato "progettare per il cambiamento" (*design for change*). Anche qui osserviamo che la modularità del sistema ne facilita la modifica.

Si distinguono i seguenti tipi di manutenzione:

**correttiva:** individuare e correggere errori;

**adattativa:** cambiamenti di ambiente operativo (*porting*) in senso stretto, cioè relativo al cambiamento di piattaforma hardware e software, ma anche in senso piú ampio, relativo a cambiamenti di leggi o procedure, linguistici e culturali (per esempio, le date si scrivono in modi diversi in diversi paesi);

**perfettiva:** aggiunte e miglioramenti.

### 2.1.7 Attività di supporto

Alcune attività vengono svolte contemporaneamente alle fasi già illustrate:

**documentazione:** La maggior parte dei deliverable è costituita da documentazione. Altra documentazione viene prodotta ed usata internamente al processo di sviluppo, come, per esempio, rapporti periodici sull'avanzamento dei lavori, linee guida per gli sviluppatori, verbali delle riunioni, e simili.

**controllo di qualità:** Oltre ai controlli finali sul prodotto, occorre controllare ed assicurare la qualità dei semilavorati ottenuti dalle fasi intermedie.

**gestione:** La gestione comprende la pianificazione del processo di sviluppo, la allocazione delle risorse (in particolare le risorse umane, con lo *staffing*), l'organizzazione dei flussi di informazione fra i gruppi di lavoro e al loro interno, ed altre attività di carattere organizzativo. Un aspetto particolare della gestione, orientato al prodotto piú che al processo,

è la *gestione delle configurazioni*, volta a mantenere i componenti del prodotto e le loro versioni in uno stato aggiornato e consistente.

## 2.2 Modelli evolutivi

Abbiamo osservato che nello sviluppo del software bisogna prevedere la necessità di cambiamenti. È quindi opportuno usare processi di sviluppo in cui la necessità di introdurre dei cambiamenti venga rilevata tempestivamente ed i cambiamenti stessi vengano introdotti facilmente.

Nei processi basati su modelli evolutivi, il software viene prodotto in modo incrementale, in passi successivi (Fig. 2.3). Ogni passo produce, a seconda delle varie strategie possibili, una parte nuova oppure una versione via via piú raffinata e perfezionata del sistema complessivo. Il prodotto di ciascun passo viene valutato ed i risultati di questa valutazione determinano i passi successivi, finché non si arriva ad un sistema che risponde pienamente alle esigenze del cliente, almeno finché non si sentirà il bisogno di una nuova versione.

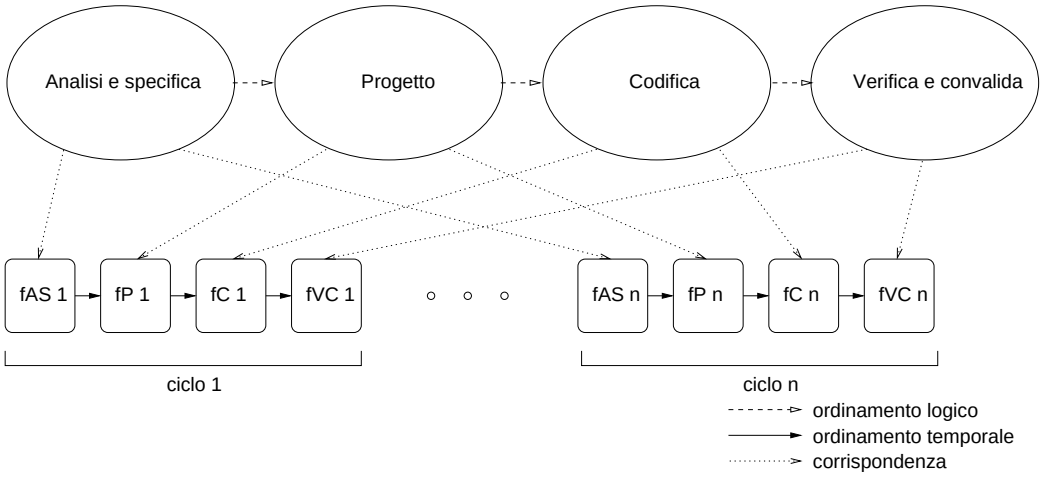


Figura 2.3: Un processo incrementale.

### 2.2.1 Prototipazione

Un *prototipo* è una versione approssimata, parziale, ma funzionante, dell'applicazione che viene sviluppata. La prototipazione, cioè la costruzione e l'uso

di prototipi, entra nei modelli evolutivi in diversi modi [8]:

- Un prototipo può essere costruito e valutato nel corso dello studio di fattibilità.
- Un prototipo *esplorativo* viene costruito e consegnato all'utente durante l'analisi dei requisiti, in modo che l'utente possa provarlo e chiarire i requisiti del sistema.
- Un prototipo *sperimentale* viene usato nella fase di progetto per studiare diverse implementazioni alternative.
- Un prototipo può essere una parte funzionante ed autonoma del sistema finale, e può essere consegnato all'utente, che comincia ad usarlo nella propria attività (strategia *early subset, early delivery*).

Quando il prototipo viene usato nell'analisi dei requisiti, bisogna scegliere se buttar via il prototipo, una volta ottenuta l'approvazione dell'utente, per passare alla progetto *ex novo* del sistema, oppure costruire il prodotto finale ampliando e perfezionando il prototipo. Nel primo caso si parla di prototipo *usa e getta* (*throw-away*), nel secondo si parla di prototipo *evolutivo* o *sistema pilota*.

Un prototipo usa e getta ha quindi una struttura interna del tutto diversa da quello che sarà il prodotto finale, e può essere realizzato senza vincoli di efficienza, per esempio usando linguaggi dichiarativi (come, p.es., il Prolog) o linguaggi di scripting (p.es., Perl o Python) che permettono tempi di sviluppo piú brevi a scapito delle prestazioni e di altre qualità desiderabili del prodotto.

Un prototipo evolutivo richiede che fin dall'inizio si facciano scelte che condizioneranno le fasi successive. Questo richiede naturalmente un maggiore sforzo di progetto. Un rischio comune nell'uso dei prototipi consiste nel "prendere sul serio" delle scelte che in fase di prototipazione erano accettabili e trascinarle fino alla realizzazione del sistema, anche se inadeguate al prodotto finale. Per esempio, durante la prototipazione si può scegliere un algoritmo inefficiente per una determinata funzione, in attesa di trovarne uno migliore; ma una volta che il prototipo è fatto e "funziona" (piú o meno bene), è facile cedere alla tentazione di prendere per buono quello che è stato fatto e lasciare nel prodotto un componente di cui fin dall'inizio era nota l'inadeguatezza.

Infine, osserviamo che una forma molto comune di prototipazione consiste nello sviluppo dell'interfaccia utente. A questo scopo sono disponibili ambienti, linguaggi e librerie che permettono di realizzare facilmente delle interfacce interattive.

### 2.2.2 Lo Unified Process

Lo Unified Process (UP) [5, 1] è un processo evolutivo concepito per lo sviluppo di software orientato agli oggetti ed è stato concepito dagli ideatori del linguaggio UML. Questo processo si può schematizzare come segue:

- l'arco temporale del processo di sviluppo è suddiviso in quattro *fasi* successive;
- ogni fase ha un obiettivo e produce un insieme di semilavorati chiamato *milestone* ("pietra miliare");
- ogni fase è suddivisa in un numero variabile di *iterazioni*;
- nel corso di ciascuna iterazione possono essere svolte tutte le attività richieste (analisi, progetto...), anche se, a seconda della fase e degli obiettivi dell'iterazione, alcune attività possono essere predominanti ed altre possono mancare;
- ciascuna iterazione produce una versione provvisoria (*baseline*) del prodotto, insieme alla documentazione associata.

#### Attività (*workflow*)

Nello UP si riconoscono cinque attività, dette *workflow* (o *flussi di lavoro*): *raccolta dei requisiti* (*requirements*), *analisi* (*analysis*), *progetto* (*design*), *implementazione* (*implementation*), e *collaudo* (*test*). Le prime due attività corrispondono a quella che abbiamo chiamato complessivamente *analisi e specifica dei requisiti*.

#### Fasi

Le quattro fasi dello UP sono:

**Inizio (*inception*):** i suoi obiettivi corrispondono a quelli visti per lo studio di fattibilità, a cui si aggiunge una analisi dei rischi di varia natura (tecnica, economica, organizzativa...) in cui può incorrere il progetto. Anche il *milestone* di questa fase è simile all'insieme di documenti e altri artefatti (per esempio, dei prototipi) che si possono produrre in uno studio di fattibilità. Un documento caratteristico dello UP, prodotto in questa fase, è il *modello dei casi d'uso*, cioè una descrizione sintetica delle possibili interazioni degli utenti col sistema, espressa mediante la notazione UML.

**Elaborazione (*elaboration*):** gli obiettivi di questa fase consistono nell'estendere e perfezionare le conoscenze acquisite nella fase precedente, e

nel produrre una *baseline architetturale eseguibile*. Questa è una prima versione, eseguibile anche se parziale, del sistema, che non si deve considerare un prototipo, ma una specie di ossatura che serva da base per lo sviluppo successivo. Il milestone della fase comprende quindi il codice che costituisce la baseline, il suo modello costituito da vari diagrammi UML, e le versioni aggiornate dei documenti prodotti nella fase di inizio.

**Costruzione (*construction*):** ha l'obiettivo di produrre il sistema finale, partendo dalla baseline architetturale e completando le attività di raccolta dei requisiti, analisi e progetto portate avanti nelle fasi precedenti. Il milestone comprende, fra l'altro, il sistema stesso, la sua documentazione in UML, una *test suite* e i manuali utente. La fase si conclude con un periodo di beta-test.

**Transizione (*transition*):** gli obiettivi di questa fase consistono nella correzione degli errori trovati in fase di beta-test e quindi nella consegna e messa in opera (*deployment*) del sistema. Il milestone consiste nella versione definitiva del sistema e dei manuali utente, e nel piano di assistenza tecnica.

### Distribuzione delle attività nelle fasi

Come risulta dai paragrafi precedenti, in ciascuna fase si possono svolgere attività diverse. Nella fase di inizio sono preponderanti le attività di raccolta e analisi dei requisiti, ma c'è una componente di progettazione per definire un'architettura iniziale ad alto livello, non eseguibile. Può esere richiesta anche l'attività di implementazione, se si realizzano dei prototipi. Nella fase di elaborazione le cinque attività tendono ad avere pesi simili nello sforzo complessivo, con la tendenza per le attività di analisi a decrescere nelle iterazioni finali in termini di lavoro impegnato, mentre le attività di progetto e implementazione crescono corrispondentemente. Nella fase di costruzione sono preponderanti le attività di progetto e implementazione, ma non sono ancora terminate quelle di analisi, essendo previsto, come in tutti i processi evolutivi, che i requisiti possano cambiare in qualunque momento. Naturalmente, un accurato lavoro di analisi dei requisiti nelle fasi iniziali renderà poco probabile l'eventualità di grandi cambiamenti nelle fasi finali, per cui ci si aspetta che in queste ultime i cambiamenti siano limitati ed abbiano scarso impatto sull'architettura del sistema. Infine, nella fase di transizione i requisiti dovrebbero essere definitivamente stabilizzati e le attività di progetto e implementazione dovrebbero essere limitate alla correzione degli ultimi errori.

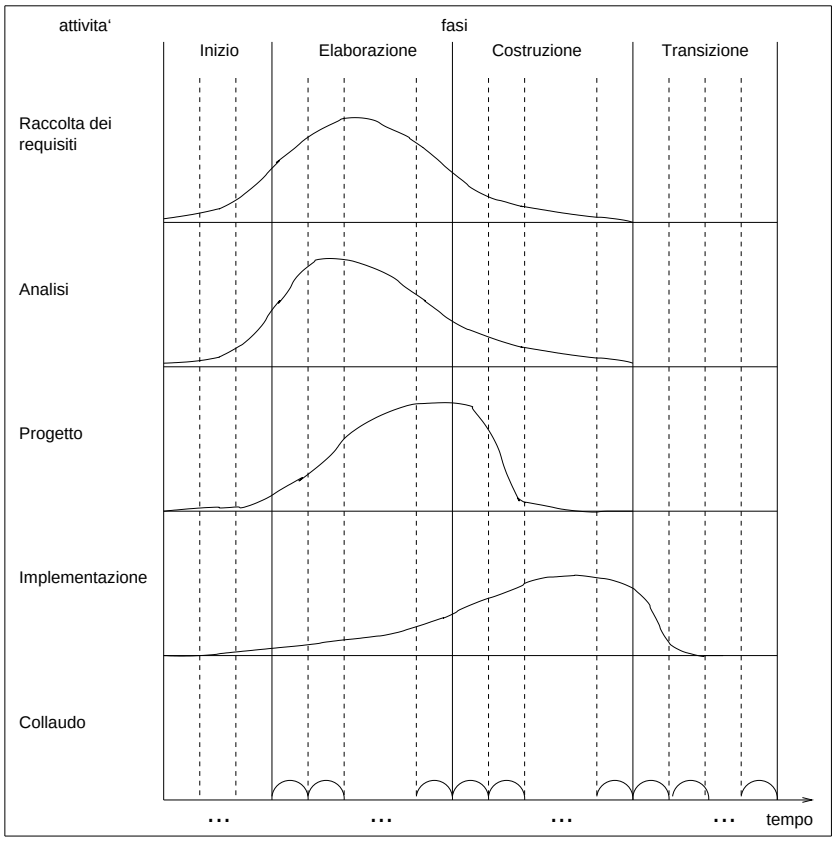


Figura 2.4: Fasi e attività nello Unified Process.

Questo andamento è schematizzato in Fig. 2.4. Le linee tratteggiate delimitano le iterazioni (che non necessariamente sono di durata uniforme come sembra mostrare la figura) e le cinque curve mostrano, in modo molto qualitativo e del tutto ipotetico, l'impegno richiesto dalle attività nel corso del processo di sviluppo.

### 2.2.3 Modelli trasformazionali

Nei modelli trasformazionali, il processo di sviluppo consiste in una serie di *trasformazioni* che, partendo da una specifica iniziale ad alto livello, producono delle descrizioni del sistema sempre piú dettagliate, fino al punto in cui si può passare alla fase di codifica. Tutte le specifiche sono *formali* cioè espresse in un linguaggio di tipo matematico, in modo che la correttezza di ciascuna versione delle specifiche rispetto alla versione precedente possa essere verificata in modo rigoroso (esclusa, naturalmente, la specifica iniziale, che pur

essendo espressa formalmente non può essere verificata rispetto ai requisiti dell'utente, necessariamente informali). Le trasformazioni da una specifica all'altra sono anch'esse formali, e quindi tali da preservare la correttezza delle specifiche.

Un processo trasformazionale è analogo alla soluzione di un'equazione:

$$f(x, y) = g(x, y)$$

$$\dots$$

$$y = h(x)$$

La formula iniziale viene trasformata attraverso diversi passaggi, usando le regole dell'algebra, fino alla forma finale.

Questa analogia, però, non è perfetta, perché nel caso delle equazioni la soluzione contiene la stesse informazioni dell'equazione iniziale (in forma esplicita invece che implicita), mentre nel processo di sviluppo del software vengono introdotte nuove informazioni nei passi intermedi, man mano che i requisiti iniziali vengono precisati o ampliati.

### Esempio

Consideriamo, per esempio, un sistema costituito da due sottosistemi, il primo dei quali, $\mathcal{S}_1$, è formato dai due processi $P$ e $Q$, e il secondo, $\mathcal{S}_2$, dai processi $R$ ed $S$. Il processo $P$ può eseguire le azioni $a$ e $b$, il processo $Q$ le azioni $c$ e $d$, il processo $R$ l'azione $a$, e il processo $S$ l'azione $c$.

Da una prima specifica risulta che nel sottosistema $\mathcal{S}_1$ i processi $P$ e $Q$ si possono evolvere in modo concorrente senza alcun vincolo di sincronizzazione reciproca, cosí come i processi $R$ ed $S$ nel sottosistema $\mathcal{S}_2$. I due sottosistemi, invece, hanno dei vincoli di sincronizzazione reciproca, dovendo eseguire insieme le azioni $a$ e $c$. La Fig. 2.5 schematizza questa descrizione.

Un sistema di questo tipo può essere descritto con un formalismo appartenente alla famiglia delle *algebre dei processi*, per esempio col linguaggio LOTOS. In questo linguaggio sono definiti numerosi operatori di composizione fra processi, fra cui quello di *interleaving*, cioè l'esecuzione concorrente senza vincoli, rappresentato da $|||$, e quello di *sincronizzazione*, rappresentato da $|[\cdots]|$. A ciascun operatore sono associate una *semantica* che definisce il risultato della composizione (in termini delle possibili sequenze di azioni eseguite dal processo risultante) e delle *regole di trasformazione* (o *di inferenza*) sulle espressioni contenenti l'operatore.

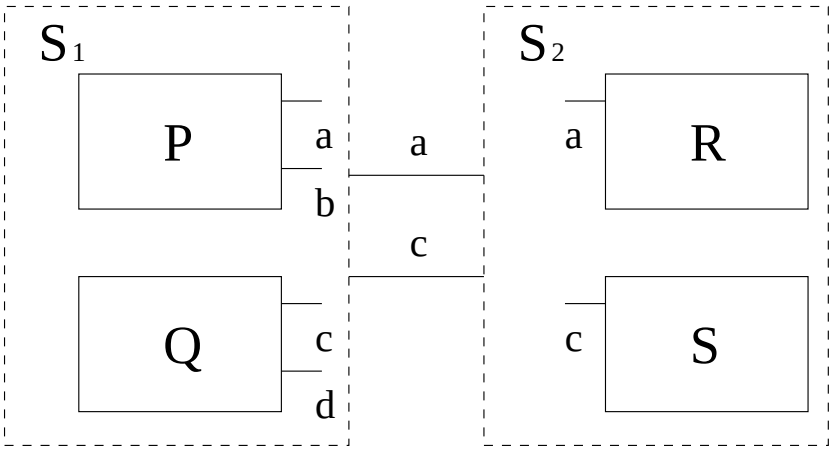


Figura 2.5: Struttura risultante dalla specifica iniziale.

Il sistema considerato viene quindi rappresentato dalla seguente espressione LOTOS:

$$(P[a,b] \;|||\; Q[c,d]) \;|[a,c]|\; (R[a] \;|||\; S[c])$$

L'espressione si può leggere cosí:

> I processi $P$ e $Q$ possono eseguire qualsiasi sequenza di azioni degli insiemi $\{a,b\}$ e $\{c,d\}$, rispettivamente, senza vincoli reciproci, e analogamente i processi $R$ ed $S$ con le azioni degli insiemi $\{a\}$ e $\{c\}$. I due sistemi $(P[a,b] \;|||\; Q[c,d])$ e $(R[a] \;|||\; S[c])$ sono sincronizzati su $a$ e $c$, cioè devono eseguire contemporaneamente ciascuna di queste azioni.

Le regole di trasformazione permettono di riscrivere questa espressione nel modo seguente:

$$(P[a,b] \;|[a]|\; R[a]) \;|||\; (Q[c,d] \;|[c]|\; S[c])$$

La nuova espressione si può leggere cosí:

> I processi $P$ ed $R$ sono sincronizzati su $a$, I processi $Q$ ed $S$ sono sincronizzati su $c$. I due sistemi $(P[a,b] \;|[a]|\; R[a])$ e $(Q[c,d] \;|[c]|\; S[c])$ possono eseguire qualsiasi sequenza permessa dai rispettivi vincoli di sincronizzazione, senza vincoli reciproci.

La seconda espressione è semanticamente equivalente alla prima, ma rappresenta una diversa organizzazione interna (e quindi una possibile implementazione) del sistema, che adesso è scomposto nei due sottosistemi $\mathcal{S}' = \{P, R\}$ e $\mathcal{S}'' = \{Q, S\}$ (Fig. 2.6). Il nuovo raggruppamento dei processi è migliore del precedente, perché mette insieme i processi che devono interagire e separa quelli reciprocamente indipendenti.

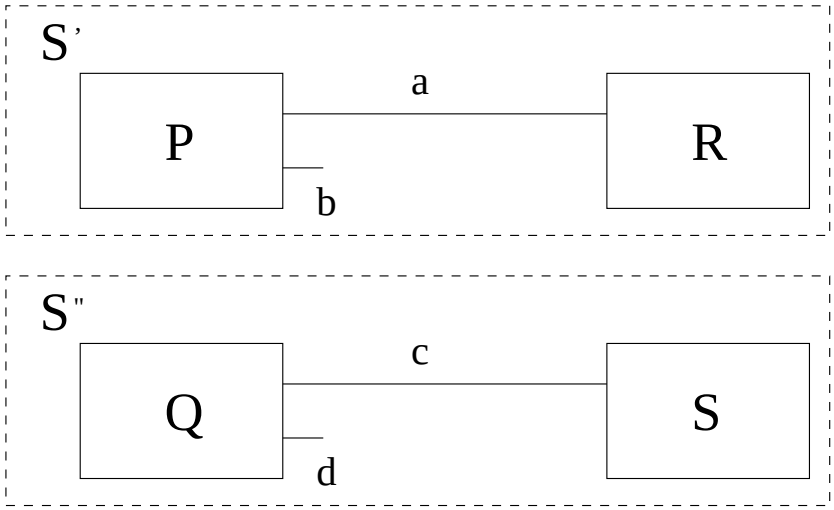


Figura 2.6: Struttura trasformata.

### Processo di sviluppo

Un processo trasformazionale generalmente è assistito da uno strumento CASE che esegue le trasformazioni scelte dal progettista e, qualora vengano introdotte nuove informazioni (per esempio, requisiti o vincoli aggiuntivi), ne verifica la consistenza rispetto a quelle già esistenti.

Lo strumente CASE può permettere di associare ad ogni scelta le relative motivazioni. Viene cosí costruito un database di progetto (*repository*) contenente tutte le informazioni relative alla storia ed allo stato attuale del progetto. Osserviamo però che l'uso del repository non è limitato ai processi di tipo trasformazionale.

L'ambiente di progetto può comprendere un sistema di gestione delle configurazioni che impedisce di modificare il codice senza prima modificare le specifiche rilevanti. Dopo che sono state aggiornate le specifiche al livello di astrazione appropriato, viene ricostruito il processo di derivazione fino ad

ottenere una nuova versione del software. Si evita cosí quella degradazione del software a cui si è accennato parlando della manutenzione.

Se le specifiche sono espresse in un linguaggio eseguibile, come avviene di norma, allora ciascuna specifica è un prototipo di tipo evolutivo. Il modello trasformazionale è quindi caratterizzato dalla prototipazione e dall'uso di linguaggi formali. È concepibile un processo di tipo trasformazionale che non usi linguaggi di specifica eseguibili, ma generalmente sono disponibili degli interpreti per i linguaggi formali di specifica, che rendono possibile la prototipazione.

Il prodotto finale può essere implementato nello stesso linguaggio delle specifiche, o in un linguaggio diverso.

### 2.2.4 Modello Cleanroom

Il modello Cleanroom [9] è un modello evolutivo concepito con l'obiettivo di produrre software privo di errori. Questo risultato si dovrebbe ottenere attraverso l'uso di specifiche formali, l'applicazione di metodi rigorosi di verifica statica, cioè basata sull'analisi del codice e non sulla sua esecuzione, e la certificazione per mezzo di collaudo statistico indipendente. Il software viene prodotto come una serie di incrementi, ciascuno dei quali viene specificato, realizzato e collaudato ripetutamente finché non supera le prove di certificazione, dopo di che l'incremento può essere consegnato all'utente che lo convalida, eventualmente chiedendo una modifica delle specifiche e ripetendo il ciclo. Quando un incremento è accettato, si passa all'incremento successivo.

Piú in dettaglio:

- Le attività di specifica, sviluppo e certificazione (verifica) sono affidate a tre gruppi distinti.
- Il software viene sviluppato in cicli successivi, ognuno dei quali produce un *incremento* del prodotto.
- In ogni ciclo vengono definiti i requisiti dell'incremento, e in base a questi requisiti il gruppo di sviluppo produce il codice ed il gruppo di certificazione prepara i dati di prova (*test cases*). Il gruppo di sviluppo *non esegue i programmi* e si affida unicamente a metodi di analisi statica per valutare la correttezza del codice. Quando il gruppo di sviluppo confida di aver prodotto del codice corretto, lo consegna al gruppo di certificazione che prova il sistema complessivo (insieme di incrementi ottenuti fino a quel momento), certificandone l'affidabilità

statisticamente e restituendo i risultati al gruppo di sviluppo. Questo modifica il software e fornisce al gruppo di certificazione una *notifica di cambiamento del progetto* (*engineering change notice*). Il ciclo viene quindi ripetuto.

- Il processo termina quando si raggiunge un'affidabiltà accettabile.

L'affidabilità viene quantificata misurando la frequenza dei guasti nel corso del collaudo. Il collaudo è una simulazione dell'uso del software basata sul *profilo operativo* dell'applicazione, cioè uno schema tipico delle interazioni degli utenti col sistema. Per esempio, se si deve valutare l'affidabilità di un sito web si può ricavare un profilo operativo dalla storia degli accessi a siti web simili: questo profilo dice che tipo di richieste vengono fatte e con quale frequenza, e con queste informazioni si può simulare il carico del nuovo sito.

## Letture

**Obbligatorie:** Cap. 7 Ghezzi, Jazayeri, Mandrioli, oppure Cap. 1 Ghezzi, Fuggetta et al., oppure Cap. 3 Pressman. Quest'ultimo espone l'argomento in modo un po' diverso, ma ovviamente valido.

**Facoltative:** Sez. 13.7.2 Pressman, Cap. 2 Arlow, Neustadt.

# Capitolo 3

# Analisi e specifica dei requisiti

In questo capitolo presentiamo alcuni linguaggi e metodi usati nella fase di analisi e specifica dei requisiti. I requisiti descrivono ciò che l'utente si aspetta dal sistema, e *specificarli* significa esprimerli in modo chiaro, univoco, consistente e completo. I requisiti si distinguono in *funzionali* e *non funzionali*.

I requisiti funzionali descrivono cosa deve fare il sistema, generalmente in termini di relazioni fra dati di ingresso e dati di uscita, oppure fra *stimoli* (dall'ambiente al sistema) e *risposte* del sistema. Questi requisiti sono generalmente esprimibili in modo formale.

I requisiti non funzionali esprimono dei *vincoli* o delle caratteristiche di qualità. Queste ultime sono piú difficili da esprimere in modo formale, in particolare è difficile esprimerle in modo quantitativo. Fra le caratteristiche di qualità del software ricordiamo le seguenti:

**sicurezza (safety):** Capacità di funzionare senza arrecare danni a persone o cose (piú precisamente, con un rischio limitato a livelli accettabili). Si usa in questo senso anche il termine *innocuità*.

**riservatezza (security):** Capacità di impedire accessi non autorizzati ad un sistema, e in particolare alle informazioni in esso contenute. Spesso il termine *sicurezza* viene usato anche in questo senso.

**robustezza:** Capacità di funzionare in modo accettabile anche in situazioni non previste, come guasti o dati di ingresso errati.

**prestazioni:** Uso efficiente delle risorse, come il tempo di esecuzione e la memoria centrale.

**usabilità:** Facilità d'uso.

**interoperabilità:** Capacità di integrazione con altre applicazioni.

I requisiti di sicurezza, riservatezza e robustezza sono aspetti del piú generale requisito di *affidabilità*, nel senso piú corrente di questo termine. Ricordiamo che le definizioni piú rigorose di questo termine si riferiscono alla probabilità che avvengano malfunzionamenti entro determinati periodi di tempo.

## 3.1 Classificazioni dei sistemi software

Nell'affrontare l'analisi dei requisiti, è utile individuare certe caratteristiche generali del sistema che dobbiamo sviluppare. A questo scopo possiamo considerare alcuni criteri di classificazione dei sistemi, che vedremo nel resto di questa sezione.

### 3.1.1 Requisiti temporali

Una prima importante classificazione delle applicazioni può essere fatta in base ai requisiti temporali, rispetto ai quali i sistemi si possono caratterizzare come:

**sequenziali:** senza vincoli di tempo;
**concorrenti:** con sincronizzazione fra processi;
**in tempo reale:** con tempi di risposta prefissati.

Nei sistemi sequenziali un risultato corretto (rispetto alla specifica del sistema in termini di relazioni fra ingresso e uscita) è accettabile qualunque sia il tempo impiegato per ottenerlo. Naturalmente è sempre desiderabile che l'elaborazione avvenga velocemente, ma la tempestività del risultato non è un requisito funzionale, bensí un requisito relativo alle prestazioni o all'usabilità. Inoltre, un sistema è sequenziale quando è visto come un singolo processo, le cui interazioni con l'ambiente (operazioni di ingresso e di uscita) avvengono in una sequenza prefissata.

I sistemi concorrenti, invece, sono visti come insiemi di processi autonomi che in alcuni momenti possono comunicare fra di loro. Le interazioni reciproche dei processi, e fra questi e l'ambiente, sono soggette a vincoli di sincronizzazione ed avvengono in sequenze non determinate a priori. Per *vincoli di sincronizzazione* si intendono delle relazioni di precedenza fra eventi, come, per esempio, "*l'azione* a *del processo* P *deve essere eseguita dopo l'azione* b *del processo* Q", oppure "*la valvola n. 2 non si deve aprire prima*

*che si sia chiusa la valvola n. 1*". In generale, un insieme di vincoli di sincronizzazione su un insieme di processi interagenti può essere soddisfatto da diverse sequenze di eventi, ma il verificarsi di sequenze che violano tali vincoli è un malfunzionamento.

Nei sistemi concorrenti il tempo impiegato per l'elaborazione, come nei sistemi sequenziali, non è un requisito funzionale. Come esempio molto semplice di sistema concorrente possiamo considerare il comando `cat` del sistema Unix combinato per mezzo di un *pipe* al comando `lpr`:

```
cat swe.txt | lpr
```

Il processo `cat` legge un file e ne scrive il contenuto sull'uscita standard, l'operatore *pipe* collega l'uscita di `cat` all'ingresso del processo `lpr`, che scrive sulla stampante. I due processi lavorano a velocità diverse, ma sono sincronizzati in modo che il processo piú veloce (presumibilmente `cat`) aspetti il piú lento. Il vincolo di sincronizzazione si può esprimere informalmente cosí: "*il processo* `lpr` *deve scrivere i caratteri nello stesso ordine in cui li riceve dal processo* `cat`".

Questo esempio rientra nel modello della coppia produttore/consumatore con controllo di flusso (*flow control*), in cui il produttore si blocca finché il consumatore non è pronto a ricevere nuove informazioni: in questo caso, i due processi hanno un vincolo di sincronizzazione reciproca (il consumatore elabora un dato solo dopo che il produttore lo ha prodotto, il produttore elabora un nuovo dato solo se il consumatore lo può ricevere), ma non esistono limiti prefissati per il tempo di esecuzione.

Un sistema in tempo reale è generalmente un sistema concorrente, e in piú deve fornire i risultati richiesti entro limiti di tempo prefissati: in questi sistemi un risultato, anche se corretto, è inaccettabile se non viene prodotto in tempo utile. Per esempio, una coppia produttore/consumatore *senza* controllo di flusso, in cui il produttore funziona ad un ritmo indipendente da quello del consumatore, è un sistema real-time, poiché in questo caso il produttore impone al consumatore un limite massimo sul tempo di esecuzione. Se il consumatore non rispetta questo limite si perdono delle informazioni. Un esempio di questo tipo di sistema è un sensore che manda informazioni a un elaboratore, dove la frequenza di produzione dei dati dipende solo dal sensore o dall'evoluzione del sistema fisico controllato.

### 3.1.2 Tipo di elaborazione

Un'altra classificazione dei sistemi si basa sul tipo di elaborazione compiuta prevalentemente. I sistemi si possono quindi caratterizzare come:

**orientati ai dati:** mantengono e rendono accessibili grandi quantità di informazioni (p.es., banche dati, applicazioni gestionali);

**orientati alle funzioni:** trasformano informazioni mediante elaborazioni complesse (p.es., compilatori);

**orientati al controllo:** interagiscono con l'ambiente, modificando il proprio stato in seguito agli stimoli esterni (p.es., sistemi operativi, controllo di processi).

Bisogna però ricordare che ogni applicazione usa dei dati, svolge delle elaborazioni, ed ha uno stato che si evolve, in modo piú o meno semplice. Nello specificare un sistema è quindi necessario, in genere, prendere in considerazione tutti questi tre aspetti.

### 3.1.3 Software di base o applicativo

Un'altra classificazione si può ottenere considerando se il software da realizzare serve a fornire i servizi base di elaborazione ad altro software (per esempio, se si deve realizzare un sistema operativo o una sua parte), oppure software intermedio fra il software di base e le applicazioni (librerie), oppure software applicativo vero e proprio.

## 3.2 Linguaggi di specifica

Normalmente i requisiti, sia funzionali che non funzionali, vengono espressi in linguaggio naturale. Questo, però, spesso non basta a specificare i requisiti con sufficiente precisione, chiarezza e concisione. Per questo sono stati introdotti numerosi *linguaggi di specifica* che possano supplire alle mancanze del linguaggio naturale.

### 3.2.1 Classificazione dei formalismi di specifica

I formalismi usati nella specifica dei sistemi privilegiano in grado diverso gli aspetti dei sistemi considerati piú sopra. Alcuni formalismi sono specializ-

zati per descrivere un aspetto particolare, mentre altri si propongono una maggiore generalità. Alcune metodologie di sviluppo si affidano ad un unico formalismo, mentre altre ne sfruttano piú d'uno.

I formalismi per la specifica, quindi, possono essere suddivisi analogamente ai tipi di sistemi per cui sono concepiti, per cui si avranno formalismi orientati ai dati, alle funzioni, e via dicendo. Inoltre, i formalismi di specifica vengono classificati anche secondo i due criteri del *grado di formalità* e dello *stile di rappresentazione*, che descriviamo di séguito.

### Grado di formalità

I linguaggi si possono suddividere in *formali*, *semiformali*, *informali*.

Un linguaggio è *formale* se la sua sintassi e la sua semantica sono definite in modo matematicamente rigoroso; il significato di questa frase sarà meglio definito nella parte dedicata alla logica, dove vedremo come la sintassi e la semantica di un linguaggio si possano esprimere per mezzo di concetti matematici elementari, come insiemi, funzioni e relazioni.

Una specifica espressa in un linguaggio formale è precisa e verificabile, e inoltre lo sforzo di traduzione dei concetti dal linguaggio naturale a un linguaggio formale aiuta la comprensione di tali concetti da parte degli analisti. Questo è dovuto al fatto che, dovendo riformulare in un linguaggio matematico un concetto espresso in linguaggio naturale, si è costretti ad eliminare le ambiguità e ad esplicitare tutte le ipotesi date per scontate nella forma originale.

Il maggiore limite dei linguaggi formali è il fatto che richiedono un certo sforzo di apprendimento e sono generalmente poco comprensibili per chi non ha una preparazione adeguata. Conviene però ricordare che la sintassi dei linguaggi formali può essere resa piú "amichevole", per esempio usando notazioni grafiche, oppure sostituendo simboli matematici con parole.

Per notazioni o linguaggi "semiformali" si intendono quelle che hanno una sintassi (spesso grafica) definita in modo chiaro e non ambiguo ma non definiscono una semantica per mezzo di concetti matematici, per cui il significato dei simboli usati viene espresso in modo informale. Nonostante questo limite, i linguaggi semiformali sono molto usati perché permettono di esprimere i concetti in modo piú conciso e preciso del linguaggio naturale, e sono generalmente piú facili da imparare ed usare dei linguaggi formali.

Fra i linguaggi semiformali possiamo includere il *linguaggio naturale strut-*

*turato*, cioè il linguaggio naturale usato con una sintassi semplificata e varie convenzioni che lo rendano piú chiaro e preciso.

I linguaggi informali non hanno né una sintassi né una semantica definite rigorosamente. I linguaggi naturali hanno una sintassi codificata dalle rispettive grammatiche, ma non formalizzata matematicamente, ed una semantica troppo ricca e complessa per essere formalizzata. Al di fuori del linguaggio naturale non esistono dei veri e propri linguaggi informali, ma solo delle notazioni grafiche inventate ed usate liberamente per schematizzare qualche aspetto di un sistema, la cui interpretazione viene affidata all'intúito del lettore ed a spiegazioni in linguaggio naturale. I disegni fatti alla lavagna durante le lezioni rientrano in questo tipo di notazioni.

### Stile di rappresentazione

Un'altra possibile suddivisione è fra linguaggi *descrittivi* e *operazionali*.

La differenza fra questi linguaggi può essere compresa considerando, per esempio, un semplice sistema costituito da un contenitore di gas con una valvola di scarico. Se la pressione $p$ del gas supera un certo valore di soglia $P$, la valvola si apre, e quando la pressione torna al di sotto del valore di soglia la valvola si chiude. Possiamo rappresentare questo sistema con una formula logica:

$$p < P \Leftrightarrow \textbf{valvola-chiusa}$$

in cui **valvola-chiusa** è una proposizione che è vera quando la valvola è chiusa. In questo caso, il sistema è rappresentato da una relazione fra le proprietà delle entità coinvolte, e si ha quindi una rappresentazione descrittiva (o *dichiarativa*). Lo stesso sistema può avere una rappresentazione operazionale, cioè in termini di una *macchina astratta*, che si può trovare in certo *stato* e passare ad un altro stato (effettuando cioè una *transizione*) quando avvengono certi *eventi*. La figura 3.1 mostra la macchina astratta corrispondente all'esempio, che si può trovare nello stato "aperto" o "chiuso" secondo il valore della pressione. Le espressioni $(p \geq P := T)$ e $(p < P := T)$ denotano gli eventi associati al passaggio della pressione a valori, rispettivamente, maggiori o minori della soglia $P$ (la prima espressione, per esempio, significa "$p \geq P$ diventa vero").

Quindi i linguaggi descrittivi rappresentano un sistema in termini di entità costituenti, delle loro proprietà, e delle relazioni fra entità, mentre i linguaggi operazionali lo rappresentano in termini di stati e transizioni che

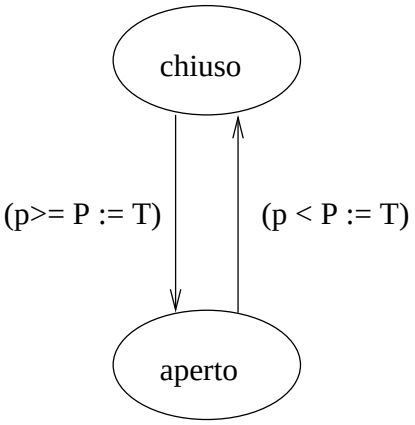


Figura 3.1: Una descrizione operazionale

ne definiscono il comportamento. Spesso una specifica completa richiede che vengano usate tutte e due le rappresentazioni.

## 3.3 Formalismi orientati ai dati

Fra formalismi orientati ai dati possiamo distinguere quelli usati per definire la *semantica* dei dati da quelli che ne definiscono la *sintassi*. I primi permettono di descrivere il significato delle informazioni associate ai dati, mentre i secondi descrivono la struttura dei dati, ovvero la forma in cui si presentano. Fra i formalismi di tipo semantico prenderemo in considerazione solo il modello Entità–Relazioni, gli altri che verranno trattati in questa sezione sono di tipo sintattico.

### 3.3.1 Modello Entità–Relazioni

Il modello Entità–Relazioni (ER) è un modello descrittivo semiformale per applicazioni orientate ai dati, di tipo semantico. Permette di descrivere la struttura concettuale dei dati, cioè le relazioni logiche fra gli oggetti del "mondo reale" rappresentati dai dati, indipendentemente sia dalle operazioni che devono essere eseguite sui dati (p.es., ricerca, ordinamento,...), che dalla loro implementazione.

Un'*entità* rappresenta un insieme di oggetti, ciascuno dei quali possiede *attributi*, che rappresentano informazioni significative dal punto di vista

dell'applicazione per caratterizzare ciascun oggetto. Alcuni attributi, detti *attributi chiave*, identificano ciascun elemento dell'insieme. Per esempio, l'insieme degli impiegati di una ditta si può rappresentare con l'entità *Impiegati*, definita dagli attributi *Matricola*, *Cognome*, *Nome*, e *Stipendio*. La *Matricola* è un attributo chiave, perché permette di identificare ciascun impiegato (anche in casi di omonimia).

Un'entità (Fig. 3.2) viene rappresentata da un rettangolo, ed i suoi attributi vengono rappresentati da ellissi, collegate al rettangolo.

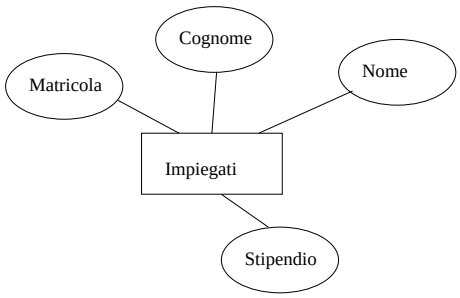


Figura 3.2: Un'entità

Gli oggetti (e quindi le entità a cui appartengono) sono legati reciprocamente da *relazioni*. Una relazione può sussistere fra elementi di una stessa entità (per esempio un impiegato *collabora* con altri impiegati), o fra elementi di due o piú entità.

Una relazione (Fig. 3.3) si rappresenta con una losanga collegata alle entità che vi partecipano. Le linee di collegamento sono decorate con frecce per rappresentare relazioni biunivoche (frecce verso due entità) o funzionali (frecce verso una sola entità). La figura 3.3 mostra tre relazioni: ogni impie-

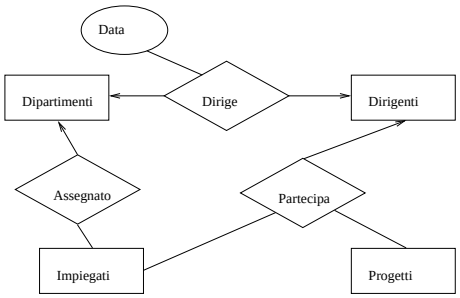


Figura 3.3: Un diagramma Entità/Relazioni

gato è assegnato a un dipartimento e partecipa ad uno o piú progetti, ogni

progetto vede la partecipazione di uno o piú impiegati e di un dirigente, e infine ogni dirigente è a capo di un dipartimento. La figura mostra anche una relazione caratterizzata da un attributo, cioè la data d'inizio dell'incarico di un dirigente alla guida di un dipartimento. La relazione fra dirigenti e dipartimenti è uno a uno.

### 3.3.2 Espressioni regolari

Un linguaggio basato sulle *espressioni regolari* permette di descrivere formalmente la sintassi dei dati elaborati da un'applicazione, e per questo le espressioni regolari verranno trattate in questa sezione dedicate ai linguaggi orientati ai dati. È bene precisare che le espressioni regolari possono specificare anche sequenze di eventi o di azioni, per cui potrebbero essere trattate anche nell'àmbito dei formalismi di tipo operazionale. Osserviamo, in particolare, che esiste una stretta relazione fra le espressioni regolari e gli automi a stati finiti (Sez. 3.5.1).

Un'espressione regolare descrive una famiglia di sequenze (o *stringhe*) di simboli appartenenti ad un alfabeto: piú precisamente, per *alfabeto* si intende un insieme finito di simboli elementari (che in pratica può coincidere con l'insieme di caratteri alfanumerici usati in un sistema di elaborazione, ma anche con altri tipi di informazione) comprendente l'elemento nullo ($\lambda$), e un'espressione regolare rappresenta un'insieme di stringhe per mezzo di operazioni applicate ad altri insiemi di stringhe.

Le espressioni regolari elementari su un alfabeto sono le stringhe formate da un unico simbolo. Per esempio, se l'alfabeto è l'insieme $\{a, b, c\}$, le espressioni regolari fondamentali sono **a**, **b** e **c**, che rappresentano rispettivamente i sottoinsiemi $\{a\}$, $\{b\}$ e $\{c\}$, ovvero le stringhe $a$, $b$ e $c$.

L'operazione di *concatenazione* di due espressioni $E_1$ ed $E_2$, rappresentata di solito con la semplice giustapposizione delle espressioni, produce l'insieme delle stringhe formate da un prefisso uguale ad una stringa appartenente a $E_1$ seguito da un suffisso uguale ad una stringa appartenente a $E_2$. Per esempio, $\mathbf{ab} = \{ab\}$, ovvero l'insieme contenente solo la stringa $ab$.

L'operazione di *scelta* fra due espressioni $E_1$ ed $E_2$, rappresentata di solito col simbolo '|', produce l'unione degli insiemi rappresentati dalle due espressioni. Per esempio, $\mathbf{a} \mid \mathbf{b} = \{a, b\}$, $\mathbf{a}(\mathbf{b} \mid \mathbf{c}) = \{ab, ac\}$.

L'operazione di *chiusura* (anche *chiusura* o *stella di Kleene*), rappresentata col simbolo '$*$', di un'espressione $E$ produce l'insieme formato dalla

stringa nulla e dalle stringhe che si possono ottenere concatenando un numero finito (ma non limitato) di stringhe appartenenti a $E$. Per esempio, $\mathbf{a}* = \lambda \mid \mathbf{a} \mid \mathbf{aa} \mid \ldots = \{\lambda, a, aa, \ldots\}$.

Un'applicazione tipica di questo formalismo è la specifica degli elementi lessicali (*token*) di un linguaggio di programmazione (costanti, identificatori, parole chiave...). Tuttavia, come già accennato, i simboli dell'alfabeto usato nelle espressioni regolari possono rappresentare altri concetti, come, per esempio, azioni o eventi. Le espressioni regolari sono quindi uno strumento di specifica molto flessibile. La natura formale delle espressioni regolari fa sí che un'espressione si possa trasformare in altre espressioni equivalenti, di verificare l'equivalenza di due espressioni, e soprattutto di verificare se una particolare stringa di simboli appartiene all'insieme definito da un'espressione.

Ricordiamo inoltre che le espressioni regolari, oltre ad essere usate come linguaggio di specifica, hanno un ruolo importante nei linguaggi di programmazione (come, per esempio, il Perl) o negli strumenti (per esempio, i programmi `sed` e `grep` nei sistemi Unix) dedicati all'elaborazione di testi.

Esistono varie notazioni per le espressioni regolari, e qui descriveremo un sottoinsieme della notazione impiegata dal programma Lex [7]. Questo programma legge un file contenente la specifica di alcune espressioni regolari, a ciascuna delle quali sono associate delle elaborazioni (*azioni*), e produce un programma in C che, leggendo un file di testo, riconosce le stringhe descritte dalle espressioni regolari e, per ogni riconoscimento, esegue le azioni associate. Anche il linguaggio in cui sono scritte le espressioni regolari si chiama Lex.

In Lex, un'espressione regolare è formata da *caratteri testuali* e da *operatori*. I caratteri testuali sono lettere, cifre, simboli aritmetici e segni d'interpunzione. Gli operatori usati nelle espressioni regolari sono i seguenti:

| | |
|---|---|
| `[ ... ]` | classi di caratteri, per esempio `[a-zA-Z0-9]` rappresenta l'insieme costituito dai caratteri alfabetici minuscoli e maiuscoli e dalle cifre decimali; |
| `.` | qualsiasi carattere escluso il carattere di nuova linea; |
| `\|` | espressioni alternative (scelta), per esempio `ab\|cd` rappresenta `ab` e `cd`; |
| `*` | zero o piú volte l'espressione precedente (chiusura di Kleene); |
| `?` | espressione opzionale, per esempio `ab?c` rappresenta `ac` e `abc`; |
| `+` | una o piú volte l'espressione precedente; |
| `{ m, n }` | ripetizione fra $m$ ed $n$ volte: `a{2}` riconosce `aa`, `a{2,4}` rappresenta `aa`, `aaa`, `aaaa`. |

La concatenazione di due espressioni si esprime semplicemente scrivendole una di seguito all'altra. Osserviamo che tutti gli operatori del Lex si possono esprimere in termini degli operatori fondamentali di concatenazione, scelta e chiusura, eventualmente introducendo la stringa nulla.

A questa notazione si aggiunge la possibilità di dare un nome a delle espressioni, e di riferirsi ad esse scrivendone il nome fra parentesi graffe.

**Esempio**

Il seguente esempio è una semplice grammatica per riconoscere i token usati in espressioni di assegnamento formate da identificatori, costanti intere ed operatori aritmetici e di assegnamento. Un identificatore è formato da uno o piú caratteri, il primo dei quali è una lettera o una sottolineatura '_' e gli altri possono essere lettere, cifre o sottolineature. Un intero è costituito da una o piú cifre.

```
[a-zA-Z_][a-zA-Z_0-9]*    return( IDENTIFIER );
[0-9]+                    return( INTEGER );
"="                       return( '=' );
"+"                       return( '+' );
"-"                       return( '-' );
"*"                       return( '*' );
"/"                       return( '/' );
```

Le azioni associate al riconoscimento di stringhe appartenenti alle diverse espressioni regolari sono, in questo caso, delle semplici istruzioni `return`. Il programma Lex, partendo da questa grammatica, produrrà una funzione di riconoscimento che, leggendo un file, eseguirà l'istruzione `return` appropriata ogni volta che riconosce un token. Questa funzione potrà essere usata da un programma piú complesso, per esempio da un compilatore.

### 3.3.3 Grammatiche non contestuali

Le *grammatiche non contestuali* (*context-free grammars*) vengono usate comunemente per specificare le sintassi dei linguaggi di programmazione, che sono troppo complesse per essere descritte dal linguaggio delle espressioni regolari. Ricordiamo che una grammatica non contestuale (o *di tipo 2*) è definita da un insieme di *simboli non terminali*, che rappresentano strutture complesse (per esempio, programmi, istruzioni, espressioni), un insieme

di *simboli terminali*, che rappresentano strutture elementari (per esempio, identificatori, costanti, parole chiave), e un insieme di *produzioni* della forma

$$A \rightarrow \alpha$$

dove $A$ è un simbolo non terminale e $\alpha$ è una sequenza di simboli, ciascuno dei quali può essere terminale o non terminale. Una produzione di questa forma dice che il simbolo $A$ può essere sostituito dalla sequenza $\alpha$[1].

Tipicamente tali grammatiche vengono descritte con una notazione basata sulla *forma normale di Backus-Naur* (*BNF*), ben nota a chiunque abbia studiato un linguaggio di programmazione. È utile tener presente che i linguaggi che possono essere specificati non si limitano ai linguaggi di programmazione, ma comprendono anche i linguaggi di comando dei sistemi operativi e delle applicazioni interattive, i formati dei file di configurazione, e in generale i formati di qualsiasi dato letto o prodotto da un'applicazione.

Per esempio, la seguente specifica in linguaggio YACC [7] descrive una sintassi per le espressioni aritmetiche:

```
espr : espr '+' espr
     | espr '-' espr
     | espr '*' espr
     | espr '/' espr
     | IDENTIFIER
     | INTEGER
     ;
```

Il programma Yacc, analogamente al programma Lex già visto, partendo da questa sintassi può produrre una funzione in linguaggio C che riconosce le espressioni aritmetiche in un file di testo.

### 3.3.4 ASN.1

La *Abstract Syntax Notation 1* è una notazione destinata a descrivere i dati scambiati su reti di comunicazioni. La ASN.1 è definita dallo standard ISO/IEC 8824-1:2002, derivato dallo standard CCITT X.409 (1984)[2].

[1] Osserviamo che la sostituzione può essere applicata qualunque sia il *contesto* di $A$, cioè i simboli che precedono e seguono $A$; questo spiega il termine "non contestuali".

[2] Si veda il sito `asn1.elibel.tm.fr`.

Per mezzo della ASN.1 si descrive una *sintassi astratta*, cioè generica e indipendente dalle diverse implementazioni delle applicazioni che si scambiano dati specificati in ASN.1. Questo permette la comunicazione fra applicazioni eseguite su architetture diverse. Una sintassi astratta è costituita da un insieme di definizioni di tipi e definizioni di valori. Queste ultime sono definizioni di costanti. I tipi sono *semplici* e *strutturati*.

### Tipi semplici

I tipi semplici sono `INTEGER`, `REAL`, `BOOLEAN`, `CHARACTER STRING`, `BIT STRING`, `OCTET STRING`, `NULL`, e `OBJECT IDENTIFIER`. Per alcuni di questi tipi si possono definire dei sottotipi esprimendo delle restrizioni di vario tipo, come nei seguenti esempi:

```
Valore ::= INTEGER
Tombola ::= INTEGER (1..90)
Positive ::= INTEGER (0< ..MAX)
Negative ::= INTEGER (MIN .. <0)
Zero ::= INTEGER (0)
NonZero :: INTEGER (INCLUDES Negative | INCLUDES Positive)
```

Il tipo `Valore` coincide con `INTEGER`, i tipi `Tombola`, `Positive` e `Negative` sono sottoinsiemi di `INTEGER`. Il tipo `Zero` viene definito per enumerazione (contiene solo l'elemento `0`), ed il tipo `NonZero` viene definito come unione di altri tipi.

Di seguito diamo alcuni esempi di definizioni di tipi o valori che usano tipi semplici.

I valori del tipo `REAL` sono terne formate da mantissa, base, esponente:

```
avogadro REAL ::= {60221367, 10, 16}
```

Quest'esempio è una definizione di un valore, che specifica il tipo e il valore della costante `avogadro`.

I `BIT STRING` rappresentano stringhe di bit:

```
BinMask BIT STRING ::= '01110011'B
HexMask BIT STRING ::= '3B'H
```

I valori `BinMask` e `HexMask` vengono espressi in notazione rispettivamente binaria ed esadecimale.

Gli `OCTET STRING` rappresentano stringhe di byte, di cui si può indicare la lunghezza:

```
Message ::= OCTET STRING (SIZE(0..255))
```

Ci sono alcui tipi predefiniti di stringhe di caratteri, come `NumericString` (cifre decimali e spazio), `PrintableString` (lettere, cifre e caratteri speciali), `GraphicString` (simboli), etc.:

```
myPhoneNumber NumericString ::= 883455
grazieGreco GraphicString ::= ευχαριστω
```

Il tipo `OBJECT IDENTIFIER` serve ad assegnare nomi unici alle varie entità definite da specifiche in ASN.1, basandosi su uno schema gerarchico di classificazione chiamato *Object Identifier Tree*:

```
AttributeType ::= OBJECT IDENTIFIER
telephoneNumber AttributeType ::= {2 5 4 20}
```

In questo esempio, `telephoneNumber` è un identificatore di oggetti il cui valore è `{2 5 4 20}`, dove 2 rappresenta il dominio degli identificatori definiti congiuntamente dall'ISO e dal CCITT, 5 rappresenta il dominio degli indirizzari (*directories*), 4 rappresenta il dominio degli attributi (di una voce in un indirizzario), e 20 rappresenta i numeri telefonici. Se un'organizzazione volesse mantenere un database contenente i nomi e i numeri di telefono dei propri dipendenti, potrebbe pubblicare lo schema logico del database etichettando il tipo `telephoneNumber` con l'identificatore (*OID*) `{2 5 4 20}`. In questo modo chiunque sviluppi del software che debba interagire col database ha un riferimento univoco alla definizione del tipo `telephoneNumber`: l'identificatore permette quindi di definire un significato standard agli elementi di un sistema informativo.

### Tipi strutturati

I tipi strutturati sono `ENUMERATED`, `SEQUENCE`, `SET`, e `CHOICE`.

I tipi `ENUMERATED` sono simili ai tipi enumerati del Pascal o del C++. È possibile assegnare un valore numerico alle costanti dell'enumerazione:

```
DaysOfTheWeek ::= ENUMERATED
{
    sunday(0), monday(1), ... saturday(6)
}
```

I tipi `SEQUENCE` sono simili ai record del Pascal o alle `struct` del C++. L'ordine dei campi nella sequenza è significativo.

```
WeatherReport ::= SEQUENCE
{
    stationNumber   INTEGER (1..99999),
    timeOfReport    UTCTime,
    pressure        INTEGER (850..1100),
    temperature     INTEGER (-100..60),
}
```

Il tipo `UTCTime` che appare in questo esempio è un tipo predefinito (uno *useful type*), i cui valori possono avere la forma YYMMDDHHMMSSZ oppure YYMMDDHHMMSS±HHMM, dove i secondi sono opzionali. Nella prima forma, 'Z' è, letteralmente, la lettera zeta.

Nei tipi `SET`, i valori dei campi si possono presentare in qualsiasi ordine:

```
TypeA ::= SET
{
    p BOOLEAN,
    q INTEGER,
    r BIT STRING
}

valA TypeA ::=
{
    p TRUE
    r '83F'H
    q -7
}

TypeB ::= SET
{
    r [0] INTEGER,
    s [1] INTEGER,
    t [2] INTEGER
}
```

I numeri fra parentesi quadre si chiamano *tag* e servono a distinguere componenti dello stesso tipo.

I tipi `CHOICE` sono simili alle `union` del C++:

```
typeC ::= CHOICE
{
    x [0] REAL,
    y [1] INTEGER,
    z [2] NumericString
}
```

### Moduli

Le definizioni di una specifica in ASN.1 possono essere raggruppate in moduli. Ogni modulo è identificato da un `OBJECT IDENTIFIER` e dichiara quali tipi e valori vengono esportati o importati.

```
WeatherReporting {2 6 6 247 7} DEFINITIONS ::=
BEGIN
    EXPORTS WeatherReport;
    IMPORTS IdentifyingString FROM StationsModule { ... };
    WeatherReport ::= SEQUENCE
    {
        idString IdentifyingString,
        stationNumber (0..99999),
        ...
    }
END
```

In questo esempio si suppone che `StationsModule` sia un modulo definito altrove, identificato dal nome e dall'`OBJECT IDENTIFIER`, qui non indicato.

### Sintassi di trasferimento

La *sintassi di trasferimento* specifica il modo in cui i dati descritti dalla sintassi astratta vengono codificati per essere trasmessi. Una sintassi di trasferimento standardizzata (ISO 8825:1987) è costituita dalle *Basic Encoding Rules* (*BER*).

## 3.4 Formalismi orientati alle funzioni

Le notazioni orientate alle funzioni servono a descrivere l'aspetto funzionale dei sistemi, cioè le elaborazioni che vengono compiute sui dati. In queste notazioni il sistema viene descritto in termini di blocchi funzionali collegati

da flussi di dati da elaborare, e questi blocchi a loro volta vengono scomposti in blocchi piú semplici. Osserviamo che questo metodo, oltre che per la specifica dei requisiti, si presta alla specifica del progetto, cioè a descrivere come viene realizzato il sistema. In questo caso, i blocchi funzionali vengono identificati con i sottosistemi che implementano le funzioni specificate.

In questi formalismi la rappresentazione grafica si basa sui *diagrammi di flusso dei dati* (DFD), nei quali il sistema da specificare viene visto come una rete di trasformazioni applicate ai dati che vi fluiscono. I DFD permettono il *raffinamento* della specifica mediante *scomposizione*: ogni nodo (*funzione*) può essere scomposto in una sottorete. Le metodologie basate sulla scomposizione di funzioni di trasformazione dei dati sono indicate col termine *Structured Analysis/Structured Design* (SA/SD). La scomposizione di ogni funzione deve rispettare il vincolo della *continuità del flusso informativo*, cioè i flussi attraversanti la frontiera della sottorete ottenuta dall'espansione di un nodo devono essere gli stessi che interessano il nodo originale.

Ogni funzione può essere scomposta in funzioni componenti. Quando non si ritiene utile scomporre ulteriormente una funzione, questa può essere specificata in linguaggio naturale, in pseudocodice[3] o in linguaggio naturale strutturato, o in un linguaggio algoritmico simile ai linguaggi di programmazione, ad alto livello e tipato.

Gli elementi della notazione sono:

**agenti esterni:** rappresentati da rettangoli, sono i produttori e i consumatori dei flussi di dati all'ingresso e all'uscita del sistema complessivo;

**funzioni:** (o *processi*) rappresentate da cerchi, sono le trasformazioni operate sui dati;

**flussi:** rappresentati da frecce, sono i dati scambiati fra le funzioni, nel verso indicato;

**depositi:** rappresentati da doppie linee, sono memorie permanenti.

Ogni elemento viene etichettato con un nome.

La Fig. 3.4 mostra il primo livello della specifica di un sistema di monitoraggio dei malati in un ospedale. Una prima espansione viene mostrata in Fig. 3.5, e in Fig. 3.6 mostriamo l'espansione della bolla `monitoraggio centrale` di Fig. 3.5.

[3]Un linguaggio simile a un linguaggio di programmazione, spesso non specificato formalmente.

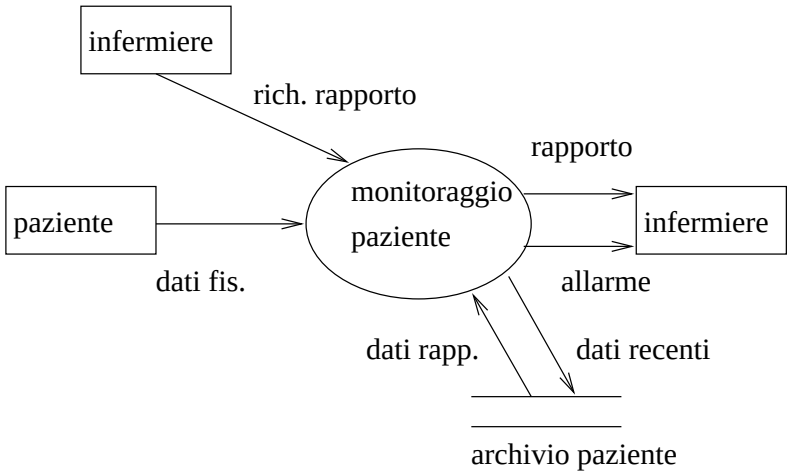


Figura 3.4: Primo livello

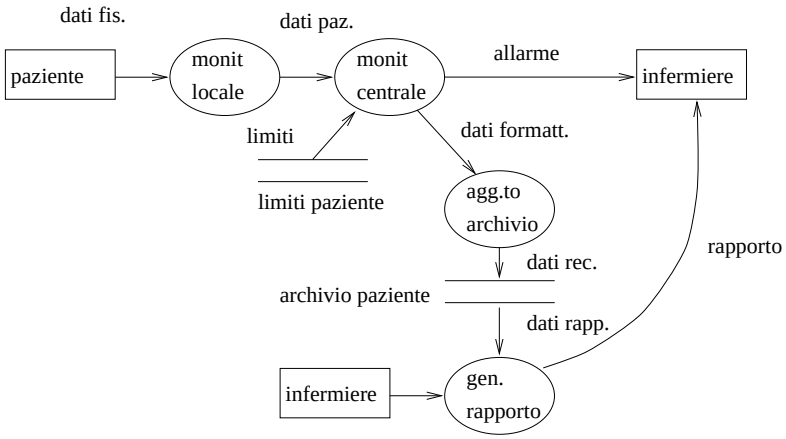


Figura 3.5: Secondo livello

## 3.5 Formalismi orientati al controllo

Questi formalismi descrivono gli aspetti dei sistemi relativi alla loro evoluzione temporale, alle possibili sequenze di eventi o di azioni, alla sincronizzazione fra le attività dei loro sottosistemi, o fra il sistema e l'ambiente in cui opera. Questi aspetti sono particolarmente importanti nei sistemi *reattivi*, che devono reagire a stimoli provenienti dall'ambiente, che si presentano in un ordine generalmente non prevedibile.

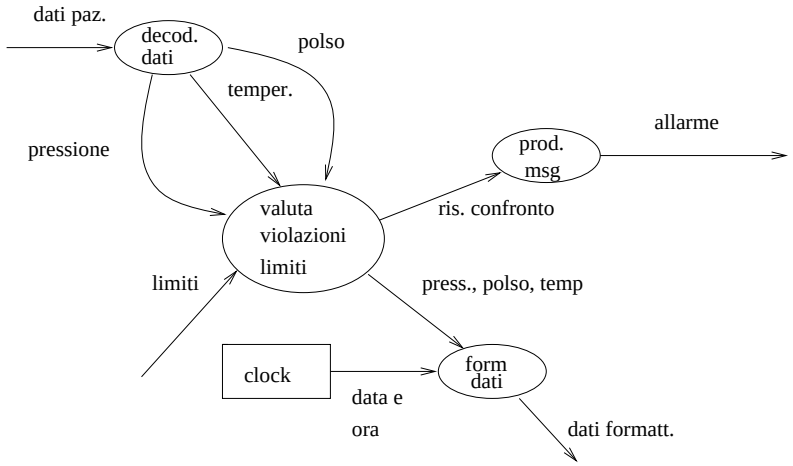


Figura 3.6: Terzo livello (parziale)

### 3.5.1 Automi a stati finiti

Col formalismo degli *automi a stati finiti* (ASF) o *macchine a stati*[4] descriviamo un sistema attraverso gli *stati* in cui si può trovare, le *transizioni*, cioè i passaggi da uno stato all'altro, gli *ingressi* (stimoli esterni) che causano le transizioni, le *uscite* del sistema, che possono essere associate alle transizioni (*macchine di Mealy*) o agli stati (*macchine di Moore*), e lo *stato iniziale* da cui parte l'evoluzione del sistema.

L'ambiente esterno agisce sul sistema rappresentato dall'automa generando una successione di ingressi discreti, e il sistema risponde a ciascun ingresso cambiando il proprio stato (eventualmente restando nello stato corrente) e generando un'uscita. L'automa definisce le regole secondo cui il sistema risponde agli stimoli esterni.

Un ASF è quindi definito da una sestupla

$$< S, I, U, d, t, s_0 >$$

con

- $S$ insieme degli stati
- $I$ insieme degli ingressi
- $U$ insieme delle uscite

[4]Precisiamo che sono stati descritti diversi tipi di macchine a stati; nel seguito faremo riferimento a uno dei tipi usati piú comunemente, i *trasduttori deterministici*.

- $d : S \times I \rightarrow S$ funzione di transizione
- $t : S \times I \rightarrow U$ funzione di uscita (macchine di Mealy)
- $s_0 \in S$ stato iniziale

Nelle macchine di Moore si ha $t : S \rightarrow U$.

Un ASF viene rappresentato con un grafo orientato, i cui nodi (cerchi o rettangoli ovalizzati) sono gli stati, e gli archi orientati, etichettati dagli ingressi e dalle uscite, descrivono la funzione di transizione e la funzione di uscita. Lo stato iniziale viene indicato da un'arco senza stato di origine.

L'automa rappresentato in Fig. 3.7 descrive l'interazione fra un utente ed un centralino che accetta chiamate interne a numeri di due cifre e chiamate esterne a numeri di tre cifre, precedute dallo zero (esempio da [4]). Nel diagramma, le transizioni sono etichettate con espressioni della forma '*ingresso*/*uscita*'; alcune transizioni non producono uscite. Un ingresso della forma '*m*:*n*' rappresenta una cifra fra *m* e *n*. Ogni stato è identificato sia da un nome che da un numero. Al numero si farà riferimento nella rappresentazione tabulare (v. oltre) dell'automa.

Un automa può essere rappresentato anche per mezzo di tabelle. Una rappresentazione possibile si basa su una matrice quadrata di ordine $n$ (numero degli stati): se esiste una transizione dallo stato $s_i$ allo stato $s_j$, l'elemento della matrice sulla riga $i$ e la colonna $j$ contiene l'ingresso che causa la transizione e la corrispondente uscita. La rappresentazione tabulare dell'ASF di Fig. 3.7 è data dalla tabella 3.5.1.

### Componibilità e scalabilità

Nella specifica di sistemi complessi, e in particolare di sistemi concorrenti, è spesso necessario o conveniente rappresentare il sistema complessivo come un aggregato di sottosistemi. Un formalismo di specifica ha la proprietà della *componibilità* se permette di costruire la rappresentazione del sistema complessivo per mezzo di semplici operazioni di composizione sulle rappresentazioni dei sottosistemi. Per *scalabilità* si intende la capacità di rappresentare un sistema in modo tale che la complessità di tale rappresentazione sia dello stesso ordine di grandezza della somma delle complessità dei singoli sottosistemi.

Nel caso degli ASF la componibilità e la scalabilità sono limitate, come mostrerà l'esempio sviluppato nei paragrafi successivi.

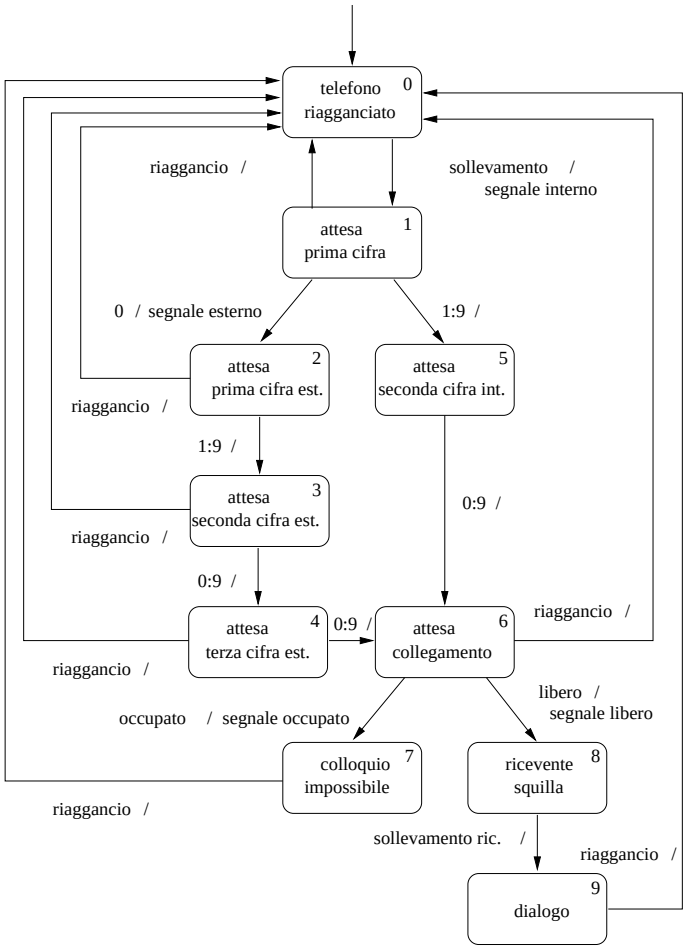


Figura 3.7: Un automa a stati finiti

La figura 3.8 mostra tre sottosistemi: un produttore, un magazzino ed un consumatore (esempio da [4]). Il produttore si trova inizialmente nello stato ($p_1$) in cui è pronto a produrre qualcosa (per esempio, un messaggio o una richiesta di elaborazione), e con la produzione passa allo stato ($p_2$) in cui attende di depositare il prodotto; col deposito ritorna allo stato iniziale. Il consumatore ha uno stato iniziale ($c_1$) in cui è pronto a prelevare qualcosa, e col prelievo passa allo stato ($c_2$) in cui attende che l'oggetto venga consumato. Il magazzino ha tre stati, in cui, rispettivamente, è vuoto ($m_1$), contiene un oggetto ($m_2$), e contiene due oggetti ($m_3$); il magazzino passa da uno stato all'altro in seguito alle operazioni di deposito e di prelievo.

Il sistema complessivo, mostrato in figura 3.9, ha dodici stati, invece dei sette usati per descrivere i sottosistemi separatamente: ciascuno dei dodici

| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | | S/int. | | | | | | | | |
| 1 | R/ | | 0/est. | | | 1:9/ | | | | |
| 2 | R/ | | | 1:9/ | | | | | | |
| 3 | R/ | | | | 0:9/ | | | | | |
| 4 | R/ | | | | | | 0:9/ | | | |
| 5 | R/ | | | | | | 0:9/ | | | |
| 6 | R/ | | | | | | | O/occ. | L/lib. | |
| 7 | R/ | | | | | | | | | |
| 8 | | | | | | | | | | Sr/ |
| 9 | R/ | | | | | | | | | |

Tabella 3.1: Rappresentazione tabulare

**R**: riaggancio; **S**: sollevamento del chiamante; **O**: ricevente occupato; **L**: ricevente libero; **Sr**: sollevamento del ricevente

stati è una possibile combinazione degli stati dei sottosistemi. In generale, componendo un numero $n$ di sottosistemi in un sistema complessivo, si ha che:

1. l'insieme degli stati del sistema complessivo è il prodotto cartesiano degli insiemi degli stati dei sottosistemi;
2. cioè, ogni stato del sistema complessivo è una $n$-upla formata da stati dei sottosistemi, per cui viene nascosta la struttura gerarchica del sistema (gli stati dei sottosistemi vengono "concentrati" nello stato globale);
3. l'evoluzione del sistema viene descritta come se ad ogni passo uno solo dei sottosistemi potesse compiere una transizione, mentre in generale è possibile che transizioni in sottosistemi distinti possano avvenire in modo concorrente;
4. il numero degli stati del sistema totale cresce esponenzialmente col numero dei sottosistemi.

Mentre i punti 1 e 4 dimostrano la poca scalabilità degli ASF, il punto 3 ne mette in evidenza un'altra caratteristica: non esprimono la concorrenza dei sottosistemi, per cui si prestano solo alla specifica di sistemi sequenziali.

Il problema dell'aumento della complessità quando si compongono sottosistemi dipende dal fatto che, nel modello di ASF qui presentato, lo stato del sistema è "globale", in quanto in un dato istante l'intero sistema viene

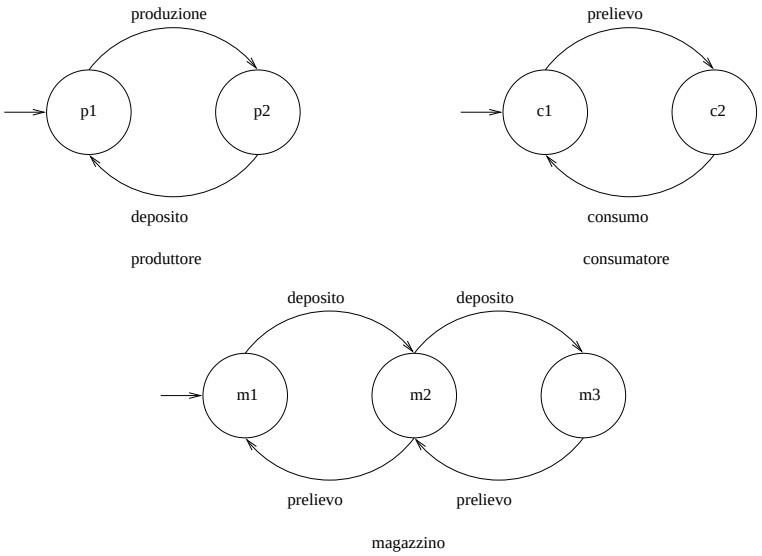


Figura 3.8: Esempio (1)

modellato da un unico stato, e "atomico", in quanto lo stato non contiene altra informazione che la propria identità e le funzioni di transizione e di uscita.

Un modo per rendere meno complesse le specifiche di sistemi per mezzo di ASF consiste nell'estendere il concetto di "stato" associandovi delle strutture dati. Nel nostro esempio, i tre stati dell'ASF che rappresenta il magazzino potrebbero essere sostituiti da un unico stato a cui è associata una variabile il cui valore è il numero di elementi immagazzinati. Le uscite associate alle operazioni di prelievo e di deposito sarebbero azioni di decremento e, rispettivamente, incremento della variabile (Fig. 3.10). Il prelievo e il deposito, a loro volta, sarebbero condizionati dal valore della variabile, cosa che comporta l'estensione del concetto di "transizione", a cui si devono aggiungere delle condizioni (*guardie*) che devono essere soddisfatte affinché una transizione possa aver luogo.

Un'altra estensione degli ASF consiste nell'introduzione di stati composti, cioè descritti, a loro volta, da macchine a stati. In questo modo un sistema complesso può essere descritto ad alto livello da un automa con pochi stati, ciascuno di quali può essere decomposto in sottostati quando serve una specifica piú dettagliata. Questo metodo è alla base degli *Statecharts*, un formalismo che esamineremo nel capitolo relativo ai metodi orientati agli oggetti.

Infine, nelle *reti di Petri* lo stato del sistema viene modellato in modo

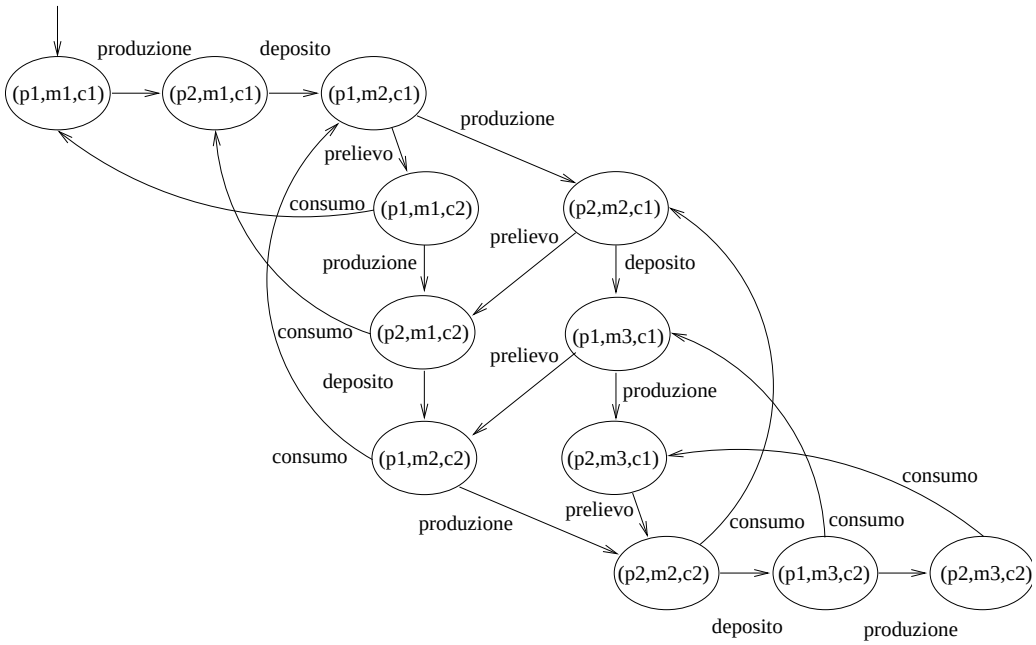


Figura 3.9: Esempio (2)

diverso, tale che la specifica renda esplicito e visibile il fatto che il sistema totale è composto da sottosistemi. Questo permette di descrivere sistemi concorrenti.

### 3.5.2 Reti di Petri

Le reti di Petri sono un formalismo operazionale per la descrizione di sistemi, le cui caratteristiche di astrattezza e generalità ne permettono l'applicazione a numerosi problemi, quali, per esempio, la valutazione delle prestazioni, i protocolli di comunicazione, i sistemi di controllo, le architetture multiprocessor e data flow, e i sistemi fault-tolerant.

Il modello delle reti di Petri è basato sui concetti di *condizioni* e *azioni*, piuttosto che di stati e transizioni. Una rete di Petri descrive un sistema specificando l'insieme delle azioni che possono essere compiute dal sistema o dall'ambiente in cui si trova, le condizioni che rendono possibile ciascuna azione, e le condizioni che diventano vere in seguito all'esecuzione di ciascuna azione. Lo stato del sistema è definito dall'insieme delle condizioni che in un dato istante sono vere, e le transizioni da uno stato all'altro sono causate

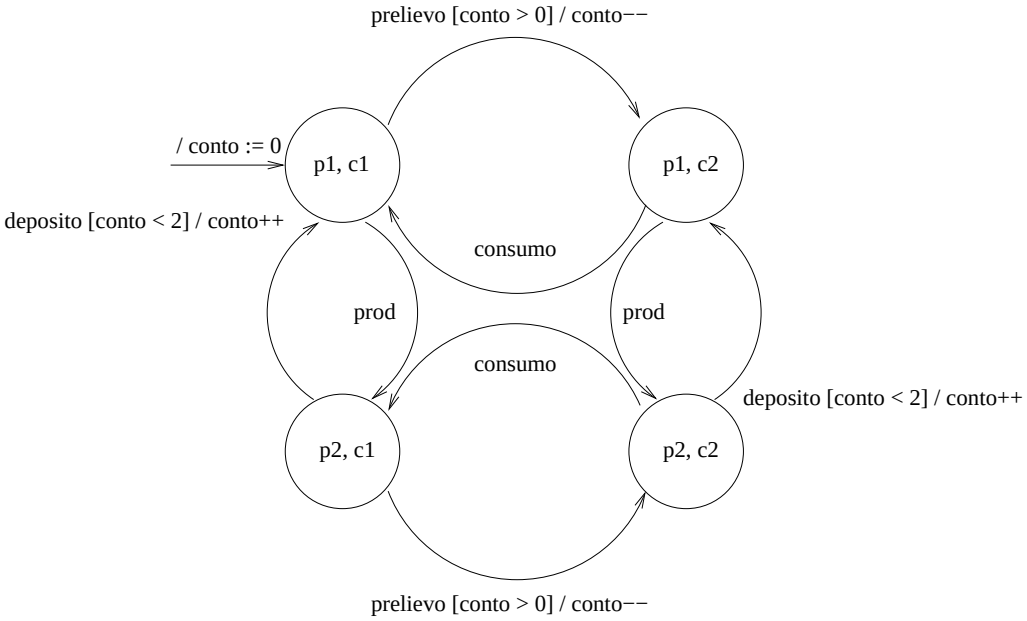


Figura 3.10: Esempio (3)

dall'esecuzione delle azioni (che anche nel vocabolario delle reti di Petri si chiamano *transizioni*).

Una rete di Petri è un grafo orientato bipartito, i cui nodi sono divisi in *posti* (*places*) e *transizioni*, e i cui archi uniscono posti a transizioni o transizioni a posti. Se un arco va da un place a una transizione il place è detto di *ingresso* alla transizione, altrimenti è di *uscita*. Ad ogni place $p$ è associato un numero intero $M(p)$, e si dice che $p$ è *marcato* con $M(p)$ token, o che *contiene* $M(p)$ token. La funzione $M$ che assegna un numero di token a ciascun place è la *marcatura* della rete; questa funzione serve a modellare le condizioni che rendono possibile lo *scatto* (cioè l'esecuzione) delle transizioni. Per esempio, un certo place potrebbe modellare la condizione "*ci sono dei messaggi da elaborare*" e la sua marcatura potrebbe modellare il numero di messaggi in attesa a un certo istante. Graficamente i place vengono rappresentati da cerchi, le transizioni da segmenti o rettangoli, gli archi da frecce, e i token contenuti in un place $p$ da $M(p)$ "pallini" dentro al place.

Piú precisamente, una rete è una terna:

$$N = \langle P, T; F \rangle$$

dove

- gli insiemi $P$ (posti) e $T$ (transizioni) sono disgiunti:

$$P \cap T = \emptyset$$

- la rete non è vuota:

$$P \cup T \neq \emptyset$$

- $F$ è una relazione che mappa posti in transizioni e transizioni in posti:

$$F \subseteq (P \times T) \cup (T \times P)$$

La *relazione di flusso* $F$ può essere espressa per mezzo di due funzioni pre() e post() che associano, rispettivamente, gli elementi di ingresso e gli elementi di uscita a ciascun elemento della rete, sia esso un posto o una transizione. Possiamo considerare separatamente le restrizioni delle due funzioni ai due insiemi di nodi, usando per semplicità lo stesso nome della rispettiva funzione complessiva:

- Pre-insieme di un posto[5]:

$$\text{pre} : P \to 2^T$$

$$\text{pre}(p) = {}^\bullet p = \{t \in T \mid \langle t, p \rangle \in F\}$$

- Pre-insieme di una transizione:

$$\text{pre} : T \to 2^P$$

$$\text{pre}(t) = {}^\bullet t = \{p \in P \mid \langle p, t \rangle \in F\}$$

- Post-insieme di un posto:

$$\text{post} : P \to 2^T$$

$$\text{post}(p) = p^\bullet = \{t \in T \mid \langle p, t \rangle \in F\}$$

- Post-insieme di una transizione:

$$\text{post} : P \to 2^P$$

$$\text{post}(y) = y^\bullet = \{z \in P \mid \langle y, z \rangle \in F\}$$

[5]la notazione $2^T$ rappresenta l'*insieme potenza* di $T$, cioè l'insieme dei suoi sottoinsiemi.

La rete cosí definita è semplicemente la descrizione di un grafo, data elencandone i nodi e gli archi col relativo orientamento (una coppia $\langle x, y\rangle$ è un arco orientato da $x$ a $y$), quindi la terna $\langle P, T; F\rangle$ rappresenta solo la struttura della rete. Per rappresentare lo stato iniziale del sistema modellato dalla rete si introduce la funzione *marcatura iniziale*, oltre a una *funzione peso* il cui significato verrà chiarito piú oltre.

Una *Rete Posti/Transizioni* è quindi una quintupla:

$$N_{P/T} = \langle P, T; F, W, M_0\rangle$$

dove

- la funzione $W$ è il *peso* (o *molteplicità*) degli archi:

$$W : F \to \mathbb{N} - \{0\}$$

- la funzione $M_0$ è la *marcatura iniziale*:

$$M_0 : P \to \mathbb{N}$$

### Abilitazione e regola di scatto

La marcatura iniziale descrive lo stato iniziale del sistema: in questo stato, alcune azioni saranno possibili e altre no. Delle azioni possibili, alcune (o tutte) verranno eseguite, in un ordine non specificato (e non specificabile in questo formalismo). L'esecuzione di queste azioni modificherà lo stato del sistema, che verrà descritto da una nuova funzione marcatura, ovvero cambierà il valore associato a ciascun place (valore chiamato convenzionalmente "numero di token").

Come già detto, le azioni sono modellate dalle transizioni. Quando l'azione corrispondente ad una transizione è possibile nello stato attuale, si dice che la transizione è *abilitata* nella marcatura attuale; quando l'azione viene eseguita si dice che la transizione *scatta*.

Una transizione $t$ è *abilitata nella marcatura* $M$ se e solo se:

$$\forall_{p \in {}^\bullet t}\ M(p) \geq W(\langle p, t\rangle)$$

cioè, per ogni posto di ingresso alla transizione, si ha che la sua marcatura è maggiore o uguale al peso dell'arco che lo unisce alla transizione. Nel caso che ${}^\bullet t$ sia vuoto, la transizione è sempre abilitata.

La funzione peso è un mezzo per esprimere semplicemente certe condizioni. Supponiamo, per esempio, che un sistema operativo, per ottimizzare l'accesso al disco, raggruppi le richieste di lettura in blocchi di $N$ operazioni: questo si potrebbe modellare con un posto la cui marcatura rappresenta il numero di richieste da servire, una transizione che rappresenta l'esecuzione delle richieste, e un arco il cui peso, uguale a $N$, fa sí che l'esecuzione avvenga solo quando c'è un numero sufficiente di richieste.

Graficamente, il peso di un arco si índica scrivendone il valore accanto all'arco. Se il peso non è indicato, è uguale a 1.

L'abilitazione di una transizione $t$ in una marcatura $M$ si índica con la formula

$$M\,[t\rangle$$

che si legge "*$t$ è abilitata in $M$*".

Dopo lo scatto di una transizione la rete assume una nuova marcatura $M'$ il cui valore è determinato dalla seguente *regola di scatto*, ove $M$ è la marcatura precedente allo scatto:

$$\begin{array}{ll}
\forall_{p\in({}^\bullet t - t^\bullet)} & M'(p) = M(p) - W(\langle p,t\rangle) \\
\forall_{p\in(t^\bullet - {}^\bullet t)} & M'(p) = M(p) + W(\langle t,p\rangle) \\
\forall_{p\in({}^\bullet t \cap t^\bullet)} & M'(p) = M(p) - W(\langle p,t\rangle) + W(\langle t,p\rangle) \\
\forall_{p\in P-({}^\bullet t \cup t^\bullet)} & M'(p) = M(p)
\end{array}$$

La regola di scatto si può cosí riassumere:

1. la marcatura di ciascun posto di ingresso viene diminuita del peso dell'arco dal posto alla transizione;
2. la marcatura di ciascun posto di uscita viene aumentata del peso dell'arco dalla transizione al posto;
3. la marcatura di ciascun posto di ingresso e di uscita varia di una quantità pari alla differenza fra i pesi dell'arco dalla transizione al posto e dell'arco dal posto alla transizione;
4. la marcatura dei posti non collegati direttamente alla transizione (quindi né d'ingresso né d'uscita) resta invariata.

Se, in una data marcatura, piú transizioni sono abilitate, ne scatta una sola, scelta (da un ipotetico esecutore della rete) in modo non deterministico.

La notazione

$$M\,[t\rangle\,M'$$

índica che la marcatura $M'$ è il risultato dello scatto della transizione $t$ abilitata in $M$.

Osserviamo che lo scatto (*firing*) di una transizione ha solo effetti locali, cambiando solo la marcatura dei posti di ingresso e di uscita della transizione stessa.

Due transizioni $t_1$ e $t_2$ possono scattare in successione se

$$M\,[t_1\rangle\,M' \wedge M'\,[t_2\rangle\,M''$$

e si dice allora che la *sequenza di scatti* $t_1t_2$ è abilitata in $M$, e si scrive

$$M\,[t_1t_2\rangle\,M''$$

In generale, rappresentiamo una sequenza di scatti di lunghezza $n$ con la seguente notazione:

$$S^{(n)} = t_1 \ldots t_n$$

e scriviamo

$$M\,[t_1 \ldots t_n\rangle\,M^{(n)}$$

ovvero

$$M\,\left[S^{(n)}\right\rangle\,M^{(n)}$$

**Esempio: produttore/consumatore**

La figura 3.11 mostra un sistema produttore/consumatore specificato per mezzo di una rete di Petri (questo esempio ed i successivi sono tratti da [4]). Il produttore ed il consumatore sono collegati da un magazzino intermedio di capacità 2. Osserviamo che, a differenza dell'esempio visto per gli automi a stati finiti, qui è possibile attribuire ciascun place ad uno dei tre sottosistemi: in questo senso la rappresentazione dello stato globale è distribuita fra i sottosistemi.

La marcatura iniziale, indicata in figura, specifica che il produttore è pronto a produrre un elemento (posto $p_1$), il consumatore è pronto a prelevare (posto $c_1$) un elemento dal magazzino (purché ci sia un elemento da prelevare), e il magazzino ha due posti liberi (posto *libero*).

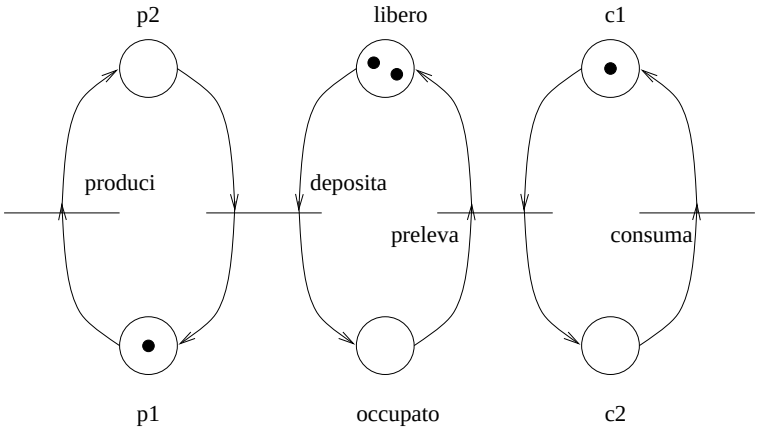


Figura 3.11: pn1

### Situazioni fondamentali

Consideriamo alcune relazioni fra transizioni, determinate sia dalla struttura della rete che dalla marcatura.

**sequenza:** due transizioni $t$ ed $u$ si dicono *in sequenza* in una marcatura $M$ se e solo se

$$M\,[t\rangle \wedge \neg(M\,[u\rangle) \wedge M\,[tu\rangle$$

cioè $u$ diventa abilitata solo dopo che è scattata $t$.

**conflitto:** due transizioni $t$ ed $u$ si dicono *in conflitto effettivo* in una marcatura $M$ se e solo se

$$M\,[t\rangle \wedge M\,[u\rangle \wedge \exists_{p \in {}^\bullet t \cap {}^\bullet u}(M(p) < W(\langle p, t\rangle) + W(\langle p, u\rangle))$$

Questa condizione significa che tutte e due le transizioni sono abilitate, ma lo scatto di una disabilita l'altra, togliendo dei token dai place di ingresso comuni alle due transizioni.

La condizione ${}^\bullet t \cap {}^\bullet u \neq \emptyset$ si dice *conflitto strutturale*, e corrisponde all'esistenza di place di ingresso in comune fra le due transizioni.

**concorrenza:** due transizioni $t$ ed $u$ si dicono *in concorrenza effettiva* in una marcatura $M$ se e solo se

$$M\,[t\rangle \wedge M\,[u\rangle \wedge \forall_{p \in {}^\bullet t \cap {}^\bullet u}(M(p) \geq W(\langle p, t\rangle) + W(\langle p, u\rangle))$$

Questa condizione significa che tutte e due le transizioni sono abilitate e possono scattare tutte e due in qualsíasi ordine, poiché non si influenzano a vicenda.

Un caso particolare di concorrenza effettiva è la *concorrenza strutturale*, data dalla condizione ${}^\bullet t \cap {}^\bullet u = \emptyset$; in questo caso le due transizioni non hanno place di ingresso in comune.

### Esempio

In questo esempio modelliamo un programma concorrente costituito da tre processi. Rappresentiamo il programma in linguaggio Ada. In Ada, un processo è costituito da un sottoprogramma di tipo *task*. Un task può contenere dei sottoprogrammi detti *entries* che possono essere invocati da altri task con delle istruzioni di *entry call*. Quando un task (cliente) invoca un'entry di un altro task (server), i due task si sincronizzano secondo il meccanismo del *rendezvous*: il corpo di una entry viene eseguito quando il server nella sua esecuzione è arrivato all'inizio dell'entry (istruzione `accept`) e un cliente è arrivato alla chiamata dell'entry. Il primo dei task che arriva al punto di sincronizzazione (rispettivamente, `accept` o entry call) aspetta che arrivi anche l'altro. Durante l'esecuzione del corpo dell'entry, il task cliente viene sospeso, e se piú clienti invocano una stessa entry, questi vengono messi in coda. Il server, con un'istruzione `select`, può accettare chiamate per piú entries: in questo caso viene eseguita la prima entry che è stata chiamata da qualche cliente.

Un'istruzione di entry call è formata dal nome del task contenente la entry, seguito da un punto, dal nome dell'entry e da eventuali parametri fra parentesi. Nel nostro esempio, i nomi dei task sono `A`, `B` e `C`, le entries sono `E1` ed `E2`, ed i corpi delle entries sono `R1` ed `R2`.

La Fig. 3.12 mostra il testo del programma, corrispondente alla rete di Fig. 3.13, in cui le transizioni rappresentano l'esecuzione delle istruzioni, ed i place rappresentano lo stato del programma fra una transizione e l'altra.

Osservando la rete, possiamo mettere in evidenza la concorrenza fra le transizioni `SA1`, `SB1` ed `SC1`, il conflitto fra `accept E1` e `accept E2`, la sequenza fra `R1` e `end E1`, e fra `R2` e `end E2`.

### Raggiungibilità delle marcature

L'analisi della *raggiungibilità* delle marcature di una rete permette di stabilire se il sistema modellato dalla rete può raggiungere determinati stati o no, e, se può raggiungerli, a quali condizioni. Permette quindi di verificare se il sistema soddisfa proprietà di *vitalità* e di *sicurezza*. La vitalità ci dice che il

```
task body A is        |  task body B is         |      task body C is
    SA1;              |      SB1;               |          SC1;
    B.E1;             |      select             |          B.E2;
    SA2;              |          accept E1 do   |          SC2;
end A;                |              R1;        |      end C;
                      |          end E1;        |
                      |      or                 |
                      |          accept E2 do   |
                      |              R2;        |
                      |          end E2;        |
                      |      end select;        |
                      |      SB2;               |
                      |  end B;                 |
```

Figura 3.12: Un programma concorrente (1).

sistema è in grado di raggiungere le configurazioni desiderate: per esempio, nel sistema di controllo di un ascensore vogliamo verificare che il sistema raggiunga lo stato in cui l'ascensore e' fermo, è al livello di un piano, e le porte sono aperte. La sicurezza ci dice che il sistema non può raggiungere confogurazioni indesiderate: per esempio, non vogliamo che le porte di un ascensore restino bloccate, oppure che si aprano mentre è in movimento o fra due piani.

La verifica delle proprietà di vitalità e sicurezza è possibile con qualunque linguaggio di tipo operazionale. Nelle reti di Petri, queste verifiche si basano sui concetti esposti di séguito.

**Raggiungibilità in un passo.** Data una rete P/T e due marcature $M$ ed $M'$, si dice che $M'$ è *raggiungibile in un passo* da $M$ se

$$\exists_{t \in T} M \, [t\rangle \, M'$$

**Insieme di raggiungibilità.** L'*insieme di raggiungibilità* $R_N(M)$ di una rete posti/transizioni $N$ con marcatura $M$ è la *chiusura transitiva* della relazione di raggiungibiltà in un passo della rete, cioè il piú piccolo insieme di marcature tale che:

$$M \in R_N(M)$$

$$M' \in R_N(M) \wedge \exists_{t \in T} M' \, [t\rangle \, M'' \Rightarrow M'' \in R_N(M)$$

ovvero, $M$ appartiene all'insieme di raggiungibilità e, se $M''$ è raggiungibile in un passo da una marcatura $M'$ appartenente all'insieme di raggiungibilità, allora $M''$ vi appartiene.

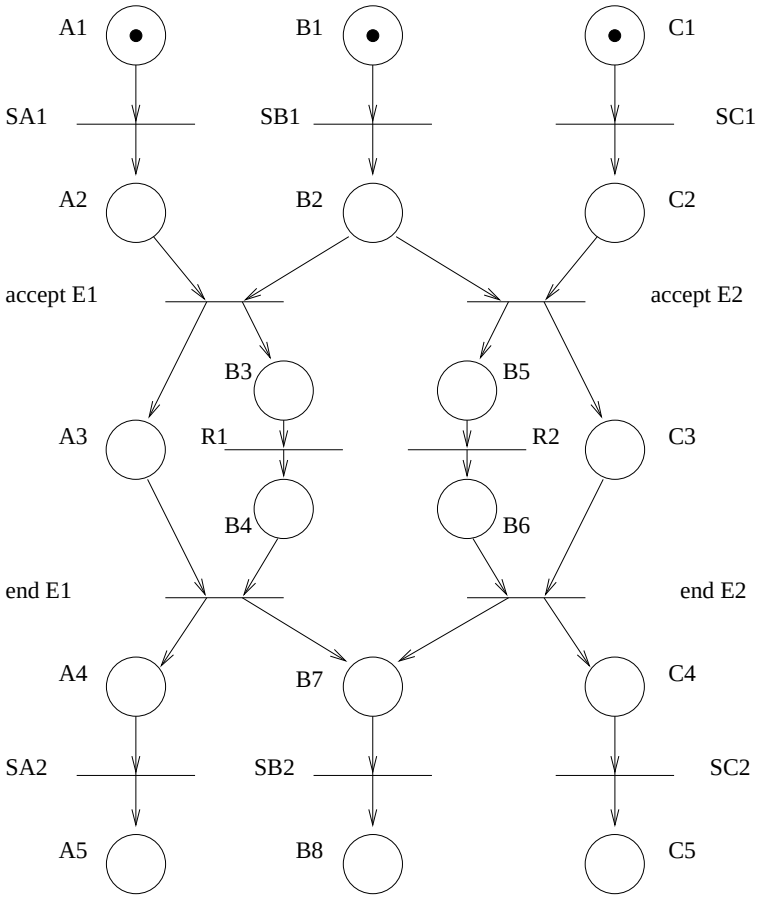


Figura 3.13: Un programma concorrente (2).

**Grafo di raggiungibilità.** Il *grafo di raggiungibilità* di una rete P/T è un grafo i cui nodi sono gli elementi dell'insieme di raggiungibilità, e gli archi sono le transizioni fra una marcatura all'altra. La rete viene cosí tradotta in un automa a stati, generalmente infinito.

### Vitalità delle transizioni

Le proprietà di vitalità di un sitema si possono definire, come abbiamo visto nel paragrafo precedente, in termini di raggiungibilità delle marcature, cioè della possibilità che il sistema assuma determinate configurazioni. Un altro modo di caratterizzare la vitalità si riferisce alla possibilità che il sistema esegua determinate azioni. La *vitalità* di una transizione descrive la possibilità che la transizione venga abilitata, ed è quindi una proprietà legata alla ca-

pacità del sistema di funzionare senza bloccarsi. Si definiscono cinque gradi di vitalità per una transizione:

**grado 0:** la transizione non può mai scattare;

**grado 1:** esiste almeno una marcatura raggiungibile in cui la transizione può scattare;

**grado 2:** per ogni intero $n$ esiste almeno una sequenza di scatti in cui la transizione scatta almeno $n$ volte; è quindi possibile, per qualsíasi $n$, scegliere una sequenza in cui la transizione scatta $n$ volte, ma poi la transizione può non scattare piú;

**grado 3:** esiste almeno una sequenza di scatti in cui la transizione scatta infinite volte;

**grado 4:** per ogni marcatura $M$ raggiungibile da $M_0$, esiste una sequenza di scatti che a partire da $M$ abilita la transizione (si dice che la transizione è *viva*);

Osserviamo che per una transizione vitale al grado 3 esiste almeno una sequenza si scatti (e quindi di marcature) in cui scatta infinite volte, però sono possibili anche sequenze di scatti in cui ciò non accade, perché portano la rete in una marcatura a partire da cui la transizione non può piú scattare. Per una transizione viva (grado 4) è sempre possibile portare la rete in una marcatura ove la transizione è abilitata.

Una rete P/T si dice *viva* se e solo se tutte le sue transizioni sono vive.

Consideriamo, per esempio, la rete di Fig. 3.14. Questa rete modella due processi $A$ e $B$ che devono usare in modo concorrente due risorse; queste ultime sono rappresentate dai place $R_1$ ed $R_2$. Dalla marcatura iniziale è possibile arrivare, per esempio attraverso la sequenza di scatti $[t_1 t_2 t'_1 t'_2]$ ad una marcatura in cui $M(p_3) = M(p'_3) = 1$ e la marcatura degli altri place è zero. In questa marcatura le transizioni $t_3$ e $t'_3$ non possono scattare e inoltre non è possibile raggiungere una marcatura in cui esse siano abilitate: hanno una vitalità di grado zero. Questo significa che ciascuno dei due processi ha acquisito una delle due risorse e nessuno può procedere, non potendo acquisire la risorsa mancante. Questa è una tipica situazione di *deadlock*.

Si può evitare il deadlock modificando il sistema in modo che ciascun processo acquisisca le due risorse in un solo passo. Questo comportamento è modellato dalla rete di Fig. 3.15. In questo caso tutte le transizioni hanno vitalità di grado 4. Ricordiamo però che l'abilitazione di una transizione non implica il suo scatto effettivo: è possibile che il conflitto fra $t_2$ e $t'_2$ venga sempre risolto a favore della stessa transizione. Questa è una situazione di *starvation*.

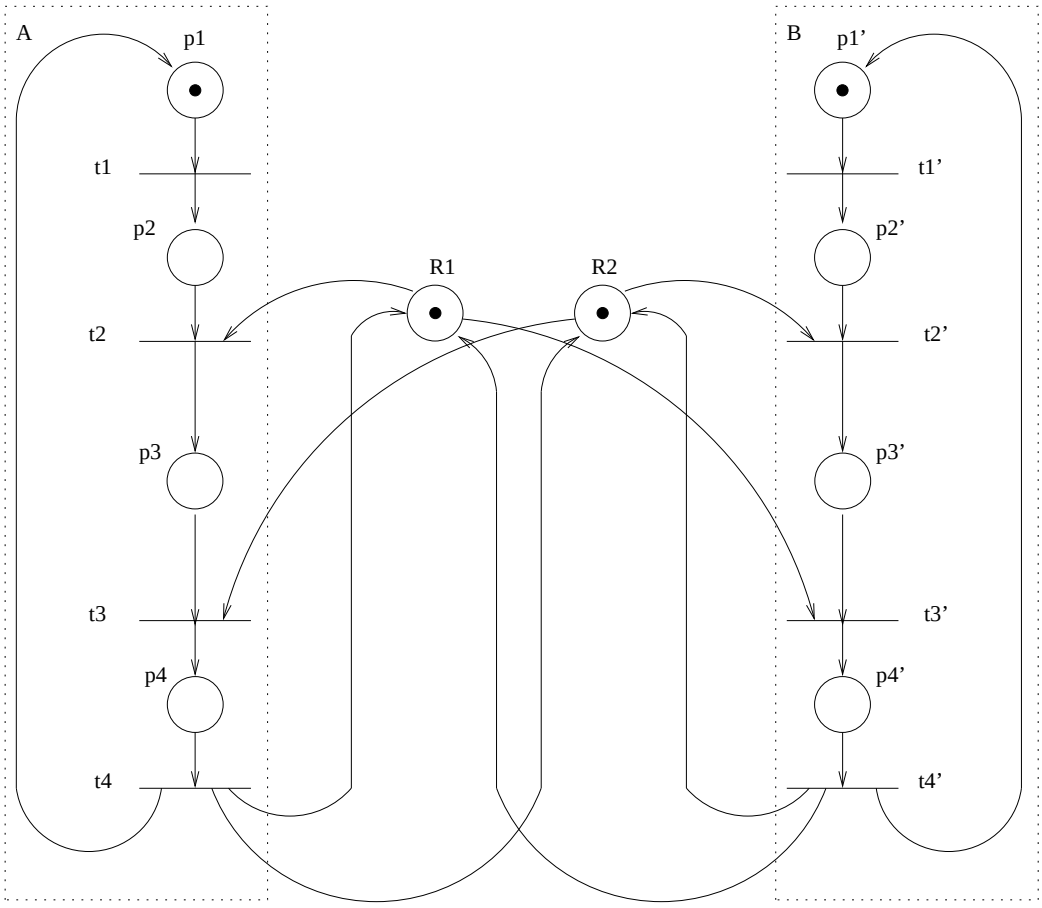


Figura 3.14: Un sistema con possibilità di deadlock.

### Reti limitate

Nel modello generale delle reti di Petri non si pongono limiti al valore della marcatura di ciascun place, ma ciascuna rete, a seconda della sua struttura e della marcatura iniziale, può essere tale che esistano limiti alla marcatura di ciascun place oppure no. Molto spesso l'assenza di limiti nella marcatura di un place índica un situazione indesiderata nel sistema modellato (per esempio, il traboccamento di qualche struttura dati, o il sovraccarico di qualche risorsa), per cui è utile formalizzare il concetto di limitatezza in modo da poter verificare questa proprietà.

Un posto di una rete P/T si dice *k-limitato* (*k-bounded*) se in qualunque marcatura raggiungibile il numero di token non supera $k$.

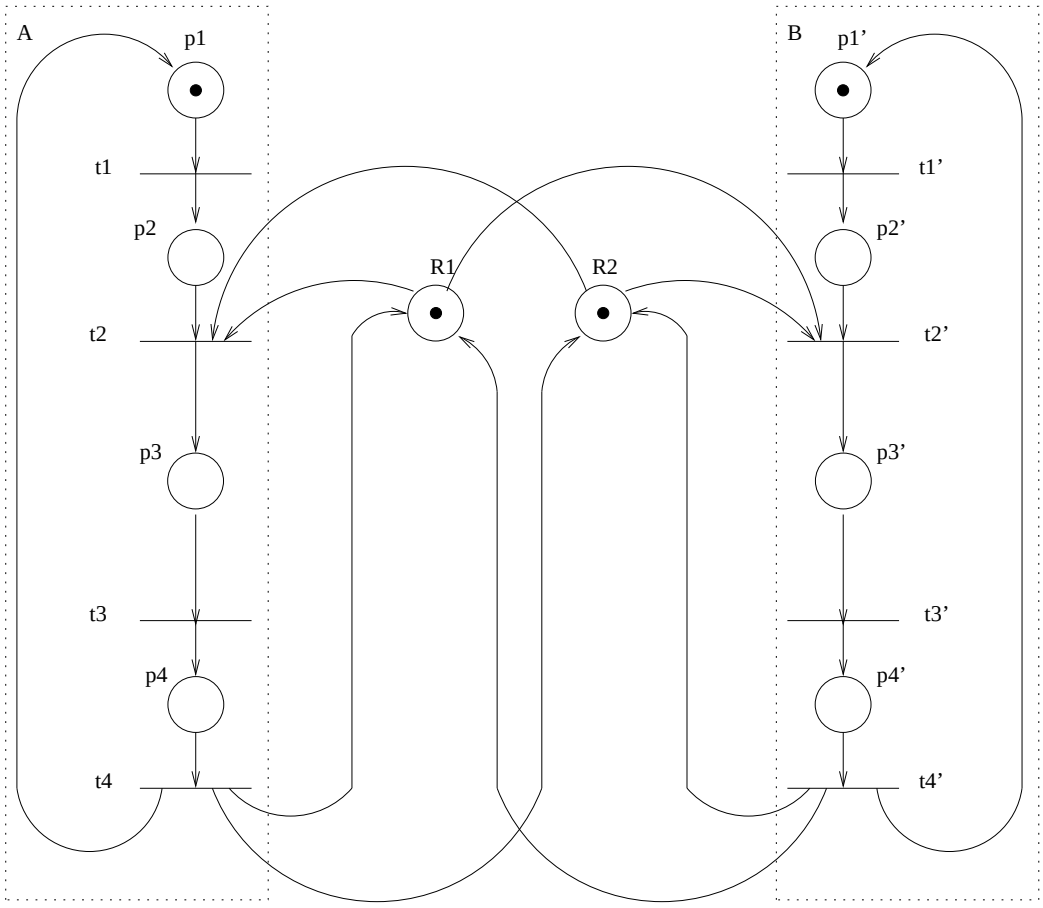


Figura 3.15: Una soluzione per evitare il deadlock.

Una rete P/N con marcatura $M$ si dice *limitata* (*bounded*) se e solo se

$$\exists_{k \in \mathbb{N}} \forall_{M' \in R_N(M)} \forall_{p \in P} M'(p) \leq k$$

cioè se esiste un intero $k$ tale che, in ogni marcatura raggiungibile da $M$, la marcatura di ciascun place sia non maggiore di $k$.

Se $k = 1$, la rete si dice *binaria*.

### Rappresentazione matriciale

È possibile rappresentare una rete di Petri, con le sue possibili marcature e le sequenze di scatti, per mezzo di matrici e vettori. Questa rappresentazione

permette di ricondurre l'analisi della rete a semplici operazioni su matrici, facilitando quindi l'uso di strumenti automatici.

Le matrici *di ingresso* $I$ e *di uscita* $O$ di una rete sono due matrici rettangolari di dimensione $|P| \times |T|$ ($|P|$ righe per $|T|$ colonne)[6]; l'elemento $I_{i,j}$ della matrice di ingresso è il peso dell'arco dal posto $p_i$ alla transizione $t_j$, e l'elemento $O_{i,j}$ della matrice di uscita è il peso dell'arco dalla transizione $t_j$ al posto $p_i$, ossia:

$$\begin{array}{ll} \forall_{\langle p_i,t_j\rangle \in F} & I_{i,j} = W(\langle p_i, t_j\rangle) \\ \forall_{\langle p_i,t_j\rangle \notin F} & I_{i,j} = 0 \\ \forall_{\langle t_j,p_i\rangle \in F} & O_{i,j} = W(\langle t_j, p_i\rangle) \\ \forall_{\langle t_j,p_i\rangle \notin F} & O_{i,j} = 0 \end{array}$$

Osserviamo che ogni colonna di queste matrici è associata ad una transizione, e che ciascun elemento della colonna rappresenta la variazione della marcatura causata dallo scatto della transizione per ogni place di ingresso o di uscita.

La marcatura $M$ viene rappresentata da un vettore colonna $m$ le cui componenti sono le marcature dei singoli posti:

$$m_i = M(p_i)$$

Lo scatto di una transizione $t_j$ nella marcatura $m$ produce la nuova marcatura[7]

$$m' = m - I_{\bullet j} + O_{\bullet j}$$

ovvero, introducendo la *matrice d'incidenza* $C = O - I$,

$$m' = m + C_{\bullet j}$$

Queste equazioni significano che, per ciascun place, la nuova marcatura è uguale alla precedente, meno i pesi degli archi uscenti dal place, piú i pesi degli archi entranti; sono cioè una riformulazione della regola di scatto.

Per ogni sequenza di scatti $S$ si definisce un vettore colonna $s$ di dimensione $|T| \times 1$, in cui ogni componente $s_j$ riporta il numero di volte che la transizione $t_j$ è scattata nella sequenza $S$. Quindi se $M\,[S\rangle\, M'$, vale l'*equazione fondamentale*

$$m' = m + Cs$$

[6]Ricordiamo che $|\mathcal{S}|$ è la cardinalità dell'insieme $\mathcal{S}$, cioè, per un insieme finito, il numero dei suoi elementi

[7]La notazione $A_{\bullet i}$ rappresenta la $i$-esima colonna della matrice $A$.

### Reti conservative

Spesso ci interessa verificare che il numero di token in una rete o in una parte di essa rimanga costante. Una rete si dice quindi *conservativa* se in un suo sottoinsieme di posti la somma dei token contenuti è costante. Il sottoinsieme è individuato da una *funzione peso* che dice quali posti escludere dal conto. Questa funzione molto spesso assegna il peso uno ai posti appartenenti all'insieme e il peso zero ai posti da escludere, ma nel caso piu' generale il peso può assumere valori nell'insieme dei numeri relativi, per modellare situazioni in cui la marcatura di certi place deve essere contata piú di una volta, oppure deve essere sottratta dal numero totale (ha quindi un peso negativo).

Data quindi una funzione peso $H$

$$H : P \to Z$$

una rete P/T con marcatura M si dice *conservativa rispetto alla funzione H* se e solo se

$$\forall_{M' \in R_N(M)} \quad \sum_{p \in P} H(p)M'(p) = \sum_{p \in P} H(p)M(p)$$

Un caso particolare è la rete *strettamente conservativa*, in cui tutti i posti hanno peso unitario, per cui si considera tutta la rete invece di un suo sottoinsieme proprio.

Se una rete è rappresentata in forma matriciale, $H$ viene rappresentata da un vettore riga (*P-vettore*) $h$ tale che $h_i = H(p_i)$. Un *P-invariante* è un P-vettore tale che il suo prodotto scalare per qualsiasi marcatura raggiungibile sia costante. Un P-invariante è quindi una particolare funzione $H$ che individua una sottorete conservativa. Per trovare un tale vettore bisogna risolvere in $h$ la seguente equazione:

$$hm' = hm$$

per qualsiasi marcatura $m'$ raggiungibile da $m$, e quindi per qualsiasi sequenza di scatti. Dall'equazione fondamentale otteniamo

$$h(m + Cs) = hm$$

$$hCs = 0$$

$$hC = \mathbf{0}$$

dove $\mathbf{0}$ è un vettore nullo.

La soluzione in $h$ di questa equazione è formata, per le reti vive, da un insieme di vettori e dalle loro combinazioni lineari.

Come applicazione di questi concetti, consideriamo due processi che devono accedere ad una risorsa in mutua esclusione. Nella rete mostrata in Fig. 3.16, i place $p_2$ e $p_3$ rappresentano le sezioni critiche dei due processi, le transizioni $t_1$ e $t_4$ rappresentano l'ingresso dei due processi nelle rispettive sezioni critiche, e le transizioni $t_2$ e $t_5$ ne rappresentano l'uscita. Il place $p_7$ rappresenta la disponibilità della risorsa.

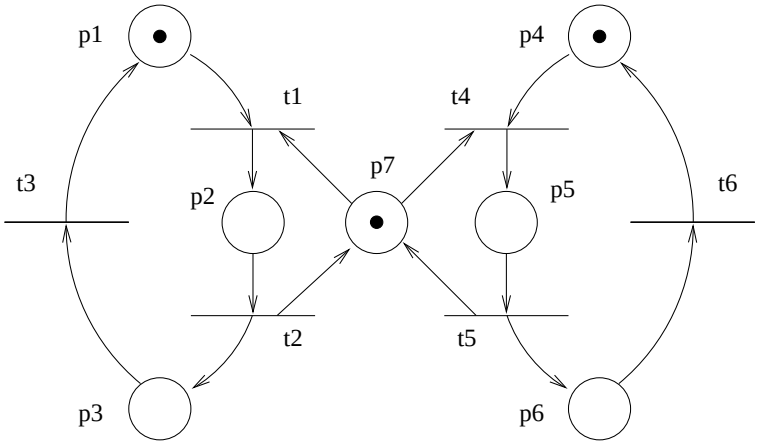


Figura 3.16: pn3

Con la numerazione dei posti e delle transizioni sú esposta, la matrice di incidenza e la marcatura iniziale sono

$$C = \begin{bmatrix} -1 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ -1 & 1 & 0 & -1 & 1 & 0 \end{bmatrix}, \; m = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

I sottoinsiemi di posti che ci interessano sono $P_1 = \{p_1, p_2, p_3\}$, $P_2 = \{p_4, p_5, p_6\}$, e $P_3 = \{p_2, p_7, p_5\}$. Le sottoreti $P_1$ e $P_2$ rappresentano, rispettivamente, gli stati dei due processi: poiché i processi sono sequenziali, il numero di token nelle sottoreti corrispondenti è necessariamente uguale a uno, altrimenti ci sarebbe un errore nella modellazione dei processi per mezzo della rete di

Petri. La sottorete $P_3$ permette di modellare la mutua esclusione: deve contenere un solo token, che appartiene o a $p_7$ (risorsa libera), oppure a uno solo dei place $p_2$ e $p_5$ (risorsa occupata da uno dei processi).

I tre sottoinsiemi sono rappresentati dai P-vettori

$$\begin{aligned} h^{(1)} &= \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \\ h^{(2)} &= \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \\ h^{(3)} &= \begin{bmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \end{aligned}$$

Possiamo verificare che questi vettori sono soluzioni del sistema omogeneo $hC = \mathbf{0}$, e pertanto i prodotti

$$h^{(i)}m, \quad i = 1, \ldots, 3$$

sono costanti al variare di $m$ sulle marcature raggiungibili, ed uguali ad uno. Questo implica, fra l'altro, che viene soddisfatta la condizione di mutua esclusione.

Dai P-invarianti si possono ricavare molte informazioni. In particolare, se $h$ è un P-invariante rispetto ad una marcatura $m$, e si ha che

$$hm' \neq hm$$

allora la marcatura $m'$ non è raggiungibile da $m$.

### Sequenze di scatti cicliche

Una sequenza di scatti si dice *ciclica* se riporta la rete alla marcatura iniziale. Come abbiamo visto, una sequenza di scatti si può rappresentare con un vettore (*T-vettore*) che associa a ciascuna transizione il numero di volte che essa scatta nella sequenza. Dall'equazione fondamentale

$$m' = m + Cs$$

si ottiene che i T-vettori soluzioni dell'equazione

$$Cs = \mathbf{0}$$

rappresentano sequenze di scatti cicliche. Tali vettori sono i *T-invarianti* della rete.

Nell'esempio sulla mutua esclusione, un $T$-invariante è

$$s = \begin{bmatrix} 3 \\ 3 \\ 3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

## 3.6 Logica

La logica serve a formalizzare il ragionamento, e in particolare a decidere in modo rigoroso ("meccanico") se certe affermazioni sono vere e se dalla verità di certe affermazioni si può dedurre la verità di altre affermazioni. È evidente che la logica è fondamentale per qualsiasi forma di ragionamento scientifico, anche quando viene condotto per mezzo del linguaggio naturale, e in qualsiasi campo di applicazione, anche al di fuori delle discipline strettamente scientifiche e tecnologiche. In particolare, la logica moderna è stata sviluppata per servire da fondamento alle discipline matematiche, fra cui rientra gran parte della scienza dell'informazione.

Al di là del suo carattere fondamentale, la logica può essere usata come linguaggio di specifica. Nell'ingegneria del software, quindi, una logica serve da linguaggio di specifica, quando si descrive un sistema in termini delle sue proprietà. Un sistema specificato per mezzo della logica può essere analizzato rigorosamente, e la specifica stessa può essere trasformata per ottenere descrizioni equivalenti ma piú vicine all'implementazione. Usando opportuni linguaggi (linguaggi di *programmazione logica*, come, per esempio, il Prolog), una specifica logica può essere eseguibile, e quindi fornire un prototipo del sistema specificato.

La logica è ovviamente importante per l'ingegneria del software anche perché, come accennato prima, è alla base di tutti i metodi formali usati nell'informatica. Inoltre la logica, come linguaggio di specifica, si può integrare con altri linguaggi, per esempio con l'aggiunta di annotazioni formali che chiariscano aspetti del sistema lasciati indeterminati da descrizioni informali.

Esistono diversi tipi di logica, come la logica proposizionale e la logica dei predicati che vedremo fra poco, ognuno dei quali quali si presta a determinati scopi e campi di applicazione. Ciascuno di questi tipi di logica permette di definire dei *sistemi formali* (o *teorie* formali), ognuno dei quali si basa su un

*linguaggio* per mezzo del quale si possono scrivere formule che rappresentano le affermazioni che ci interessano. Un linguaggio viene descritto dalla propria *sintassi*. Il significato che attribuiamo alle formule è dato dalla *semantica* del linguaggio. La semantica associa i simboli del linguaggio alle entità di cui vogliamo parlare; queste entità costituiscono il *dominio* o *universo di discorso* del linguaggio.

Dato un linguaggio e la sua semantica, un insieme di *regole di inferenza* permette di stabilire se una formula può essere derivata da altre formule, ovvero se una formula è un *teorema* di un certo sistema formale, ovvero se esiste una *dimostrazione* di tale formula.

Le regole di inferenza si riferiscono alla sintassi del linguaggio: possiamo applicare una regola di inferenza ad un insieme di formule senza analizzare il loro significato, e in particolare senza sapere se sono vere o false. Un sistema formale è *corretto* (*sound*) se tutte le formule ottenibili dalle regole d'inferenza sono vere rispetto alla semantica del linguaggio, e *completo* se tutte le formule vere rispetto alla semantica sono ottenibili per mezzo delle regole d'inferenza. Naturalmente la correttezza è un requisito indispensabile per un sistema formale.

Un sistema formale è *decidibile* se esiste un algoritmo che può decidere in un numero finito di passi se una formula è vera (la logica del primo ordine non è decidibile).

Un sistema formale è quindi un apparato di definizioni e regole che ci permette di ragionare formalmente su un qualche settore della conoscenza, Precisiamo che spesso un particolare sistema formale viene indicato come una *logica*. Questa sovrapposizione di termini non crea problemi, perché dal contesto si capisce se si parla *della logica* in generale (la scienza del ragionamento formale) o *di una logica* particolare (un determinato sistema formale).

Le definizioni che verranno date nel séguito sono tratte principalmente da **??**, con varie semplificazioni e probabilmente qualche imprecisione.

### 3.6.1 Calcolo proposizionale

Il calcolo proposizionale è la logica piú semplice. Gli elementi fondamentali del suo linguaggio[8] (non ulteriormente scomponibili) sono le *proposizioni*,

[8]Piú precisamente, dei linguaggi di tipo proposizionale. Per semplicità, diremo genericamente "*il linguaggio* proposizionale".

cioè delle affermazioni che possono essere vere o false. Il fatto che le proposizioni non siano scomponibili, cioè siano *atomiche*, significa che in una frase come "*il tempo è bello*", il linguaggio del calcolo proposizionale non ci permette di individuare il soggetto e il predicato, poiché questo linguaggio non contiene dei simboli che possano nominare oggetti, proprietà o azioni: in un linguaggio proposizionale possiamo nominare soltanto delle frasi intere. Questo comporta anche che non è possibile mettere in evidenza la struttura comune di certi insiemi di frasi, come, per esempio, "*Aldo è bravo*", "*Beppe è bravo*", e "*Carlo è bravo*". Pertanto, qualsiasi proposizione può essere rappresentata semplicemente da una lettera dell'alfabeto, come $T$ per "*il tempo è bello*, $A$ per "*Aldo è bravo*", eccetera.

Le proposizioni vengono combinate per mezzo di alcuni operatori per formare frasi piú complesse: gli operatori (chiamati *connettivi*) sono simili alle congiunzioni del linguaggio naturale, da cui hanno ereditato i nomi. Ai connettivi sono associate delle regole (*tabelle* o *funzioni di verità*) che permettono di stabilire la verità delle frasi complesse a partire dalle proposizioni.

Infine, un calcolo proposizionale ha delle *regole d'inferenza* che permettono di derivare alcune frasi a partire da altre frasi. Le regole d'inferenza sono scelte in modo che le frasi derivate per mezzo di esse siano vere, se sono vere le frasi di partenza. Si fa in modo, cioè, che i sistemi formali di tipo proposizionale siano corretti.

### Sintassi

Nel calcolo proposizionale un linguaggio è formato da:

- un insieme numerabile $\mathcal{P}$ di *simboli proposizionali* ($A$,$B$,$C$, ...);
- un insieme finito di *connettivi*; per esempio: $\neg$ (*negazione*), $\wedge$ (*congiunzione*), $\vee$ (*disgiunzione*), $\Rightarrow$ (*implicazione*[9]), $\Leftrightarrow$ (*equivalenza* o *coimplicazione*);
- un insieme di *simboli ausiliari* (parentesi e simili);
- un insieme $\mathcal{W}$ di *formule* (dette anche *formule ben formate*, *well-formed formulas*, o *wff*), definito dalle seguenti regole:
  1. un simbolo proposizionale è una formula;
  2. se $\mathcal{A}$ e $\mathcal{B}$ sono formule, allora sono formule anche $(\neg\mathcal{A})$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$, ...

[9]L'operazione logica associata a questo connettivo si chiama anche *implicazione materiale*, per distinguerla dall'*implicazione logica* che verrà introdotta piú oltre. Ricordiamo anche che a volte viene usato il simbolo $\rightarrow$ per l'implicazione materiale e il simbolo $\Rightarrow$ per l'implicazione logica.

3. solo le espressioni costruite secondo le due regole precedenti sono formule.

Alle regole sintattiche si aggiungono di solito delle regole ulteriori che permettono di semplificare la scrittura delle formule indicando le priorità dei connettivi, in modo simile a ciò che avviene nella notazione matematica. I connettivi vengono applicati in quest'ordine:

$$\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$$

per cui, ad esempio, la formula

$$A \Leftrightarrow \neg B \vee C \Rightarrow A$$

equivale a

$$A \Leftrightarrow (((\neg B) \vee C) \Rightarrow A)$$

Vediamo quindi che una formula del calcolo proposizionale, analogamente a un'espressione aritmetica, ha una struttura gerarchica: si parte da simboli elementari (i simboli proposizionali) che vengono combinati con operatori unari o binari ottenendo formule che si possono a loro volta combinare, e cosí via.

### Semantica

La semantica di un calcolo proposizionale stabilisce le regole che associano un *valore di verità* a ciascuna formula. Il calcolo di questo valore avviene con un metodo di *ricorsione strutturale*: una formula complessa viene scomposta nelle sue formule componenti, fino ai simboli proposizionali; si assegnano i rispettivi valori di verità a ciascuno di essi (per mezzo della *funzione di valutazione*, vedi oltre), sulla base di questi valori si calcolano i valori delle formule di cui fanno parte i rispettivi simboli (per mezzo delle *funzioni di verità*, vedi oltre), e cosí via fino ad ottenere il valore della formula complessiva.

La semantica è quindi data da:

- l'insieme *booleano* $\mathbb{B} = \{\mathbf{T}, \mathbf{F}\}$, contenente i valori *vero* ($\mathbf{T}$, *true*) e *falso* ($\mathbf{F}$, *false*).
- una *funzione di valutazione* $v : \mathcal{P} \to \mathbb{B}$;
- le *funzioni di verità* di ciascun connettivo

$$\begin{array}{lcl} H_\neg & : & \mathbb{B} \to \mathbb{B} \\ H_\wedge & : & \mathbb{B}^2 \to \mathbb{B} \\ & \ldots & \end{array}$$

- una *funzione di interpretazione* $S_v : \mathcal{W} \to \mathbb{B}$ cosí definita:

$$\begin{aligned} S_v(A) &= v(A) \\ S_v(\neg\mathcal{A}) &= H_\neg(S_v(\mathcal{A})) \\ S_v(\mathcal{A} \wedge \mathcal{B}) &= H_\wedge(S_v(\mathcal{A}), S_v(\mathcal{B})) \\ &\dots \end{aligned}$$

  dove $A \in \mathcal{P}$, $\mathcal{A} \in \mathcal{W}$, $\mathcal{B} \in \mathcal{W}$.

La funzione di valutazione assegna un valore di verità a ciascun simbolo proposizionale. Osserviamo che questa funzione è arbitraria, nel senso che viene scelta da chi si vuole servire di un linguaggio logico per rappresentare un certo dominio, in modo da riflettere ciò che si considera vero in tale dominio.

Per esempio, consideriamo le proposizioni $A$ (*Aldo è bravo*), $S$ (*Aldo passa l'esame di Ingegneria del Software*) e $T$ (*il tempo è bello*). A ciascuna proposizione si possono assegnare valori di verità scelti con i criteri ritenuti piú adatti a rappresentare la situazione, per esempio la valutazione di $A$ può essere fatta in base a un giudizio soggettivo sulle capacità di Aldo, la valutazione di $S$ e $T$ può essere fatta in base all'osservazione sperimentale o anche assegnata arbitrariamente, considerando una situazione ipotetica.

Le funzioni di verità dànno il valore di verità restituito dall'operatore logico rappresentato da ciascun connettivo, in funzione dei valori di verità delle formule a cui viene applicato l'operatore. Di solito queste funzioni vengono espresse per mezzo di tabelle di verità come le seguenti:

| $x$ | $y$ | $H_\neg(x)$ | $H_\wedge(x,y)$ | $H_\vee(x,y)$ | $H_\Rightarrow(x,y)$ | $H_\Leftrightarrow(x,y)$ |
|---|---|---|---|---|---|---|
| **T** | **T** | **F** | **T** | **T** | **T** | **T** |
| **T** | **F** | **F** | **F** | **T** | **F** | **F** |
| **F** | **T** | **T** | **F** | **T** | **T** | **F** |
| **F** | **F** | **T** | **F** | **F** | **T** | **T** |

Le funzioni di verità, a differenza della funzione di valutazione, sono parte integrante del linguaggio: una volta che si è scelto un certo linguaggio per rappresentare un dominio, non si è piú liberi di scegliere le funzioni di verità.

Un insieme di connettivi si dice *completo* se è sufficiente ad esprimere tutte le funzioni di verità possibili. Alcuni insiemi completi sono:

$$\{\neg, \wedge\}$$
$$\{\neg, \vee\}$$
$$\{\neg, \Rightarrow\}$$

La funzione d'interpretazione, infine, calcola il valore di verità di qualsiasi formula, in base alla funzione di valutazione, alle funzioni di verità, ed alla struttura della formula stessa.

### Soddisfacibilità e validità

Una formula, in generale, può essere vera o falsa a seconda della funzione di valutazione, cioè del modo in cui vengono assegnati valori di verità ai simboli proposizionali. Esistono però delle formule il cui valore di verità *non dipende* dalla funzione di valutazione, cioè sono vere (o false) per qualsiasi scelta dei valori di verità dei simboli proposizionali. Ovviamente queste formule sono particolarmente utili perché sono di applicabilità piú generale. Nei paragrafi seguenti esprimeremo piú precisamente la dipendenza del valore di verità di una formula dalla funzione di valutazione.

Se, per una formula $\mathcal{F}$ ed una valutazione $v$, si ha che $S_v(\mathcal{F}) = \mathbf{T}$, si dice che $v$ *soddisfa* $\mathcal{F}$, e si scrive $v \models \mathcal{F}$.

**Osservazione.** Il simbolo $\models$ non appartiene al linguaggio del calcolo proposizionale, poiché non serve a costruire delle formule, ma è solo un'abbreviazione della frase "soddisfa", che appartiene al metalinguaggio, cioè al linguaggio di cui ci serviamo per parlare del calcolo proposizionale.

Una formula si dice *soddisfacibile* o *consistente* se esiste almeno una valutazione che la soddisfa. Per esempio:

$$A \Rightarrow \neg A \quad \text{per } v(A) = \mathbf{F}$$

Una formula si dice *insoddisfacibile* o *inconsistente* se non esiste alcuna valutazione che la soddisfi. Si dice anche che la formula è una *contraddizione*. Per esempio:

$$A \wedge \neg A$$

Una formula soddisfatta da tutte le valutazioni è una *tautologia*, ovvero è *valida*. La verità di una tautologia non dipende quindi dalla verità delle singole proposizioni che vi appaiono, ma unicamente dalla struttura della formula. Esempi di tautologie sono:

$$A \vee \neg A$$

$$A \Rightarrow A$$
$$\neg\neg A \Rightarrow A$$
$$\neg(A \wedge \neg A)$$
$$(A \wedge B) \Rightarrow A$$
$$A \Rightarrow (A \vee B)$$

Se $\mathcal{A} \Rightarrow \mathcal{B}$ (ove $\mathcal{A}$ e $\mathcal{B}$ sono formule) è una tautologia, si dice che $\mathcal{A}$ *implica logicamente* $\mathcal{B}$, ovvero che $\mathcal{B}$ è *conseguenza logica* di $\mathcal{A}$.

Conviene osservare la differenza fra *implicazione* (o *implicazione materiale*) e *implicazione logica*. Infatti, la formula $\mathcal{A} \Rightarrow \mathcal{B}$ si legge "$\mathcal{A}$ implica $\mathcal{B}$", e questa frase significa che il valore di verità della formula viene calcolato secondo la funzione di verità dell'operatore di implicazione a partire dai valori di $\mathcal{A}$ e $\mathcal{B}$, che a loro volta dipendono, in generale, dalla struttura di queste ultime formule e dalla particolare valutazione. La frase "$\mathcal{A}$ implica logicamente $\mathcal{B}$", invece, significa che la formula $\mathcal{A} \Rightarrow \mathcal{B}$ è vera per ogni valutazione, ed è quindi una affermazione piú forte.

Analogamente, se $\mathcal{A} \Leftrightarrow \mathcal{B}$ è una tautologia si dice che $\mathcal{A}$ e $\mathcal{B}$ sono *logicamente equivalenti*.

### 3.6.2 Teorie formali

Nel calcolo dei predicati è sempre possibile accertarsi se una data formula è valida oppure no: basta calcolare il suo valore di verità per tutte le valutazioni possibili, che per ciascuna formula costituiscono un insieme finito, essendo finiti l'insieme dei simboli proposizionali nella formula, delle loro occorrenze, e dei loro possibili valori di verità. Questo metodo diretto di verifica della validità fa riferimento alla semantica del linguaggio usato. I metodi basati sulla semantica del linguaggio sono indicati comunemente col termine *model checking*, e sono impiegati per molti linguaggi formali di specifica nell'ingegneria del software.

Se il numero di simboli proposizionali in una formula è grande, la verifica diretta può essere impraticabile. Inoltre la verifica diretta è generalmente impossibile nelle logiche piú potenti del calcolo proposizionale (come la logica dei predicati, che vedremo fra poco), la cui semantica può comprendere insiemi infiniti di valori. Si pone quindi il problema di *dedurre* la verità di una formula non attraverso il calcolo diretto del suo valore di verità, ma attraver-

so certe relazioni, basate sulla sua sintassi, con altre formule. Informalmente, il meccanismo di deduzione si può descrivere in questo modo:

1. si sceglie un insieme di formule che consideriamo valide *a priori*, senza necessità di dimostrazione, oppure che verifichiamo direttamente per mezzo della semantica;
2. si definiscono delle regole (dette *d'inferenza*) che, date alcune formule con una certa struttura, considerate vere, permettono di scrivere nuove formule;
3. a partire dalle formule introdotte al punto 1, si costruiscono delle catene di formule applicando ripetutamente le regole d'inferenza;
4. ogni formula appartenente a una delle catene cosí costruite si considera dimostrata sulla base delle formule che la precedono nella catena.

### Regole di inferenza

Una *regola di inferenza* è una relazione fra formule; per ogni $n$-upla di formule che soddisfa una regola di inferenza $R$, una determinata formula della $n$-upla viene chiamata *conseguenza diretta* delle altre formule della $n$-upla *per effetto di* $R$. Per esempio, una regola di inferenza potrebbe essere l'insieme delle triple aventi la forma $\langle \mathcal{B}, \mathcal{A}, \mathcal{A} \Rightarrow \mathcal{B} \rangle$. Questa regola si scrive generalmente in questa forma:

$$\frac{\mathcal{A} \qquad \mathcal{A} \Rightarrow \mathcal{B}}{\mathcal{B}}$$

che si legge "$\mathcal{B}$ *è conseguenza diretta di* $\mathcal{A}$ *e di* $\mathcal{A} \Rightarrow \mathcal{B}$". Meno sinteticamente: "Se $\mathcal{A}$ è vera ed $\mathcal{A}$ implica $\mathcal{B}$, allora possiamo dedurre che $\mathcal{B}$ è vera".

### Dimostrazioni e teoremi

Possiamo ora introdurre il concetto di *teoria* (o *sistema*) *formale*, nell'àmbito del quale si definiscono i concetti di *dimostrazione* e di *teorema*.

Una teoria formale è data da

1. un linguaggio $\mathcal{L}$;
2. un insieme $\mathbf{A}$ (eventualmente infinito) di formule di $\mathcal{L}$ chiamate *assiomi*;
3. un insieme finito di regole di inferenza fra formule di $\mathcal{L}$.

Se $\Gamma$ è un insieme di formule, dette *ipotesi* o *premesse*, $\mathcal{F}$ è una formula da dimostrare, e $\Delta$ è una sequenza di formule, allora si dice che $\Delta$ è una

*dimostrazione* (o *deduzione*) *di* $\mathcal{F}$ *da* $\Gamma$ se l'ultima formula di $\Delta$ è $\mathcal{F}$ e ciascun'altra: o (i) è un assioma, o (ii) è una premessa, o (iii) è conseguenza diretta di formule che la precedono nella sequenza $\Delta$.

Si dice quindi che $\mathcal{F}$ *segue da* $\Gamma$, o è *conseguenza* di $\Gamma$, e si scrive

$$\Gamma \vdash \mathcal{F}$$

Se l'insieme $\Gamma$ delle premesse è vuoto, allora la sequenza $\Delta$ è una **dimostrazione di** $\mathcal{F}$, ed $\mathcal{F}$ è un *teorema*, e si scrive

$$\vdash \mathcal{F}$$

Quindi, in una teoria formale, una dimostrazione è una sequenza di formule tali che ciascuna di esse o è un assioma o è conseguenza diretta di alcune formule precedenti, e un teorema è una formula che si può dimostrare ricorrendo solo agli assiomi, senza ipotesi aggiuntive.

**Osservazione.** Notare la differenza fra *dimostrazione di* $\mathcal{F}$ e *dimostrazione di* $\mathcal{F}$ *da* $\Gamma$.

### Una teoria formale per il calcolo proposizionale

Una semplice teoria formale per il calcolo proposizionale può essere definita come segue [10]:

- Il linguaggio $\mathcal{L}$ è costituito dalle formule ottenute a partire dai simboli proposizionali, dai connettivi $\neg$ e $\Rightarrow$, e dalle parentesi.
- Gli assiomi sono tutte le espressioni date dai seguenti *schemi*:

$$\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{A}) \tag{3.1}$$
$$(\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{C})) \Rightarrow ((\mathcal{A} \Rightarrow \mathcal{B}) \Rightarrow (\mathcal{A} \Rightarrow \mathcal{C})) \tag{3.2}$$
$$(\neg\mathcal{B} \Rightarrow \neg\mathcal{A}) \Rightarrow ((\neg\mathcal{B} \Rightarrow \mathcal{A}) \Rightarrow \mathcal{B}) \tag{3.3}$$

  Questa teoria formale ha quindi un insieme infinito di assiomi; osserviamo anche qualsiasi assioma ottenuto da questi schemi è una tautologia.
- L'unica regola d'inferenza è il *modus ponens* (MP): una formula $\mathcal{B}$ è conseguenza diretta di $\mathcal{A}$ e $\mathcal{A} \Rightarrow \mathcal{B}$. Si scrive anche

$$\frac{\mathcal{A} \qquad \mathcal{A} \Rightarrow \mathcal{B}}{\mathcal{B}}$$

Dimostriamo, per esempio, che $F \Rightarrow F$, per ogni formula $F$[10]:

$$
\begin{array}{rll}
(F \Rightarrow ((F \Rightarrow F) \Rightarrow F)) & \Rightarrow & ((F \Rightarrow (F \Rightarrow F)) \Rightarrow (F \Rightarrow F)) \\
 & & \text{da 3.2, } \mathcal{A} = \mathcal{C} = F,\ \mathcal{B} = F \Rightarrow F \qquad (3.4) \\
F \Rightarrow ((F \Rightarrow F) \Rightarrow F) & & \text{da 3.1, } \mathcal{A} = F,\ \mathcal{B} = F \Rightarrow F \qquad (3.5) \\
(F \Rightarrow (F \Rightarrow F)) \Rightarrow (F \Rightarrow F) & & \text{da 3.4 e 3.5 per MP,} \\
 & & \mathcal{A} = (F \Rightarrow ((F \Rightarrow F) \Rightarrow F)), \\
 & & \mathcal{B} = (F \Rightarrow (F \Rightarrow F)) \Rightarrow (F \Rightarrow F) \qquad (3.6) \\
F \Rightarrow (F \Rightarrow F) & & \text{da 3.1, } \mathcal{A} = F,\ \mathcal{B} = F \qquad (3.7) \\
F \Rightarrow F & & \text{da 3.6 e 3.7 per MP,} \\
 & & \mathcal{A} = F \Rightarrow (F \Rightarrow F), \\
 & & \mathcal{B} = F \Rightarrow F \qquad (3.8)
\end{array}
$$

Un utile risultato dimostrabile su questa teoria formale è il *Teorema della deduzione*: *se* $\Gamma$ *è un insieme di wff,* $\mathcal{A}$ *e* $\mathcal{B}$ *sono wff, e* $\Gamma \cup \{\mathcal{A}\} \vdash \mathcal{B}$*, allora* $\Gamma \vdash \mathcal{A} \Rightarrow \mathcal{B}$. In particolare, se $\{\mathcal{A}\} \vdash \mathcal{B}$, allora $\vdash \mathcal{A} \Rightarrow \mathcal{B}$. Cioè, si può affermare che $\mathcal{A}$ implica logicamente[11] $\mathcal{B}$ se $\mathcal{B}$ è dimostrabile da $\mathcal{A}$, coerentemente col comune modo di dimostrare i teoremi in matematica.

**Osservazione.** Il Teorema della deduzione è, propriamente, un metateorema, perché afferma una proprietà di un sistema formale, mentre un teorema è una formula dimostrabile nell'àmbito del sistema stesso.

Infine, si può dimostrare che questa teoria formale è corretta e completa, cioè che ogni teorema di questa teoria è una tautologia, e viceversa.

### 3.6.3 Logica del primo ordine

La logica del primo ordine (anche *FOL*, *first order logic*) permette di formare delle frasi in cui è possibile riferirsi a entità individuali (*oggetti* o *individui*), sia specificamente, per mezzo di simboli (*costanti* e *funzioni*) che denotano particolari entità, sia genericamente, per mezzo di simboli (*variabili*) che possono riferirsi a diversi individui. Le frasi del linguaggio, inoltre, possono essere *quantificate* rispetto alle variabili che vi compaiono, cioè si può indicare

[10] Scriviamo $F$ invece di $\mathcal{F}$ per evidenziare graficamente la formula

[11] Se $\mathcal{A} \Rightarrow \mathcal{B}$ è un teorema, è una tautologia.

se determinate proprietà valgono per tutti i valori di tali variabili o solo per alcuni.

Le frasi piú semplici che possiamo esprimere nel linguaggio della logica del primo ordine sono del tipo "gli oggetti $a$, $b$, $c$, ... sono legati dalla relazione $p$" (o, come caso particolare, "l'oggetto $a$ gode della proprietà $p$"). Queste formule vengono combinate per costruire frasi piú complesse, usando i connettivi ed i quantificatori.

Per esempio, consideriamo queste frasi:

1. "*Aldo è bravo*", "*Beppe è bravo*", "*Carlo è bravo*".
2. "*Aldo passa Ingegneria del Software*", "*Beppe passa Sistemi Operativi e Reti*", "*Carlo passa Epistemologia Generale*".

Le frasi del gruppo 1 dicono che certi individui godono di una certa proprietà, ovvero che Aldo, Beppe e Carlo appartengono all'insieme degli studenti bravi. Le frasi del gruppo 2 dicono che esiste una certa relazione fra certi individui e certi altri individui (non si fa distinzione fra entità animate e inanimate), ovvero che le coppie (Aldo, Ingegneria del Software), (Beppe, Sistemi Operativi e Reti), e (Carlo, Epistemologia Generale) appartengono all'insieme di coppie ordinate tali che l'individuo nominato dal primo elemento della coppia abbia superato l'esame nominato dal secondo elemento.

Sia le frasi del primo gruppo che quelle del secondo si chiamano *predicati* e si possono scrivere sinteticamente come:

$$b(A),\ b(B),\ b(C),\ p(A, I),\ p(B, S),\ p(C, E)$$

dove $b$ sta per "è bravo", $p$ per "passa l'esame di", $A$ per "Aldo", e cosí via.

Queste frasi sono istanze particolari delle formule $b(x)$ e $p(x, y)$, dove le variabili $x$ e $y$ sono dei simboli segnaposto che devono essere sostituiti da nomi di individui per ottenere delle frasi cui si possa assegnare un valore di verità. È bene sottolineare che la formula $b(x)$ non significa "*qualcuno è bravo*": la formula non significa niente, non è né vera né falsa, finché il simbolo $x$ non viene sostituito da un'espressione che denoti un particolare individuo, per esempio "Aldo" o anche "il padre di Aldo". Dopo questa sostituzione, la formula può essere vera o falsa, ma ha comunque un valore definito.

Se vogliamo esprimere nella logica del primo ordine la frase "*qualcuno è bravo*", dobbiamo usare un nuovo tipo di simbolo, il quantificatore esistenziale: $\exists x b(x)$, cioè "esiste qualcuno che è bravo". Questa formula ha significato

cosí com'è, perché afferma una proprietà dell'insieme degli studenti bravi (la proprietà di non essere vuoto).

Una formula come $b(x)$, dove, cioè, appaiono variabili non quantificate, si dice *aperta*, mentre una formula come $\exists x b(x)$, con le variabili quantificate, si dice *chiusa*.

### Sintassi

Quanto sopra viene espresso piú formalmente dicendo che un linguaggio del primo ordine è costituito da:

- un insieme numerabile $\mathcal{V}$ di *variabili* ($x$, $y$, $z$, ...);
- un insieme numerabile $\mathcal{F}$ di *simboli di funzione* ($f$, $g$, $h$, ...); a ciascun simbolo di funzione è associato il numero di argomenti (*arietà*) della funzione: $\mathcal{F} = \mathcal{F}_0 \cup \mathcal{F}_1 \cup \mathcal{F}_2 \cup \ldots$, dove $\mathcal{F}_n$ è l'insieme dei simboli $n$-ari di funzione; le funzioni di arietà zero sono chiamate *costanti* ($a$, $b$, $c$, ...);
- un insieme $\mathcal{T}$ di *termini* ($t_1$, $t_2$, $t_3$, ...), definito dalle seguenti regole:
  1. una variabile è un termine;
  2. se $f$ è un simbolo $n$-ario di funzione e $t_1, \ldots, t_n$ sono termini, allora $f(t_1, \ldots, t_n)$ è un termine; in particolare, un simbolo $c$ di arietà nulla è un termine (*costante*);
  3. solo le espressioni costruite secondo le due regole precedenti sono termini.
- un insieme numerabile $\mathcal{P}$ di *simboli di predicato* ($p$, $q$, $r$, ...); a ciascun simbolo di predicato è associato il numero di argomenti (*arietà*) del predicato: $\mathcal{P} = \mathcal{P}_0 \cup \mathcal{P}_1 \cup \mathcal{P}_2 \cup \ldots$, dove $\mathcal{P}_n$ è l'insieme dei simboli $n$-ari di predicato; i predicati di arietà zero corrispondono ai simboli proposizionali del calcolo proposizionale;
- un insieme finito di *connettivi*; per esempio:

$$\neg, \wedge, \vee, \Rightarrow, \ldots$$

- un insieme finito di *quantificatori*; per esempio:

$$\forall, \exists$$

- un insieme di *simboli ausiliari* (parentesi e simili);
- un insieme $\mathcal{W}$ di *formule*, definito dalle seguenti regole:

1. se $p$ è un simbolo $n$-ario di predicato e $t_1, \ldots, t_n$ sono termini, allora $p(t_1, \ldots, t_n)$ è una formula (detta *atomica*, in questo caso);
2. se $\mathcal{A}$ e $\mathcal{B}$ sono formule, allora sono formule anche $(\neg\mathcal{A})$, $(\mathcal{A} \wedge \mathcal{B})$, $(\mathcal{A} \vee \mathcal{B})$, ...
3. se $\mathcal{A}$ è una formula e $x$ è una variabile, allora sono formule anche $(\forall x\mathcal{A})$ e $(\exists x\mathcal{A})$. Le formule $(\forall x\mathcal{A})$ e $(\exists x\mathcal{A})$ sono dette *formule quantificate* ed $\mathcal{A}$ è il *campo* del rispettivo quantificatore.
4. solo le espressioni costruite secondo le regole precedenti sono formule.

Le priorità nell'ordine di applicazione dei connettivi è la stessa vista per il calcolo proposizionale, salvo che la priorità dei quantificatori è intermedia fra quella di $\vee$ e quella di $\Rightarrow$.

### Semantica

La semantica di una logica del primo ordine è data da:

- l'insieme $\mathbb{B} = \{\mathbf{T}, \mathbf{F}\}$;
- le *funzioni di verità* di ciascun connettivo;
- un insieme non vuoto $\mathcal{D}$, detto *dominio* dell'interpretazione;
- una *funzione di interpretazione delle funzioni* $\Phi : \mathcal{F} \to \mathcal{F}_\mathcal{D}$, dove $\mathcal{F}_\mathcal{D}$ è l'insieme delle funzioni su $\mathcal{D}$. $\Phi$ assegna a ciascun simbolo $n$-ario di funzione una funzione $\mathcal{D}^n \to \mathcal{D}$.
- una *funzione di interpretazione dei predicati* $\Pi : \mathcal{P} \to \mathcal{R}_\mathcal{D}$, dove $\mathcal{R}_\mathcal{D}$ è l'insieme delle relazioni su $\mathcal{D}$. $\Pi$ assegna a ciascun simbolo $n$-ario di predicato una funzione $\mathcal{D}^n \to \mathbb{B}$.
- un *assegnamento di variabili* $\xi : \mathcal{V} \to \mathcal{D}$;
- un *assegnamento di termini* $\Xi : \mathcal{T} \to \mathcal{D}$ cosí definito:

$$\begin{aligned} \Xi(x) &= \xi(x) \\ \Xi(f(t_1, \ldots, t_n)) &= \Phi(f)(\ \Xi(t_1), \ldots, \Xi(t_n)\ ) \end{aligned}$$

dove $x \in \mathcal{V}$, $t_i \in \mathcal{T}$, $f \in \mathcal{F}$;

- una *funzione di interpretazione* $S_{I,\xi} : \mathcal{W} \to \mathbb{B}$ cosí definita:

$$\begin{array}{rcl} S_{I,\xi}(p(t_1,\ldots,t_n)) & = & \Pi(p)(\ \Xi(t_1),\ldots,\Xi(t_n)\ ) \\ S_{I,\xi}(\neg\mathcal{A}) & = & H_\neg(S_{I,\xi}(\mathcal{A})) \\ S_{I,\xi}(\mathcal{A}\wedge\mathcal{B}) & = & H_\wedge(S_{I,\xi}(\mathcal{A}), S_{I,\xi}(\mathcal{B})) \\ & \ldots & \\ S_{I,\xi}(\exists x\mathcal{A}) & = & \mathbf{T} \\ & & \text{se e solo se esiste un } d \in \mathcal{D} \text{ tale che} \\ & & [S_{I,\xi}]_{x/d}(\mathcal{A}) = \mathbf{T} \\ S_{I,\xi}(\forall x\mathcal{A}) & = & \mathbf{T} \\ & & \text{se e solo se per ogni } d \in \mathcal{D} \text{ si ha} \\ & & [S_{I,\xi}]_{x/d}(\mathcal{A}) = \mathbf{T} \end{array}$$

  dove $p \in \mathcal{P}$, $t_i \in \mathcal{T}$, $\mathcal{A},\mathcal{B} \in \mathcal{W}$, $I = (\mathcal{D}, \Phi, \Pi)$ è detta *interpretazione* del linguaggio, $x \in \mathcal{V}$, e $[S_{I,\xi}]_{x/d}$ è la funzione di interpretazione uguale a $S_{I,\xi}$ eccetto che assegna alla variabile $x$ il valore $d$.

Il dominio è l'insieme degli oggetti di cui vogliamo parlare. Per esempio, se volessimo parlare dell'aritmetica, il dominio sarebbe l'insieme dei numeri interi.

La funzione d'interpretazione delle funzioni stabilisce il significato dei simboli che usiamo nel nostro linguaggio per denotare le funzioni. Nell'esempio dell'aritmetica, una funzione di interpretazione delle funzioni potrebbe associare la funzione aritmetica *somma* ($somma \in \mathcal{F}_\mathcal{D}$) al simbolo di funzione $f$ ($f \in \mathcal{F}$), cioè $\Phi(f) = somma$. In particolare, $\Phi$ stabilisce il significato delle costanti, per esempio possiamo associare al simbolo $a$ il numero *zero*, al simbolo $b$ il numero *uno*, e cosí via. Di solito, quando si usa la logica per parlare di un argomento ove esiste una notazione tradizionale, si cerca di usare quella notazione, a meno che non si voglia sottolineare la distinzione fra linguaggio e dominio (come stiamo facendo qui). Quindi in genere è possibile usare il simbolo '+' per la somma, il simbolo '1' per il numero *uno*, eccetera.

La funzione d'interpretazione dei predicati stabilisce il significato dei simboli che denotano le relazioni. Per esempio, una funzione di interpretazione dei predicati può associare la relazione *minore o uguale* al simbolo di predicato $p$.

L'assegnamento di variabili stabilisce il significato delle variabili. Per esempio, possiamo assegnare il valore *tre* alla variabile $x$.

L'assegnamento di termini stabilisce il significato dei termini. Per esempio, dato il termine $f(b,x)$, e supponendo che $\Phi(f) = somma$, $\Phi(b) = uno$, $\xi(x) = tre$, allora $\Xi(f(b,x)) = quattro$.

| simbolo | funzione | interpretazione | simbolo comune |
|---|---|---|---|
| $f$ | $\Phi$ | somma | $+$ |
| $b$ | $\Phi$ | uno | 1 |
| $x$ | $\xi$ | tre | 3 |
| $p$ | $\Pi$ | minore o uguale | $\leq$ |

Tabella 3.2: Esempio di interpretazione.

Infine, la funzione d'interpretazione (delle formule) stabilisce il significato delle formule di qualsivoglia complessità. Come nel calcolo proposizionale, la funzione di interpretazione ha una definizione ricorsiva e si applica analizzando ciascuna formula nelle sue componenti. Le formule piú semplici (formule atomiche) sono costituite da simboli di predicato applicati ad $n$-uple di termini, quindi il valore di verità di una formula atomica (fornito dalla funzione $\Pi$) dipende dall'assegnamento dei termini.

La definizione della funzione di interpretazione per le formule quantificate rispecchia la comune nozione di quantificazione esistenziale ed universale, a dispetto della notazione un po' oscura qui adottata. L'espressione $[S_{I,\xi}]_{x/d}$, come abbiamo visto, è definita come "la funzione di interpretazione uguale a $S_{I,\xi}$ eccetto che assegna alla variabile $x$ il valore $d$". Questo significa che, per vedere se una formula quantificata sulla variabile $x$ è vera, non ci interessa *il* valore attribuito a $x$ dal particolare assegnamento $\xi$ , ma *l'insieme* dei possibili valori di $x$: per la quantificazione esistenziale vediamo se almeno uno dei valori possibili rende vero il campo del quantificatore, per la quantificazione universale vediamo se tutti i valori possibili lo rendono vero. In ambedue i casi, per le altre variabili si considerano i valori assegnati da $\xi$.

Come esempio di interpretazione, data la formula $p(f(b,x),x) \wedge p(b,x)$, si ha $S_{I,\xi}(p(f(b,x),x) \wedge p(b,x)) = \mathbf{F}$, se l'interpretazione $I$ e l'assegnamento di variabili $\xi$ sono compatibili con le interpretazioni ed assegnamenti visti sopra. Si verifica facilmente che $I$ e $\xi$ trasformano la formula logica considerata nell'espressione $1 + 3 \leq 3 \wedge 1 \leq 3$. La Tab. 3.2 riassume le interpretazioni dei vari simboli.

### Soddisfacibilità e validità

Nel calcolo proposizionale l'interpretazione di una formula dipende solo dalla sua struttura e, in generale, dalla funzione di valutazione. Nella logica dei predicati le formule sono piú complesse e il loro valore di verità dipende,

oltre che dalla struttura della formula, anche dal dominio di interpretazione, dalle funzioni d'interpretazione dei simboli di funzione e di predicato, e dalla funzione di assegnamento di variabili. Il dominio e le funzioni di interpretazione costituiscono, come abbiamo visto, l'interpretazione del linguaggio e ne definiscono l'aspetto strutturale, invariabile. Anche nella logica dei predicati ci interessa studiare la dipendenza del valore di verità delle formule dall'interpretazione del linguaggio e dall'assegnamento di variabili. Di séguito le definizioni relative.

Una formula $\mathcal{A}$ è *soddisfacibile in un'interpretazione* $I$ se e solo se esiste un assegnamento di variabili $\xi$ tale che $S_{I,\xi}(\mathcal{A}) = \mathbf{T}$. Si dice allora che l'interpretazione $I$ *soddisfa* $\mathcal{A}$ con assegnamento di variabili $\xi$, e si scrive $I \overset{\xi}{\models} \mathcal{A}$.

Una formula $\mathcal{A}$ è *soddisfacibile* (tout-court) se e solo se esiste un'interpretazione $I$ in cui $\mathcal{A}$ è soddisfacibile.

Una formula $\mathcal{A}$ è *valida in un'interpretazione* (o *vera in un'interpretazione*) $I$ se e solo se $S_{I,\xi}(\mathcal{A}) = \mathbf{T}$ per ogni un assegnamento di variabili $\xi$. Si dice quindi che $I$ è un *modello* di $\mathcal{A}$, e si scrive $I \models \mathcal{A}$.

Una formula $\mathcal{A}$ è (*logicamente*) *valida* se e solo se è valida per ogni interpretazione $I$ e si scrive $\models \mathcal{A}$.

I concetti di validità e soddisfacibilità si estendono a insiemi di formule: un insieme di formule è equivalente alla loro congiunzione.

Una formula $\mathcal{A}$ *implica logicamente* $\mathcal{B}$, ovvero $\mathcal{B}$ è *conseguenza logica* di $\mathcal{A}$ se e solo se $\mathcal{A} \Rightarrow \mathcal{B}$ è valida.

Una formula $\mathcal{A}$ *è logicamente equivalente* a $\mathcal{B}$, se e solo se $\mathcal{A} \Leftrightarrow \mathcal{B}$ è valida.

### Una teoria formale per la FOL

Una semplice teoria formale per la FOL può essere definita come segue [10]:

- il linguaggio $\mathcal{L}$ è un linguaggio del primo ordine che usa i connettivi $\neg$ e $\Rightarrow$, il quantificatore $\forall$, e le parentesi[12];
- gli assiomi sono divisi in *assiomi logici* ed *assiomi propri* (o *non logici*);

[12]La quantificazione esistenziale si può esprimere usando il quantificatore universale, se poniamo $\exists x\mathcal{A}$ equivalente a $\neg(\forall x(\neg\mathcal{A}))$.

gli assiomi logici sono tutte le espressioni date dai seguenti *schemi*:

$$\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{A}) \tag{3.9}$$

$$(\mathcal{A} \Rightarrow (\mathcal{B} \Rightarrow \mathcal{C})) \Rightarrow ((\mathcal{A} \Rightarrow \mathcal{B}) \Rightarrow (\mathcal{A} \Rightarrow \mathcal{C})) \tag{3.10}$$

$$(\neg\mathcal{B} \Rightarrow \neg\mathcal{A}) \Rightarrow ((\neg\mathcal{B} \Rightarrow \mathcal{A}) \Rightarrow \mathcal{B}) \tag{3.11}$$

$$\forall x \mathcal{A}(x) \Rightarrow \mathcal{A}(t) \tag{3.12}$$

$$\forall x(\mathcal{A} \Rightarrow \mathcal{B}) \Rightarrow (\mathcal{A} \Rightarrow \forall x \mathcal{B}) \tag{3.13}$$

- le regole d'inferenza sono:

  **modus ponens** come nel calcolo proposizionale;
  **generalizzazione**

$$\frac{\mathcal{A}}{\forall x \mathcal{A}}$$

Negli assiomi, $\mathcal{A}(x)$ rappresenta una formula contenente la variabile $x$ (ed eventualmente altre variabili), ed $\mathcal{A}(t)$ è la formula ottenuta da $\mathcal{A}(x)$ sostituendo $x$ col termine $t$.

Allo schema di assioma 3.12 bisogna aggiungere la restrizione che $t$ sia *libero per $x$ in $\mathcal{A}$*, cioè che non contenga variabili che in $\mathcal{A}$ abbiano un quantificatore $\forall$ il cui campo d'azione includa $x$; questo vincolo impedisce che variabili libere vengano sostituite da variabili quantificate. Supponiamo, per esempio, che $\mathcal{A}(x)$ sia la formula $\neg\forall y(x = y)$ e $t$ sia il termine $y$. Nella formula $\mathcal{A}(x)$ la variabile $y$ è quantificata da $\forall y$, e il campo d'azione di questo quantificatore è la formula $(x = y)$, che contiene $x$, quindi il termine $t$ non è libero per $x$ in $\mathcal{A}(x)$. Se nonostante questo volessimo applicare lo schema 3.12, otterremmo la formula $\forall x(\neg\forall y(x = y)) \Rightarrow \neg(\forall y(y = y))$, che è falsa: infatti questa formula significa che "*se per ogni $x$ esiste un $y$ diverso da $x$, allora non è vero che ogni $y$ è uguale a se stesso*".

Allo schema di assioma 3.13 bisogna aggiungere la restrizione che $\mathcal{A}$ non contenga occorrenze libere di $x$. Questo serve ad impedire che qualche variabile "perda" la sua quantificazione. Per esempio, se $\mathcal{A}$ e $\mathcal{B}$ corrispondono tutte e due alla formula $p(x)$, dove $p(x)$ viene interpretato come "$x$ è pari", l'applicazione dell'assioma 3.13 porterebbe alla formula $\forall x(p(x) \Rightarrow p(x)) \Rightarrow (p(x) \Rightarrow \forall x\ p(x))$, che è falsa: l'antecedente dell'implicazione principale $(\forall x(p(x) \Rightarrow p(x)))$ significa che "*per ogni $x$ la parità implica la parità*", e il conseguente $((p(x) \Rightarrow \forall x\ p(x)))$ significa che "*la parità di un numero implica la parità di tutti i numeri*".

Gli assiomi propri sono specifici di ciascuna particolare teoria (p.es., teoria dei gruppi, dell'aritmetica, ...) e possono mancare. Un sistema formale privo di assiomi propri si dice *calcolo dei predicati*.

In ogni calcolo dei predicati, tutti i teoremi sono formule logicamente valide e viceversa (Teorema di completezza di Gödel).

### 3.6.4 Logiche tipate

Nelle logiche tipate, il dominio è ripartito in *tipi* (*types*, *sorts*). Esiste un quantificatore per ciascun tipo, per ogni predicato viene fissato il tipo di ciascun argomento, e per ogni funzione vengono fissati i tipi degli argomenti e del risultato. Le logiche tipate sono equivalenti alle logiche non tipate, nel senso che qualsiasi espressione di una logica tipata può essere tradotta in un'espressione non tipata equivalente, ma permettono di esprimere le specifiche in modo piú naturale.

### 3.6.5 Esempio di specifica e verifica formale

Consideriamo il problema dell'ordinamento di un vettore (esempio tratto da [4]). Vogliamo specificare la relazione fra un vettore arbitrario $x$ di $N > 2$ elementi ed il vettore $y$ ottenuto ordinando $x$ in ordine crescente, supponendo che gli elementi del vettore abbiano valori distinti. Possiamo esprimere questa relazione cosí:

$$\begin{aligned} \mathit{ord}(x,y) &\Leftrightarrow \mathit{permutazione}(x,y) \wedge \mathit{ordinato}(y) \\ \mathit{permutazione}(x,y) &\Leftrightarrow \forall k((1 \le k \wedge k \le N) \Rightarrow \\ &\quad \exists i(1 \le i \wedge i \le N \wedge y_i = x_k) \wedge \\ &\quad \exists j(1 \le j \wedge j \le N \wedge x_j = y_k)) \\ \mathit{ordinato}(x) &\Leftrightarrow \forall k(1 \le k \wedge k < N \Rightarrow x_k \le x_{k+1}) \end{aligned}$$

Possiamo verificare la correttezza di un programma che ordina un vettore di $N$ elementi rispetto alla specifica. Trascureremo la parte di specifica relativa alla permutazione di un vettore. L'ordinamento di un vettore `v` può essere ottenuto col seguente frammento di programma, che implementa l'algoritmo bubblesort, dove $M = N - 1$ (ricordiamo che in C++ gli indici di un array di $N$ elementi vanno da 0 a $N - 1$):

```
for (i = 0; i < M; i++) {           // 0 ≤ i < M
    for (j = 0; j < M-i; j++) {     // 0 ≤ j < M − i
        if (v[j] > v[j+1]) {
            t = v[j];
```

```
            v[j] = v[j+1];
            v[j+1] = t;
        } // v_j ≤ v_{j+1}                                    (i)
    } // ∀k(M − i − 1 ≤ k < M ⇒ v_k ≤ v_{k+1})                (ii)
} // ∀k(0 ≤ k < M ⇒ v_k ≤ v_{k+1})                            (iii)
```

L'asserzione (i) vale dopo l'esecuzione dell'istruzione `if`, per il valore corrente di `j`. Le asserzioni (ii) e (iii) valgono, rispettivamente, all'uscita del loop interno (quando sono stati ordinati gli ultimi $i+2$ elementi) e del loop esterno (quando sono stati ordinati tutti gli elementi). Le altre due asserzioni esprimono gli intervalli dei valori assunti da `i` e `j`. La forma $A \leq B < C$ è un'abbreviazione di $A \leq B \wedge B < C$.

L'asserzione (i) può essere verificata informalmente considerando le operazioni svolte nel corpo dell'istruzione `if` (una verifica formale richiede una definizione della semantica dell'assegnamento).

Possiamo verificare la formula (ii) per induzione sul valore di $i$, che controlla il numero di iterazioni del ciclo esterno.

Per $i = 0$, la (ii) diventa

$$\forall k(M - 1 \leq k < M \Rightarrow v_k \leq v_{k+1})$$

La variabile $k$ assume solo il valore $M-1$, e $j$ varia fra 0 ed $M-1$, pertanto all'uscita del ciclo si ha dalla (i) che $v_{M-1} \leq v_M$, e la (ii) è verificata per $i = 0$.

Per un $\bar{\imath}$ arbitrario purché minore di $M-1$, la (ii) (che è vera per l'ipotesi induttiva) assicura che gli ultimi $\bar{\imath}+2$ elementi del vettore sono ordinati. Sono state eseguite $M - \bar{\imath}$ iterazioni del ciclo interno, e (dalla (i)) $v_{M-\bar{\imath}-1} < v_{M-\bar{\imath}}$. Per $i = \bar{\imath} + 1$, il ciclo interno viene eseguito $M - \bar{\imath} - 1$ volte, e all'uscita del ciclo $v_{M-\bar{\imath}-2} < v_{M-\bar{\imath}-1}$. Quindi gli ultimi $i+2$ elementi sono ordinati. La (ii) è quindi dimostrata.

La formula (iii) si ottiene dalla (ii) ponendovi $i = M - 1$, essendo $M-1$ l'ultimo valore assunto da $i$. Risulta quindi che all'uscita del ciclo esterno il vettore è ordinato, q.e.d.

### 3.6.6 La notazione Z

La notazione Z [14] è uno dei piú noti formalismi di specifica basati sulla logica. Le basi di questo formalismo sono:

- una logica del primo ordine, con teoria dell'uguaglianza;
- la teoria degli insiemi (di Zermelo-Fraenkel);
- un sistema deduttivo basato sulla *deduzione naturale*;
- un *linguaggio di schemi* che permette di strutturare una specifica in un insieme di definizioni raggruppate in moduli.

La sintassi della logica usata da Z è leggermente diversa da quella usata in queste dispense. La teoria degli insiemi comprende i concetti relativi alle relazioni e alle funzioni, nonché all'aritmetica, ed usa un grande numero di operatori che permettono di esprimere sinteticamente le proprietà delle entità specificate.

Il sistema di deduzione naturale, a differenza dei sistemi minimi visti finora, sfrutta molte regole di inferenza associate agli operatori logici e matematici del linguaggio. Questo permette di costruire dimostrazioni piú facilmente.

La notazione Z comprende varie convenzioni per facilitare la specifica dei sistemi di calcolo. In particolare, lo stato di un sistema viene descritto da uno schema (gruppo di definizioni e formule logiche), e le operazioni che lo modificano vengono descritte da schemi che definiscono le variazioni dello stato in seguito alle operazioni. Opportune notazioni permettono di esprimere sinteticamente queste variazioni.

La metodologia di sviluppo basata sulla notazione Z prevede che un sistema venga descritto in termini matematici (usando concetti quali *insieme*, *sequenza*, *funzione*...), e che la specifica iniziale venga trasformata progressivamente fino ad ottenere una descrizione abbastanza dettagliata dell'implementazione. Il sistema deduttivo permette di verificare determinate relazioni o proprietà, oltre alla consistenza fra le diverse versioni della specifica ottenute nei vari passi del processo di sviluppo.

### 3.6.7 La Real-Time Logic

La Real-Time Logic (RTL) [6] è una logica tipata del primo ordine orientata all'analisi di proprietà temporali dei sistemi in tempo reale.

La semantica della RTL è un modello che riassume i concetti fondamentali dei sistemi in tempo reale:

**azioni**
**predicati di stato**

**eventi**
**vincoli temporali**

Le azioni possono essere *semplici* o *composte*, possono essere eseguite in sequenza o in parallelo, e possono essere soggette a vincoli di sincronizzazione.

I predicati di stato sono asserzioni sullo stato del sistema fisico di cui fa parte il software da analizzare.

Gli eventi sono suddivisi in *eventi esterni* (causati dall'ambiente esterno), *eventi iniziali* che segnano l'inizio di un'azione, *eventi finali* che ne segnano la fine, *eventi di transizione* che segnano i cambiamenti in qualche attributo del sistema.

I vincoli temporali sono asserzioni sui valori assoluti degli istanti in cui si verificano gli eventi. I vincoli possono essere *periodici* se prescrivono che le occorrenze di un evento si verifichino con una certa frequenza, *sporadici* se prescrivono che un'occorrenza di un evento avvenga entro un certo tempo dopo l'occorrenza di un altro evento.

La sintassi dell'RTL comprende delle costanti che denotano eventi, azioni, e numeri interi, ed operatori per rappresentare la struttura delle azioni composte. La *funzione di occorrenza* $@(e, i)$ denota l'istante dell'$i$-esima occorrenza dell'evento $e$. I predicati di stato vengono espressi specificando l'intervallo di tempo in cui sono (o devono essere) veri. Il linguaggio dell'RTL include i comuni connettivi logici e gli operatori aritmetici e relazionali.

La parte deduttiva della RTL si basa su una serie di assiomi che descrivono le proprietà fondamentali di eventi e azioni, ricorrendo alla funzione di occorrenza. Per esempio, la funzione di occorrenza stessa soddisfa, per ogni evento $E$, i seguenti assiomi:

$$\forall i @(E, i) = t \Rightarrow t \geq 0$$
$$\forall i \forall j (@(E, i) = t \wedge @(E, j) = t' \wedge i < j) \Rightarrow t < t'$$

Questi assiomi formalizzano due semplici nozioni: ogni occorrenza di un evento avviene in un istante successivo all'istante $t = 0$ preso come riferimento, e due occorenze successive di uno stesso evento avvengono in istanti successivi.

Un sistema viene specificato attraverso la sua descrizione in RTL: questa descrizione, insieme agli assiomi, costituisce una teoria rispetto alla quale si possono verificare delle proprietà del sistema, in particolare le *asserzioni di sicurezza* che rappresentano i vincoli temporali che devono essere rispettati dal sistema. Per esempio, la seguente formula specifica che un'operazione di

lettura (*READ*) deve durare meno di 10 unità di tempo:

$$\forall i @(\uparrow READ, i) + 10 \geq @(\downarrow READ, i)$$

Nella formula precedente, i simboli $\uparrow READ$ e $\downarrow READ$ rappresentano, rispettivamente, l'evento iniziale e l'evento finale dell'azione *READ*. Un'altro esempio riguarda la risposta $A$ del sistema ad ogni occorrenza dell'evento sporadico $E$, nell'ipotesi che le occorrenze successive di $E$ siano separate da un intervallo minimo $s$ e che $A$ debba terminare entro la scadenza (*deadline*) $d$:

$$\forall i @(E, i) + s \leq @(E, i + 1)$$
$$\forall i \exists j @(E, i) \leq @(\uparrow A, j) \wedge @(\downarrow A, j) \leq @(E, i) + d$$

### 3.6.8 Logiche modali e temporali

Le logiche modali arricchiscono il linguaggio della logica permettendo di esprimere aspetti del ragionamento che nel linguaggio naturale possono essere espressi dai modi grammaticali (indicativo, congiuntivo, condizionale). Per esempio, la frase "se piovesse non uscirei" suggerisce che non sta piovendo nel momento in cui viene pronunciata, *oltre* ad esprimere il fatto che la pioggia e l'azione di uscire sono legate da un'implicazione; questa sovrapposizione di significati non si ha nella frase "se piove non esco". In particolare, una logica modale può distinguere una verità necessaria (p.es., "tutti i corpi sono soggetti alla legge di gravità", nel mondo fisico in cui viviamo) da una verità contingente ("una particolare mela sta cadendo"). Le logiche temporali sono una classe di logiche modali, e servono ad esprimere il fatto che certe formule sono vere o false a seconda del periodo di tempo in cui si valutano.

Le logiche modali presuppongono che le loro formule vengano valutate rispetto a piú "mondi possibili" (universi di discorso), a differenza della logica non modale che si riferisce ad un solo mondo. Il fatto che una formula sia necessariamente vera corrisponde, per la semantica delle logiche modali, al fatto che la formula è vera in tutti i mondi considerati possibili in una particolare logica.

La sintassi delle logiche modali è quella dello logiche non modali, a cui si aggiungono *operatori modali*:

- operatore di necessità $\Box$;
- operatore di possibilità $\Diamond$;

L'insieme delle formule ben formate viene quindi esteso con questa regola:

- se $\mathcal{F}$ è una formula, allora sono formule anche $\Box\mathcal{F}$ e $\Diamond\mathcal{F}$.

La semantica di una logica modale è data da una *terna di Kripke* $\langle \mathbf{W}, \mathcal{R}, V \rangle$, dove:

- l'insieme $\mathbf{W}$ è l'insieme dei mondi (o interpretazioni);
- $\mathcal{R}$ è la *funzione di accessibilità* (o *visibilità*) : $\mathcal{R} : \mathbf{W} \to \mathbf{W}$;
- $V$ è la *funzione di valutazione* $V : \mathcal{W} \times \mathbf{W} \to \mathbb{B}$;

La funzione di accessibilità assegna una struttura all'insieme dei mondi possibili, determinando quali mondi sono considerati possibili a partire dal "mondo attuale": p.es., in una logica temporale, i mondi possibili in un dato istante sono quelli associati agli istanti successivi. Si suppone che la funzione di accessibilità sia riflessiva e transitiva.

La funzione di valutazione è tale che:

- $V(\Box\mathcal{F}, w) = \mathbf{T}$ se e solo se, per ogni $v \in \mathbf{W}$ tale che $\mathcal{R}(w, v)$, si ha $V(\mathcal{F}, v) = \mathbf{T}$;
- $V(\Diamond\mathcal{F}, w) = \mathbf{T}$ se e solo se esiste un $v \in \mathbf{W}$ tale che $\mathcal{R}(w, v)$ e $V(\mathcal{F}, v) = \mathbf{T}$;

Seguono alcuni assiomi per la logica modale:

$$\Box\neg\mathcal{P} \Leftrightarrow \neg\Diamond\mathcal{P} \tag{3.14}$$

$$\Box(\mathcal{P} \Rightarrow \mathcal{Q}) \Rightarrow (\Box\mathcal{P} \Rightarrow \Box\mathcal{Q}) \tag{3.15}$$

$$\Box\mathcal{P} \Rightarrow \mathcal{P} \tag{3.16}$$

I tre schemi di assiomi si possono leggere, rispettivamente, cosí:

- (3.14): $\mathcal{P}$ è necessariamente falso se e solo se $\mathcal{P}$ non può essere vero.
- (3.15): se $\mathcal{P}$ implica necessariamente $\mathcal{Q}$, allora la necessità di $\mathcal{P}$ implica la necessità di $\mathcal{Q}$.
- (3.15): la necessità di $\mathcal{P}$ implica $\mathcal{P}$.

### Logica temporale

Nella logica temporale, i mondi possibili rappresentano *stati* del sistema su cui vogliamo ragionare, corrispondenti a diversi istanti del tempo. La relazione di accessibilità descrive la struttura del tempo: per esempio, se ammettiamo che ad uno stato possano seguire due o piú stati alternativi, abbiamo un tempo ramificato. Si possono quindi considerare dei tempi lineari, o ramificati, o circolari (sistemi periodici), o ancora piú complessi. Inoltre il tempo può essere continuo o discreto. Nel seguito ci riferiremo ad un tempo lineare e discreto. L'insieme dei mondi possibili è quindi una successione di stati $s_i$, ove la relazione di accessibilità è tale che $\mathcal{R}(s_i, s_j)$ se e solo se $i \leq j$.

Nelle logiche temporali i due operatori modali principali assumono questi significati:

- $\Box$ *henceforth*, d'ora in poi;
- $\Diamond$ *eventually*, prima o poi;

I due operatori sono legati dalle relazioni

$$\Box\mathcal{F} \Leftrightarrow \neg\Diamond\neg\mathcal{F}$$
$$\Diamond\mathcal{F} \Leftrightarrow \neg\Box\neg\mathcal{F}$$

Da questi operatori se ne possono definire altri:

- $\circ$ *next*, prossimo istante;
- **U** *until*, finché ($p\mathbf{U}q$ se e solo se esiste $k$ tale che $q$ è vero in $s_k$ e $p$ è vero per tutti gli $s_i$ con $i \leq k$);
- $\blacksquare$ *always in the past*, sempre in passato;
- $\blacklozenge$ *sometime in the past*, qualche volta in passato;

Nell'esempio seguente usiamo una logica temporale per modellare un semplice sistema a scambio di messaggi (potrebbe far parte della specifica di un protocollo di comunicazione):

$$\Box(\text{send(m)} \Rightarrow \text{state} = \textbf{connected})$$
$$\Box(\text{send\_ack(m)} \Rightarrow \blacklozenge\ \text{receive(m)})$$
$$\Box(\text{B.receive(m)} \Rightarrow \blacklozenge\ \text{A.send(m)})$$

$$\Box(\text{A.send(m)} \Rightarrow \Diamond\text{B.receive(m)})$$
$$\Box(\Box\Diamond\text{A.send(m)} \Rightarrow \Diamond\text{B.receive(m)})$$
$$\Box(\text{B.receive(m)} \Rightarrow \Diamond\text{B.send\_ack(m)})$$

Le prime tre formule esprimono proprietà di *sicurezza* (non succede niente di brutto):

- se viene spedito un messaggio $m$, la connessione è attiva;
- se viene mandato un acknowledgement per un messaggio, il messaggio è già stato ricevuto;
- se il processo $B$ riceve un messaggio, il messaggio è già stato spedito;

Le formule successive esprimono proprietà di *vitalità* (prima o poi succede qualcosa di buono):

- se viene spedito un messaggio $m$, prima o poi viene ricevuto;
- se un messaggio $m$ viene spedito piú volte, prima o poi viene ricevuto;
- se il processo $B$ riceve un messaggio, prima o poi restituisce un acknowledgement.

## 3.7 Linguaggi orientati agli oggetti

In questa sezione vengono esposti i concetti fondamentali dell'analisi e specifica orientata agli oggetti, facendo riferimento ad un particolare linguaggio di specifica, l'UML (*Unified Modeling Language*).

Nell'analisi orientata agli oggetti, la descrizione un sistema parte dall'identificazione di *oggetti* (cioè elementi costitutivi) e di *relazioni* fra oggetti. Questo è molto simile a quanto abbiamo visto nel modello entità-relazioni, ma nel modello orientato agli oggetti vengono rappresentate esplicitamente le *operazioni* che si possono eseguire sugli oggetti, e che ne definiscono il comportamento.

Un oggetto, quindi, viene descritto sia dai propri attributi, cioè da un'insieme di proprietà che lo caratterizzano, che dalle operazioni che l'oggetto può compiere interagendo con altri oggetti; queste operazioni possono avere dei parametri di ingresso forniti dall'oggetto che *invoca* un'operazione e restituire risultati in funzione dei parametri d'ingresso e degli attributi dell'oggetto

che *esegue* l'operazione. Gli attributi, oltre a rappresentare proprietà degli oggetti, possono rappresentarne lo *stato*, e le operazioni possono modificare lo stato. Per chiarire la differenza fra *proprietà* e *stato*, possiamo pensare di rappresentare un'automobile con due attributi, vel_max (velocità massima) e marcia: il primo rappresenta una proprietà statica, mentre il secondo rappresenta i diversi stati di funzionamento della trasmissione (potrebbe assumere i valori -1, 0, 1, 2... per "retromarcia", "folle", "prima"...). Possiamo completare la descrizione dell'automobile con le operazioni imposta_vel_max(), marcia_alta() e marcia_bassa(): la prima permette di modificare una proprietà (per esempio dopo una modifica meccanica), le altre a cambiare lo stato di funzionamento. Quindi il modello orientato agli oggetti unifica i tre punti di vista dei linguaggi di specifica: la descrizione dei dati, delle funzioni, e del controllo.

Generalmente, in un sistema esistono piú oggetti simili, cioè delle entità distinte che hanno gli stessi attributi (con valori eventualmente, ma non necessariamente, diversi) e le stesse operazioni. Una *classe* è una descrizione della struttura e del comportamento comuni a piú oggetti. Una specifica orientata agli oggetti consiste essenzialmente in un insieme di classi e di relazioni fra classi.

Un'altra caratteristica del modello orientato agli oggetti è il ruolo che ha in esso il concetto di *generalizzazione*. Questo concetto permette di rappresentare il fatto che alcune classi hanno in comune una parte della loro struttura o del loro comportamento.

I concetti fondamentali del modello orientato agli oggetti si possono cosí riassumere:

**oggetti:** Un oggetto corrisponde ad un'entità individuale del sistema che vogliamo modellare. Gli oggetti hanno *attributi*, che ne definiscono le proprietà o lo stato, e *operazioni*, che ne definiscono il comportamento. Inoltre, ogni oggetto ha una *identità* che permette di distinguerlo dagli altri oggetti.

**legami:** Oggetti diversi possono essere in qualche relazione fra loro: tali relazioni fra oggetti sono dette *legami* (*links*).

**classi:** Gli oggetti che hanno la stessa struttura e comportamento sono raggruppati in classi, ed ogni oggetto è un'*istanza* di qualche classe. Una classe quindi descrive un insieme di oggetti che hanno la stessa struttura (cioè lo stesso insieme di attributi) e lo stesso comportamento, ma sono distinguibili l'uno dall'altro, e in generale, ma non necessariamente, hanno diversi valori degli attributi.

**associazioni:** Un'associazione sta ad un legame come una classe sta ad un oggetto: un'associazione è un insieme di link simili per struttura e significato.

**altre relazioni:** Si possono rappresentare altri tipi di relazioni, fra cui la generalizzazione. Osserviamo che questa è una relazione fra classi (corrispondente all'inclusione nella teoria degli insiemi), non fra oggetti.

### 3.7.1 L'UML

Lo UML [12] è una notazione grafica orientata agli oggetti, largamente diffusa nell'industria. Questa notazione è applicabile dalla fase di analisi e specifica dei requisiti alla fase di codifica, anche se non vincola quest'ultima fase all'uso di linguaggi orientati agli oggetti: un sistema progettato in UML può essere implementato con linguaggi non orientati agli oggetti, sebbene questi ultimi, ovviamente, non siano la scelta piú naturale. Per il momento studieremo alcuni aspetti dell'UML relativi alla fase di analisi e specifica.

Il linguaggio UML, la cui standardizzazione è curata dall'OMG (Object Management Group)[13], è passato attraverso una serie di revisioni. La versione attuale è la 2.1.2. In queste dispense si vuol dare soltanto un'introduzione ai concetti fondamentali di questo linguaggio, rinunciando sia alla completezza che al rigore, per cui la maggior parte delle nozioni qui esposte è applicabile a qualsiasi versione dell'UML. Ove sia necessario mettere in evidenza qualche differenza di concetti, di notazioni o di terminologia, si userà il termine "UML2" per la versione attuale, e "UML1" per le versioni precedenti.

In UML, un sistema viene descritto da varî punti di vista, o *viste* (*views*), e le rappresentazioni corrispondenti a tali punti di vista sono realizzate da diversi tipi di diagrammi, ciascuno rivolto a determinati aspetti del sistema. Ogni diagramma è formato da *elementi di modello*, o meglio dalle loro rappresentazioni grafiche: un elemento di modello è un'insieme di informazioni che descrive una particolare caratteristica del sistema, e queste informazioni vengono rappresentate graficamente nei diagrammi. Gli elementi di modello costituiscono il "vocabolario" di cui si serve lo sviluppatore per definire un modello.

Per il momento ci occuperemo dei punti di vista piú rilevanti nella fase di analisi: il punto di vista dei *casi d'uso* (*use cases*), il punto di vista *statico*,

[13] v. `www.omg.org`

e quello *dinamico*. Il punto di vista fondamentale è quello statico[14], cioè la descrizione delle classi e delle relazioni nel sistema modellato.

La notazione UML permette di rappresentare gli elementi di modello a varî livelli di dettaglio, in modo che si possa specificare un sistema con diversi gradi di astrazione o di approssimazione. Molti elementi hanno una forma minima (per esempio, un semplice simbolo) e delle forme estese, piú ricche di informazioni. Per esempio, una classe può essere rappresentata da un rettangolo contenente solo il nome della classe, oppure da un rettangolo diviso in tre scompartimenti, contenenti nome, attributi ed operazioni, che a loro volta possono essere specificati in modo piú o meno dettagliato.

In UML esistono tre meccanismi di estensione che permettono di adattare il linguaggio a esigenze specifiche: *vincoli*, *stereotipi* e *valori etichettati*.

I *vincoli* sono annotazioni che descrivono determinate condizioni imposte al sistema specificato, come, per esempio, l'insieme di valori ammissibili per un attributo, o il fatto che certe relazioni fra oggetti siano mutuamente esclusive. I vincoli possono venire espressi in linguaggio naturale, oppure nell'*Object Constraint Language* (OCL) dell'UML, o in qualsiasi linguaggio appropriato. La sintassi UML richiede che i vincoli vengano scritti fra parentesi graffe.

Gli *stereotipi* sono dei nuovi elementi di modello ottenuti specializzando elementi già esistenti. Per esempio, se in una certa applicazione è frequente l'uso di classi che rappresentano dei controllori di dispositivi, si può creare lo stereotipo ≪controller≫ per identificare le classi usate in tale modo; il nome di uno stereotipo viene sempre scritto fra i caratteri ≪ e ≫, e può essere accompagnato o sostituito da un'icona.

I *valori etichettati* (*tagged values*) sono delle proprietà, espresse da un nome e un valore, che si possono associare ad elementi di modello. Per esempio, ad un elemento si può associare la proprietà date il cui valore è la data piú recente in cui quell'elemento è stato modificato: {date = 1 apr 1955}. Se una proprietà ha un valore logico, si scrive solo il nome della proprietà nei casi in cui è vera, altrimenti si omette: {approvato}. Osserviamo che un valore etichettato è una proprietà dell'elemento di modello, non dell'entità reale modellata. Anche i valori etichettati si scrivono fra parentesi graffe.

Infine, i diagrammi possono contenere delle *note*, cioè dei commenti, che si presentano come una raffigurazione stilizzata di un biglietto con un "orec-

[14] In alcune metodologie, però, si dà un ruolo centrale al punto di vista dei casi d'uso.

chio" ripiegato, e possono essere collegate ad un elemento di modello con una linea tratteggiata.

I meccanismi di estensione si possono applicare a qualsiasi elemento di modello, per cui la possibilità della loro presenza generalmente verrà sottintesa nelle descrizioni dei vari tipi di elementi.

Come si è detto piú sopra, i meccanismi di estensione permettono di adattare il linguaggio UML alle esigenze di particolari campi di applicazione o processi di sviluppo. Un *profilo* è un insieme coerente e documentato di estensioni. Numerosi profili sono stati standardizzati dall'OMG (Object Management Group).

### 3.7.2 Diagrammi dei casi d'uso

Un diagramma dei casi d'uso schematizza il comportamento del sistema dal punto di vista degli utenti, o, piú in generale, di altri sistemi che interagiscono col sistema specificato. Un *attore* rappresenta un'entità esterna al sistema, un *caso d'uso* rappresenta un servizio offerto dal sistema. Ciascun servizio viene espletato attraverso sequenze di messaggi scambiati fra gli attori ed il sistema.

Attori e casi d'uso sono legati da associazioni che rappresentano comunicazioni. I casi d'uso *non rappresentano sottosistemi*, per cui non possono interagire, cioè scambiarsi messaggi, fra di loro, ma possono essere legati da relazioni di *inclusione*, *estensione* e *generalizzazione*.

Il comportamento richiesto al sistema per fornire il servizio rappresentato da un caso d'uso può essere specificato in vari modi, per esempio per mezzo di macchine a stati o di descrizioni testuali. In fase di analisi può essere sufficiente una descrizione in linguaggio naturale, oppure una descrizione piú formale delle di sequenze di interazioni (*scenari*) fra attori e sistema previste per lo svolgimento del servizio.

La Fig. 3.17 mostra un semplice diagramma di casi d'uso, relativo a un sistema di pagamento POS (*Point Of Sale*). Gli attori sono il cassiere e il cliente, i servizi forniti dal sistema sono il pagamento, il rimborso e il login; quest'ultimo coinvolge solo il cassiere. Il servizio di pagamento ha due possibili estensioni, cioè comportamenti aggiuntivi rispetto a quello del caso d'uso fondamentale.

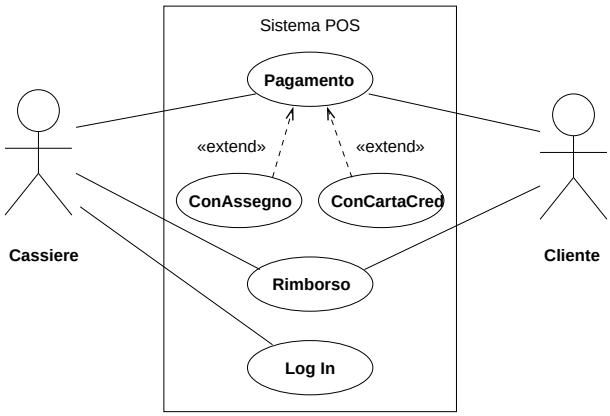


Figura 3.17: Una diagramma di casi d'uso.

### 3.7.3 Classi e oggetti

Come abbiamo visto, una classe rappresenta astrattamente un insieme di oggetti che hanno gli stessi attributi, operazioni, relazioni e comportamento. In un modello di analisi una classe si usa per rappresentare un concetto pertinente al sistema che viene specificato, concetto che può essere concreto (per esempio, un controllore di dispositivi) o astratto (per esempio, una transazione finanziaria). In un modello di progetto una classe rappresenta un'entità software introdotta per implementare l'applicazione.

Una classe viene rappresentata graficamente da un rettangolo contenente il nome della classe e, opzionalmente, l'elenco degli attributi e delle operazioni.

Se la classe ha uno stereotipo, il nome dello stereotipo viene scritto sopra al nome della classe, oppure si mette l'icona dello stereotipo nell'angolo destro in alto del rettangolo.

La rappresentazione minima di una classe consiste in un rettangolo contenente solo il nome ed eventualmente lo stereotipo. Se la classe è uno stereotipo e questo è rappresentabile da un'icona, la rappresentazione minima consiste nell'icona e nel nome.

#### Attributi

Ogni attributo ha un *nome*, che è l'unica informazione obbligatoria. Le altre informazioni associate agli attributi sono:

**tipo:** può essere uno dei tipi fondamentali predefiniti dall'UML (corrispondenti a quelli usati comunemente nei linguaggi di programmazione), un tipo definito in un linguaggio di programmazione, o una classe definita (in UML) dallo sviluppatore;

**visibilità:** *privata*, *protetta*, *pubblica*, *package*; quest'ultimo livello di visibilità significa che l'attributo è visibile da tutte le classi apparteneti allo stesso package (Sez. 4.4.1);

***scope*:** *istanza*, se l'attributo appartiene al singolo oggetto, *statico* se appartiene alla classe, cioè è condiviso fra tutte le istanze della classe, analogamente ai campi statici del C++ o del Java;

**molteplicità:** índica se l'attributo può essere replicato, cioè avere piú valori (si può usare per rappresentare un array);

**valore iniziale:** valore assegnato all'attributo quando viene creato l'oggetto.

La sintassi UML per gli attributi è la seguente:

```
<visibilita‘> <nome> <molteplicita‘> : <tipo> = <val-iniziale>
```

Se un attributo ha scope statico, viene sottolineato.

La visibilità si rappresenta con i seguenti simboli:

| | |
|---|---|
| `+` | pubblica |
| `#` | protetta |
| `~` | package |
| `-` | privata |

La molteplicità si índica con un numero o un intervallo numerico fra parentesi quadre, come nei seguenti esempi:

| | |
|---|---|
| `[3]` | tre valori |
| `[1..4]` | da uno a quattro valori |
| `[1..*]` | uno o piú valori |
| `[0..1]` | zero o un valore (attributo opzionale) |

### Operazioni

Ogni operazione viene identificata da una *segnatura* (*signature*) costituita dal nome della funzione e dalla *lista dei parametri*, eventualmente vuota.

Per ciascun parametro si specifica il nome e, opzionalmente, le seguenti informazioni:

**direzione:** *ingresso* (in), *uscita* (out), *ingresso e uscita* (inout);
**tipo;**
**valore default:** valore passato al metodo che implementa la funzione, se l'argomento corrispondente al parametro non viene specificato.

Inoltre le operazioni, analogamente agli attributi, hanno scope, visibilità e tipo; quest'ultimo è il tipo dell'eventuale valore restituito. Anche queste informazioni sono opzionali.

La sintassi per le operazioni è la seguente:

```
<visibilita‘> <nome> (<lista-parametri>) : <tipo>
```

dove ciascun parametro della lista ha questa forma:

```
<direzione> <nome> : <tipo> = <val-default>
```

Solo il nome del parametro è obbligatorio, ed i parametri sono separati da virgole.

È utile osservare la distinzione fra *operazione* e *metodo*. Un'operazione è la specifica di un comportamento, specifica che si può ridurre alla semplice descrizione dei parametri o includere vincoli (per esempio, precondizioni, invarianti e postcondizioni) e varie annotazioni. Un metodo è l'implementazione di un'operazione. L'UML non ha elementi di modello destinati a descrivere i metodi, che comunque non interessano in fase di specifica. Se bisogna descrivere l'implementazione di un'operazione, per esempio nella fase di codifica, si possono usare delle note associate all'operazione.

### Oggetti

Un oggetto viene rappresentato da un rettangolo contenente i nomi dell'oggetto e della classe d'appartenenza, sottolineati e separati dal carattere ':', ed opzionalmente gli attributi con i rispettivi valori. Il nome della classe o quello dell'oggetto possono mancare. In questo caso, il nome della classe viene preceduto dal carattere ':'. È possibile esprimere uno stereotipo come nella rappresentazione delle classi.

La forma minima di un oggetto consiste in un rettangolo col nome dell'oggetto e/o della classe e l'eventuale stereotipo, oppure, se è il caso, nell'icona dello stereotipo col nome dell'oggetto.

Nell'UML la rappresentazione esplicita di oggetti è piuttosto rara, in quanto il lavoro di analisi, specifica e progettazione si basa essenzialmente sulle classi e le loro relazioni, però gli oggetti possono essere usati per esemplificare delle situazioni particolari o tipiche. La figura 3.18 mostra la rappresentazione di una classe e di una sua istanza.

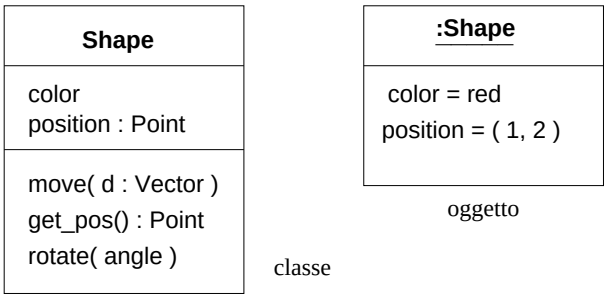


Figura 3.18: Una classe ed un oggetto

### 3.7.4 Associazioni e link

Un'associazione (o un link), nella forma piú semplice, si rappresenta come una linea fra le classi (o oggetti) coinvolte, etichettata col nome della relazione (o link). Se un'associazione coinvolge piú di due classi, si rappresenta con una losanga unita da linee alle classi coinvolte (analogamente per i link).

Alcuni semplici modi di rappresentare le associazioni sono mostrati in Fig. 3.19: la classe **User** rapresenta gli utenti di un sistema di calcolo, ciascuno contraddistinto da un numero identificatore (uid), e la classe **Account** rappresenta i relativi account.

La *molteplicità* di una classe in un'associazione, cioè il numero di istanze di una classe che possono essere in relazione con istanze di un'altra classe, può essere indicata con intervalli numerici alle estremità delle linee che rappresentano la relazione: per esempio, `1`, `0..1` (zero o uno), `1..*` (uno o piú), `0..*` (zero o piú), `*` (equivalente a `0..*`).

Alle estremità di una linea si possono specificare anche i rispettivi *ruoli* delle classi coinvolte nell'associazione. Un ruolo è un nome che serve a specificare la relazione fra le istanze della classe etichettata dal ruolo e le istanze

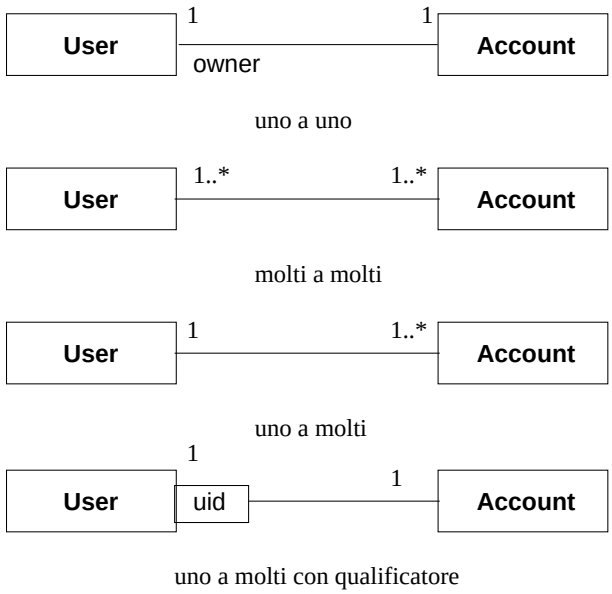


Figura 3.19: Associazioni

della classe associata. Per esempio, in Fig. 3.25 c'è un'associazione fra la classe **Docente** e la classe **Dipartimento**. All'estremità di questa associazione corrispondente alla classe *Docente* si trova il nome di ruolo presidente: questo significa che un docente può assumere il ruolo di presidente di un dipartimento. Invece di indicare il ruolo, in questo caso si potrebbe dare un nome all'associazione, per esempio presiede. Questa scelta non comporterebbe ambiguità nell'interpretazione dell'associazione, essendo abbastanza ovvio che è un docente a presiedere un dipartimento, e non viceversa, ma in generale l'uso dei ruoli è piú chiaro. In ogni modo è possibile usare insieme il nome dell'associazione e i nomi dei ruoli. L'uso dei ruoli è inoltre utile quando si hanno associazioni riflessive, come in Fig. 3.20.

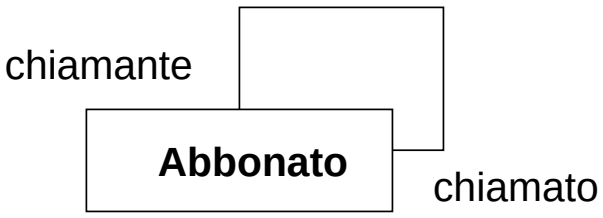


Figura 3.20: Ruoli in un'associazione riflessiva.

A un'estremità di un'associazione può apparire anche un *qualificatore*, rappresentato da un rettangolo avente un lato combaciante con un lato della

classe. Il qualificatore è un attributo dell'associazione, che distingue i diversi oggetti di una classe che possono essere in relazione con oggetti dell'altra. Il qualificatore rappresenta cioè un'informazione che permette di individuare una particolare istanza, fra molte, di una classe associata. Nella Fig. 3.19, la penultima associazione mostra che un utente può avere piú account, e nell'associazione successiva questa molteplicità viene ridotta a uno, grazie al qualificatore uid che identifica i singoli account.

Un'associazione può avere degli attributi e delle operazioni: si parla in questo caso di *classe associazione*. Una classe associazione viene rappresentata collegando alla linea dell'associazione il simbolo della classe (in forma estesa o ridotta), mediante una linea tratteggiata.

### Aggregazione

La *aggregazione* è un'associazione che lega un'entità complessa (aggregato) alle proprie parti componenti. Nei diagrammi di classi e di oggetti, una relazione di aggregazione viene indicata da una piccola losanga all'estremità dell'associazione che si trova dalla parte della classe (o dell'oggetto) che rappresenta l'entità complessa.

La differenza fra un'associazione pura e semplice e un'aggregazione non è netta. L'aggregazione è un'annotazione aggiuntiva che esprime il concetto di "appartenenza", "contenimento", "ripartizione", o in generale di una forma di subordinazione strutturale non rigida. Per esempio, nella Fig. 3.21 la relazione fra una banca e i suoi clienti viene modellata da una semplice associazione, poiché i clienti non "fanno parte" della loro banca, mentre la relazione fra una squadra e i suoi giocatori si può modellare piú accuratamente con una aggregazione. In questo secondo caso, però, sarebbe stata accettabile anche una semplice associazione.

Un'istanza di una classe può appartenere a piú d'una aggregazione. La Fig. 3.22 mostra che la classe **Studente** partecipa, come componente, alle aggregazioni con le classi **Squadra** e **Coro**. Il diagramma di oggetti nella stessa figura mostra una possibile configurazione di istanze compatibile col diagramma delle classi: la studentessa anna appartiene a due istanze della classe **Squadra**, lo studente beppe appartiene ad un'istanza della classe **Squadra** e ad una della classe **Coro**, lo studente carlo appartiene a un'istanza della classe **Coro**.

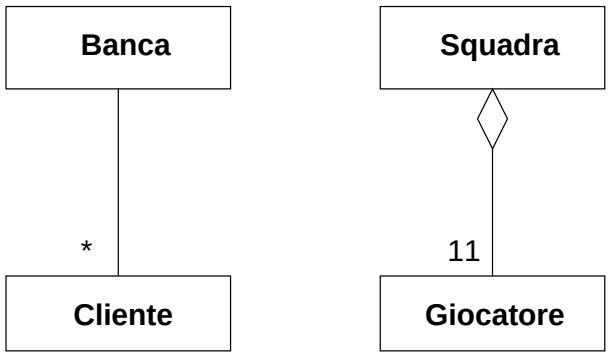


Figura 3.21: Associazioni e aggregazioni.

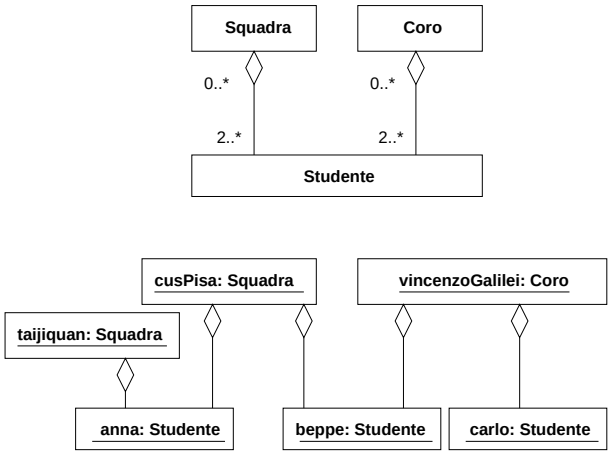


Figura 3.22: Aggregazioni multiple.

### 3.7.5 Composizione

La *composizione* modella una subordinazione strutturale rigida, tale cioè che l'aggregato abbia il completo e unico controllo delle parti componenti. Mentre nell'aggregazione le parti sono indipendenti dall'aggregato (esistono anche al di fuori dell'aggregazione), nella composizione esiste una dipendenza stretta fra composto e componenti. Spesso la composizione rappresenta una situazione in cui l'esistenza dei componenti coincide con quella del composto, per cui la creazione e la distruzione del composto implicano la creazione e la distruzione dei componenti. In ogni caso, un componente può appartenere ad un solo composto, e il composto è il "padrone" del componente.

La composizione si rappresenta con una losanga nera dalla parte dell'entità complessa, oppure si possono disegnare i componenti all'interno dell'en-

tità stessa. La Fig. 3.23 mostra queste due notazioni.

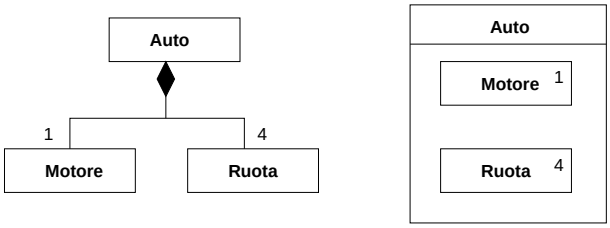


Figura 3.23: Composizione.

Una classe può appartenere come componente a piú di una composizione, ma un'istanza può appartenere ad una sola istanza. Nella Fig. 3.24 la classe **Motore** è in relazione di composizione con **Nave** e **Auto**, ma una qualsiasi sua istanza può essere componente di una sola istanza di una delle due classi.

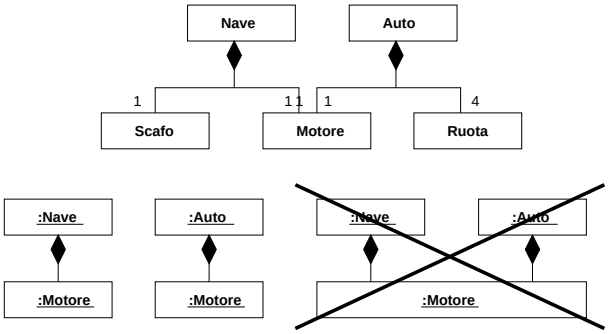


Figura 3.24: Composizione fra classi e fra oggetti.

Infine, la Fig. 3.25 mostra un semplice diagramma delle classi con associazioni, aggregazioni e composizioni.

### 3.7.6 Generalizzazione

Una classe (che chiameremo *classe base* o *superclasse*) *generalizza* altre classi (che chiameremo *classi derivate* o *sottoclassi*) quando ne riassume alcune caratteristiche comuni (attributi, operazioni, associazioni, vincoli...). Le caratteristiche di tale classe, cioè, sono tali da definire un insieme di oggetti che include l'unione degli insiemi di oggetti definiti dalle sottoclassi. Un oggetto appartenente ad una classe derivata, quindi, appartiene anche alla classe

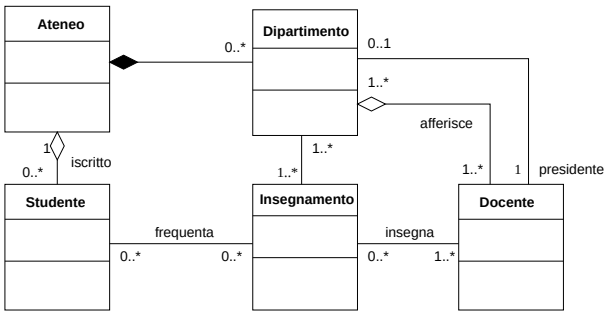


Figura 3.25: Un diagramma delle classi.

base. È possibile che ogni oggetto appartenente alla classe base appartenga ad almeno una delle classi derivate, oppure che alcuni oggetti appartengano solo alla classe base. Osserviamo anche che si parla di generalizzazione anche quando una classe base ha una sola classe derivata.

La relazione di generalizzazione si può anche chiamare *specializzazione*, cambiando il punto di vista ma restando immutato il significato: una classe è una specializzazione di un'altra classe se aggiunge delle caratteristiche alla sua struttura o al suo comportamento (è quindi un'*estensione*) oppure vi aggiunge dei vincoli (si ha quindi una *restrizione*). Per esempio, data una classe "ellisse" con gli attributi "asse maggiore" e "asse minore", la classe "ellisse colorata" è un'estensione, poiché aggiunge l'attributo "colore", mentre la classe "cerchio" è una restrizione, poiché pone il vincolo che gli attributi "asse maggiore" e "asse minore" abbiano lo stesso valore. Osserviamo, a scanso di equivoci, che anche un'estensione è un sottoinsieme della superclasse.

Il fatto che le istanze di una classe derivata siano un sottoinsieme delle istanze della classe base si esprime anche col *principio di sostituzione* (di Liskov): un'istanza della classe derivata può sostituire un'istanza della classe base. Questo principio è un utile criterio per valutare se una certa classe può essere descritta come una specializzazione di un'altra. È particolarmente importante verificare l'applicabilità del principio di Liskov nel caso in cui la classe derivata venga ottenuta per restrizione dalla classe base, cioè aggiungendo dei vincoli. Se, per esempio, supponiamo che la classe "ellisse" abbia due operazioni che permettono di modificare gli assi separatamente, allora la classe "cerchio" deve implementare queste operazioni in modo che si rispetti il vincolo di uguaglianza fra i due assi: questo fa sí che le operazioni caratteristiche di "ellisse" siano applicabili a "cerchio".

Poiché una sottoclasse ha le stesse caratteristiche della superclasse, si dice che tali caratteristiche vengono *ereditate*. In particolare, vengono ereditate

le operazioni, di cui si può effettuare anche una *ridefinizione* (*overriding*) nella classe derivata. In una classe derivata, cioè, si può avere un'operazione che ha la stessa *segnatura* (nome dell'operazione, tipo del valore restituito, numero, tipo e ordine degli argomenti) di un'operazione della classe base, ma una diversa implementazione. Naturalmente la ridefinizione è utile se la nuova implementazione è compatibile col significato originario dell'operazione, altrimenti non varrebbe il principio di sostituzione. Per esempio, consideriamo una classe **Studente** con l'operazione iscrivi(), che rappresenta l'iscrizione dello studente secondo la procedura normale. Se una categoria di studenti, per esempio quelli già in possesso di una laurea, richiede una procedura diversa, si può definire una classe **StudenteLaureato**, derivata da **Studente**, che ridefinisce opportunamente l'operazione iscrivi().

Una classe derivata può essere ulteriormente specializzata in una o piú classi, e una classe base può essere ulteriormente generalizzata (finché non si arriva alla classe universale che comprende tutti i possibili oggetti). Quando si hanno delle catene di generalizzazioni, può essere utile distinguere le classi basi o derivate *dirette* da quelle *indirette*, a seconda che si ottengano "in un solo passo" o no, a partire da un'altra classe. Le classi base e derivate indirette si chiamano anche, rispettivamente, *antenati* e *discendenti*.

La generalizzazione si rappresenta con una freccia terminante in un triangolo vuoto col vertice che tocca la superclasse.

### Classi astratte e concrete

Una classe è *concreta* se esistono degli oggetti che siano istanze *dirette* di tale classe, cioè appartengano ad essa e non a una classe derivata. Se consideriamo due o piú classi concrete che hanno alcune caratteristiche in comune, possiamo generalizzarle definendo una classe base che riassume queste caratteristiche. Può quindi accadere che non possano esistere delle istanze dirette di questa classe base, che si dice allora *astratta*.

Per esempio, consideriamo delle classi come **Uomo**, **Lupo**, **Megattera** eccetera[15]. Gli animali appartenenti a queste classi hanno delle caratteristiche in comune (per esempio, sono omeotermi e vivipari), che si possono riassumere nella definizione di una classe base *Mammifero*. Questa classe è astratta perché non esiste un animale che sia un mammifero senza appartenere anche a una delle classi derivate.

[15]Ovviamente stiamo usando il termine "classe" in modo diverso da come viene usato nelle scienze naturali.

In UML l'astrattezza di una classe (e di altri elementi di modello) si rappresenta con la proprietà abstract. Per convenzione, nei diagrammi i nomi delle entità astratte si scrivono in corsivo; se non è pratico scrivere in corsivo (per esempio in un diagramma fatto a mano) si può scrivere la parola abstract fra parentesi graffe sotto al nome della classe.

### Eredità multipla

Si parla di *eredità multipla* quando una classe derivata è un sottoinsieme di due o piú classi che non sono in relazione di generalizzazione/specializzazione fra di loro. In questo caso, le istanze della classe derivata ereditano le caratteristiche di tutte le classi base.

### Aggregazione ricorsiva

L'uso combinato della generalizzazione e dell'aggregazione permette di definire strutture ricorsive[16]. Nella Fig. 3.26 il diagramma delle classi specifica la struttura delle operazioni aritmetiche, e il diagramma degli oggetti ne mostra una possibile realizzazione. Si osservi che, mentre la specifica è ricorsiva, la realizzazione è necessariamente gerarchica.

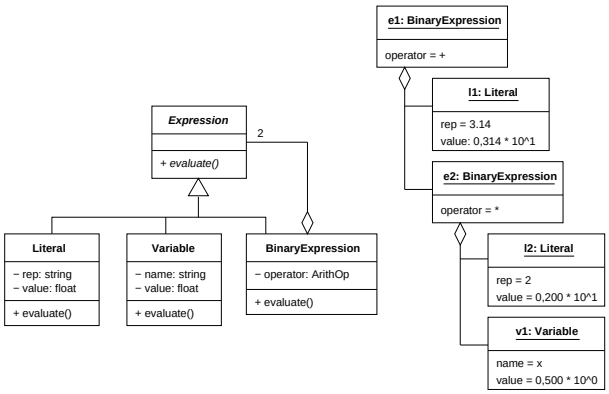


Figura 3.26: Aggregazione ricorsiva fra classi.

[16]Osserviamo che si parla di strutture *ricorsive*, non *cicliche*. L'aggregazione è una relazione gerarchica, cioè priva di cicli.

### Insiemi di generalizzazioni

Nei modelli di analisi la relazione di generalizzazione viene usata spesso per classificare le entità del dominio analizzato, mettendone in evidenza le reciproche affinità e differenze. Per esempio, può essere utile classificare i prodotti di un'azienda, i suoi clienti o i suoi dipendenti.

Per descrivere precisamente una classificazione, l'UML mette a disposizione il concetto di *insieme di generalizzazioni*. Informalmente, un insieme di generalizzazioni è un modo di raggruppare le sottoclassi di una classe base, cioè un insieme di sottoinsiemi, a cui si può dare un nome che descriva il criterio con cui si raggruppano le sottoclassi. Un insieme di generalizzazioni è *completo* se ogni istanza della classe base appartiene ad almeno una delle sottoclassi, e *disgiunto* se le sottoclassi sono disgiunte (l'intersezione di ciascuna coppia di sottoclassi è vuota). Per default, un insieme di generalizzazioni è incompleto e disgiunto.

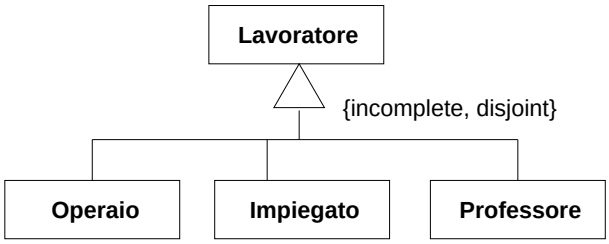


Figura 3.27: Insieme di generalizzazioni incompleto e disgiunto.

Nell'esempio di Fig. 3.27 si suppone che una persona possa fare un solo lavoro; l'insieme di generalizzazioni è quindi disgiunto, ed è incompleto perché evidentemente esistono molte altre categorie di lavoratori. In Fig. 3.28, invece, si suppone che una persona possa praticare piú di uno sport, quindi l'insieme di generalizzazioni è overlapping (non disgiunto).

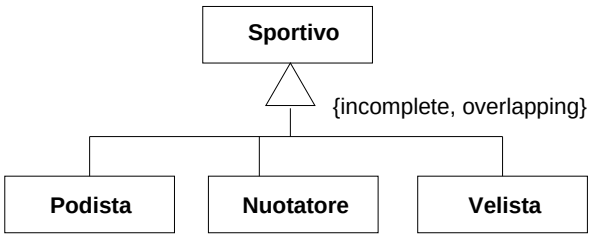


Figura 3.28: Insieme di generalizzazioni incompleto e non disgiunto.

Nell'esempio di Fig. 3.29 l'insieme dei dipendenti di un'azienda viene

classificato secondo due criteri ortogonali, retribuzione e mansione. Si hanno quindi due insiemi di generalizzazioni, ognuno completo e disgiunto. Ogni dipendente è un'istanza sia di una classe di retribuzione (**Classe1** eccetera) che di una classe di mansioni (**Tecnico** o **Amministrativo**). Poiché nessun dipendente può appartenere *solo* a una classe di retribuzione o a una classe di mansioni, tutte le classi della Fig. 3.29 (a) sono astratte. Nella Fig. 3.29 (b) si mostra una classe concreta costruita per eredità multipla da una classe di retribuzione e una classe di mansioni.

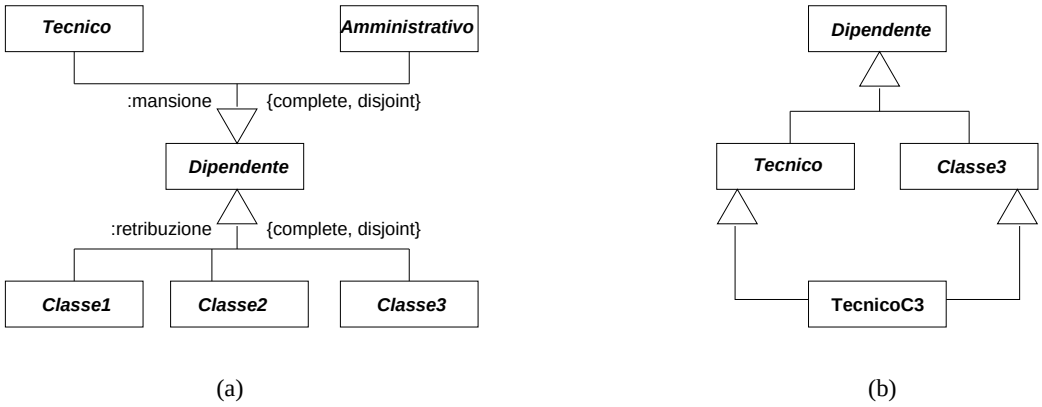


Figura 3.29: Due insiemi di generalizzazioni.

Nelle versioni dell'UML precedenti alla 2.0 c'era il concetto di *discriminante*, che aveva un significato simile a quello degli insiemi di generalizzazioni, anche se era definito in modo diverso.

**Osservazione.** L'esempio mostrato in Fig. 3.29 serve a chiarire il significato degli insiemi di generalizzazioni, ma non è necessariamente il modo migliore per modellare la situazione presa ad esempio. Si può osservare, infatti, che un dipendente rappresentato da un'istanza di **TecnicoC3**, secondo questo modello, non può cambiare mansione né classe di retribuzione, poiché la relazione di generalizzazione è statica e fissa rigidamente l'appartenza delle istanze di **TecnicoC3** alle superclassi *Tecnico* e *Classe3*. Si avrebbe un modello piú realistico considerando la retribuzione e la mansione come concetti *associati* ai dipendenti, come mostra la Fig. 3.30. I link fra istanze possono essere creati e distrutti dinamicamente, per cui le associazioni permettono di costruire strutture logiche piú flessibili rispetto alla generalizzazione.

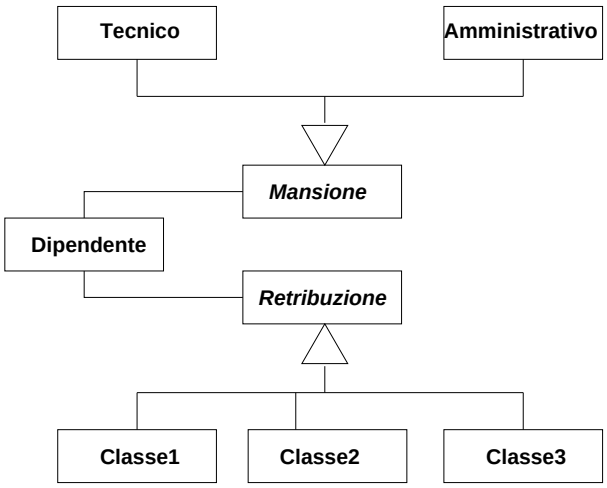


Figura 3.30: Associazioni invece di generalizzazioni.

### 3.7.7 Diagrammi di stato

Il *diagramma di stato* è uno dei diagrammi che fanno parte del modello dinamico di un sistema, e rappresenta una *macchina a stati* associata a una classe, a una *collaborazione* (interazione di un insieme di oggetti), o a un metodo. La macchina a stati descrive l'evoluzione temporale degli oggetti e la loro risposta agli stimoli esterni.

Le macchine a stati impiegate nell'UML sono basate sul formalismo degli automi a stati finiti visto in precedenza, che qui, però, viene esteso notevolmente in modo da renderlo piú espressivo e sintetico. Nell'UML viene usata la notazione degli *Statecharts*, che permette una descrizione strutturata degli stati di un automa. Questa notazione è ricca di concetti e di varianti grafiche, e ne saranno illustrati solo gli aspetti principali.

I concetti fondamentali rappresentati dal diagramma di stato sono:

**occorrenze:** avvenimenti associati a istanti nel tempo e (spesso implicitamente) a punti nello spazio. Per esempio, se in un dato momento viene premuto un tasto di un terminale, questa è un'occorrenza. Quando lo stesso tasto (o un altro) viene premuto, è un'altra occorrenza.

**eventi:** insiemi di occorrenze di uno stesso tipo. Per esempio, l'evento **TastoPremuto** potrebbe rappresentare l'insieme di tutte le possibili occorrenze della pressione di un tasto. Nel séguito, spesso si userà il termine "evento" al posto di "occorrenza".

**stati:** situazioni in cui un oggetto soddisfa certe condizioni, esegue delle attività, o semplicemente aspetta degli eventi.

**transizioni:** conseguentemente al verificarsi di un evento, un oggetto può modificare il proprio stato: una transizione è il passaggio da uno stato ad un altro, che si può modellare come se fosse istantaneo.

**azioni:** possono essere eseguite in corrispondenza di transizioni, e sono non interrompibili, per cui si possono modellare come se fosero istantanee.

**attività:** possono essere associate agli stati, possono essere interrotte dal verificarsi di eventi che causano l'uscita dallo stato, ed hanno una durata non nulla.

Un evento può essere una *chiamata di operazione*, un *cambiamento*, un *segnale*, o un *evento temporale*.

Un evento di chiamata avviene quando viene ricevuta una richiesta di esecuzione di un'operazione.

Un evento di cambiamento è il passaggio del valore di una condizione logica da falso a vero, e si rappresenta con la parola **when** seguita da un'espressione booleana. Per esempio, l'espressione **when** $(p > P)$ rappresenta l'evento "*p supera la soglia P*".

I segnali sono delle entità che gli oggetti si possono scambiare per comunicare fra di loro, e possono essere strutturati in una gerarchia di generalizzazione: per esempio, un ipotetico segnale **Input** può essere descritto come generalizzazione dei segnali **Mouse** e **Keyboard**, che a loro volta possono essere ulteriormente strutturati. Una gerarchia di segnali si rappresenta graficamente in modo simile ad una gerarchia di classi. Un segnale si rappresenta come un rettangolo contenente lo stereotipo ≪signal≫, il nome del segnale ed eventuali attributi.

Un evento temporale si verifica quando il tempo assume un particolare valore assoluto (per esempio, il 31 dicembre 1999), oppure quando è trascorso un certo periodo da un determinato istante (per esempio, dieci secondi dall'arrivo di un segnale). Gli eventi temporali del primo tipo si rappresentano con la parola **when** seguita da una condizione booleana (per esempio, **when** (date = 2002-12-31)), quelli del secondo tipo con la parola **after** seguita da un intervallo di tempo.

Se un evento si verifica nel corso di una transizione, non ha influenza sull'eventuale azione associata alla transizione (ricordiamo che le azioni non sono interrompibili) e viene messo in coda per essere considerato nello stato successivo.

Se, mentre un oggetto si trova in un certo stato, si verificano degli eventi che non innescano transizioni associate a quello stato, l'oggetto si può comportare in due modi: i) questi eventi vengono "consumati" e quindi non potranno piú influenzare l'oggetto (è come se non fossero mai accaduti), oppure ii) questi eventi vengono marcati come *differiti* (*deferred*) e memorizzati finché l'oggetto non entra in uno stato in cui tali eventi non sono piú marcati come differiti. In questo nuovo stato, gli eventi cosí memorizzati o innescano una transizione, o vengono perduti definitivamente. Gli eventi differiti in uno stato vengono dichiarati come tali nel simbolo dello stato, con la parola defer (o deferred, in UML1).

Gli stati si rappresentano come rettangoli ovalizzati contenenti opzionalmente il nome dello stato, un'eventuale attività (preceduta dalla parola do) ed altre informazioni che vedremo piú oltre. In particolare, uno stato può contenere dei sottostati. Uno stato può contenere la parola entry seguita dal nome di un'azione (separato da una barra). Questo significa che l'azione deve essere eseguita ogni volta che l'oggetto entra nello stato in questione. Analogamente, la parola exit etichetta un'azione da eseguire all'uscita dallo stato. Una coppia *evento/azione* entro uno stato significa che al verificarsi dell'evento viene eseguita l'azione corrispondente, e l'oggetto resta nello stato corrente (*transizione interna*). In questo caso non vengono eseguite le eventuali azioni di entry o di exit.

Quando bisogna indicare uno stato iniziale, si usa una freccia che parte da un cerchietto nero e raggiunge lo stato iniziale. Uno stato finale viene rappresentato da un cerchio contenente un cerchietto annerito (un "bersaglio"). Se un oggetto ha un comportamento ciclico, non viene indicato uno stato finale.

Le transizioni sono rappresentate da frecce fra gli stati corrispondenti, etichettate col nome dell'evento causante la transizione, con eventuali attributi dell'evento, con una condizione (*guardia*) necessaria per l'abilitazione della transizione (racchiusa fra parentesi quadre), e con un'azione da eseguire, separata dalle informazioni precedenti per mezzo di una barra obliqua. Ciascuna di queste tre informazioni è opzionale. Se manca l'indicazione dell'evento, la transizione avviene al termine dell'attività associata allo stato di partenza: si tratta di una transizione *di completamento*.

Un'azione può inviare dei segnali, e in questo caso si usa la parola **send** seguita dal nome e da eventuali parametri del segnale. L'invio di un segnale si può rappresentare anche graficamente, mediante una figura a forma di cartello indicatore, etichettata col nome e i parametri del segnale.

### Macchine a stati gerarchiche

La descrizione del modello dinamico generalmente è gerarchica, cioè articolata su diversi livelli di astrazione, in ciascuno dei quali alcuni elementi del livello superiore vengono raffinati ed analizzati.

Un'attività associata ad uno stato può quindi essere descritta a sua volta da una macchina a stati. Questa avrà uno stato iniziale ed uno o piú stati finali. Il sottodiagramma che descrive l'attività può sempre essere disegnato all'interno dello stato che la contiene. Se non ci sono transizioni che attraversano il confine del sottodiagramma, questo può essere disegnato separatamente.

In generale, qualsiasi stato (superstato) può essere decomposto in sottostati, che ereditano le transizioni che coinvolgono il superstato.

Consideriamo, per esempio, la macchina a stati associata all'interazione fra l'utente e un centralino (Fig. 3.31). Si suppone che l'utente possa comporre numeri di tre cifre, oppure premere un tasto che seleziona un numero memorizzato. Il diagramma di questa macchina a stati non sfrutta la possibilità di composizione gerarchica offerta dagli Statechart, per cui le transizioni causate dagli eventi riaggancio devono essere mostrate per ciascuno stato successivo a quello iniziale.

Il diagramma si semplifica se raggruppiamo questi stati in un superstato (**Attivo**) e ridisegnamo le transizioni come in Fig. 3.32. La transizione in ingresso al superstato **Attivo** porta la macchina nel sottostato iniziale (**Attesa1**) di quest'ultimo, mentre la transizione di completamento fra i due stati ad alto livello avviene quando la sottomacchina dello stato **Attivo** termina il proprio funzionamento. La transizione attivata dagli eventi riaggancio viene ereditata dai sottostati: questo significa che, in qualsiasi sottostato di **Attivo**, il riaggancio riporta la macchina nello stato **Inattivo**.

### Stati concorrenti

Uno stato può essere scomposto anche in *regioni concorrenti*, che descrivono attività concorrenti nell'àmbito dello stato che le contiene. Queste attività, a loro volta, sono descritte da macchine a stati. La Fig. 3.33 mostra il funzionamento di un termoventilatore, in cui il controllo della velocità e quello della temperatura sono indipendenti.

Gli stati concorrenti possono interagire attraverso *eventi condivisi*, *scam-*

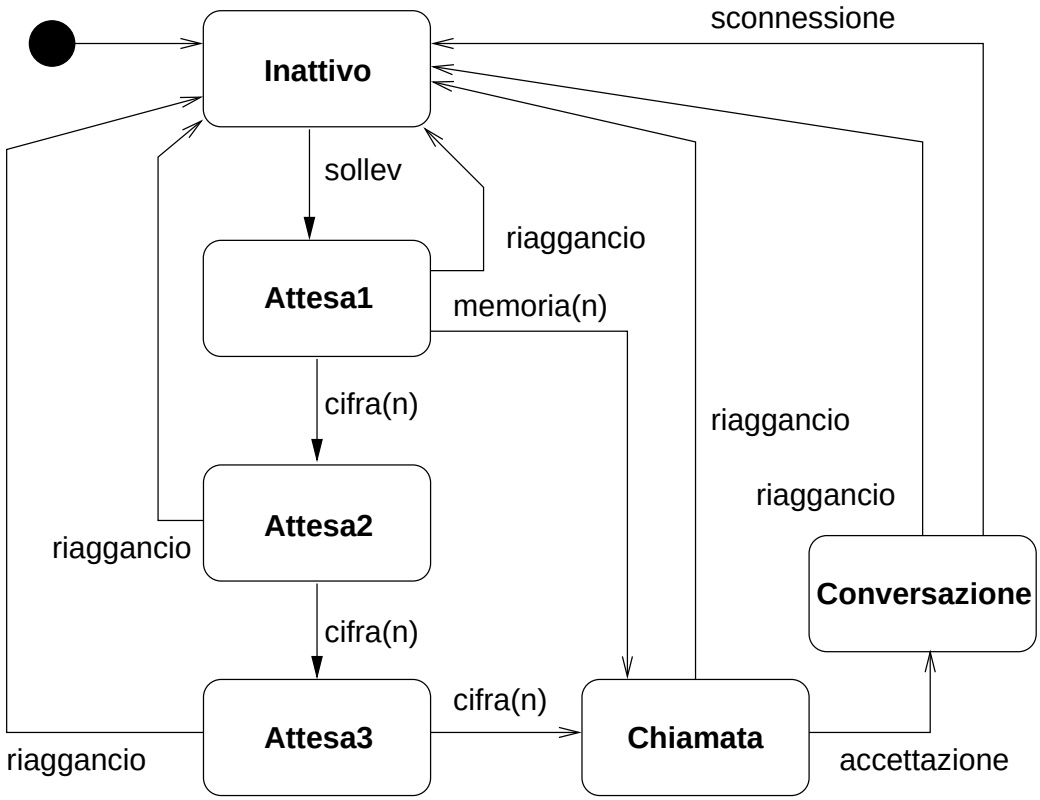


Figura 3.31: Una macchina a stati non gerarchica.

*bio di segnali*, *parametri* degli eventi o dei segnali, e *attributi* dell'oggetto a cui appartiene la macchina a stati. Le interazioni avvengono, oltre che con lo scambio di eventi, anche attraverso la valutazione delle espressioni che costituiscono le guardie e le azioni associate alle transizioni. L'esecuzione delle azioni può modificare gli attributi condivisi fra stati concorrenti, però è bene evitare, finché possibile, di modellare in questo modo l'interazione fra stati. Questo meccanismo di interazione, infatti, è poco strutturato e poco leggibile, e rende piú probabili gli errori nella specifica o nella realizzazione del sistema. La valutazione delle guardie, invece, non può avere effetti collaterali. In una guardia si può verificare se un oggetto si trova in un certo stato, usando l'operatore logico `oclInState` del linguaggio OCL.

È possibile descrivere attività concorrenti anche senza ricorrere alla scomposizione in regioni concorrenti, quando tali attività sono eseguite da oggetti diversi, a cui sono associate macchine a stati distinte. La Fig. 3.34 mostra il comportamento di un produttore e di un consumatore che si sincronizzano scambiandosi segnali.

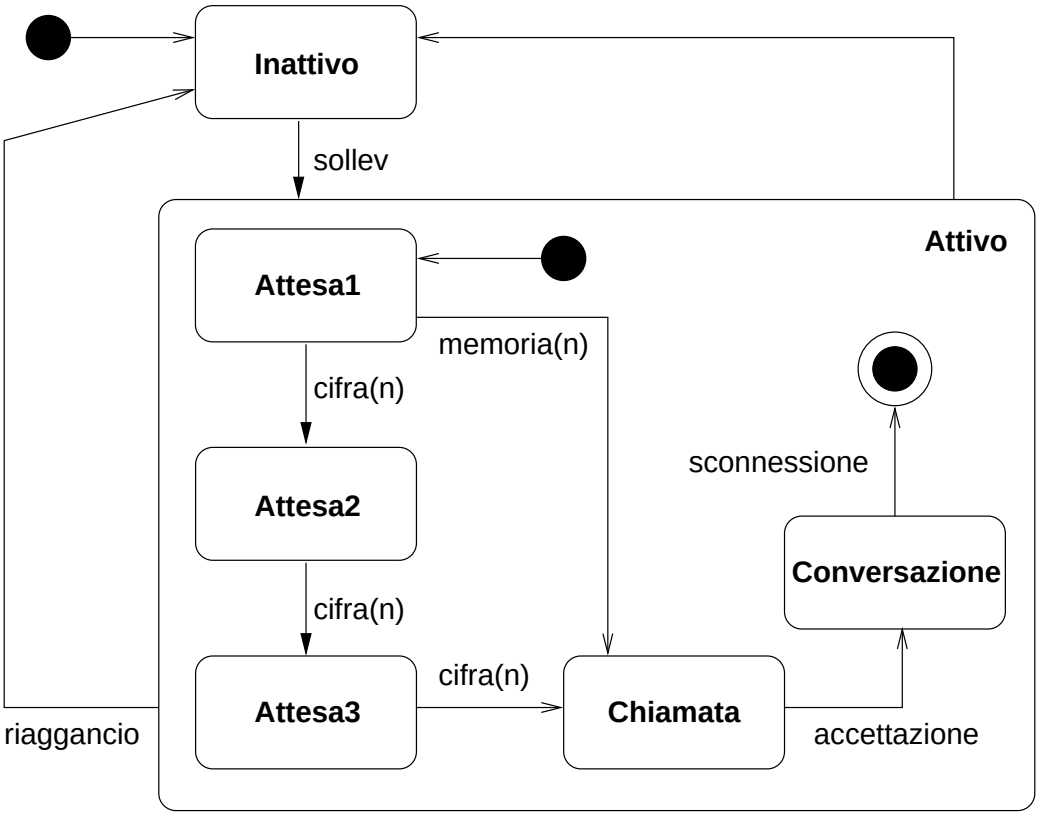


Figura 3.32: Una macchina a stati gerarchica.

### 3.7.8 Diagrammi di interazione

I *diagrammi di interazione* mostrano gli scambi di messaggi fra oggetti. Esistono due tipi di diagrammi di interazione: i diagrammi *di sequenza* e quelli *di comunicazione*.

Un diagramma di sequenza descrive l'interazione fra piú oggetti mettendo in evidenza il flusso di messaggi scambiati e la loro successione temporale. I diagrammi di sequenza sono quindi adatti a rappresentare degli *scenari* possibili nell'evoluzione di un insieme di oggetti. È bene osservare che ciascun diagramma di sequenza rappresenta esplicitamente una o piú istanze delle possibili sequenze di messaggi, mentre un diagramma di stato definisce implicitamente tutte le possibili sequenze di messaggi ricevuti (eventi) o inviati (azioni **send**) da un oggetto interagente con altri.

Un diagramma di sequenza è costituito da simboli chiamati *lifeline*, che rappresentano i diversi ruoli degli oggetti coinvolti nell'interazione. Sotto

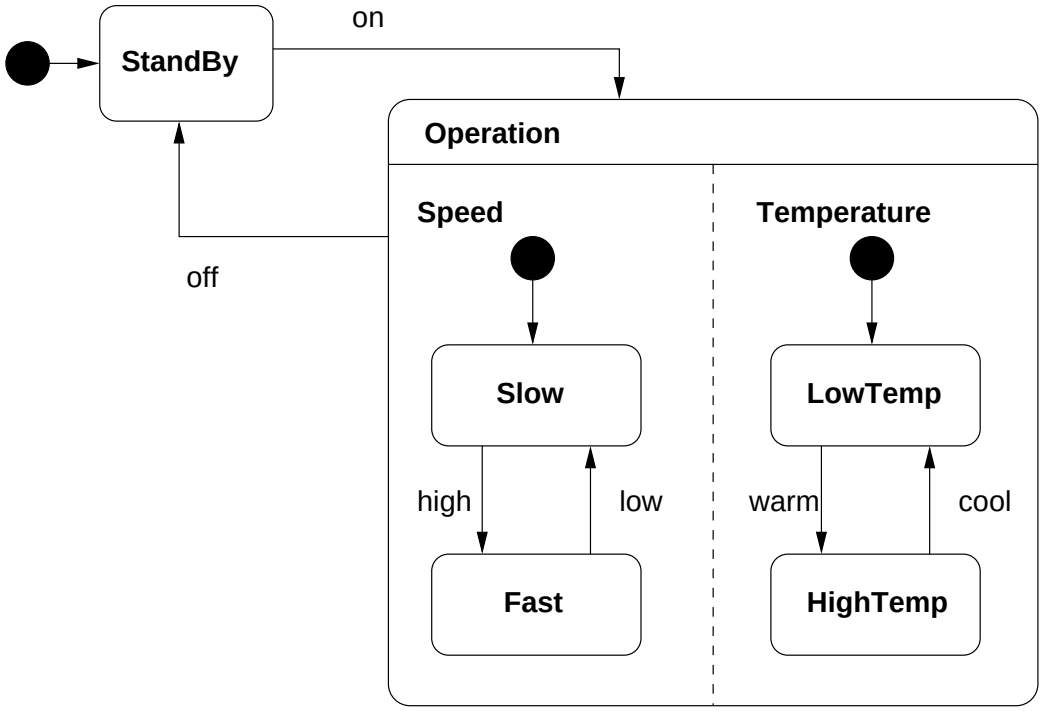


Figura 3.33: Una macchina a stati gerarchica con regioni concorrenti.

ogni lifeline c'è una linea verticale che rappresenta l'evoluzione temporale di ciascun oggetto. Lo scambio di un messaggio, o la chiamata di un'operazione, si rappresenta con una freccia dalla linea verticale dell'oggetto sorgente a quella del destinatario. L'ordine dei messaggi lungo le linee verticali ne rispecchia l'ordine temporale. Si può disegnare anche un asse dei tempi, parallelo alle lifeline, su cui evidenziare gli eventi, etichettando gli istanti corrispondenti con degli identificatori o con dei valori temporali, che possono essere usati per specificare vincoli di tempo. I periodi in cui un oggetto è coinvolto in un'interazione possono essere messi in evidenza sovrapponendo una striscia rettangolare alla linea verticale.

La Fig. 3.35 mostra un semplice diagramma di sequenza che descrive l'interazione di due utenti con un centralino telefonico.

Un diagramma di comunicazione (chiamato *diagramma di collaborazione* in UML1) mette in evidenza l'aspetto strutturale di un'interazione, mostrando esplicitamente i legami (istanze di associazioni) fra gli oggetti, e ricorrendo a un sistema di numerazione strutturato per indicare l'ordinamento temporale dei messaggi.

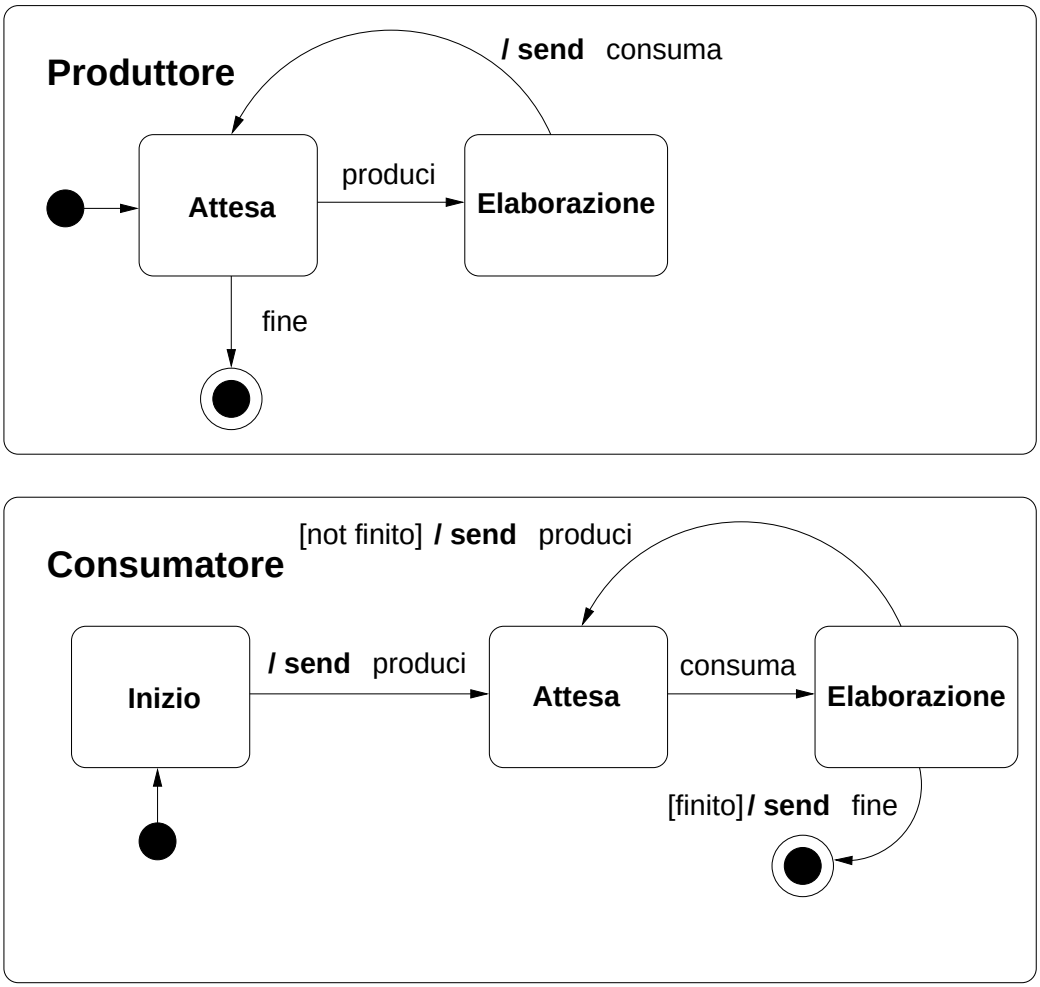


Figura 3.34: Due macchine a stati interagenti.

La Fig. 3.36 mostra il diagramma di comunicazione corrispondente al diagramma di sequenza di Fig. 3.35.

### 3.7.9 Diagrammi di attività

I *diagrammi di attività* servono a descrivere il flusso di controllo e di informazioni dei processi. In un modello di analisi si usano spesso per descrivere i processi del dominio di applicazione, come, per esempio, le procedure richieste nella gestione di un'azienda, nello sviluppo di un prodotto, o nelle transazioni economiche. In un modello di progetto possono essere usati per descrivere algoritmi o implementazioni di operazioni. Si può osservare che,

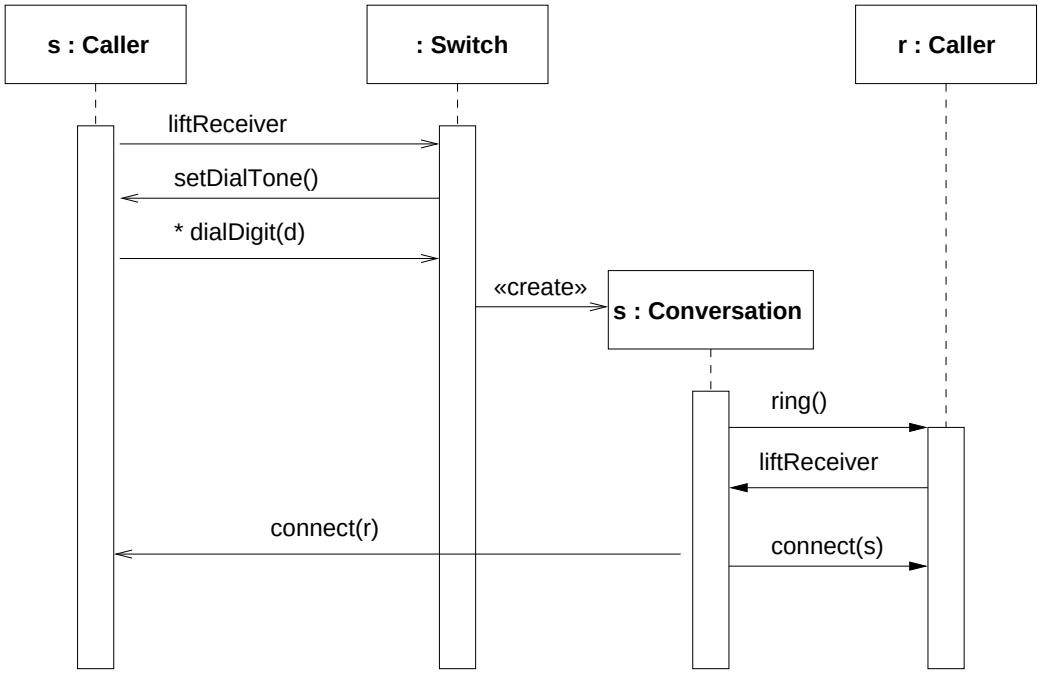


Figura 3.35: Un diagramma di sequenza.

nella loro forma piú semplice, i diagrammi di attività sono molto simili ai tradizionali *diagrammi di flusso* (*flowchart*).

Un diagramma di attività è formato da *nodi* e *archi*. I nodi rappresentano *attività* svolte in un processo, *punti di controllo* del flusso di esecuzione, o *oggetti* elaborati nel corso del processo. Gli archi collegano i nodi per rappresentare i flussi di controllo e di informazioni.

I diagrammi di attività possono descrivere attività svolte da entità differenti, raggruppandole graficamente. Ciascuno dei gruppi cosí ottenuti è una *partizione*, detta anche *corsia* (*swimlane*).

La Fig. 3.37 mostra un diagramma di attività che descrive (in modo molto semplificato) il processo di sviluppo di un prodotto, mostrando quali reparti di un'azienda sono responsabili per le varie attività.

Le due linee orizzontali spesse rappresentano, rispettivamente, un nodo di controllo di tipo *fork* (diramazione del flusso di controllo in attività parallele) e un nodo di tipo *join* (ricongiungimento di attività parallele). Il nodo Specification è un nodo oggetto, in questo caso il documento di specifica prodotto dall'attività Design.

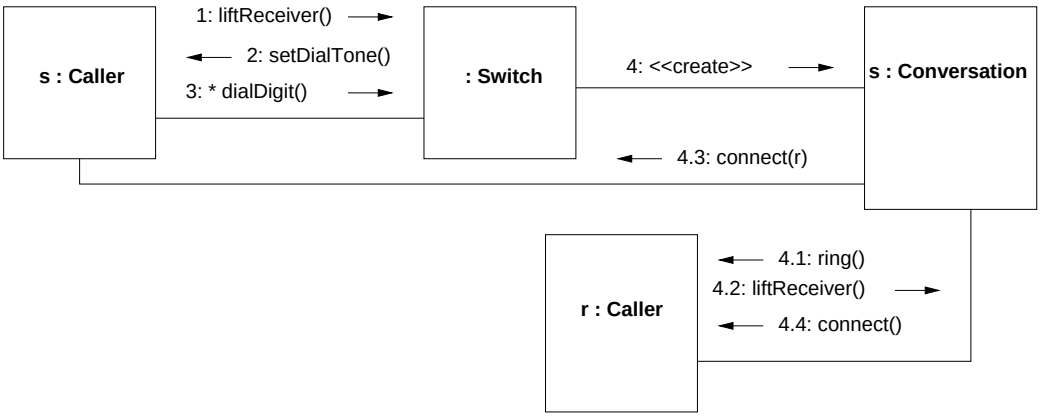


Figura 3.36: Un diagramma di comunicazione.

# Letture

**Obbligatorie:** Cap. 5 e Sez. 4.6 Ghezzi, Jazayeri, Mandrioli, oppure Cap. 2 (esclusi 2.5.1, 2.5.2, pagg. 90–100 di 2.6.2, 2.6.4, 2.7.3) Ghezzi, Fuggetta et al., oppure Cap. 7 (esclusi 7.4.2, 7.6.2, 7.7), Sez. 8.4–8.9, Sez. 23.1, Sez. 24.1 Pressman.

**Facoltative:** Cap. 23 Pressman. Sulla logica, Cap. 1 e 2 Quine.

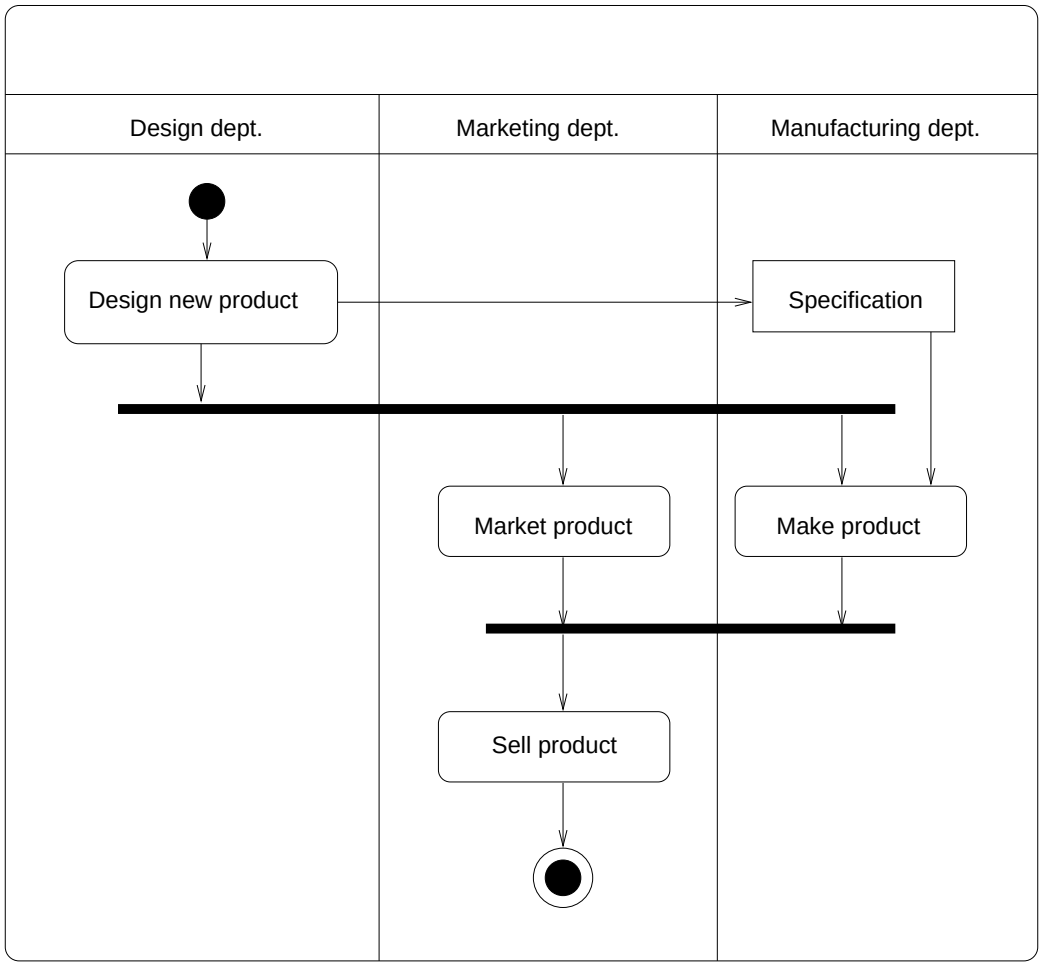


Figura 3.37: Un diagramma di attività.

# Capitolo 4

# Il progetto

L'obiettivo dell'attività di progetto è produrre una *architettura software*, cioè una descrizione della struttura del sistema da realizzare, espressa in termini dei suoi *moduli*, cioè dei suoi componenti e delle loro relazioni reciproche.

Questa descrizione deve rispondere a due esigenze contrastanti: deve essere abbastanza astratta da permettere una facile comprensione del sistema (e quindi la verifica dell'adeguatezza del progetto rispetto alle specifiche), ed abbastanza dettagliata da fornire una guida alla successiva fase di codifica.

L'attività di progettazione è un processo iterativo attraverso una serie di approssimazioni successive a partire da un primo progetto ad alto livello, che índica una suddivisione del sistema in pochi grandi blocchi, fino ad ottenere una struttura "abbastanza" dettagliata. Generalmente la fase di progetto viene considerata come composta di due sottofasi, il *progetto architetturale*, o *di sistema*, ed il *progetto dettagliato*.

È difficile stabilire a priori a quale livello di dettaglio si debba fermare la fase di progetto. Inoltre, l'uso di metodi formali, di tecniche di prototipazione e di linguaggi ad alto livello (adatti sia al progetto che alla codifica) rende sfumata la distinzione fra progetto e codifica.

Neppure la distinzione fra la fase di analisi e specifica dei requisiti e la fase di progetto è completamente netta, poiché nella fase di progetto spesso si rilevano delle incompletezze e inconsistenze delle specifiche, che devono quindi essere riconsiderate e riformulate. Per questo sono sempre piú diffusi i processi di sviluppo che alternano ciclicamente fasi di analisi e fasi di progetto.

Infine, ricordiamo che il progetto dell'architettura software è legato al-

l'architettura hardware, che pone dei vincoli alle scelte del progettista. Il progettista del software deve anche specificare le relazioni fra architettura software ed architettura hardware, e in particolare l'assegnamento dei vari componenti software ai componenti hardware che li devono eseguire.

## 4.1 Obiettivi della progettazione

Data una specifica, esistono molti modi per realizzarla. Naturalmente la scelta fra le diverse possibiltà non è arbitraria, ma è guidata sia da vincoli di carattere economico, sia dalla necessità di conseguire un'adeguata qualità del prodotto o del progetto stesso. Per qualità del progetto si intendono quelle caratteristiche che, pur non essendo "visibili" all'utente, determinano la qualità del prodotto, e quelle caratteristiche che offrono un vantaggio economico al produttore del software, rendendo piú efficace il processo di sviluppo.

Fra le caratteristiche che determinano la qualità del prodotto o del progetto, citiamo l'*affidabilità*, la *modificabilità* (ricordiamo il principio "*design for change*"), la *comprensibilità* e la *riusabilità*. L'esperienza accumulata finora dimostra che queste proprietà dipendono fortemente da un'altra, la *modularità*. Un sistema è *modulare* se è composto da un certo numero di sottosistemi, ciascuno dei quali svolge un compito ben definito e dipende dagli altri in modo semplice. È chiaro che un sistema cosí strutturato è piú comprensibile di uno la cui struttura venga oscurata dalla mancanza di una chiara suddivisione dei compiti fra i suoi componenti, e dalla complessità delle dipendenze reciproche. La comprensibilità a sua volta, insieme alla semplicità delle interdipendenze, rende il sistema piú facile da verificare, e quindi piú affidabile. Inoltre, il fatto che ciascun sottosistema sia il piú possibile indipendente dagli altri ne rende piú facili la modifica e il riuso.

### 4.1.1 Struttura e complessità

Per spiegare in modo piú concreto il concetto di "struttura", e il nesso fra struttura e complessità, consideriamo due modi diversi di organizzare un programma in C, i cui componenti potrebbero essere le funzioni che formano il programma. In un primo caso il programma (che supponiamo composto da un migliaio di istruzioni) consiste di un'unica funzione (`main()`), e nell'altro consiste di dieci o venti funzioni. Nel primo caso si ha poca struttura poiché tutte le istruzioni del programma stanno in un unico "calderone", ove ciascuna di esse può interagire con le altre, per esempio attraverso i valori di

variabili globali; di conseguenza la complessità è alta. Nel secondo caso si ha piú struttura, poiché il campo di azione di ciascuna istruzione è limitato alla funzione a cui appartiene, e quindi ogni funzione "nasconde" le proprie istruzioni. Se lo scopo di ciascuna funzione è chiaro, il programma può essere descritto e compreso in termini delle relazioni fra le funzioni componenti, ed è quindi meno complesso. Se consideriamo un terzo caso, in cui il programma è formato da cento o duecento funzioni, vediamo che la complessità torna ad aumentare a causa delle numerose interazioni fra le funzioni. Il grado di strutturazione di un sistema dipende quindi dal giusto valore di *granularità*, cioè dalle dimensioni dei componenti considerati, che non deve essere né troppo grossa né troppo fine. Dal momento che i componenti elementari del software sono le singole istruzioni del linguaggio di programmazione usato, che hanno una granularità finissima, un software ben strutturato ha un'organizzazione gerarchica che lo suddivide in vari strati con diversi livelli di astrazione: il sistema complessivo è formato da un numero ridotto di sottosistemi, ciascuno dei quali è diviso in un certo numero di moduli, divisi a loro volta in sottomoduli, e cosí via. In questo modo, entro ciascun livello di astrazione si può esaminare una parte del sistema in termini di pochi elementi costitutivi. La Fig. 4.1 cerca di visualizzare intuitivamente questi concetti; nello schema di destra nella figura, i cerchietti rappresentano le interfacce dei moduli, di cui parleremo piú oltre.

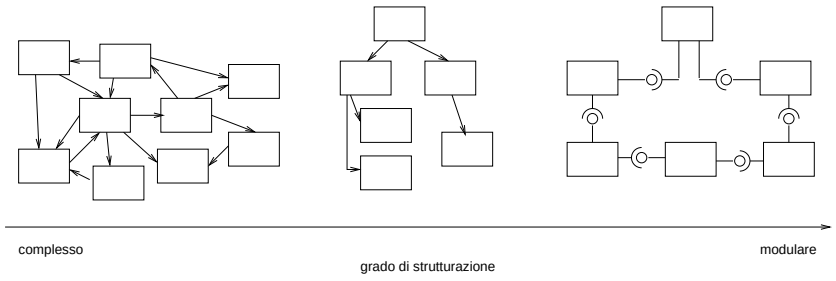


Figura 4.1: Complessità e struttura.

La modularità di un sistema influenza anche la pianificazione e la gestione dell'attività di progetto e di codifica, poiché rende possibile la ripartizione del lavoro fra diversi gruppi di sviluppatori, i quali possono lavorare in modo tanto piú indipendente (anche dal punto di vista organizzativo: si pensi a progetti sviluppati da consorzi di aziende diverse, magari situate in paesi diversi) quanto piú i sottosistemi sono logicamente indipendenti.

È quindi fondamentale, nel progettare un sistema, conoscere e comprendere i principî ed i metodi che permettono di realizzare architetture modulari.

## 4.2 Moduli

Il termine *modulo* può assumere significati diversi in contesti diversi, e in questo corso adotteremo il significato piú generale: una porzione di software che contiene e fornisce risorse o servizi[1], e a sua volta può usare risorse o servizi offerti da altri moduli. Un modulo viene quindi caratterizzato dalla sua *interfaccia*, cioè dall'elenco dei servizi offerti (o *esportati*) e richiesti (o *importati*). Queste due parti dell'interfaccia si chiamano *interfaccia offerta* e *interfaccia richiesta*, ma spesso la parola "interfaccia" viene usata in riferimento alla sola parte offerta.

Le risorse offerte da un modulo possono essere strutture dati, operazioni, e definizioni di tipi. L'interfaccia di un modulo può specificare un *protocollo*, cioè un insieme di vincoli sulle possibili sequenze di scambi di messaggi (o chiamate di operazioni) fra il modulo e i suoi clienti. Alle operazioni si possono associare *precondizioni* e *postcondizioni* (v. oltre). Infine, l'interfaccia può specificare le *eccezioni*, cioè condizioni anomale che si possono verificare nell'uso del modulo. Riguardo alla possibilità che un modulo possa offrire l'accesso diretto (cioè senza la mediazione di procedure apposite) ad alcune strutture dati, vedremo fra poco che questa pratica viene sconsigliata nei metodi attuali di progetto e di programmazione.

L'interfaccia di un modulo è una specifica, che viene realizzata dall'*implementazione* del modulo. Possiamo distinguere fra implementazioni *composte*, in cui l'interfaccia del modulo viene implementata per mezzo di piú sottomoduli, e implementazioni *semplici*, in cui l'implementazione (costituita da definizioni di dati e operazioni in un qualche linguaggio di programmazione) appartiene al modulo stesso. La Fig. 4.2 riassume schematicamente la definizione di modulo qui esposta.

L'interfaccia e l'implementazione definiscono un *modulo logico*, cioè un'entità astratta capace di interagire con altre entità. Nella fase di codifica vengono prodotti dei file contenenti codice sorgente nel linguaggio di programmazione adottato. Chiameremo *moduli fisici* o *artefatti* sia i file sorgente, sia i file collegabili ed eseguibili ottenuti per compilazione e collegamento. I moduli fisici contengono (in linguaggio sorgente o in linguaggio macchina)

[1] Useremo questi due termini come sinonimi

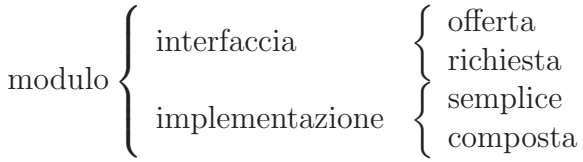


Figura 4.2: Modulo logico.

le definizioni delle entità che realizzano i moduli logici, cioè strutture dati e sottoprogrammi, e la corrispondenza fra moduli logici e fisici dipende sia dal linguaggio e dall'ambiente di programmazione che da scelte fatte dagli sviluppatori. Questa corrispondenza non è necessariamente biunivoca, in quanto un modulo fisico può contenere le definizioni relative a piú moduli logici, oppure le definizioni appartenenti a un modulo logico possono essere suddivise fra piú moduli fisici. Ai fini della mantenibilità e della modificabilità del sistema, è importante che la corrispondenza fra struttura logica e struttura fisica sia chiara e razionale. Una buona organizzazione dei moduli fisici permette un uso piú efficace di strumenti automatici di programmazione (come, per esempio il programma `make`) e di gestione delle configurazioni (per esempio, i sistemi `SCCS`, `RCS` e `CVS`).

La maggior parte dell'attività di progetto è rivolta alla definizione della struttura logica del sistema, ma nelle fasi piú "basse", cioè piú vicine all'implementazione, si può definire, almeno a grandi linee, anche la struttura fisica. Nel seguito ci riferiremo essenzialmente alla struttura logica.

### 4.2.1 Interfaccia e implementazione

Per progettare l'interfaccia di un modulo bisogna compiere un insieme di scelte guidate dai principî della *divisione delle responsabilità* e dell'*information hiding* (occultamento dell'informazione).

Il principio della divisione delle responsabilità dice che bisogna cercare di suddividere il lavoro svolto dal sistema (e ricorsivamente da ciascun sottosistema) fra i vari moduli, in modo che a ciascuno di essi venga affidato un compito ben definito e limitato (ricordiamo il vecchio motto del sistema operativo Unix: "*do one thing well*"). I moduli progettati secondo questo criterio hanno la proprietà della *coesione*, cioè di offrire un insieme omogeneo di servizi. Per esempio, un modulo che calcola dei valori e provvede anche a scriverli sull'output è poco coeso: poiché la rappresentazione dei risultati è indipendente dal procedimento di calcolo (per esempio, un insieme di

risultati potrebbe essere visualizzato con una tabella oppure un grafico, o non essere visualizzato affatto, essendo destinato ad ulteriori elaborazioni), è meglio che la funzione di produrre un output venga affidata a un modulo specializzato, in modo da avere un sistema piú strutturato e flessibile.

La divisione di responsabilità comporta anche la necessità di specificare gli obblighi reciproci dei moduli. Un modulo fornitore di un servizio garantisce ad ogni modulo cliente che in conseguenza dell'espletamento del servizio (per esempio, l'esecuzione di una procedura) saranno vere certe relazioni logiche (*postcondizioni*) sullo stato del sistema. Il modulo fornitore, però, si aspetta che all'atto dell'invocazione del servizio siano vere altre relazioni (*precondizioni*). Per esempio, un modulo che deve calcolare la radice quadrata di un numero $x$ ha come precondizione un vincolo sul segno del valore di $x$ ($x \geq 0$), e come postcondizione una relazione fra $x$ e il risultato $y$ ($x = y^2$). Le post-condizioni e le pre-condizioni quindi rappresentano, rispettivamente, le responsabilità del modulo fornitore e quelle del modulo cliente riguardo a un dato servizio. È buona prassi scrivere le pecondizioni e postcondizioni di ciascuna operazione nella documentazione di progetto.

Un altro aspetto importante della divisione delle responsabilità è la gestione degli errori e delle situazioni anomale. Se si prevede che nello svolgimento di un servizio si possano verificare delle situazioni anomale, bisogna decidere se tali situazioni possono essere gestite nel modulo fornitore, nascondendone gli effetti ai moduli clienti, oppure se il modulo fornitore deve limitarsi a segnalare il problema, delegandone la gestione ai moduli clienti.

Il principio dell'*information hiding* afferma che bisogna rendere inaccessibile dall'esterno tutto ciò che non è strettamente necessario all'interazione con gli altri moduli, in modo che vengano ridotte al minimo le dipendenze e quindi sia possibile progettare, implementare e collaudare ciascun modulo indipendentemente dagli altri.

Dopo che è stato individuato il compito di un modulo, si vede che tale compito, per essere svolto, ha bisogno di varie strutture dati e di operazioni (o di altre risorse, magari piú astratte, come tipi di dati o politiche di gestione di certe risorse). Il progettista deve decidere quali di queste entità devono far parte dell'interfaccia, ed essere cioè accessibili dall'esterno. Le entità che non sono parte dell'interfaccia servono unicamente a implementare le altre, e non devono essere accessibili. Questa scelta non è sempre facile e immediata, poiché spesso può sembrare che certe parti dell'implementazione (per esempio, una procedura) possano essere utili ad altri moduli. Comunque, quando si è deciso che una certa entità fa parte dell'implementazione bisogna che questa sia effettivamente nascosta, poiché la dipendenza di un modulo

cliente dall'implementazione di un modulo fornitore fa sí che quest'ultimo non possa venire modificato senza modificare il cliente. In un sistema di media complessità il costo di qualsiasi cambiamento può diventare proibitivo, se le dipendenze reciproche costringono gli sviluppatori a propagare i cambiamenti da un modulo all'altro.

In particolare, conviene nascondere nell'implementazione le strutture dati, a cui si dovrebbe accedere soltanto per mezzo di sottoprogrammi. Anche le politiche di accesso a una risorsa gestita da un modulo (per esempio, ordinamento FIFO oppure LIFO delle richieste di accesso), di norma devono essere nascoste ai moduli clienti, in modo che questi non dipendano da tali politiche.

In conclusione, nello specificare un'architettura software si cerca di ottenere il massimo *disaccoppiamento* fra i moduli.

### 4.2.2 Relazioni fra moduli

Come già detto, un'architettura software è la specifica di un insieme di moduli e delle loro relazioni. Fra queste, hanno un'importanza particolare le relazioni di *composizione* e di *uso*.

La relazione di composizione sussiste fra due moduli quando uno è parte dell'altro ed è necessariamente gerarchica, cioè viene descritta da un grafo orientato aciclico.

La relazione di uso sussiste quando il corretto funzionamento di un modulo richiede la presenza e generalmente[2] il corretto funzionamento di un altro modulo. Questa relazione permette l'esistenza di cicli nel grafo ad essa associato, però si cerca di evitare i cicli poiché i moduli che si trovano su un ciclo non possono venire esaminati e verificati isolatamente.

Le relazioni di composizione e uso sono il minimo indispensabile per definire un'architettura, ma ne esistono e se ne possono concepire molte, che possono essere piú o meno utili a seconda del tipo di sistema progettato, della metodologia di progetto, e del livello di dettaglio richiesto in ciascuna fase della progettazione. Per esempio, nei metodi orientati agli oggetti è particolarmente importante la relazione di generalizzazione o eredità. Un'altra

[2]Il corretto funzionamento del modulo usato non è sempre richiesto, poiché un modulo in certi casi contiene solo risorse "statiche", come strutture dati o definizioni di tipi, per le quali non si può parlare di "funzionamento"

relazione importante è la *comunicazione* fra moduli. In generale, si mettono in evidenza vari tipi di *dipendenza*, di cui l'uso è un caso particolare.

### 4.2.3 Tipi di moduli

Il concetto di modulo visto finora è molto generale. In questa sezione esamineremo alcuni concetti relativi a diversi tipi di componenti modulari del software.

Le *astrazioni procedurali*, il tipo di moduli piú semplice, sono moduli che non gestiscono strutture dati, ma forniscono soltanto delle funzioni. Un esempio tipico di astrazioni procedurali sono le tradizionali librerie matematiche.

Un modulo che gestisce una struttura dati è chiamato *oggetto astratto*: la struttura dati è astratta in quanto vi si può accedere soltanto attraverso le operazioni definite dall'interfaccia, e la sua realizzazione concreta è nascosta. Gli oggetti astratti permettono di controllare gli accessi alle informazioni contenute in una struttura dati che devono essere condivise fra piú moduli.

Un altra categoria di moduli è quella dei *tipi di dati astratti* (TDA), che definiscono dei tipi da cui si possono istanziare oggetti. Un tipo è quindi la definizione di un insieme di oggetti che hanno la stessa struttura e le stesse operazioni.

Se la struttura o l'algoritmo di un oggetto astratto (o tutti e due) possono essere resi parametrici rispetto a qualche caratteristica del sistema, si ottiene un *oggetto generico*. Generalmente la caratteristica che si rende generica è il tipo di qualche componente dell'oggetto. Un oggetto generico è la definizione di un insieme di oggetti che hanno strutture ed operazioni simili, ma specializzate secondo alcuni parametri. Per esempio, una tabella generica può rappresentare una famiglia di tabelle che differiscono per il tipo degli elementi o per il loro numero, o per tutte e due le caratteristiche.

Anche un tipo di dati astratto può essere reso generico, ottenendo un *tipo di dati astratto generico*.

Gli *oggetti* e le *classi* sono concetti che abbiamo già incontrato nell'àmbito delle metodologie di specifica orientate agli oggetti (sez. 3.7), dove vengono usati per rappresentare entità del dominio di applicazione. Nel progetto del software, gli oggetti sono i componenti software che implementano l'applicazione. Alcuni di questi oggetti simulano le entità del dominio implementan-

done direttamente attributi e operazioni, altri partecipano al funzionamento dell'applicazione fornendo i meccanismi necessari.

Ogni oggetto è un'istanza di una *classe*, che è quindi simile a un tipo di dati astratto. La differenza fra una classe ed un TDA è sottile: i tipi di dato astratti permettono al programmatore di definire dei tipi che possano essere usati come i tipi predefiniti dei linguaggi di programmazione, mentre le classi permettono di "simulare" delle entità. Un classico esempio di TDA è il tipo dei numeri complessi costruito a partire dai numeri in virgola mobile. In pratica, i linguaggi di programmazione non permettono di fare questa distinzione di natura pragmatica, che serve solo a chiarire il modo in cui si useranno i diversi tipi di moduli.

La relazione di eredità fra classi offre al progettista altre possibilità di strutturazione dell'architettura, oltre a quelle offerte dalle relazioni di uso e di composizione.

Esistono dei moduli il cui scopo è di raggruppare altri elementi (che possono essere moduli a loro volta) e di presentarli come un'unità logica. Generalmente si usa il termine *package* per riferirsi a un insieme di entità logicamente correlate (per esempio un gruppo di classi o di sottoprogrammi). Il concetto di "package" di solito implica quello di *spazio di nomi* (*namespace*): uno spazio di nomi serve a identificare un insieme di entità e ad evitare conflitti di nomenclatura fra tali entità e quelle appartenenti ad un altro spazio di nomi: due entità aventi lo stesso nome ma appartenenti a spazi di nomi diversi sono entità distinte.

Un *sottosistema* è una parte di un sistema piú grande, relativamente autonoma e caratterizzata da una propria funzione, e quindi da un'interfaccia. I sottosistemi si possono modellare per mezzo dei package o in altri modi che vedremo piú oltre.

Infine, accenniamo al concetto di *componente*. Questo viene generalmente inteso come un elemento software interamente definito dalle interfacce offerte e richieste, che possa essere sostituito da un'altro componente equivalente anche se diversamente implementato, e che possa essere riusato in contesti diversi, analogamente ai componenti elettronici. Secondo questa definizione, qualsiasi modulo realizzato rispettando rigorosamente il principio dell'information hiding e capace di offrire servizi di utilità abbastanza generale si potrebbe considerare un componente, però spesso si usa questo termine in modo piú restrittivo, richiedendo che i componenti si possano sostituire a tempo di esecuzione: questo impone dei particolari requisiti sugli ambienti di programmazione e di esecuzione, e richiede che i componenti vengano

realizzati rispettando determinate convenzioni. Una terza definizione, intermedia fra le precedenti, richiede che i componenti seguano certe convenzioni, ma non che siano necessariamente istanziabili a tempo di esecuzione.

## 4.3 Linguaggi di progetto

La fase di progetto produce alcuni documenti che descrivono l'architettura del sistema. Tali documenti sono scritti in parte in linguaggio naturale (possibilmente secondo qualche standard di documentazione che stabilisca la struttura dei documenti) e in parte mediante notazioni di progetto testuali e grafiche. Le notazioni testuali descrivono ciascun modulo usando un linguaggio simile ai linguaggi di programmazione, arricchito da costrutti che specificano le relazioni fra i moduli e descrivono precisamente le interfacce, e spesso privo di "istruzioni eseguibili". Le notazioni grafiche rappresentano in forma di diagrammi i grafi[3] definiti dalle relazioni. I metodi di progetto piú usati nella pratica sono basati su linguaggi grafici integrati da linguaggi testuali.

### 4.3.1 Unified Modeling Language

Il linguaggio UML, già visto nel capitolo relativo ai linguaggi di specifica orientati agli oggetti, è anche un linguaggio di progetto. Nella descrizione di un'architettura si possono usare gli stessi concetti (e le relative notazioni) usate in fase di specifica dei requisiti, ma con un piú fine livello di dettaglio: per esempio, in fase di progetto si devono specificare precisamente il numero e i tipi degli argomenti di operazioni di cui, in fase di specifica, si era dato solo il nome. Inoltre l'UML fornisce delle notazioni adatte a esprimere dei concetti propri della descrizione architetturale, come la struttura fisica dello hardware e del software. Vedremo nel seguito alcuni esempi di applicazione dell'UML.

### 4.3.2 CORBA Interface Description Language

Come esempio di notazione testuale di progetto (anche se di espressività alquanto ristretta), consideriamo (oltre a quella mostrata nel libro di testo),

[3]Ricordiamo che un *grafo* è un concetto matematico, mentre un *grafico* è un oggetto fisico (che può anche rappresentare un grafo).

il linguaggio CORBA IDL[4], orientato alla descrizione di sistemi CORBA (*Common Object Request Broker Architecture*). L'architettura CORBA si basa su un modello computazionale orientato agli oggetti, un'infrastruttura software di comunicazione, e un insieme di librerie di componenti software, e serve a sviluppare applicazioni distribuite.

Nel modello CORBA i componenti fondamentali di un'applicazione sono oggetti che comunicano in modo indipendente sia dalla posizione fisica che dall'implementazione delle rispettive controparti. Un oggetto può richiedere servizi ad un altro oggetto "ignorando" su quale calcolatore si trova, attraverso quale protocollo avviene la comunicazione, in che modo e con quale linguaggio è stato implementato. In questo modo il funzionamento di un sistema distribuito è interamente svincolato dall'implementazione dei singoli oggetti.

Chi progetta un oggetto CORBA ne specifica l'interfaccia usando il linguaggio IDL, e lo implementa nel linguaggio di programmazione prescelto, con l'aiuto di un compilatore IDL che traduce la specifica IDL nel linguaggio di programmazione (per esempio, Ada o C++). Il compilatore IDL produce anche del codice che serve a collegare l'applicazione al supporto run-time che provvede alla comunicazione fra oggetti.

Chi progetta un cliente di un oggetto CORBA si serve del compilatore IDL per tradurre la specifica IDL di tale oggetto nel linguaggio di programmazione in cui viene implementato il cliente. Anche in questo caso il compilatore IDL produce del codice ausiliario per integrare l'applicazione (la parte cliente) col sistema di comunicazione.

L'IDL è un linguaggio orientato agli oggetti puramente dichiarativo, essendo privo di istruzioni procedurali, come assegnamenti, cicli e simili. Ha un controllo statico dei tipi, anche se permette di dichiarare parametri e risultati il cui tipo non è noto a tempo di compilazione, ed offre un insieme di tipi base analoghi a quelli dei linguaggi di programmazione, dai quali si possono costruire tipi derivati, come le strutture (nel senso del C++, chiamate *record* in altri linguaggi), gli array a dimensione fissa e le sequenze (array a dimensione variabile). La definizione di un'interfaccia è analoga (anche sintatticamente) alla definizione di una classe in C++, con alcune differenze fra cui le piú importanti sono queste:

- tutti membri sono pubblici;

[4]Oltre al CORBA IDL, che d'ora in poi chiameremo semplicemente IDL, esistono altri linguaggi chiamati IDL.

- si possono dichiarare solo membri funzione;
- non si possono definire le implementazioni dei membri funzione;
- non si possono definire interfacce all'interno di interfacce.

A proposito dell'impossibilità di dichiarare membri dato, osserviamo che si possono dichiarare degli *attributi* di un'interfaccia, che sintatticamente sembrano membri dato, ma in realtà servono solo a dichiarare implicitamente dei membri funzione che leggono o assegnano valori. Sono, cioè, un costrutto linguistico per mettere in evidenza un uso particolare di certi membri funzione. Il compilatore IDL non genera alcuna istruzione per riservare memoria a tali pseudovariabili.

Il seguente esempio mostra l'interfaccia di un oggetto che gestisce un vettore i cui elementi sono sequenze di interi: la dimensione iniziale dell'oggetto può essere un valore qualsiasi, fra zero e `MAX_SIZE`, scelto dall'implementazione. La costante `MAX_SIZE` viene resa disponibile ai clienti. Un cliente può conoscere la dimensione e modificarla per mezzo delle operazioni, rispettivamente, `get_size()` e `resize()`. Nell'esecuzione di alcune operazioni si possono verificare le eccezioni `SizeExc` o `RangeExc`. I parametri delle operazioni sono in sola lettura, come specifica la parola chiave `in`.

```
interface Vector {
    typedef sequence<long> VElem;

    exception SizeExc {};
    exception RangeExc {};

    const short MAX_SIZE = 3200;

    short get_size();
    void resize(in short n)
        raises (SizeExc);
    VElem get_elem(in short n)
        raises (RangeExc);
    void set_elem(in short n, in VElem e)
        raises (RangeExc);
}
```

Si possono raggruppare piú interfacce, insieme a definizioni di tipo e di costanti, usando il costrutto `module`:

```
module Linear {
    interface Vector {
        // ...
    }
```

```
        interface Matrix {
            // ...
        }
    }
```

Un modulo CORBA è uno spazio di nomi, e può contenere altri moduli.

Il linguaggio IDL è stato esteso in modo da specificare dei componenti, in termini di interfacce offerte e richieste, e di collegamenti fra componenti.

## 4.4 Moduli nei linguaggi di progetto e di programmazione

Ogni linguaggio di progetto ha un suo vocabolario e le sue convenzioni per esprimere i concetti usati nella progettazione del software. I linguaggi di programmazione, a loro volta, hanno dei costrutti linguistici che si prestano piú o meno bene a rappresentare tali concetti. Per esempio, il C non ha dei costrutti che definiscano esplicitamente un modulo, però permette di definire degli oggetti astratti sfruttando le regole di visibilità e di collegamento. In C, un modulo può essere implementato da una unità di compilazione, l'interfaccia di un modulo può essere definita da un file *header*, la relazione di uso può essere rappresentata dalle direttive `#include`. Non si possono definire dei veri e propri dati astratti, ma i programmatori possono attenersi ad una "*disciplina*" nell'uso di tipi derivati, cioè accedere alle strutture dati che implementano un tipo solo attraverso le operazioni previste per quel tipo, rinunciando a sfruttare la possibilità, offerta dal linguaggio, di accedere ai dati direttamente.

I linguaggi di programmazione piú evoluti permettono una rappresentazione piú diretta dei concetti relativi alla programmazione modulare. In fase di progetto si può decidere in quale linguaggio (o in quali linguaggi, poiché si possono usare linguaggi diversi per sottosistemi diversi) deve essere implementato il sistema, e in questo caso la maggiore o minore facilità con cui il linguaggio adottato permette di esprimere certi concetti può influenzare le scelte di progetto. Se invece si progetta in modo indipendente dal linguaggio di programmazione, bisogna evitare l'uso di concetti (come, per esempio, l'eredità) che possono essere di difficile rappresentazione con alcuni linguaggi, oppure scegliere il linguaggio, verso la fine della progettazione, tenendo conto dei tipi di moduli e di relazioni usati nel progetto. In ogni caso, la conoscenza dei costrutti linguistici con cui si rappresentano i vari concetti è necessaria

per verificare la correttezza dell'implementazione rispetto al progetto e per comprendere la corrispondenza fra architettura logica ed architettura fisica.

Per questo, dopo aver introdotto la notazione UML per i concetti fondamentali della progettazione orientata agli oggetti, esamineremo alcuni esempi di moduli scritti in C++ o in Java, mettendo in evidenza soltanto quelle caratteristiche dei linguaggi che sono rilevanti per gli scopi di questo corso. Naturalmente bisogna rivolgersi ai testi specifici per conoscere tutti gli aspetti sintattici e semantici dei costrutti a cui accenneremo in seguito.

### 4.4.1 Incapsulamento e raggruppamento

Come già osservato, un programma consiste in un insieme di istruzioni che, in un certo linguaggio, definiscono numerose entità elementari, come variabili, funzioni, e tipi. Ovviamente il programma non deve essere una massa amorfa e indifferenziata di definizioni: queste definizioni devono essere raggruppate in modo da corrispondere ai moduli previsti dall'architettura, quindi un linguaggio di programmazione deve fornire, prima di tutto, dei mezzi per separare e raggruppare le entità appartenenti ai moduli. Inoltre in ciascun modulo bisogna separare l'interfaccia dall'implementazione, cioè la parte visibile da quella invisibile, quindi i linguaggi devono fornire anche i mezzi per specificare quali dichiarazioni sono visibili all'esterno e quali no. Il termine *incapsulamento* si usa spesso per riferirsi alla separazione, al raggruppamento e all'occultamento selettivo delle entità che formano un modulo: definire un modulo è come chiudere i suoi componenti in una capsula (o un pacchetto) che li protegge da un uso incontrollato.

#### Incapsulamento e raggruppamento in UML

In questa sezione introduciamo gli elementi di modello UML che rappresentano i concetti di incapsulamento e raggruppamento.

**Classi** Come già accennato, le classi in UML si possono usare per modellare dei componenti software, che possono facilmente essere implementati con i corrispondenti costrutti dei linguaggi di programmazione orientati agli oggetti. In fase di progetto si specificano completamente gli attributi e le operazioni (se non si è già fatto nella fase di analisi), indicando tipi, argomenti, valori restituiti e visibilità. Le operazioni da specificare sono quelle che definiscono l'interfaccia, e quindi hanno visibilità *pubblica* (interfaccia

verso ogni altra classe) o *protetta* (interfaccia verso le classi derivate). Le operazioni con visibilità *privata* generalmente vengono aggiunte nella fase di codifica, occasionalmente nella fase di progetto dettagliato.

In fase di progetto non si definisce l'implementazione delle classi, ma in generale si sottintende che ogni operazione debba essere implementata da un metodo, sia cioè *concreta*. Spesso però è utile dichiarare in una classe delle operazioni che non devono essere implementate nella classe stessa, ma in classi derivate; queste operazioni sono dette *operazioni astratte*. Una classe che contiene almeno una operazione astratta non può, ovviamente, essere istanziata direttamente: una tale classe si dice *classe astratta* e può solo servire da base di altre classi. Le classi derivate (direttamente o indirettamente) da una classe astratta forniscono le implementazioni (*metodi*) delle operazioni astratte, fino ad ottenere delle classi *concrete*, cioè fornite di un'implementazione completa e quindi istanziabili. In UML le operazioni e le classi astratte sono caratterizzate dalla proprietà abstract. Graficamente, questa proprietà può essere evidenziata scrivendo in caratteri obliqui il nome dell'elemento interessato.

**Interfacce** In UML, un'interfaccia si può rappresentare separatamente dall'entità che la implementa (o *dalle* entità, poiché un'interfaccia può essere implementata da moduli diversi). Un'interfaccia UML è un elemento di modello costituito da un elenco di operazioni, ovviamente astratte e con visibilità pubblica, a cui si possono aggiungere dei vincoli e un protocollo. Un'interfaccia può anche contenere delle dichiarazioni di attributi: questo significa che le classi che realizzano l'interfaccia devono rendere disponibili i valori di tali attributi, ma non necessariamente implementarli direttamente sotto forma di attributi (i loro valori, per esempio, potrebbero essere calcolati di volta in volta).

Con la notazione completa, un'interfaccia si rappresenta con un simbolo simile a quello delle classi, etichettato con parola chiave[5] ≪interface≫, come nell'esempio illustrato in Fig. 4.3.

In questo esempio la classe **HashTable** rappresenta una tabella hash. Ricordiamo che questo tipo di tabella memorizza gli elementi (coppie $\langle chiave, valore \rangle$) in un vettore, e l'índice associato alla chiave di ciascun elemento viene calcolato in funzione del valore della chiave stessa per mezzo di una funzione detta di *hash*. Poiché la funzione di hash può associare lo stesso índice a elementi

[5]Le parole chiave servono a distinguere diversi tipi di elementi di modello, mentre gli stereotipi sono specializzazioni di altri elementi di modello. Hanno la stessa sintassi.

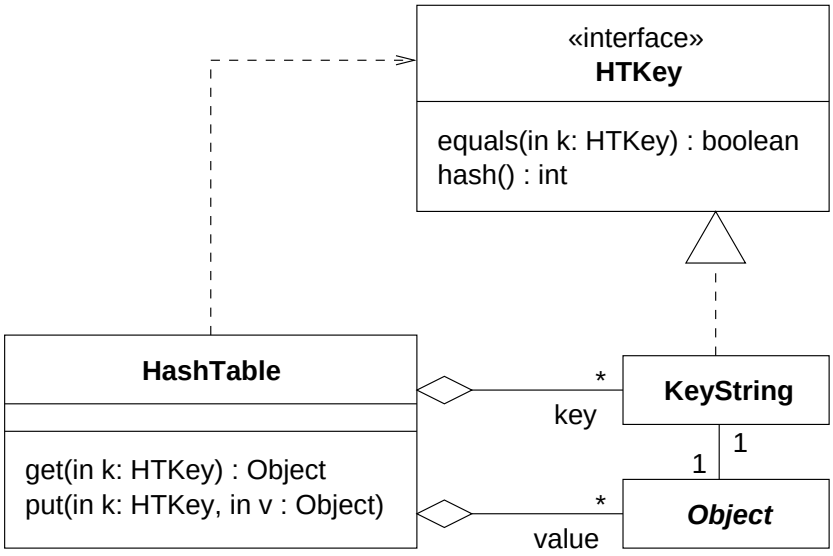


Figura 4.3: Interfacce in UML (1)

distinti, a ciascun índice corrisponde in generale non un solo elemento, ma una lista di elementi, o *bucket*. L'inserimento o la ricerca di un elemento richiedono quindi una scansione della lista individuata dall'índice restituito dalla funzione di hash. Questa ricerca a sua volta richiede un'operazione di confronto.

La funzione di hash e l'operazione di confronto dipendono dalla struttura degli elementi da memorizzare: si hanno diverse implementazioni a seconda che gli elementi siano, per esempio, stringhe di caratteri o immagini digitali. Se vogliamo far sí che la classe **HashTable** sia indipendente dal tipo di elementi che deve contenere, si ricorre al principio di divisione delle responsabilità: alla tabella di hash spetta il compito di ordinare e ricercare gli elementi in base ai risultati della funzione di hash e dell'operazione di confronto, mentre agli elementi spetta il compito di fornire l'implementazione di queste due operazioni.

L'insieme delle operazioni richieste da **HashTable** per il suo funzionamento (interfaccia richiesta) è definito dall'interfaccia **HTKey**, contenente le operazioni equals (per confrontare due elementi) e hash (per calcolare la funzione di hash). Questa relazione fra **HashTable** e **HTKey** è una *dipendenza*, rappresentata dalla freccia tratteggiata. Gli elementi che vogliamo memorizzare possono appartenere a qualsiasi classe che *realizzi* tali operazioni, cioè che comprenda le operazioni di **HTKey** nella propria interfaccia offer-

ta, insieme ad altre eventuali operazioni. Nell'esempio, la classe **KeyString** implementa queste due operazioni, e la relazione fra **KeyString** e **HTKey** è una *realizzazione*, rappresentata da una linea tratteggiata terminante con una punta a triangolo equilatero.

Le interfacce si rappresentano in forma ridotta per mezzo di un cerchio, come mostrato in Figura 4.4. In questo caso la classe (o altro elemento di modello) che realizza l'interfaccia viene collegata a quest'ultima per mezzo di una semplice linea non tratteggiata, mentre una classe che richiede l'interfaccia viene collegata all'interfaccia con una dipendenza (freccia tratteggiata), o col simbolo di *assembly* introdotto nell'UML2.

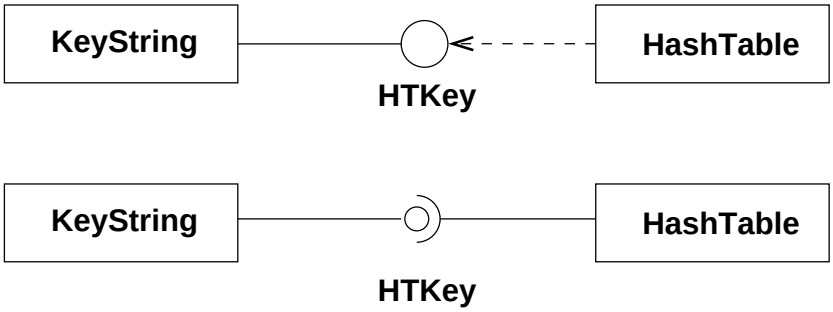


Figura 4.4: Interfacce in UML (2)

**Package** L'UML dispone di un elemento di modello, chiamato *package*, per esprimere il raggruppamento di altri elementi di modello. Questi elementi possono essere di qualsiasi genere, anche interi diagrammi, e l'organizzazione in package di un modello UML può non avere una corrispondenza diretta con la struttura dell'architettura software (per esempio, il progettista potrebbe usare un package per raccogliere tutti i diagrammi di classi, un altro per i diagrammi di stato, e cosí via), anche se molto spesso i package vengono usati in modo da rispecchiare tale struttura.

L'interfaccia di un package consiste nell'insieme dei suoi elementi *esportati*, cioè resi visibili. La Fig. 4.5 mostra il package **Simulator**, che contiene gli elementi pubblici (cioè esportati) **SignalGenerators** e **Filters** (che in questo caso supponiamo essere dei package a loro volta, ma potrebbero essere classi o altri elementi di modello), e l'elemento privato **Internals**. I caratteri '+' e '-' denotano, rispettivamente, gli elementi pubblici e quelli privati.

Un package costituisce lo spazio dei nomi degli elementi contenuti. Ogni

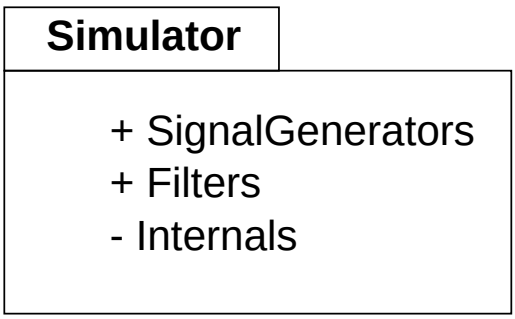


Figura 4.5: Package in UML (1)

elemento appartenente ad un package ha un *nome qualificato*, formato dal nome del package e dal nome semplice dell'elemento, separati dai caratteri ::. Se il package è contenuto in un altro package, il nome del package esterno precede quello del package interno. Se, nell'esempio di Fig. 4.6, supponiamo che il package **Simulator** contenga una classe **Solver**, il nome qualificato di quest'ultima sarebbe **Simulator::Solver**. Questo nome qualificato può essere usato nelle classi appartenenti ad altri package, per esempio nella dichiarazione di parametri. Per usare soltanto il nome semplice, un altro package (nell'esempio, **UserInterface** deve *importare* il package **Simulator**, cioè inserire lo spazio di nomi di **Simulator** nel proprio. La figura mostra i tre package con le rispettive dipendenze di ≪import≫. Un'altra dipendenza mostrata comunemente è quella di uso, che si ha quando qualche elemento di un package usa elementi dell'altro.

**Componenti** In UML1 il termine *componente* (*component*) designava un elemento di modello destinato a rappresentare i moduli fisici di un sistema e la loro corrispondenza con i moduli logici implementati. In UML2, invece, un componente è un modulo logico definito da una o piú interfacce offerte e da una o piú interfacce richieste, sostituibile e riusabile. Faremo riferimento a quest'ultima definizione.

Un componente può realizzare piú di una interfaccia: per esempio, un componente destinato a controllare un impianto stereo potrebbe avere un'interfaccia per il controllo della radio (con le operazioni scegli_banda(), scegli_canale()...), una per il lettore CD (con scegli_traccia(), successivo()...) e cosí via (Fig. 4.7 (a)). Altrettanto vale per le interfacce richieste. La possibilità di suddividere la specifica di un componente in diverse interfacce permette di esprimere piú chiaramente le dipendenze fra componenti.

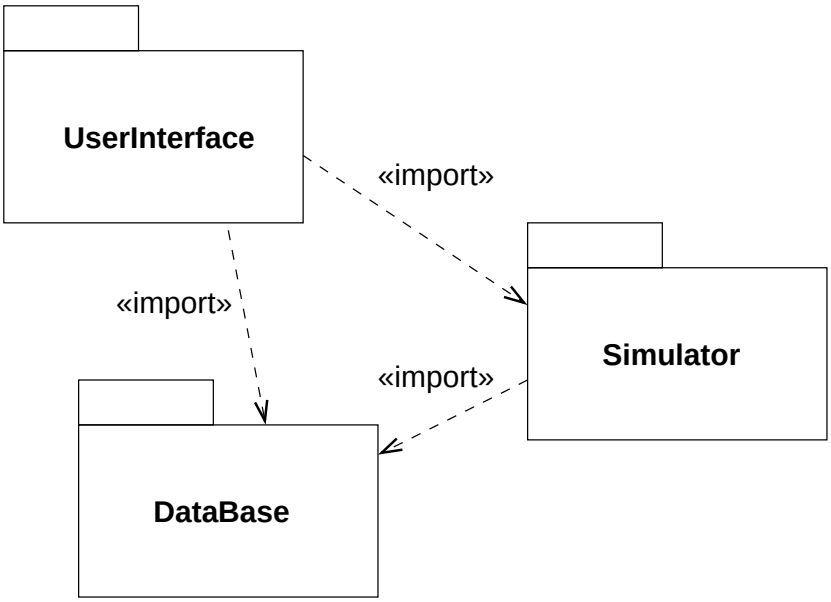


Figura 4.6: Package in UML (2)

Generalmente le interfacce vengono associate a dei *port*, definiti come punti di interazione del componente con l'ambiente esterno. Un port è quindi caratterizzato da una o piú interfacce coinvolte in una determinata interazione. Nella Fig. 4.7 (b), le interfacce PwrVolume, Tuner e Player, destinate all'interazione con l'utente, sono raggruppate nel port User, mentre le coppie di interfacce TunerREQ e TunerRPY, e PlayerREQ e PlayerRPY, usate rispettivamente per comunicare con la radio e col lettore di CD, sono assegnate ai port Radio e CdPlay. Gli altri due port riguardano l'interazione con l'alimentazione e l'amplificatore.

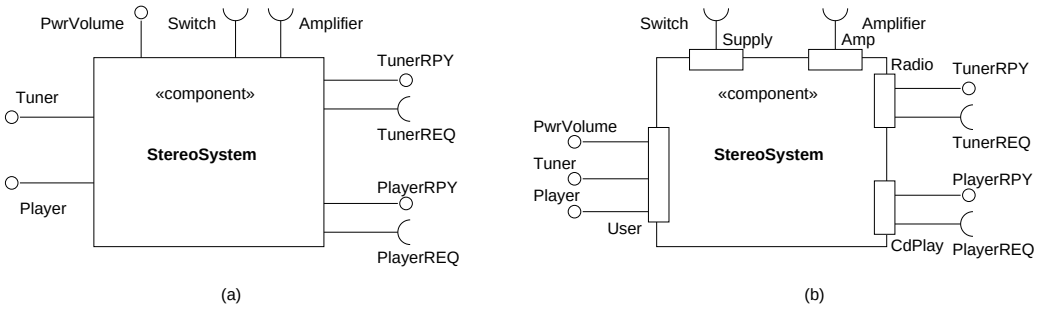


Figura 4.7: Componenti in UML.

Un componente può essere realizzato *direttamente* da un'istanza di una classe, o *indirettamente* da istanze di piú classi o componenti cooperanti; nel secondo caso è possibile mostrare la struttura interna del componente, come mostra la Fig. 4.8. In questo esempio supponiamo che il componente venga implementato dalle istanze di alcune classi. Le frecce fra i port e le interfacce delle classi rappresentano la relazione di *delega*, ed è possibile etichettarle esplicitamente con la parola chiave ≪delegate≫. Questa relazione mostra la provenienza o la destinazione delle comunicazioni (chiamate di operazioni e trasmissioni di eventi) passanti attraverso i port. La linea fra l'istanza di **Amplif** e quella di **Tuning** rappresenta l'interazione fra queste due parti del componente.

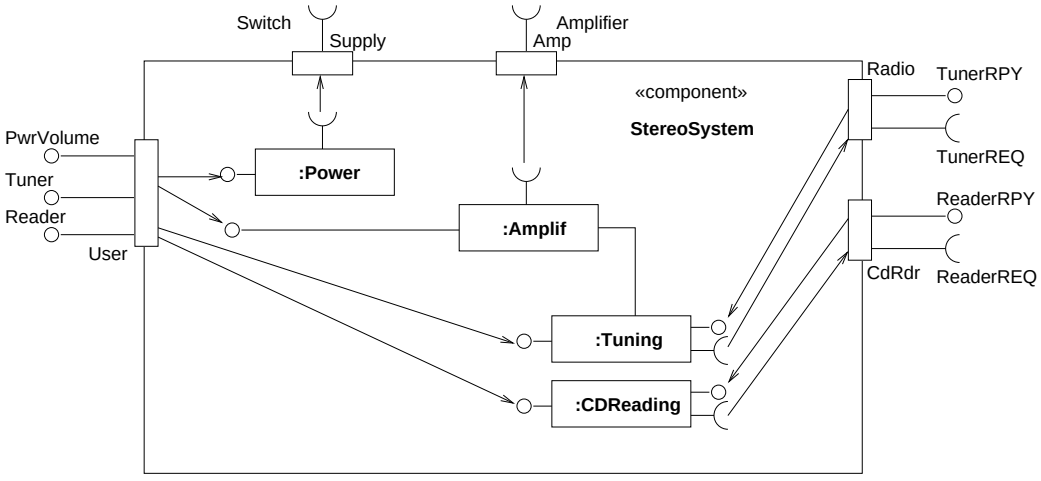


Figura 4.8: Componenti strutturati in UML.

### Incapsulamento e raggruppamento in C++

Nel linguaggio C++ l'incapsulamento e il raggruppamento si ottengono con i costrutti `class` e `namespace`.

**Classi e interfacce** In questo corso si presuppone la conoscenza del linguaggio C++ almeno nei suoi aspetti fondamentali, per cui non ci soffermeremo sul concetto di *classe* come definito in questo linguaggio, che è molto vicino a quello definito nell'UML. Ricordiamo soltanto che in C++ le operazioni virtuali pure corrispondono alle operazioni astratte dell'UML, e che una classe con almeno un'operazione virtuale pura è una classe astratta.

Una classe senza membri dato, che abbia solo operazioni virtuali pure con visibilità pubblica, permette di esprimere in C++ il concetto di *interfaccia*. Ricordando che un'interfaccia UML può contenere dichiarazioni di attributi, osserviamo che un attributo di un'interfaccia UML può essere rappresentato in C++ per mezzo di una coppia di operazioni virtuali pure, una per assegnare un valore all'attributo, e una per leggerlo.

**Namespace** Una dichiarazione `namespace` definisce uno spazio di nomi, offrendo un meccanismo di raggruppamento piú semplice e generale delle classi. Infatti una classe, oltre a raggruppare delle entità, definisce le proprietà comuni di un insieme di oggetti che si possono istanziare, mentre uno spazio di nomi si limita ad organizzare un insieme di definizioni. Uno spazio di nomi può contenere definizioni di qualsiasi genere, incluse classi ed altri spazi di nomi. Questo costrutto può essere usato per rappresentare, al livello di astrazione del codice sorgente, concetti architetturali piú astratti, come package e componenti.

Il seguente esempio mostra la dichiarazione di un `namespace` (contenuta in un file di intestazione `Linear.h`) che raggruppa due classi:

```
// Linear.h

namespace Linear {
    class Vector {
        double* p;
        int  size;
    public:
        enum { max = 32000 };
        Vector(int s);
        double& operator[](int i);
        // ...
    };

    class Matrix {
        // ...
    };
}
```

L'implementazione del modulo la cui interfaccia è definita dallo spazio di nomi `Linear` può essere definita in un altro file, `Linear.cc`:

```
// Linear.cc

#include "Linear.h"
```

```
using namespace Linear;

Vector::Vector(int s)
{
    p = new double[size = s];
}

double&
Vector::operator[](int i)
{
    return p[i];
}

// altre definizioni per Vector e Matrix
// ...
```

La direttiva `#include`, interpretata dal preprocessore, serve a includere il file `Linear.h`, rendendo cosí visibili al compilatore le dichiarazioni contenute in tale file. In questo caso, al modulo logico specificato dal `namespace Linear` corrisponde un modulo fisico costituito dall'unità di compilazione formata da `Linear.cc` e `Linear.h`. La direttiva `using namespace`, interpretata dal compilatore, dice che quando il compilatore trova dei nomi non dichiarati, deve cercarne le dichiarazioni nello spazio di nomi `Linear`. Questa direttiva, quindi, rappresenta approssimativamente la relazione ≪import≫ fra i package. Alternativamente all'uso della direttiva `using namespace` si possono usare delle dichiarazioni `using`, una per ciascun nome appartenente a `Linear` usato nell'unità di compilazione:

```
using Linear::Vector;
```

Oppure, si può qualificare esplicitamente ogni occorrenza di nomi appartenenti a `Linear`. Per esempio, la definizione del costruttore della classe `Vector` può essere scritta cosí:

```
int&
Linear::Vector::operator[](int i)
{
    // ...
}
```

Un modulo cliente deve includere il file di intestazione e usare i nomi in esso dichiarati, con le convenzioni linguistiche già viste:

```
// main.cc

#include <iostream>
#include "Linear.h"

using namespace Linear;

main()
{
    Vector v(4);

    v[0] = 0.0;
    // ...
}
```

La direttiva `using namespace` può essere usata per comporre spazi di nomi, come nel seguente esempio:

```
// Linear.h

namespace Linear {
    class Vector {
        // ...
    };

    class Matrix {
        // ...
    };
}
.............................
// Trig.h

namespace Trig {
    double sin(double x);
    double cos(double x);
    // ...
}
.............................
// Math.h

#include "Linear.h"
#include "Trig.h"

namespace Math {
    using namespace Linear;
    using namespace Trig;
}
```

### 4.4.2 Moduli generici

È molto comune che un algoritmo si possa applicare a tipi di dato diversi (per esempio, un algoritmo di ordinamento si può applicare a una sequenza di numeri interi, o di reali, o di stringhe), oppure che diversi tipi di dati, ottenuti per composizione di tipi piú semplici, abbiano la stessa struttura (per esempio, liste di interi, di reali, o di stringhe). I moduli generici permettono di "mettere a fattor comune" la struttura di algoritmi e tipi di dato, mettendo in evidenza la parte variabile che viene rappresentata dai parametri del modulo.

Questo è utile dal punto di vista della comprensibilità del progetto, in quanto rende esplicito il fatto che certi componenti del sistema sono fatti allo stesso modo, e da quello della facilità di programmazione, in quanto permette di riusare facilmente dei componenti che altrimenti verrebbero riprogettati daccapo o adattati manualmente da quelli già esistenti, col rischio di introdurre errori. Inoltre in molti casi l'uso di moduli generici permette di realizzare programmi piú efficienti, poiché generalmente vengono istanziati (cioè tradotti in moduli concreti, con valori definiti dei parametri) a tempo di compilazione.

#### Moduli generici in UML

In UML si possono rappresentare elementi di modello generici, e in particolare classi e package. La rappresentazione grafica di una classe generica è simile a quella di una classe non generica, con un rettangolo tratteggiato contenente i nomi ed eventualmente i tipi dei parametri, come in Fig. 4.9. I parametri possono essere nomi di tipi (classi o tipi base), valori, operazioni, e si possono indicare dei valori di default.

La relazione fra una classe generica ed una sua istanza si può rappresentare esplicitamente con una dipendenza etichettata dallo stereotipo ≪bind≫ accompagnato dai valori dei paramentri (Fig. 4.10) o in modo implicito, in cui la classe istanza viene etichettata dal nome della classe generica seguito dai valori dei parametri (Fig. 4.11).

Precisiamo che un'istanza di una classe generica è una classe, *non* un oggetto. Istanziare una classe generica e istanziare una classe sono due operazioni molto diverse, ma purtroppo nella terminologia corrente si usa la stessa parola.

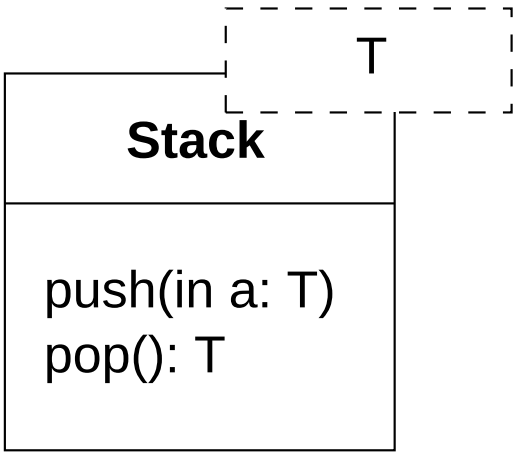


Figura 4.9: Classi generiche in UML (1).

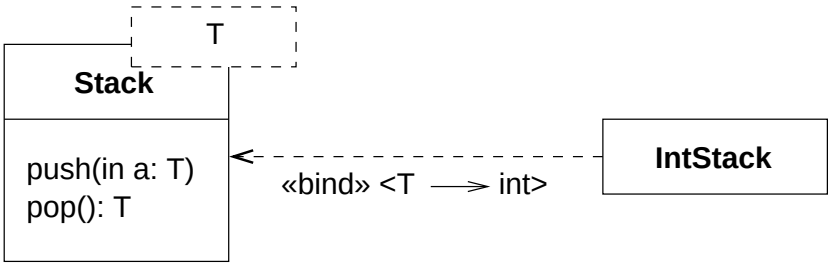


Figura 4.10: Classi generiche in UML (2).

### Moduli generici in C++

In C++ i moduli generici si realizzano col meccanismo dei template, o classi modello. Di seguito mostriamo una possibile definizione di uno stack di elementi di tipo generico:

```
// Stack.h

template<class T>
class Stack {
    T* v;
    T* p;
    int size;
public:
    Stack(int s);
    ~Stack();
    void push(T a);
```

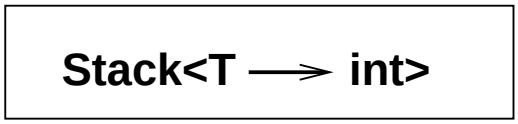


Figura 4.11: Classi generiche in UML (3).

```
    T pop();
};

template<class T>
Stack<T>::Stack(int s)
{
    v = p = new T[size = s];
}

// altre definizioni
// ...
```

Questo modulo generico può essere usato in un modulo cliente come questo:

```
#include <iostream.h>
#include "Stack.h"

main()
{
    Stack<int> si(10);

    si.push(3);
    si.push(4);
    cout << si.pop() << endl << si.pop() << endl;
}
```

In C++, una classe generica viene istanziata implicitamente quando si dichiara un oggetto. Spetta all'ambiente di programmazione generare il codice nel modo opportuno (possibilmente senza duplicazioni).

Una classe generica può essere parametrizzata anche rispetto a valori costanti e a funzioni. Nell'esempio seguente la classe `VectOp` fornisce due operazioni su array di elementi generici: la funzione `elementwise()` applica a due array, elemento per elemento, un'operazione binaria generica `Binop()`, mentre la funzione `fold()` applica la stessa operazione, in sequenza, a tutti gli elementi di un array, "accumulando" il risultato (questo modo di applicare

un'operazione agli elementi di una sequenza si chiama *folding* nel campo della programmazione funzionale). La funzione `elementwise()` è analoga alla somma vettoriale, mentre la `fold()` è analoga alla sommatoria di una successione finita di numeri. La classe `VectOp` è quindi generica rispetto a tre parametri: il tipo `T` degli elementi delle sequenze (rappresentate come array) su cui si opera, la dimensione `Size` delle sequenze, e l'operazione `Binop` da impiegare nei due algoritmi `elementwise()` e `fold()`[6].

```
// VectOp.h

template<class T, int Size, class T (*Binop)(class T, class T)>
class VectOp {
public:
    static void elementwise(T*, T*, T*);
    static void fold(T*, T&);
};

template<class T, int Size, class T (*Binop)(class T, class T)>
inline void VectOp<T,Size,Binop>::elementwise(T* a, T* b, T* r)
{
    for (int i = 0; i < Size; i++)
        r[i] = Binop(a[i], b[i]);
}

template<class T, int Size, class T (*Binop)(class T, class T)>
inline void VectOp<T,Size,Binop>::fold(T* a, T& r)
{
    for (int i = 0; i < Size; i++)
        r = Binop(r, a[i]);
}
```

Questa classe può essere usata come nel seguente esempio:

```
#include <iostream>
#include <string>
#include "VectOp.h"

string concatenate(string s1, string s2)
{
    return s1 + s2;
}

main()
{
    const int SIZE = 3;
    VectOp<string, SIZE, concatenate> vo;
```

[6]Si confronti questo esempio con le funzioni modello `accumulate()`, `inner_product()`, `partial_sum()` e `adjacent_differences()` della libreria standard del C++.

```
        string seq1[] = { "blue ", "white ", "green ", };
        string seq2[] = { "sky", "snow", "forest", };

        string res_seq[SIZE];
        vo.elementwise(seq1, seq2, res_seq);
        for (int i = 0; i < SIZE; i++)
            cout << res_seq[i] << endl;

        string res_string;
        vo.fold(seq1, res_string);
        cout << res_string << endl;
    }
```

Osserviamo che la libreria standard del C++ (*Standard Template Library*, STL) offre numerose classi e algoritmi generici, che permettono di risolvere molti problemi di programmazione in modo affidabile ed efficiente.

### 4.4.3 Eccezioni

Come già accennato, la gestione degli errori e delle situazioni anomale è un aspetto della suddivisione di responsabilità fra moduli. I linguaggi che offrono il meccanismo delle *eccezioni* permettono di esprimere in modo chiaro e ben strutturato le soluzioni di questo problema. Il problema consiste nel decidere in quale modulo si deve trattare una data situazione eccezionale, dopodiché bisogna specificare in quali moduli si può verificare tale situazione, come viene riconosciuta, come viene segnalata, e infine come il controllo viene passato al modulo responsabile di gestire l'errore.

Nei linguaggi dotati di gestione delle eccezioni, i moduli in cui si può verificare un'eccezione contengono delle istruzioni che *sollevano*, cioè segnalano tale eccezione. I moduli preposti a gestire l'eccezione contengono dei sottoprogrammi, detti *gestori* o *handler*, esplicitamente designati a tale scopo. A tempo di esecuzione, quando in un modulo si scopre il verificarsi di una situazione eccezionale (cosa che richiede dei controlli espliciti, tipicamente con istruzioni condizionali), l'istruzione che solleva l'eccezione causa l'interruzione dell'esecuzione del modulo ed il trasferimento del controllo al gestore "piú vicino" nella catena di chiamate che ha portato all'invocazione del modulo che ha sollevato l'eccezione.

Per fissare le idee, supponiamo che un modulo `A` chiami una funzione definita in un modulo `B`, che a sua volta chiama una funzione di un modulo `C`. Supponiamo inoltre che nell'esecuzione di `C` si possa verificare un'eccezione,

e che nel modulo `A` (ma non in `B`) ci sia un gestore per tale eccezione. Se, quando si esegue `C`, l'eccezione viene sollevata, allora vengono interrotte le esecuzioni di `C` e `B`, e il controllo passa al gestore appartenente ad `A`. In questo gestore, se necessario, si può risollevare l'eccezione, delegandone la gestione ad altri moduli ancora, di livello piú alto. Questo può accadere se nel modulo `A` si scopre che l'eccezione non può essere gestita, oppure se `A` può eseguire solo una parte delle azioni richieste.

### Eccezioni in UML

In UML1 le eccezioni si modellavano come oggetti che possono essere spediti (come segnali) da un oggetto all'altro. In questo modo, le classi che rappresentano eccezioni hanno lo stereotipo ≪exception≫. La dipendenza stereotipata ≪send≫ associa le eccezioni alle operazioni che le possono sollevare, come in Fig. 4.12.

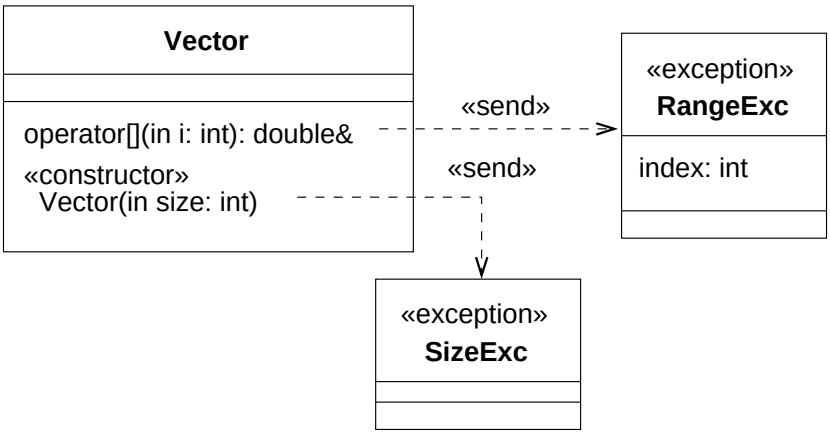


Figura 4.12: Eccezioni in UML.

In UML2 le eccezioni non sono viste come segnali, ma come oggetti che vengono creati quando l'esecuzione di un'azione incontra una situazione anomala e vengono passati come parametri d'ingresso ad un'altra azione, il gestore di interruzioni. Questo meccanismo viene modellato nel diagramma di attività, con apposite notazioni che non tratteremo. Nel diagramma delle classi, le eccezioni che si possono sollevare nel corso di una operazione possono essere elencate come valori della proprietà exceptions associata all'operazione.

### Eccezioni in C++

In C++ un'eccezione viene sollevata mediante un'istruzione `throw`, che ha un operando di tipo qualsiasi. Uno handler è costituito da un'istruzione `catch`, formata dalla parola `catch` seguita da una dichiarazione di argomento formale (analogamente alle dichiarazioni di funzione) e da un'istruzione. L'eccezione sollevata da un'istruzione `throw` viene gestita da uno handler il cui argomento formale ha lo stesso tipo dell'operando della `throw`. Conviene definire delle classi destinate a rappresentare i diversi tipi di eccezione (non esistono eccezioni predefinite dal linguaggio, ma alcune eccezioni sono definite nella libreria standard). Nel seguente esempio vengono definite le classi `RangeExc` e `SizeExc`, che rappresentano rispettivamente le eccezioni sul valore dell'indice di un elemento e sulla dimensione del vettore. Le due classi sono vuote (cosa permessa dalla sintassi del linguaggio), poiché in questo caso non ci interessa associare informazioni alle eccezioni.

Nel file di intestazione della classe `Vector`, le eccezioni sollevate da ciascuna funzione vengono dichiarate esplicitamente per mezzo delle clausole `throw` (diverse dalle *istruzioni* `throw`):

```
// Vector.h

class Vector {
    double* p;
    int  size;
public:
    enum { max = 32000 };
    class RangeExc {};
    class SizeExc {};
    Vector(int s) throw (SizeExc);
    double& operator[](int i) throw (RangeExc);
    // ...
};
```

Le clausole `throw` vengono ripetute anche nelle definizioni, in modo che il compilatore possa verificare la coerenza fra dichiarazioni e definizioni:

```
// Vector.cc

#include "Vector.h"

Vector::Vector(int s) throw (SizeExc)
{
    if (s < 0 || max < s)
        throw SizeExc();
```

```
    p = new double[size = s];
}

double&
Vector::operator[](int i) throw (RangeExc)
{
    if (0 <= i && i < size)
        return p[i];
    else
        throw RangeExc();
}
```

In questo esempio gli operandi delle istruzioni `throw` sono i costruttori delle due classi: tali costruttori producono oggetti vuoti.

Le istruzioni la cui esecuzione può sollevare un'eccezione che vogliamo gestire vengono raggruppate in un blocco `try`. Questo blocco viene seguito dagli handler:

```
#include <iostream>
#include "Vector.h"

double
g(Vector v, int i)
{
    return v[i];
}

main()
{
                                    // eccezioni non gestite
    Vector v(4);

    for (int i = 0; i < 4; i++)
        v[i] = i * 2.3;
    try {                           // eccezioni catturate
        double x = g(v, 5);
        cout << x << endl;
    } catch (Vector::RangeExc) {
        cerr << "Range Exception" << endl;
    } catch (Vector::SizeExc) {
        cerr << "Size Exception" << endl;
    }
}
```

Se si verificano eccezioni nell'esecuzione di istruzioni esterne al blocco `try`, queste eccezioni vengono propagate lungo la catena delle chiamate finché non si trova un gestore. Se non ne vengono trovati, il programma termina.

Si possono associare informazioni alle eccezioni, usando classi non vuote:

```
class Vector {
    // ...
public:
    enum { max = 32000 };
    class RangeExc {
    public:
        int index;
        RangeExc(int i) : index(i) {};
    };
    class SizeExc {};
    Vector(int s) throw (SizeExc);
    double& operator[](int i) throw (RangeExc);
    // ...
};

double&
Vector::operator[](int i) throw (RangeExc)
{
    if (0 <= i && i < size)
        return p[i];
    else
        throw RangeExc(i);
}
```

L'informazione associata all'eccezione può quindi essere usata in uno handler, se questo ha un argomento formale che denota l'eccezione lanciata:

```
main()
{
    // ...
    try {
        // ...
    } catch (Vector::RangeExc r) {
        cerr << "bad index: " << r.index << endl;
    } catch (Vector::SizeExc) {
        // ...
    }
}
```

Le eccezioni possono essere raggruppate usando l'eredità: per esempio, in un modulo che esegue dei calcoli numerici si possono definire delle eccezioni che rappresentano errori aritmetici generici (`MathError`), e da queste si possono derivare eccezioni piú specifiche, come il traboccamento (`Ovfl`) o la divisione per zero (`ZeroDiv`).

```
class MathError {};
class Ovfl : public MathError {};
class ZeroDiv : public MathError {};
```

Un modulo cliente può trattare le eccezioni in modo differenziato:

```
try {
    // ...
} catch (Ovfl) {
    // ...
} catch (MathError) {
    // ...
} catch ( ... ) {
    // ...
}
```

Nell'ultimo handler, i punti di sospensione sono la sintassi usata per specificare che lo handler gestisce qualsiasi eccezione. Quando viene sollevata un'eccezione, il controllo passa al primo gestore, in ordine testuale, il cui argomento formale ha un tipo uguale a quello dell'eccezione, o un tipo piú generale. Da questa regola segue che gli handler piú generali devono seguire quelli piú specializzati, altrimenti questi ultimi non possono essere eseguiti.

# Letture

**Obbligatorie:** Cap. 3 e 4 Ghezzi, Jazayeri, Mandrioli.

# Capitolo 5

# Progetto orientato agli oggetti

*We need to go beyond the condemnation of spaghetti code to the active encouragement of ravioli code.*
*– Raymond J. Rubey, SoftTech, Inc.*

Come abbiamo già osservato parlando della fase di analisi e specifica dei requisiti, nelle metodologie orientate agli oggetti un sistema viene visto come un insieme di oggetti interagenti e legati fra loro da vari tipi di relazioni. In fase di analisi e specifica dei requisiti si costruisce un modello in cui il sistema software (cioè il prodotto che vogliamo realizzare) viene visto nel contesto dell'ambiente in cui deve operare. Questo modello, in cui le entità appartenenti al dominio dell'applicazione (detto anche *spazio del problema*) sono preponderanti rispetto al sistema software, è il punto di partenza per definire l'architettura software, la cui struttura, negli stadi iniziali della progettazione, ricalca quella del modello di analisi.

La fase di progetto parte quindi dalle classi e relazioni definite in fase di analisi, a cui si aggiungono classi definite in fase di progetto di sistema e relative al dominio dell'implementazione (o *spazio della soluzione*). Nelle fasi successive del progetto questa struttura di base viene rielaborata, riorganizzando le classi e le associazioni, introducendo nuove classi, e definendone le interfacce e gli aspetti piú importanti delle implementazioni.

### 5.0.4 Un esempio

Immaginiamo di progettare un sistema per la gestione di una biblioteca. La Fig. 5.1 dà un'idea estremamente semplificata di un possibile modello d'ana-

lisi. Ci sono due diagrammi di casi d'uso, di cui il secondo tiene conto del requisito che solo il bibliotecario interagisca direttamente col sistema. Si è mantenuta la versione originale del diagramma, che mostra il ruolo dell'utente, per ricordare che i servizi della biblioteca sono destinati all'utente, e lasciare spazio a future estensioni in cui l'utente potrebbe accedere al sistema direttamente. La dipendenza ≪trace≫ si usa per rappresentare la relazione storica fra diverse versioni di un modello o parti di esso.

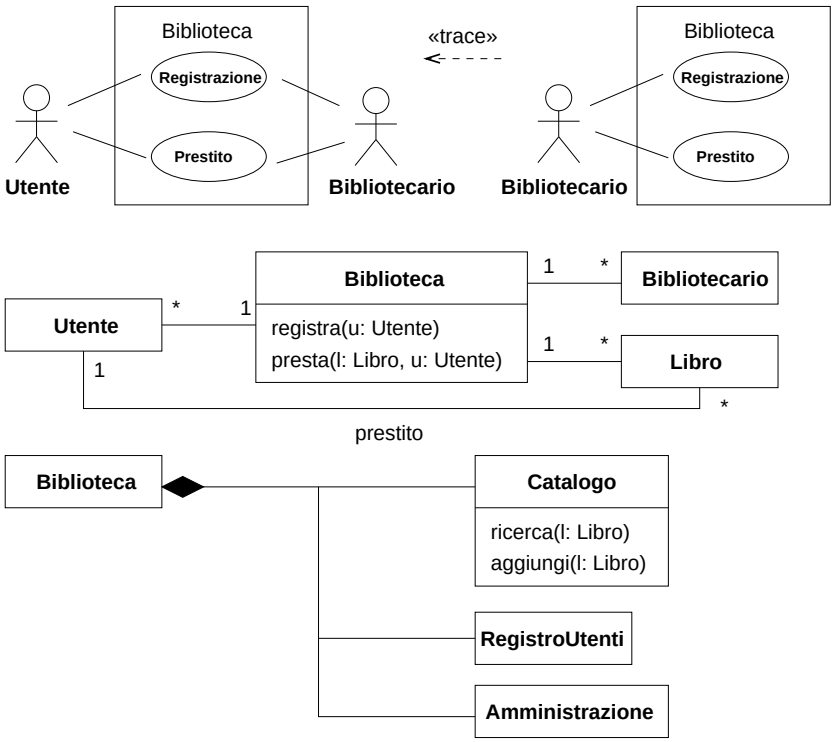


Figura 5.1: Esempio: un progetto OO (1).

La Fig. 5.2 mostra un primo abbozzo del modello di progetto. È stata introdotta una classe (che in questo stadio sta per un intero sottosistema) che rappresenta l'interfaccia presentato dal sistema al bibliotecario. Questa classe ha lo stereotipo ≪boundary≫ appunto per mostrare questo suo ruolo. È stato introdotto conseguentemente un registro dei bibliotecari. Le classi **Libro**, **Utente** e **Bibliotecario** hanno lo stereotipo ≪entity≫, comunemente usato per le classi destinate a contenere dati persistenti.

La Fig. 5.3 mostra un'ulteriore evoluzione del modello, in cui si implementa l'associazione logica prestito per mezzo di una classe apposita e di un registro dei prestiti incluso nel sottosistema di amministrazione della biblio-

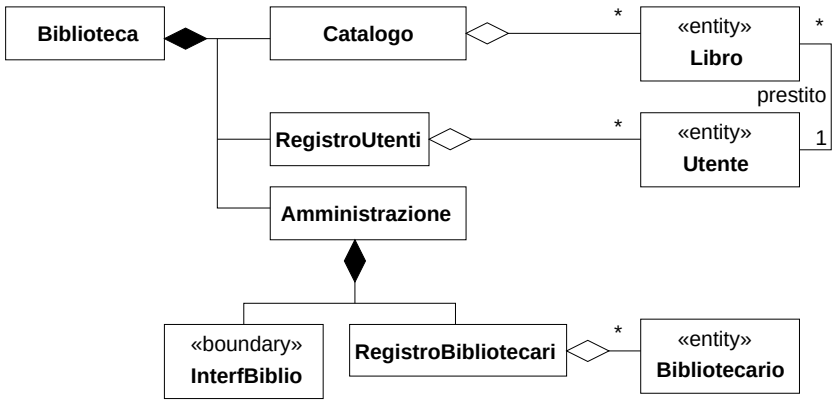


Figura 5.2: Esempio: un progetto OO (2).

teca. Inoltre si è scelto di implementare le varie aggregazioni per mezzo di classi istanziate dal template **list**, e si è aggiunto un sottosistema di persistenza, che deve provvedere a memorizzare i vari cataloghi e registri in un database.

## 5.1 Eredità e polimorfismo

In questa sezione tratteremo alcune tecniche di programmazione che sono alla base della progettazione orientate agli oggetti: l'*eredità*, il *polimorfismo* e il *binding dinamico*.

### 5.1.1 Eredità

Nella fase di analisi dei requisiti, l'eredità permette di modellare la relazione di generalizzazione (e quindi di specializzazione) fra entità del dominio di applicazione. Di solito, alle entità del dominio dell'applicazione devono corrispondere delle entità del sistema software. Nella fase di progetto viene definita un'architettura software in cui vengono mantenute le relazioni fra entità del dominio di applicazione, fra cui la generalizzazione, come mostra il seguente esempio (in C++):

```
class Person {
    char* name;
    char* birthdate;
```

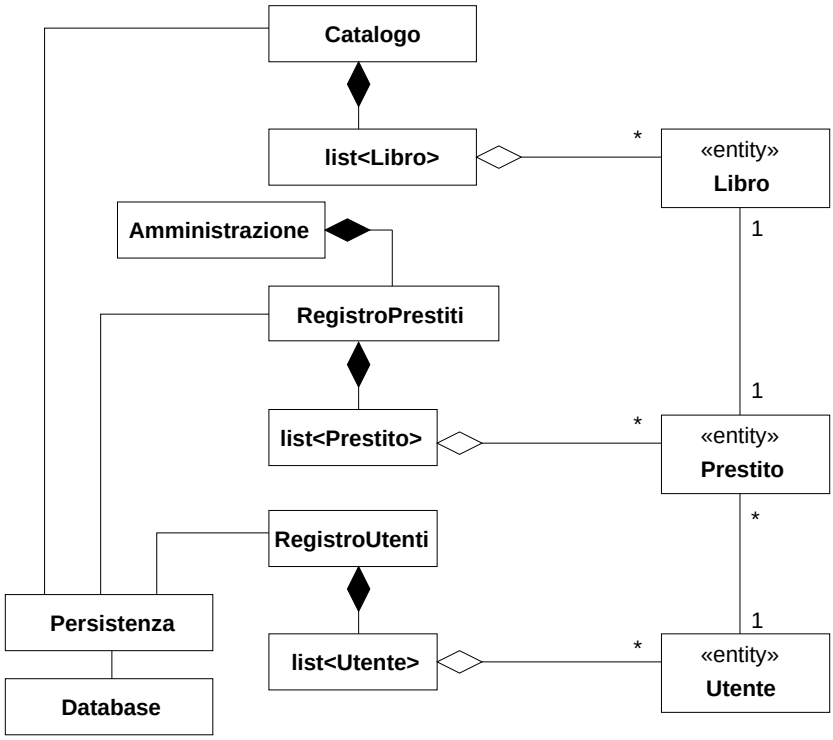


Figura 5.3: Esempio: un progetto OO (3).

```
public:
    char* getName();
    char* getBirthdate();
};

class Student : public Person {
    char* student_number;
public:
    char* getStudentNumber();
};
```

In questo caso il fatto che la classe `Student` erediti dalla classe `Person` corrisponde al fatto che nel dominio dell'applicazione (per esempio, il database dei dipendenti e degli studenti di una scuola) gli studenti sono un sottoinsieme delle persone. La classe derivata ha gli attributi e le operazioni della classe base, a cui aggiunge ulteriori attributi e operazioni.

In fase di progetto si può usare l'eredità come tecnica di riuso, cioè per riusare le operazioni di una classe base nel contesto di una classe derivata:

per esempio, se si dispone di una classe **Shape** con le operazioni per disegnare e spostare una figura geometrica, si può definire una classe derivata **ColoredShape** che aggiunge le operazioni per colorare una figura, e riusa le operazioni della classe base per disegnarla e spostarla. Questo è conveniente quando la classe derivata ha una semantica (cioè uno scopo e un significato) analoga a quella della classe base, rispettando cosí il principio di Liskov. In molti casi, però, il meccanismo della composizione (usare un'istanza di una classe, o un puntatore ad essa, come attributo di un'altra) è piú flessibile dell'eredità. È comunque da evitare l'uso dell'eredità per costruire una classe derivata che non abbia una parentela logica con la classe base.

### 5.1.2 Polimorfismo e binding dinamico

Il polimorfismo ed il binding dinamico sono due concetti che nei linguaggi orientati agli oggetti sono strettamente legati fra di loro e con il concetto di eredità.

Il *polimorfismo* è la possibilità che un riferimento (per esempio un identificatore o un puntatore) denoti oggetti o funzioni di tipo diverso. Esistono diversi tipi di polimorfismo, e la forma di polimorfismo tipica dei linguaggi object oriented è quella basata‘ sull'eredità: se una classe $D$ deriva da una classe $B$, allora ogni istanza di $D$ è anche un'istanza di $B$, per cui qualsiasi riferimento alla classe $D$ è anche un riferimento alla classe $B$. Questo tipo di polimorfismo si chiama *polimorfismo per inclusione*. In C++, per esempio, un oggetto di classe $D$ può essere assegnato ad un oggetto di classe $B$ ed un valore di tipo "puntatore a $D$" o "riferimento a $D$" può essere assegnato, rispettivamente, ad una variabile di tipo "puntatore a $B$" o "riferimento a $B$". Osserviamo che in questo e in altri linguaggi esistono altre forme di polimorfismo, come l'overloading, che non sono legate all'eredità.

Il *binding* è il legame fra un identificatore (in particolare un identificatore di funzione) ed il proprio valore. Si ha un *binding* dinamico quando il significato di una chiamata di funzione (cioè il codice eseguito dalla chiamata) è noto solo a tempo di esecuzione: il binding dinamico è quindi il meccanismo che rende possibile il polimorfismo. In C++, le funzioni che vengono invocate con questo meccanismo sono chiamate *virtuali*. Se si chiama una funzione virtuale di un oggetto attraverso un puntatore, questa chiamata è polimorfica (per inclusione). Consideriamo questo (classico) esempio:

```
class Shape {
    Point position;
```

```
public:
    virtual void draw() {};
    virtual void move(Point p) { position = p; };
};

class Circle : public Shape {
    //...
public:
    void draw() { cout << ''drawing Circle\n''; };
};

class Square : public Shape {
    //...
public:
    void draw() { cout << ''drawing Square\n''; };
};

void drawall(Shape** shps)
{
    for (int i = 0; i < 2; i++)
        shps[i]->draw();
}

main()
{
    Shape* shapes[2];
    shapes[0] = new Circle;
    shapes[1] = new Square;
    drawall(shapes);
}
```

La struttura del programma può essere rapresentata in UML come in Fig. 5.4, dove lo stereotipo ≪utility≫ índica che una classe contiene solo operazioni o dati globali (non è quindi una vera classe, ma un espediente per rappresentare costrutti non orientati agli oggetti).

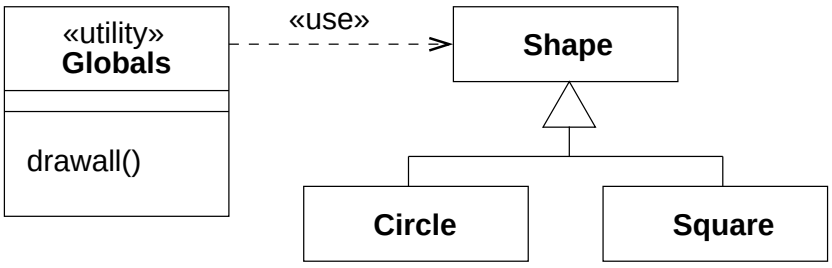


Figura 5.4: Esempio: polimorfismo (1)

Nella funzione `drawall()` il tipo dell'oggetto a cui viene applicata la fun-

zione `draw()` (che deve disegnare una figura sullo schermo) è noto soltanto a tempo di esecuzione: si sa solo che l'oggetto apparterrà alla classe `Shape` o a una classe derivata da questa. La chiamata a `draw()` in questa funzione è quindi polimorfica, ed ha un binding dinamico. Il fatto che la funzione `drawall()` ignori il tipo degli oggetti su cui deve operare ne migliora grandemente la modularità e la riusabilità rispetto ad un'implementazione che invece richieda una conoscenza statica dei tipi. Questa funzione è completamente disaccoppiata dall'implementazione delle classi `Circle` e `Square`, e inoltre continua a funzionare, immutata, anche se si aggiungono altre classi derivate da `Shape`.

### 5.1.3 Classi astratte e interfacce

Osserviamo che nel nostro esempio l'operazione `draw()` della classe `Shape` ha un'implementazione banale, è un'operazione nulla, poiché non si può disegnare una forma generica: il concetto di "forma" (*shape*) è astratto, e si possono disegnare effettivamente solo le sue realizzazioni concrete, come "cerchio" e "quadrato". La classe `Shape` è in realtà un cattivo esempio di programmazione, poiché non rappresenta adeguatamente il concetto reale che dovrebbe modellare, e questa inadeguatezza porta alla realizzazione di software poco affidabile. Infatti è possibile istanziare (contro la logica dell'applicazione) un oggetto `Shape` a cui si potrebbe applicare l'operazione `draw()`, ottenendo un risultato inconsistente.

Una prima correzione a questo errore di progetto consiste nel rendere esplicito il fatto che la classe rappresenta un concetto astratto. In C++, questo si ottiene specificando che `draw` è un'operazione virtuale pura, per mezzo dello specificatore '`= 0`':

```
class Shape {
    Point position;
public:
    virtual void draw() = 0;
    virtual void move(Point p) { position = p; };
    //...
};
```

La classe Shape è ora una classe astratta, cioè non istanziabile a causa dell'incompletezza della sua implementazione (Fig. 5.5).

Una struttura ancor piú modulare si può ottenere rappresentando esplicitamente l'interfaccia, separandola dall'implementazione.

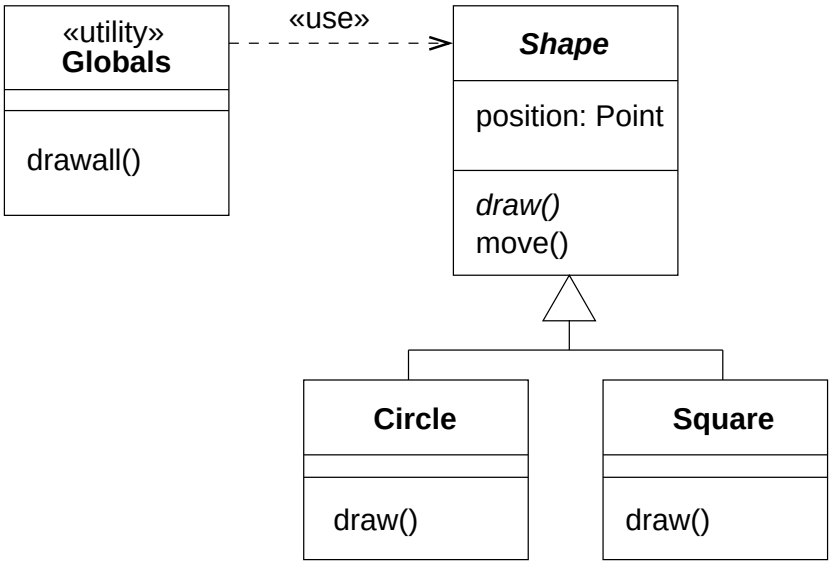


Figura 5.5: Esempio: polimorfismo (2)

```
class IShape {
public:
    virtual void draw() = 0;
    virtual void move(Point p) = 0;
};

class Shape : public IShape {
    Point position;
public:
    virtual void draw() = 0;
    virtual void move(Point p) { position = p; };
};

class Circle : public Shape {
    //...
public:
    void draw() { cout << ''drawing Circle\n''; };
};

class Square : public Shape {
    //...
public:
    void draw() { cout << ''drawing Square\n''; };
};

void drawall(IShape** shps)
{
    for (int i = 0; i < 2; i++)
        shps[i]->draw();
```

```
}

main()
{
    IShape* shapes[2];
    shapes[0] = new Circle;
    shapes[1] = new Square;
    drawall(shapes);
}
```

In questa versione la classe `IShape` definisce l'interfaccia comune a tutti gli oggetti che possono essere disegnati o spostati, la classe `Shape` definisce la parte comune delle loro implementazioni (il fatto di avere una posizione e un'operazione che modifica tale posizione), e le classi rimanenti definiscono concretamente i metodi per disegnare le varie figure. Questa struttura si può schematizzare come in Fig. 5.6.

### 5.1.4 Eredità multipla

L'eredità multipla si ha quando una classe eredita "in parallelo" da due o piú classi. Questa possibilità è particolarmente utile per ottenere un'interfaccia come composizione di altre interfacce. Le interfacce delle classi base rappresentano diversi aspetti delle entità modellate dalla classe derivata. I clienti della classe derivata possono trattare separatamente questi diversi aspetti, ottenendo cosí un buon disaccopiamento fra le varie classi.

Nel seguente esempio (in Java) un videogioco deve simulare dei velivoli[1]. Ciascuno di questi viene visto da due punti di vista: la simulazione del comportamento (`Aircraft`) e la raffigurazione grafica (`Drawable`). Questi due aspetti vengono gestiti da due sottosistemi distinti, rispettivamente `AirTrafficCtrl` e `DisplayMgr`. Il sistema è schematizzato in Fig. 5.7.

```
abstract class Aircraft {
    public double speed;
    public abstract void fly(AirTrafficCtrl atc);
}

interface Drawable {
    void draw();
}

class JetLiner extends Aircraft implements Drawable {
```

[1] In effetti, come risulta dal codice sorgente, viene simulato un solo velivolo.

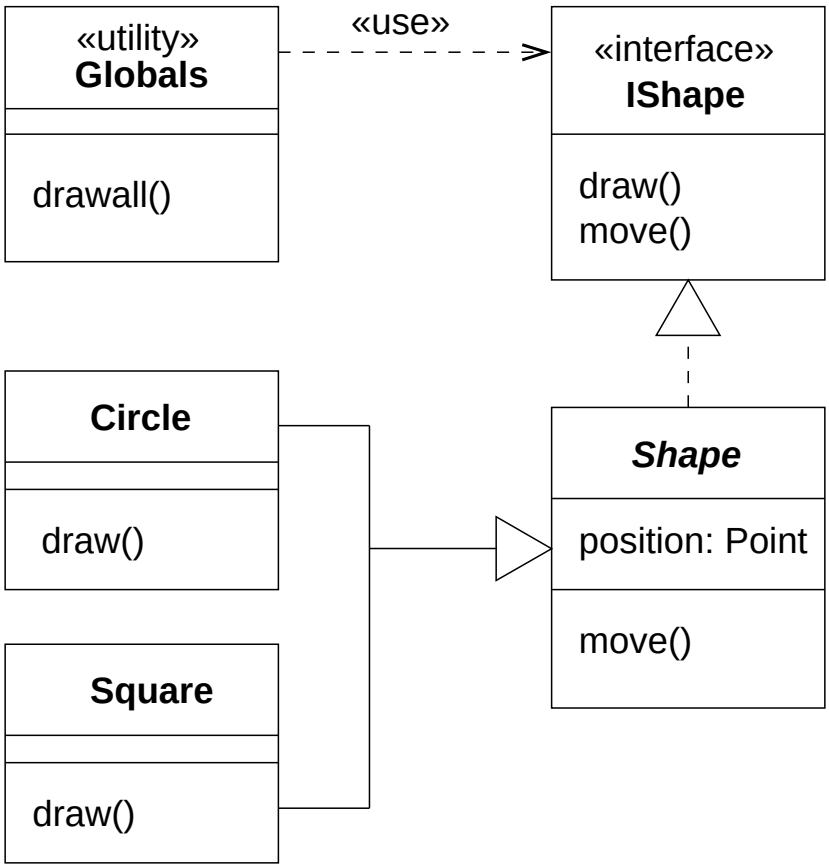


Figura 5.6: Esempio: polimorfismo (3)

```
    public JetLiner() { speed = 850.0; }
    public void fly(AirTrafficCtrl atc) { /* flight simulation */ }
    public void draw() { /* graphic rendering */ }
}

class DisplayMgr {
    private Drawable d;
    public DisplayMgr(Drawable dd) { d = dd; }
    public void display() { d.draw(); }
}

class AirTrafficCtrl {
    private Aircraft c;
    private DisplayMgr m;
    public AirTrafficCtrl(Aircraft ac, DisplayMgr dm)
       { c = ac; m = dm; }
    public void simulate() { c.fly(c); }
}
```

```
public class VideoGame {
    public static void main(String[] args)
    {
        JetLiner tu204 = new JetLiner();
        DisplayMgr dm = new DisplayMgr(tu204);
        AirTrafficCtrl atc = new AirTrafficCtrl(tu204, dm);
        dm.display();
        atc.simulate();
    }
}
```

In Java, come già osservato, un `interface` corrisponde ad una classe virtuale pura del C++. Le parole chiave `extends` e `implements` denotano l'eredità rispettivamente da classi ordinarie (eventualmente astratte) e da classi virtuali pure. Il Java permette di ereditare direttamente da una sola classe ordinaria e da piú classi virtuali pure.

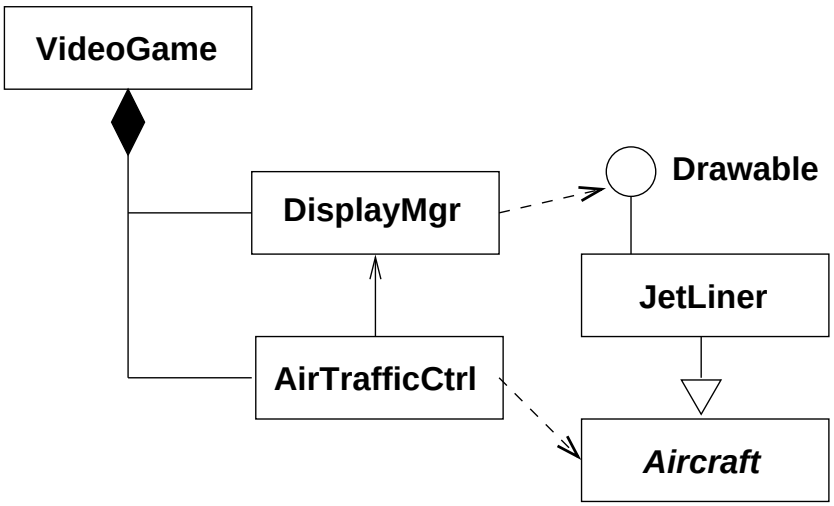


Figura 5.7: Esempio: polimorfismo (4)

## 5.2 Progetto di sistema

In questa sezione consideriamo alcune linee guida per il progetto di sistema, nel quale si fanno delle scelte di carattere fondamentale che indirizzano le attività successive. Le scelte principali riguardano la struttura generale del

sistema, di cui si individuano i componenti principali ed i meccanismi di interazione.

Altre scelte riguardano la gestione delle condizioni al contorno: bisogna cioè stabilire il comportamento del sistema nelle situazioni di inizializzazione, terminazione, ed errore. Bisogna anche stabilire delle priorità fra obiettivi contrastanti (per esempio, velocità di esecuzione e risparmio di memoria), in modo da risolvere conflitti fra soluzioni alternative che si possono presentare nel corso del progetto. Queste priorità, però, non dovranno esere applicate meccanicamente: per esempio, l'aver stabilito che la velocità di esecuzione è prioritaria rispetto al risparmio di memoria non giustifica un sovradimensionamento della memoria, se questo ha un costo eccessivo rispetto al guadagno di velocità ottenibile.

### 5.2.1 Ripartizione in sottosistemi

Il primo passo nel progetto di sistema consiste nell'individuare i sottosistemi principali. Riprendendo l'esempio del software per la gestione di una biblioteca, la Fig. 5.8 mostra la struttura di tale software a livello di sottosistemi.

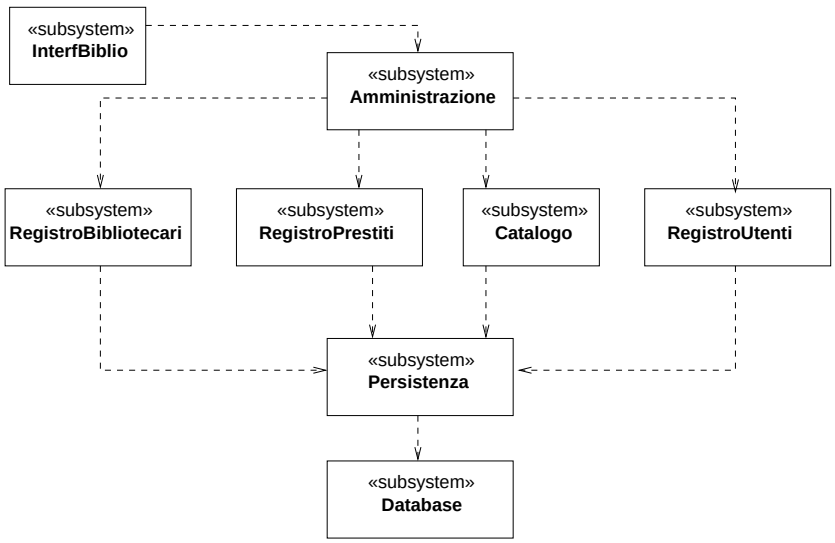


Figura 5.8: Esempio: progetto di sistema (1).

In questo esempio si sono mostrate solo le relazioni di dipendenza fra i moduli. Un passo successivo consiste nell'individuare le interfacce, mostrate

nel diagramma di Fig. 5.9, in cui si vede anche che si è scelto di implementare il database con un pacchetto software esterno, MySQL. La Fig. 5.10 mostra un'alra versione, in cui si tralascia la rappresentazione esplicita delle interfacce.

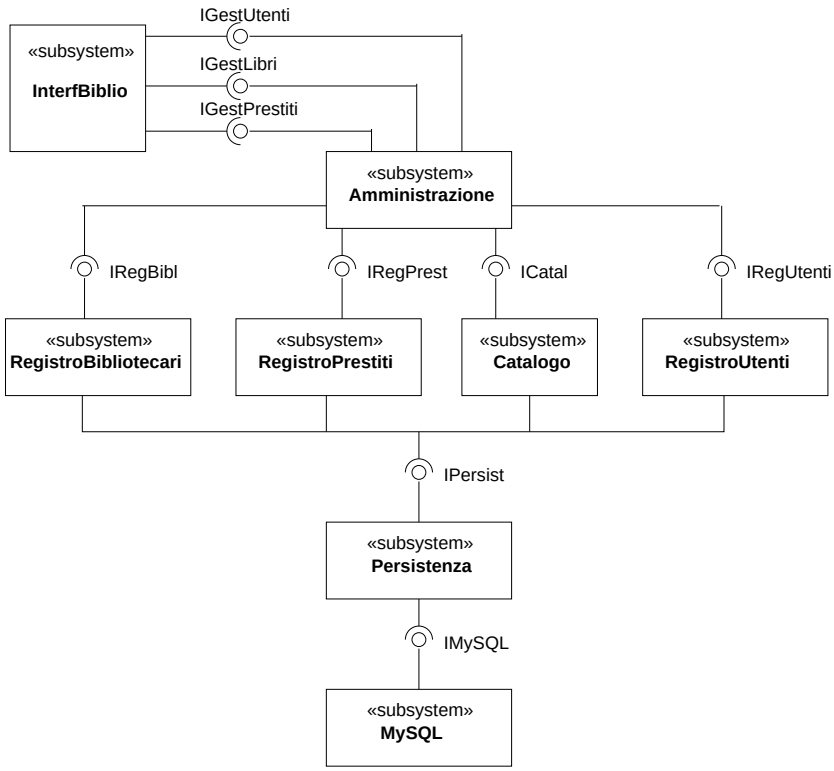


Figura 5.9: Esempio: progetto di sistema (2).

### Architetture standard

Esistono alcuni schemi di interconnessione (o *topologie*) fra sottosistemi che, insieme a determinati schemi di interazione, vengono usati comunemente e spesso sono caratteristici di certi tipi di applicazioni. Di solito conviene scegliere uno di questi schemi come punto di partenza del progetto,

Esempi di tali schemi sono le architetture a *pipeline*, *client-server* a uno o due livelli, a *repository*, mostrate in Fig. 5.11. Alcune applicazioni che usano tali architetture sono, rispettivamente, i compilatori, i servizi web piú semplici, i servizi web per l'accesso a basi di dati, gli strumenti CASE.

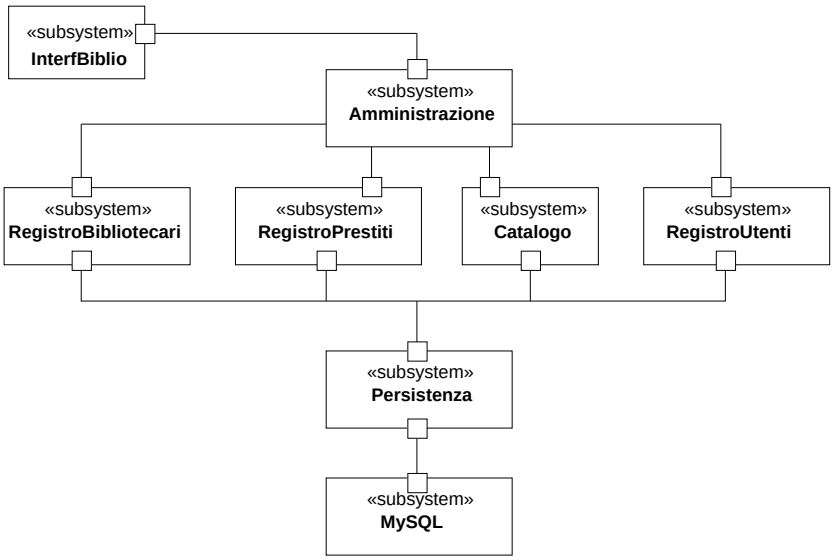


Figura 5.10: Esempio: progetto di sistema (3).

Osserviamo che il modello client-server ed il modello repository hanno la stessa topologia, ma una diversa interazione fra i sottosistemi: nel modello client-server i clienti sono indipendenti, mentre nel modello repository i vari componenti accedono al repository per scambiare dati e collaborare.

Altre categorie di applicazioni per cui si tende a usare delle architetture standard sono i sistemi interattivi, i simulatori, i sistemi real-time, i sistemi a transazioni. Quando s'intraprende il progetto di un nuovo sistema conviene verificare se esso appartiene ad una delle categorie note, e prendere a modello lo schema generale di architettura tipico di tale categoria.

### Scomposizione in strati e partizioni

Un metodo di applicabilità generale per organizzare un'architettura si basa sulla scomposizione per *strati* e per *partizioni*. Nella scomposizioni in strati ogni strato è un sottosistema che offre dei servizi ai sottosistemi di livello superiore e li implementa attraverso i servizi offerti dai sottosistemi a livello inferiore. Nella scomposizione in partizioni ogni partizione è un sottosistema che realizza una funzione del sistema complessivo. I due criteri generalmente vengono applicati insieme, poiché una partizione può essere stratificata ed uno strato può essere diviso in partizioni.

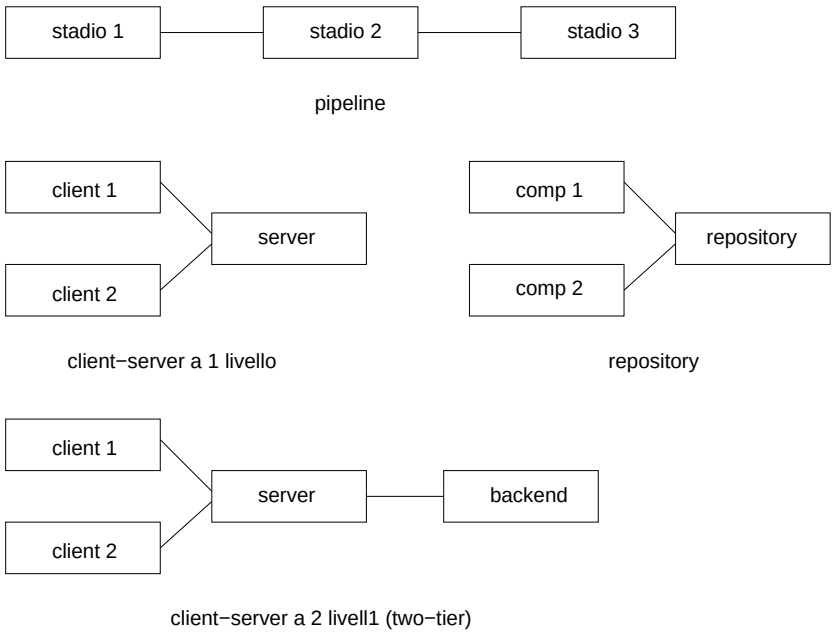


Figura 5.11: Architetture standard.

L'individuazione delle partizioni, cioè dei sottosistemi responsabili di realizzare le diverse funzioni dell'applicazione, viene guidata principalmente dalle informazioni raccolte nei documenti di specifica dei requisiti, per esempio dai diagrammi dei casi d'uso.

Gli strati vengono individuati in base ai diversi livelli di astrazione dei servizî richiesti per realizzare il sistema: esempî tipici di architetture a strati sono i sistemi operativi ed i protocolli di comunicazione.

Un esempio di scomposizione per strati è mostrato in Fig. 5.12, dove lo schema a componenti dell'esempio sul software per una biblioteca viene ridisegnato raggruppando i sottosistemi in cinque strati, rappresentati per mezzo di package. I quattro sottosistemi dello strato **Registri** sono partizioni dello stesso, e anch'esse avrebbero potuto essere messe in evidenza con dei package.

Un altro modo di rappresentare una struttura a strati e partizioni viene mostrato in Fig. 5.13. Questo stile non fa parte del linguaggio UML, ma è molto diffuso, assieme alla variante in cui i vari strati sono disegnati come corone circolari concentriche.

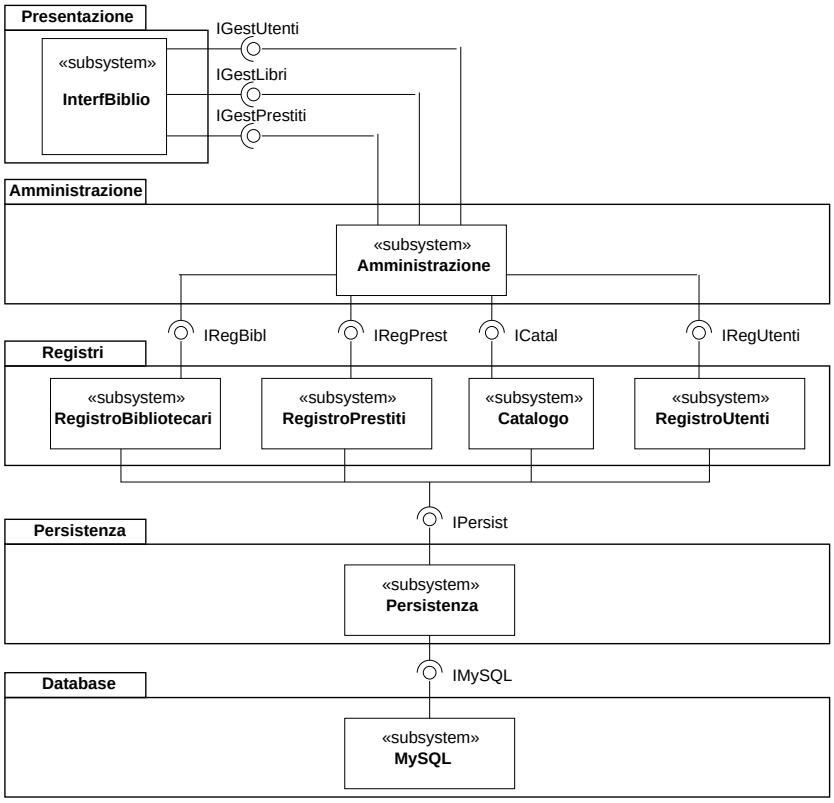


Figura 5.12: Scomposizione per strati e partizioni

### 5.2.2 Librerie e framework

Durante la scomposizione in sottosistemi bisogna infine considerare la disponibilità di librerie e di *framework*. Precisiamo che qui usiamo il termine "libreria" in un significato diverso da quello comune: di solito si intende con questo termine un componente fisico, cioè un file contenente dei moduli collegabili (p.es, i file `.a` e `.so` sui sistemi Unix, `.lib` e `.dll` sui sistemi Windows), o analoghi moduli precompilati per linguaggi interpretati (come i file `.jar` in ambiente Java). Qui invece parliamo di librerie e framework dal punto di vista logico, ricordando che tutti e due i tipi di componenti vengono realizzati e distribuiti come librerie fisiche.

Una libreria logica è una raccolta di componenti che offrono servizî ad un livello di astrazione piuttosto basso. Un framework contiene invece dei componenti ad alto livello di astrazione che offrono uno schema di soluzione preconfezionato per un determinato tipo di problema. Le librerie, quindi, si

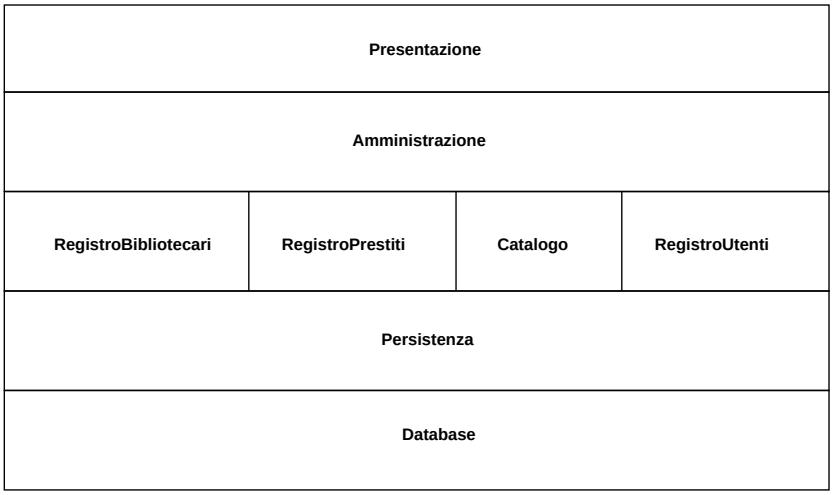


Figura 5.13: Scomposizione per strati e partizioni

usano "dal basso verso l'alto", assemblando componenti semplici e predefiniti per ottenere strutture complesse specializzate, mentre i framework si usano "dall'alto verso il basso", riempiendo delle strutture complesse predefinite (delle "intelaiature") con dei componenti semplici specializzati.

Le librerie e i framework sono componenti pronti all'uso (o, come si dice, *off the shelf*), generalmente forniti da produttori esterni e non modificabili. Nella maggior parte dei casi l'uso di tali componenti è molto vantaggioso sia dal punto di vista del processo di sviluppo che della qualità, e in particolare dell'affidabilità, del prodotto finale. Ovviamente una scelta accurata fra i componenti disponibili ha un'importanza cruciale per il successo del progetto, e questa scelta dipende sia da fattori tecnici, come le funzioni e prestazioni dei componenti, sia da fattori economici ed organizzativi, come le licenze d'uso e l'assistenza tecnica. Se fra i componenti pre-esistenti non se ne trovano di adatti allo scopo, naturalmente resta la possibilità di svilupparli *ex novo*.

Infine, conviene ricordare che una libreria o un framework possono essi stessi essere progettati e venduti come prodotto finale.

### Librerie

Una libreria può essere una raccolta di sottoprogrammi globali (cioè non appartenenti a classi) o di classi (Fig. 5.14). In tutti e due i casi l'insieme delle operazioni disponibili viene chiamato *interfaccia di programmazione* o

*API* (*Application Programming Interface*). Nel primo caso, la libreria si può rappresentare in UML come una classe, un componente o un package con lo stereotipo ≪utility≫. Nel progetto architetturale non è necessario mostrare la struttura interna della libreria, e di solito non serve nemmeno elencare le operazioni dell'interfaccia di programmazione. Nemmeno nel progetto in dettaglio è necessario, generalmente, mostrare queste informazioni, anche se in certi casi può essere utile esplicitare quali parti dell'interfaccia di programmazione vengono usate, e con quali classi della libreria si interagisce.

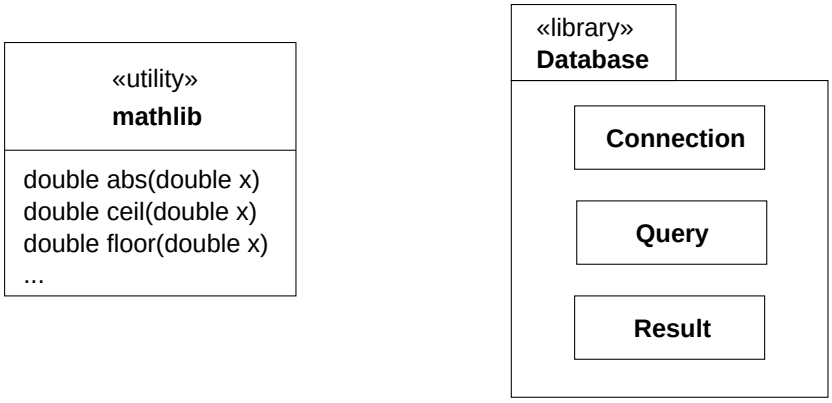


Figura 5.14: Due tipi di librerie.

### Framework

Come abbiamo detto, un framework offre uno schema di soluzione preconfezionato per un determinato tipo di problema, e viene usato in un'applicazione particolare specializzandone alcune caratteristiche. Questa specializzazione avviene generalmente per mezzo del polimorfismo: il framework offre al progettista delle classi e delle interfacce il cui comportamento viene specializzato derivando altre classi in cui si possono ridefinire i metodi delle classi e interfacce originarie, ed aggiungere ulteriori funzioni.

Un esempio (evidentemente semplificato) di framework viene mostrato in Fig. 5.15. Un'interfaccia grafica è formata da elementi (**Widget**) che possono essere composti o semplici. Gli elementi composti sono, per esempio, finestre, pannelli e menú, e tutti hanno l'operazione add(), ereditata da **Compound**, che permette di aggiungere un widget all'elemento composto. Gli elementi semplici sono, per esempio, bottoni e campi per l'inserimento di testi, ed hanno varie operazioni specifiche di ciascuna classe. In particolare, la classe

**Button** è astratta, avendo nella propria interfaccia l'operazione astratta action() che rappresenta, una volta implementata, l'azione da eseguire quando un particolare bottone viene "cliccato".

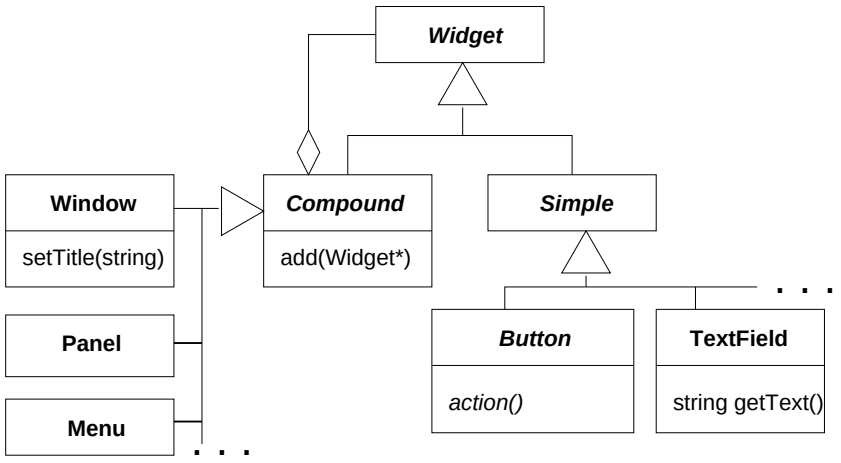


Figura 5.15: Un semplice framework.

Supponiamo di voler realizzare un'interfaccia in cui una finestra contiene un bottone e un campo testo, dove il bottone serve a scrivere sull'uscita standard il contenuto del campo testo preventivamente riempito dall'utente. Tutti i meccanismi a basso livello per la gestione del mouse, della tastiera, della visualizzaizone grafica sono forniti dal framework, oltre alla struttura logica ed ai meccanismi di interazione fra i vari elementi. Il programmatore deve soltanto scrivere una classe (p.es., **MyButton**) derivata da **Button** che implementi l'operazione action() nel modo desiderato, quindi scrivere il programma principale dell'operazione, in cui si istanzia una finestra e vi si aggiungono un'istanza di **TextField** e una di **MyButton**. La Fig. 5.16 mostra la struttura risultante.

Un framework può essere rappresentato in UML come un componente o un package. Nemmeno per i framework, come già visto per le librerie, è generalmente richiesta una rappresentazione della struttura interna.

### 5.2.3 Gestione dei dati

Nel definire la struttura globale del sistema bisogna anche scegliere come memorizzare e gestire i dati permanenti: una delle scelte principali è fra l'uso

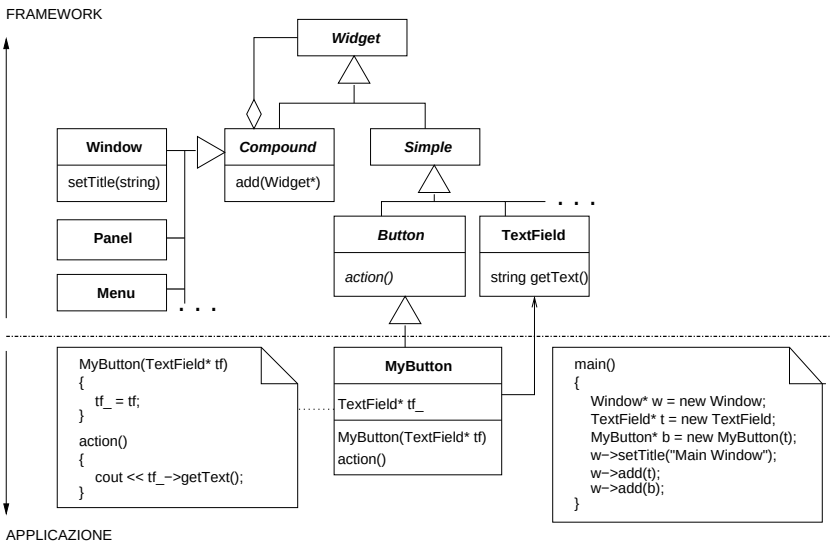


Figura 5.16: Uso di un framework in un'applicazione.

del filesystem offerto dal sistema operativo e l'uso di un sistema di gestione di database. Quest'ultima soluzione è generalmente preferibile quando la gestione di dati costituisca una parte importante, per dimensioni e complessità, delle funzioni offerte dal sistema. La scelta di usare un gestore di database comporta scelte ulteriori, principalmente quella fra sistemi relazionali e sistemi orientati agli oggetti.

### 5.2.4 Sistemi concorrenti

Il comportamento dei sistemi complessi generalmente si può descrivere come un insieme di attività concorrenti. Il progettista deve individuare tali attività e stabilire quali sottosistemi le devono svolgere, quindi analizzare le possibili interazioni fra i sottosistemi, come scambi di messaggi ed accessi a risorse condivise, e le interazioni fra il sistema e l'ambiente esterno.

In questa sezione tratteremo sistemi concorrenti *non distribuiti*, cioè eseguiti su un singolo calcolatore, rimandando ad un'altra parte del corso i sistemi distribuiti.

Un concetto fondamentale nell'analisi dei sistemi concorrenti è quello di *flusso di controllo* (*thread of control*). Un flusso di controllo è la sequenza di azioni svolte da un agente autonomo.

In un progetto orientato agli oggetti si definiscono *attivi* quegli oggetti che hanno la capacità di attivare un flusso di controllo. Il flusso di controllo di un oggetto attivo è costituito principalmente da chiamate di operazioni su oggetti (detti *passivi*) privi di un flusso di controllo indipendente, ma in un sistema concorrente il flusso di controllo di un oggetto attivo può interagire anche con quelli di altri oggetti attivi: un oggetto attivo può invocare operazioni di altri oggetti attivi (direttamente o inviando messaggi) o condividere oggetti passivi con altri oggetti attivi.

In UML gli oggetti attivi e le relative classi sono caratterizzati dalla proprietà isActive e vengono rappresentati con i lati verticali raddoppiati (in UML 1 si rappresentavano col bordo esterno piú spesso). I messaggi ed i segnali vengono rappresentati per mezzo di frecce nei diagrammi di sequenza e di collaborazione, ed è possibile usare notazioni diverse per esprimere diversi modi di sincronizzazione associati ai messaggi. Per esempio, si possono distinguere i messaggi *sincroni*, in corrispondenza dei quali il mittente resta in attesa che il ricevente completi l'azione richiesta dal messaggio, e i messaggi *asincroni*, in cui il mittente continua la sua attività indipendentemente da quella del ricevente.

Per individuare i flussi di controllo indipendenti (almeno i piú importanti, in fase di progetto di sistema) bisogna analizzare le informazioni che si ricavano dal modello dinamico, che in un progetto orientato agli oggetti è rappresentato dai diagrammi di stato, di sequenza, di collaborazione e di attività (o da informazioni analoghe se si usano notazioni diverse dall'UML).

### Esempio

La Fig. 5.17 mostra, come sempre in modo semplificato, un modello di analisi per un sistema di controllo che deve rilevare due condizioni pericolose in un impianto: incendio o perdita di alimentazione elettrica (esempio adattato da [1]). Il diagramma delle classi mostra i sottosistemi principali: un controllore centrale, due sottosistemi di monitoraggio, rispettivamente per la temperatura e l'alimentazione, un certo numero di sensori per la temperatura e per la tensione di rete, e un sottosistema per la gestione delle situazioni di emergenza. Il diagramma di sequenza è rappresentativo di una serie di diagrammi che mostrano il comportamento del sistema in varie situazioni; questo particolare diagramma mostra una possibile sequenza in cui si rileva un incendio. Inoltre, dalle specifiche risulta che l'incendio è piú pericoloso della mancanza di alimentazione, per cui una eventuale attività di gestione di una mancanza di elettricità deve essere interrotta in caso di incendio.

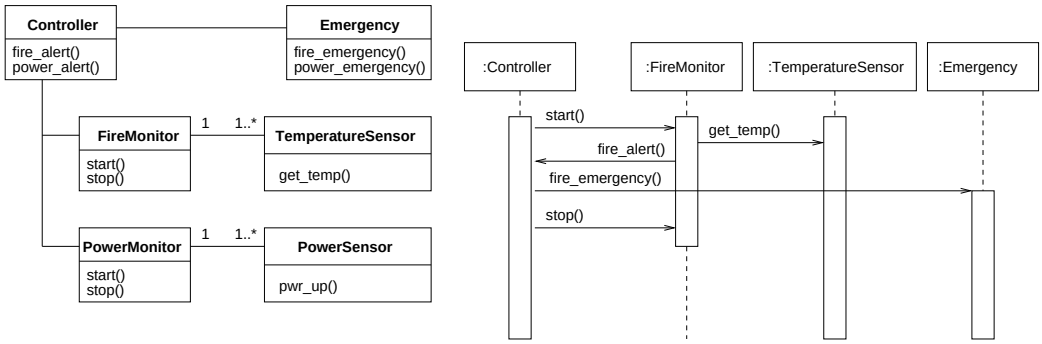


Figura 5.17: Un sistema concorrente (1).

Da quanto detto risulta che:

1. il controllore centrale dà inizio alle attività degli altri sottosistemi, e deve restare in ascolto di eventuali messaggi di allarme dai sottosistemi di monitoraggio;
2. i due sottosistemi di monitoraggio, dopo l'attivazione, devono funzionare continuamente e in parallelo fra di loro, interrogando periodicamente (in modo *polling*) i sensori;
3. i sensori e il sottosistema per le emergenze entrano in funzione solo quando ricevono messaggi dai sottosistemi di monitoraggio o dal controllore.

Possiamo quindi concludere che nel sistema ci sono tre flussi di controllo, appartenenti al controllore e ai due sottosistemi di monitoraggio, e iniziare il modello di progetto col diagramma di Fig. 5.18, in cui abbiamo individuato gli oggetti attivi e quelli passivi.

### Rappresentazione dei flussi di controllo

Il linguaggio UML permette di rappresentare i flussi di controllo concorrenti in diversi modi. La Fig. 5.19 mostra i flussi di controllo dell'esempio precedente per mezzo di un'estensione ai diagrammi di sequenza introdotta nell'UML 2. Questa estensione consiste nella possibilità di indicare dei segmenti di interazioni fra oggetti (cioè sottosequenze di scambi di messaggi), detti *frammenti*, che possono essere ripetuti, o eseguiti condizionalmente. In questo modo è possibile rappresentare graficamente i vari costrutti di con-

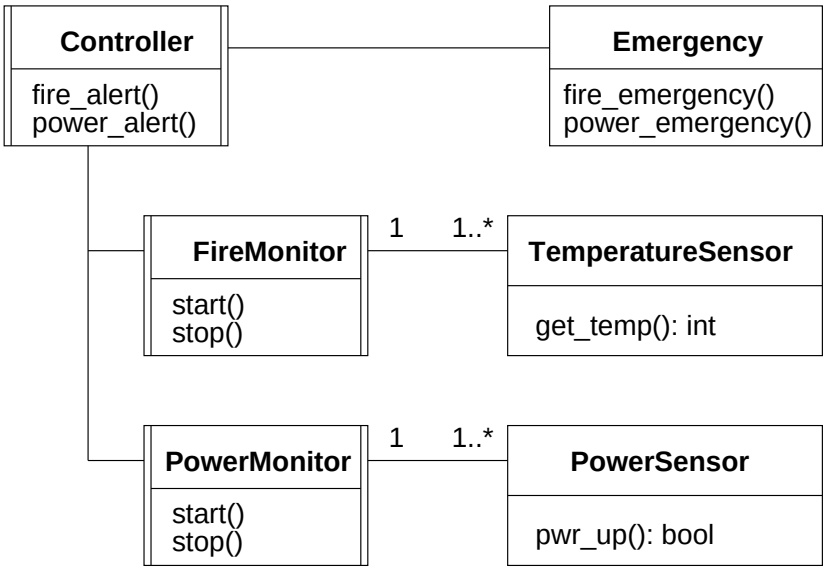


Figura 5.18: Un sistema concorrente (2).

trollo usati nei linguaggi di programmazione e anche specificare vari tipi di vincoli sulle interazioni possibili.

Nella Fig. 5.19, in cui per semplicità si suppone che ci sia un solo sensore di temperatura e un solo sensore di tensione, vediamo prima di tutto un blocco etichettato par e diviso in due regioni separate da una linea tratteggiata: questo significa che le due regioni contengono attività che avvengono in parallelo. Infatti le due regioni contengono, rispettivamente, le attività dei due sottosistemi di monitoraggio.

Nella prima regione, il controllore manda un messaggio start() al monitor antincendio. Questo esegue la sequenza specificata dal blocco loop, la cui espressione di controllo dice che il contenuto del blocco deve essere eseguito almeno una volta e poi essere ripetuto finché il valore t restituito da get_temp() è minore di una temperatura di soglia T. Se la temperatura di soglia viene superata, l'esecuzione del blocco loop termina, ed inizia una sequenza racchiusa in un blocco critical. Questo significa che tale sequenza (invio di un allarme al controllore, inizio della procedura di emergenza, e arresto del monitor) è una *sezione critica*, cioè un'attività non interrompibile. Questo rispecchia il requisito che il trattamento di un incendio abbia una priorità piú alta rispetto all'altra situazione di rischio.

Nella seconda regione, il controllore attiva il monitor dell'alimentazione,

che esegue un blocco loop condizionato sia dal valore restituito da pwr_up() che da quello di get_temp(), poiché il ciclo deve terminare anche in caso di incendio. All'uscita dal ciclo c'è un blocco opt che viene eseguito soltanto se non è stato rilevato un incendio.

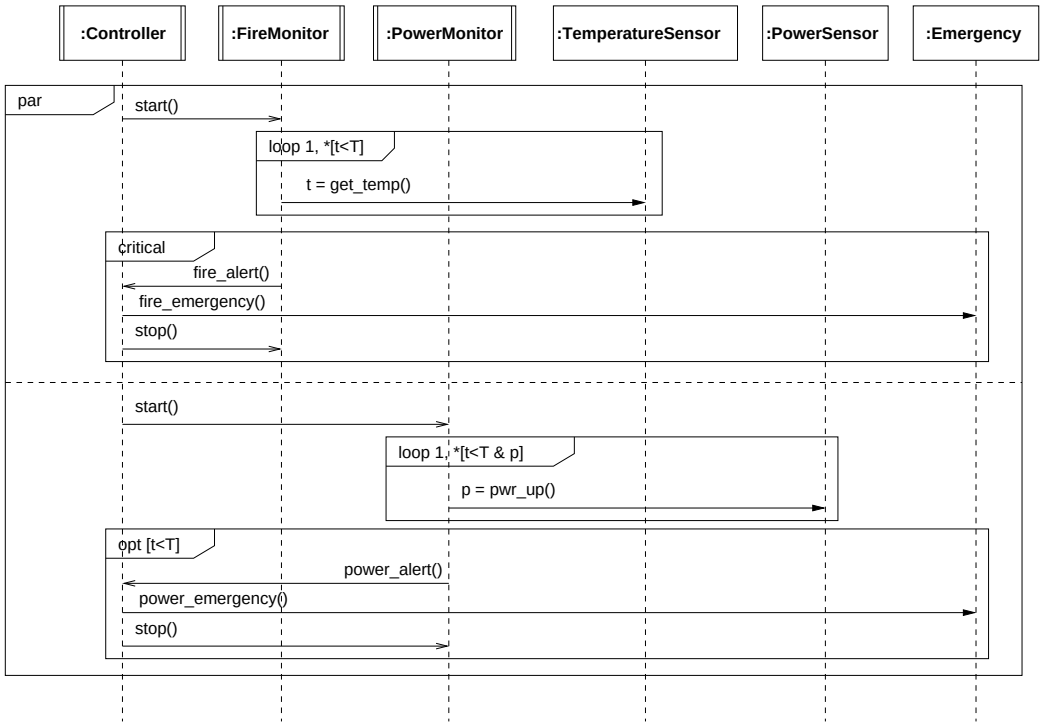


Figura 5.19: Un sistema concorrente (3).

Esistono vari altri operatori che qui non tratteremo. Osserviamo infine che abbiamo usato due tipi di frecce per due diversi tipi di messaggi: i messaggi *sincroni* sono destinati a oggetti passivi (i sensori e il gestore delle emergenze) e sono rappresentati da frecce con la punta piena, i messaggi *asincroni* sono destinati a oggetti attivi e sono rappresentati da frecce con la punta aperta.

### Condivisione delle risorse

Nell'esempio precedente ciascun oggetto passivo viene toccato dal flusso di cotrollo di un solo oggetto attivo: in questo modo gli oggetti attivi non condividono risorse. I rispettivi flussi di controllo non sono indipendenti perché, come abbiamo visto, l'attività di gestione di un incendio ha la priorità

sulla gestione di una perdita di alimentazione, però questo vincolo viene facilmente soddisfatto programmando opportunamente il controllore, che, grazie al suo ruolo centrale, può coordinare le attività dei sottosistemi ad esso subordinati.

Se consideriamo, invece, l'architettura alternativa mostrata in Fig. 5.20, vediamo che il sottosistema **Emergency** è interessato dai flussi di controllo dei due monitor (cioè, ciascuno dei due monitor può invocare operazioni di **Emergency**) ed è quindi *condiviso*. Nel progetto del sistema bisogna tener conto della condivisione di risorse per evitare problemi come il *blocco* (o *deadlock*) o l'*inconsistenza* delle informazioni causata da accessi concorrenti a dati condivisi. In questo caso particolare, bisogna rispettare il vincolo costituito dalla priorità dell'attività fire_emergency() rispetto all'altra.

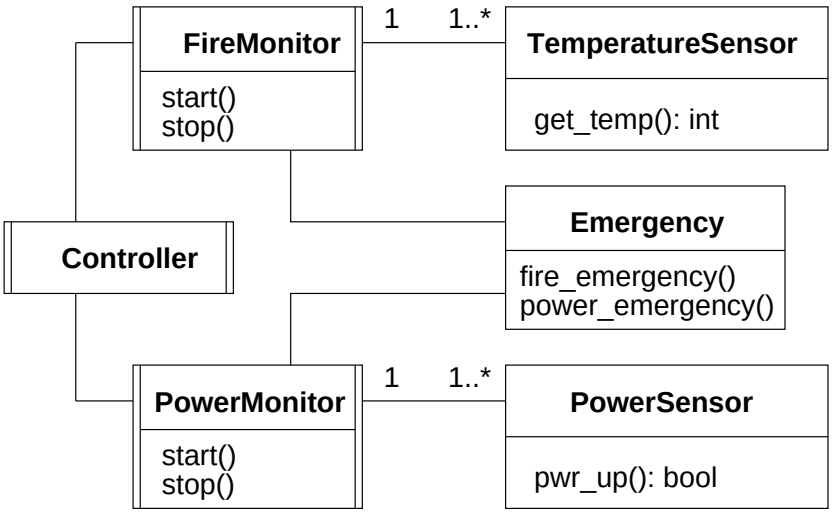


Figura 5.20: Un sistema concorrente (4).

### Meccanismi per la concorrenza

Per trattare, anche se a livello introduttivo, il progetto di sistemi concorrenti, conviene richiamare alcuni concetti fondamentali, rimandando uno studio piú approfondito ad altri insegnamenti.

Lo stato di un programma sequenziale (cioè avente un solo flusso di controllo) in qualsiasi passo delle sua esecuzione è definito dal valore del contatore di programma e dal contenuto dello stack, della memoria di lavoro (cioè l'insieme delle variabili definite nel programma), e delle strutture dati

usate dal sistema operativo che si riferiscono al programma. L'insieme di queste informazioni[2], insieme al codice macchina del programma, costituisce il *processo* associato ad un'istanza di esecuzione. Fra queste informazioni, il contatore di programma e lo stack definiscono il flusso di controllo del processo.

Il programmatore usa delle chiamate di sistema, come le primitive `fork()` ed `exec()` nei sistemi Unix, per creare i processi. Altre primitive (oppure chiamate di libreria a livello piú alto delle primitive di nucleo) permettono la *comunicazione interprocesso* e la *sincronizzazione* fra processi. Sempre riferendoci ai sistemi Unix, fra i meccanismi di comunicazione interprocesso citiamo i *pipe*, i *socket*, e la *memoria condivisa*; fra i meccanismi di sincronizzazione la chiamata `wait()` e i *semafori*.

La maggior parte dei sistemi operativi permette a un processo di eseguire in modo concorrente piú flussi di controllo. Ciascun flusso di controllo viene realizzato da un'entità chiamata *thread* o *processo leggero* (*lightweight process*), definita da uno stack e un contatore di programma. In un processo con piú flussi di controllo, detto *multithreaded*, i thread condividono la memoria di lavoro e il codice eseguibile. In generale ogni thread può eseguire un proprio segmento di codice, ma tutti i segmenti di codice associati ai thread fanno parte del testo del programma complessivo.

In pratica, il programmatore deve scrivere in un sottoprogramma le azioni che devono essere eseguite da un thread, e poi, usando apposite chiamate di sistema, chiedere al sistema operativo di creare il thread. Anche per i thread ci sono delle chiamate di sistema per il controllo dell'esecuzione e per i meccanismi di sincronizzazione. Lo standard adottato dalla maggior parte dei sistemi Unix è l'interfaccia *Posix Threads*[3].

### Strumenti per la programmazione concorrente

L'implementazione degli oggetti attivi definiti nel progetto architetturale riguarda principalmente le fasi di progetto dettagliato e di codifica, ma certe scelte relative all'implementazione devono essere fatte nel progetto architetturale.

Una prima decisione riguarda la scelta fra l'uso di processi o di thread. Un programma concorrente eseguito su un singolo calcolatore può essere composto da un solo processo multithreaded o da piú processi, ciascuno dei

[2] Piú precisamente, la successione dei valori assunti da tale insieme.

[3] v. `www.yolinux.com/TUTORIALS/LinuxTutorialPosixThreads.html`

quali può essere multithreaded o no. Poiché la creazione e la gestione dei processi è piú costosa, in termini di tempo e risorse, di quella dei thread, generalmente si scelgono questi ultimi se non ci sono vincoli contrari.

Un'altra scelta riguarda gli strumenti da usare per realizzare i meccanismi di concorrenza richiesti. Alcuni linguaggi di programmazione, come Java e Ada, hanno dei costrutti che permettono di definire degli oggetti attivi (p.es., oggetti di classe `Thread` in Java, o moduli `task` in Ada), di controllarne l'esecuzione e di sincronizzarli. Se uno di questi linguaggi (in base a requisiti non attinenti alla concorrenza) è stato scelto per l'implementazione, ovviamente conviene sfruttare gli strumenti che offre. Se la scelta del linguaggio è ancora aperta, si possono studiare i meccanismi di concorrenza usati dal linguaggio e valutare la loro adeguatezza al caso specifico.

Altri linguaggi, come il C e il C++, non hanno istruzioni o tipi predefiniti per la concorrenza[4]. Sviluppando in questi linguaggi, si presentano queste possibilità:

1. usare direttamente le chiamate di sistema;
2. incapsulare le chiamate di sistema in classi sviluppate *ad hoc*;
3. usare librerie di classi che implementano i meccanismi di concorrenza offrendo un'intefaccia orientata agli oggetti;
4. usare librerie che implementano modelli di programmazione concorrente (o parallela) a livello piú alto dei meccanismi di base della concorrenza.

L'uso diretto delle chiamate di sistema è complesso e prono agli errori, per cui si preferiscono le altre alternative, a meno che il sistema da sviluppare non sia molto semplice, o non sia necessario avere un controllo molto fine sull'esecuzione.

Per incapsulare le chiamate di sistema bisogna scrivere delle classi le cui interfacce siano piú semplici dell'interfaccia di programmazione di sistema. Per esempio, la creazione di un thread Posix richiede l'esecuzione di un certo numero di chiamate che impostano vari parametri del thread, incluso il sottoprogramma da eseguire. Si può allora definire una classe `Thread` il cui costruttore invoca tali chiamate, per cui la creazione di un thread si riduce

[4]Non si tratta di una dimenticanza. I creatori del C e del C++ seguono il principio che un linguaggio deve offrire un nucleo di funzioni base flessibili e potenti che si possano implementare in modo efficiente su qualsiasi piattaforma hardware e software. Le funzioni piú sofisticate, come la concorrenza o la *garbage collection*, possono quindi essere implementate in modi diversi secondo le esigenze delle diverse applicazioni.

a istanziare questa classe. Procedendo in questo modo, si crea uno strato di software fra l'applicazione ed il sistema operativo, che semplifica la programmazione e rende l'applicazione piú modulare e portabile.

Questa tecnica permette anche di semplificare l'accesso in mutua esclusione alle risorse condivise. A ciascuna di queste risorse bisogna associare un semaforo, cioè un meccanismo del sistema operativo che permette l'accesso a un solo thread (o processo) per volta. Un thread deve *acquisire* il semaforo, usare la risorsa, e quindi *rilasciare* il semaforo permettendo agli altri thread di accedere alla risorsa. Usando direttamente le chiamate di sistema c'è il rischio che, per un errore di programmazione, il rilascio non avvenga, con un conseguente blocco del sistema. Se il semaforo viene incapsulato in una classe, l'acquisizione viene fatta nel costruttore e il rilascio nel distruttore, cosicché il rilascio avviene automaticamente quando l'istanza del semaforo esce dal suo spazio di definizione.

Invece di implementare delle classi *ad hoc* per incapsulare i meccanismi di concorrenza, si può usare una libreria pre-esistente. Questa è probabilmente la soluzione migliore nella maggior parte dei casi. Esistono varie librerie di questo tipo, che offrono servizi simili ma con interfacce alquanto diverse e con diverse dipendenze da altre librerie e dai sistemi operativi. Citiamo, in particolare, la libreria *Boost.Thread*[5], che probabilmente verrà inclusa nella prossima versione delle librerie standard del C++. In questa libreria, la classe `thread` ha un costruttore a cui si passa come argomento il sottoprogramma da eseguire. L'interfaccia di `thread` è molto semplice: gli operatori di uguaglianza e disuguaglianza, l'operazione `join()` che permette a un thread di aspettare che termini l'esecuzione di un altro thread, l'operazione `sleep()` che permette al thread di sospendersi per un periodo di tempo determinato, e l'operazione `yield()` che permette al thread di cedere il proprio turno di esecuzione ad altri processi. Altre classi implementano vari meccanismi, come, per esempio, i alcuni tipi di semafori.

Infine, si può considerare l'uso di librerie ed ambienti di esecuzione che offrano un livello di astrazione piú alto. Queste librerie sono basate generalmente sul modello di programmazione a scambio di messaggi e permettono di applicare in modo abbastanza semplice numerose tecniche di programmazione parallela, in cui però non ci addentreremo. Di solito queste librerie presentano un'interfaccia costituito da una raccolta di numerose funzioni globali (in C o in Fortran), per cui anche in questo caso si può considerare l'opportunità di incapsularle in classi che offrano un'interfaccia piú struttu-

[5] v. `www.boost.org/doc/html/thread.html`

rata. Esempi di queste librerie sono la *Parallel Virtual Machine* (*PVM*)[6] ed il *Message Passing Interface* (*MPI*)[7], destinate allo sviluppo di applicazioni parallele (cioè eseguite su processori multipli), ma applicabili anche per sistemi concorrenti non distribuiti.

### Tecniche di implementazione

Il progetto di sistemi concorrenti è un argomento complesso che richiede la conoscenza di problemi e tecniche specifiche dei vari campi di applicazioni, e in questa sezione possiamo dare solo alcune indicazioni generiche.

Come si è visto, il primo passo del progetto consiste nell'individuare gli oggetti attivi, dopo di che bisogna analizzare le interazioni fra questi oggetti, considerando i vincoli di sincronizzazione, le comunicazioni reciproche, e l'uso delle risorse.

**Sincronizzazione** I vincoli di sincronizzazione spesso si possono esprimere in termini di eventi, stati e azioni: per esempio, si può richiedere che un oggetto attivo reagisca con una certa azione (ed eventualmente con un cambiamento del proprio stato) ad un evento causato da un altro oggetto attivo, oppure si può richiedere che un oggetto attivo resti in attesa di un evento prima di procedere nella sua esecuzione. Questo tipo di informazioni si rappresentano in modo naturale con macchine a stati associate agli oggetti attivi. A volte le macchine a stati sono già state definite nel modello di analisi, che nel modello di progetto possono essere completate o rese piú dettagliate. Nell'esempio di Fig. 3.34 (pag. 128), riportato in forma semplificata nella Fig. 5.21, il produttore ed il consumatore si sincronizzano attraverso gli eventi produci e consuma.

Questo tipo di sincronizzazione si può implementare con i *semafori di sincronizzazione*. Per realizzare questo esempio, i due eventi potrebbero essere rappresentati da due variabili booleane `produci` e `consuma`, associate a due semafori e incapsulate in una classe `Buffer`, che offre i le operazioni `signal_produci()`, `wait_produci()`, `signal_consuma()` e `wait_consuma()`. Un thread che deve attendere un evento, per esempio `consuma`, esegue la corrispondente operazione del buffer (`wait_consuma()`), che legge (in mutua esclusione) la variabile logica corrispondente al segnale (`consuma`) e,

[6] v. `www.csm.ornl.gov/pvm/`

[7] v. `www.llnl.gov/computing/tutorials/mpi/`

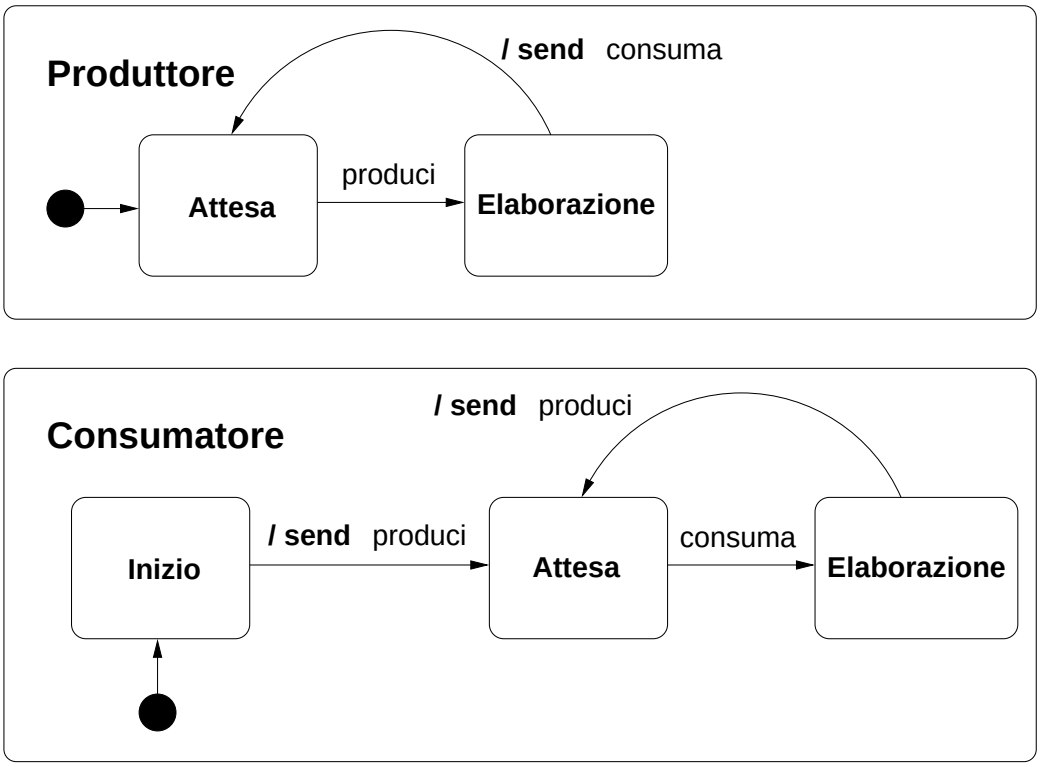


Figura 5.21: Due oggetti attivi interagenti.

fintantoché la variabile non è vera, si mette in attesa sul semaforo associato. Per inviare un evento, un thread esegue l'operazione corrispondente (`signal_consuma()`), che asserisce (in mutua esclusione) la variabile logica e segnala al semaforo che la variabile è stata cambiata. Il semaforo quindi risveglia il thread che era in attesa.

Nella libreria *Boost.Thread* i semafori di sincronizzazione sono chiamati *variabili di condizione*. Mostriamo di seguito una possibile dichiarazione della classe `Buffer` descritta sopra, seguita dalla definione di una variabile globale `buf`:

```
class Buffer {
    boost::mutex pmutex;
    boost::mutex cmutex;
    boost::condition prod;
    boost::condition cons;
    bool produci;
    bool consuma;
public:
    Buffer() : produci(false), consuma(false) {};
```

```
    void signal_produci()
    void wait_produci()
    void signal_consuma()
    void wait_consuma()
};

Buffer buf;
```

I membri `pmutex` e `cmutex` sono semafori di mutua esclusione, i membri `prod` e `cons` di sincronizzazione. La variabile `buf` è stata dichiarata con visibilità globale in modo che sia visibile alle procedure dei thread, ma questa soluzione ovviamente[8] serve solo a semplificare questo esempio. In un caso reale si userebbero altre tecniche che qui non ci interessano.

L'operazione `wait_consuma()` è implementata come segue:

```
wait_consuma()
{
    boost::mutex::scoped_lock lock(cmutex); // acquis.
    while (!consuma) {
        cons.wait(lock);                    // rilascio
    }                                       // riacquis.
    consuma = false;
                                            // rilascio
}
```

dove `lock` è un'istanza della classe `scoped_lock` il cui costruttore acquisisce un semaforo di mutua esclusione che viene poi rilasciato dal distruttore. L'operazione `wait()` rilascia il semaforo e mette il thread in attesa. Quando il thread viene risvegliato riacquisisce il semaforo.

L'operazione `signal_consuma()` è implementata come segue:

```
signal_consuma()
{
    boost::mutex::scoped_lock lock(cmutex);
    consuma = true;
    cons.notify_one();
}
```

dove l'operazione `notify_one()` fa risvegliare uno dei thread in attesa.

L'evoluzione dei due thread è definita da questi due sottoprogrammi:

[8]Nella filosofia orientata agli oggetti le variabili globali sono *Il Male*.

```
void produttore()                | void consumatore()
{                                | {
                                 |     buf.signal_produci();
    for (;;) {                   |     for (;;) {
        buf.wait_produci();      |         buf.wait_consuma();
        elabora_prod();          |         elabora_cons();
        buf.signal_consuma();    |         buf.signal_produci();
    }                            |     }
}                                | }
```

E infine, i thread vengono creati ed attivati nel seguente modo:

```
int main(int argc, char* argv[])
{
    boost::thread consum(&consumatore);
    boost::thread produt(&produttore);
    consum.join();
    produt.join();
    return 0;
}
```

dove le operazioni `join()` fanno sí che il programma principale attenda la fine dell'esecuzione dei due thread.

**Risorse** Analizzando le interazioni fra sottosistemi bisogna anche stabilire quali oggetti passivi vengono interessati da un solo flusso di controllo e quali sono condivisi. I primi si possono considerare come parte del corrispondente oggetto attivo, col quale possono essere messi in relazione di composizione nei diagrammi UML. I secondi devono essere protetti mediante semafori di mutua esclusione. La classe `Buffer` vista precedentemente è un esempio di oggetto passivo protetto da semafori: gli oggetti di questo tipo sono chiamati *monitor*. Un cliente di un monitor può accedere ai suoi dati solo attraverso le operazioni del monitor, e queste provvedono ad acquisire i semafori (spesso uno solo), mettendo in attesa il chiamante se la risorsa è occupata. Queste operazioni sono dette *operazioni con guardia* o *sincronizzate*. Nel linguaggio Java un'operazione può essere dichiarata `synchronized`, e il compilatore le associa un semaforo e la implementa inserendo le operazioni di acquisizione e rilascio.

**Comunicazione** La comunicazione fra oggetti attivi si può considerare come una forma particolare di sincronizzazione, in cui gli oggetti si scambiano dati. Bisogna stabilire quali comunicazioni sono *sincrone* e quali *asincrone*.

In una comunicazione sincrona il mittente del messaggio si blocca finché non arriva una risposta dal destinatario, mentre in una comunicazione asincrona il mittente prosegue l'esecuzione senza aspettare la risposta, che può anche non esserci. Quando è richiesta una comunicazione sincrona, il blocco del mittente avviene automaticamente se la comunicazione avviene per mezzo di una normale chiamata di procedura, altrimenti è il meccanismo di comunicazione che deve sospendere il mittente. In ogni caso, si cerca di usare un meccanismo di comunicazione fornito dal linguaggio di programmazione, da librerie o dal sistema operativo, ma in certi casi è necessario o conveniente realizzare dei meccanismi *ad hoc*. Questi sono basati su strutture dati condivise, come code o liste, che quindi richiedono le tecniche viste precedentemente.

### 5.2.5 Architettura fisica

Per *architettura fisica* si intende l'insieme dell'*architettura software (fisica)* e dell'*architettura hardware*. L'architettura software è costituita dai *componenti software fisici*, o *artefatti* che sono il prodotto finale del processo di sviluppo, cioè i file necessari per il funzionamento del sistema sviluppato. L'architettura hardware è costituita dai *nodi* che eseguono i componenti software.

Gli artefatti possono essere file eseguibili, librerie, file sorgente, file di configurazione, pagine web, o altre cose ancora. Vengono rappresentati in UML 2 come rettangoli con lo stereotipo ≪artifact≫. Poiché gli artefatti contengono l'implementazione di elementi logici, come classi e componenti, è utile esprimere la relazione fra questi elementi logici e gli artefatti: si dice che un artefatto *manifesta* certi elementi logici, e questo si rappresenta graficamente con una dipendenza diretta dall'artefatto agli elementi manifestati, etichettata con lo stereotipo ≪manifest≫ (Fig. 5.22).

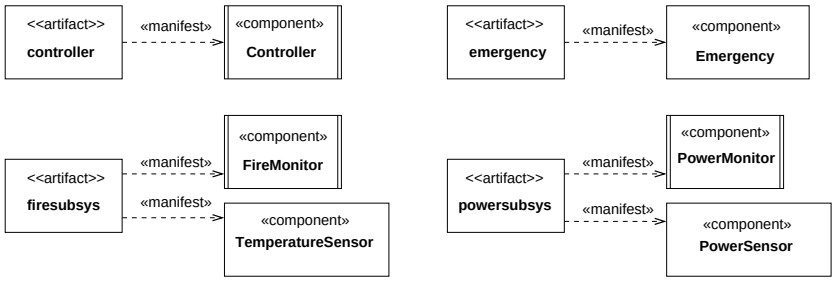


Figura 5.22: Artefatti e componenti.

I nodi modellano, generalmente, dei singoli calcolatori che possono essere collegati in rete. Se fosse necessario specificare dettagliatamente la struttura del calcolatore, allora un nodo può rappresentare parti di un calcolatore, come la CPU o le unità periferiche, ma questo non è comune. Un nodo considerato solo come hardware viene chiamato *dispositivo* e viene rappresentato come in Fig. 5.23, etichettato con lo stereotipo ≪device≫. Oltre a modellare la parte hardware di un sistema, un nodo può rappresentare anche un *ambiente di esecuzione*, cioè un sistema software, esterno all'applicazione sviluppata, entro cui viene eseguita l'applicazione. Un esempio ovvio di ambiente di esecuzione è il sistema operativo, che però, generalmente, non viene modellato esplicitamente. Un sistema che generalmente viene modellato esplicitamente come ambiente di esecuzione è un server web, visto come piattaforma su cui vengono eseguiti programmi *plug-in*, cioè sviluppati separatamente ed eseguiti dal server su richiesta. Gli ambienti di esecuzione si rappresentano come nodi con lo stereotipo ≪execution environment≫.

I *diagrammi di deployment* descrivono l'architettura fisica mostrando i nodi, i loro collegamenti, e gli artefatti istallati sui nodi. L'istallazione di un artefatto su un nodo può essere raffigurata disegnando l'artefatto dentro al nodo, o usando la dipendenza ≪deploy≫ diretta dall'artefatto al nodo (Fig. 5.23).

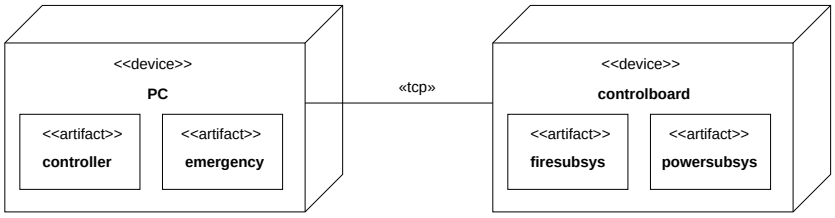


Figura 5.23: Architettura fisica.

## 5.3 I *Design pattern*

Nel progetto dei sistemi software esistono dei problemi che si presentano entro svariati campi di applicazione con diverse forme e varianti, ma con una struttura comune. Per questi problemi esistono delle soluzioni tipiche la cui utilità è stata verificata con l'esperienza di numerosi sviluppatori nell'àmbito di progetti diversi. Tali soluzioni sono chiamate *design pattern* ([3]), ed esiste

una letteratura abbastanza vasta che elenca numerosi pattern, fornendo allo sviluppatore un ricco armamentario di strumenti di progetto.

Per esempio, le strutture di Fig. 5.24 sono riconducibili allo schema di Fig. 5.25, che risolve il problema di organizzare un sistema formato da elementi diversi ma derivati da un'interfaccia comune, che possono contenere altri elementi derivati dalla stessa interfaccia. Questo schema è noto come pattern *Composite*.

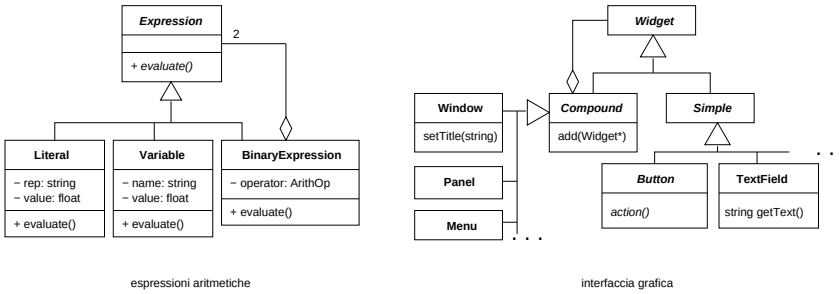


Figura 5.24: Due problemi simili.

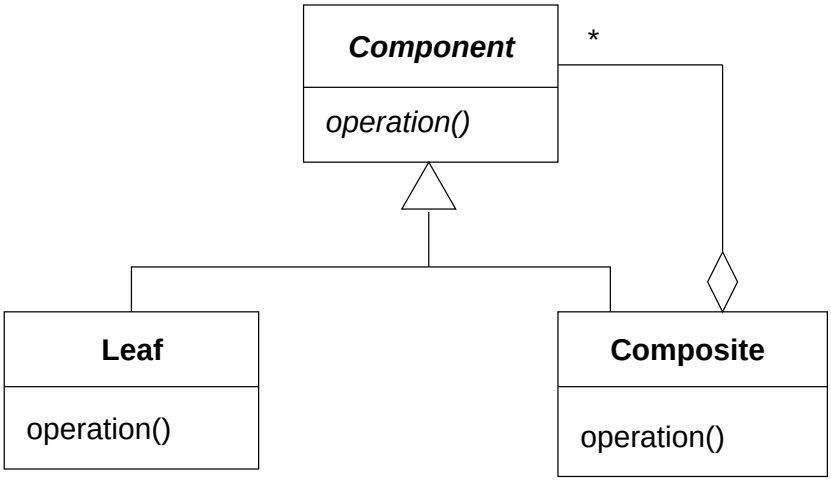


Figura 5.25: Un design pattern.

Un pattern consiste nella descrizione sintetica di un problema e della relativa soluzione. Questa viene descritta in forma sia grafica che testuale specificandone gli elementi strutturali (classi, componenti, interfacce...) con le relazioni reciproche e il loro modo di interagire. Questa descrizione viene integrata con una discussione dei vantaggi e svantaggi della soluzione,

delle condizioni di applicabilità e delle possibili tecniche di implementazione. Normalmente vengono presentati anche dei casi di studio.

Un design pattern *non* è un componente software, ma solo uno schema di soluzione per un particolare aspetto del funzionamento di un sistema. Gli elementi strutturali di un pattern rappresentano dei *ruoli* che saranno interpretati dalle entità effettivamente realizzate. Ciascuna di queste entità può interpretare un ruolo diverso in diversi pattern, poiché in un singolo sistema (o sottosistemi) si devono risolvere diversi problemi con diversi pattern. Inoltre, è bene tener presente che un pattern non è una ricetta rigida da applicare meccanicamente, ma uno schema che deve essere adattato alle diverse situazioni, anche con un pò d'inventiva.

I design pattern tendono ad essere usati in una fase del progetto, detta *progetto dei meccanismi* (*mechanistic design*), intermedia fra il progetto di sistema e quello in dettaglio, però molti pattern sono applicabili in tutto l'arco della fase di progetto, e inoltre esistono dei pattern concepiti espressamente per il progetto di sistema, fra cui le architetture citate nella Sez. 5.2.1.

In questo corso, i design pattern verranno trattati nelle ore di laboratorio.

## 5.4 Progetto dettagliato

In questa sezione consideriamo alcune linee guida per la fase di progetto in dettaglio. In questa fase vengono precisati i dettagli delle singole classi e associazioni, e vengono introdotte classi ed associazioni ausiliarie.

### 5.4.1 Progetto delle classi

Nella fase di progetto architetturale le classi definite nel modello di analisi vengono raggruppate in componenti, eventualmente insieme a qualche nuova classe introdotta in questa fase. Generalmente queste classi non sono state definite completamente, quindi nel progetto dettagliato bisogna arrivare prima di tutto al completamento della loro definizione. Per questo occorre:

1. specificare completamente attributi ed operazioni già presenti, indicando visibilità, modificabilità, tipo e direzione dei parametri;
2. aggiungere operazioni implicite nel modello, per esempio costruttori e distruttori;
3. aggiungere operazioni ausiliarie, se necessario.

Oltre a completare e raffinare le definizioni delle singole classi, si possono fare delle operazioni di ristrutturazione, per esempio scomponendo una classe in classi piú piccole, oppure riunendo piú classi in una, o ridistribuendo fra piú classi le rispettive operazioni, sempre cercando la massima coesione entro ciascuna classe. È possibile anche riorganizzare le gerarchie di generalizzazione.

### 5.4.2 Progetto delle associazioni

Le associazioni rappresentano i percorsi logici attraverso cui si propagano le interazioni fra i vari oggetti. Nel progetto delle associazioni si cerca di ottimizzare tali percorsi, ristrutturandoli se necessario, per esempio aggiungendo associazioni ausiliarie che permettono un accesso piú efficiente, o anche eliminando associazioni ridondanti.

Le associazioni devono poi essere ristrutturate in modo da permettere la loro implementazione nei linguaggi di programmazione. Questi, infatti, non hanno dei concetti primitivi che corrispondano direttamente alle associazioni, che devono quindi essere tradotte in concetti a piú basso livello. A questo scopo bisogna prima di tutto che le associazioni definite nel modello di analisi (ed eventualmente nel modello architetturale) vengano specificate completamente, indicandone la navigabilità, le molteplicità ed i nomi dei ruoli, oppure il nome dell'associazione. Avendo precisato queste informazioni, si possono considerare i casi seguenti (Fig. 5.26).

#### Associazioni da uno a uno e da molti a uno

In questo caso l'associazione può essere implementata come un semplice puntatore (e in questo caso si lascia immutata nel modello) o anche come un attributo, se la classe associata è molto semplice (p.es., `string`) oppure se si vuole modellare una composizione.

#### Associazioni da uno a molti

Nell'associazione da uno a molti un'istanza della classe all'origine dell'associazione è associata ad un gruppo di istanze dell'altra classe. In questo caso conviene introdurre una classe contenitore, possibilmente scelta fra le numerose classi (o template) di libreria disponibili per i vari linguaggi. La scelta della classe contenitore dipende dalle proprietà del gruppo di istanze,

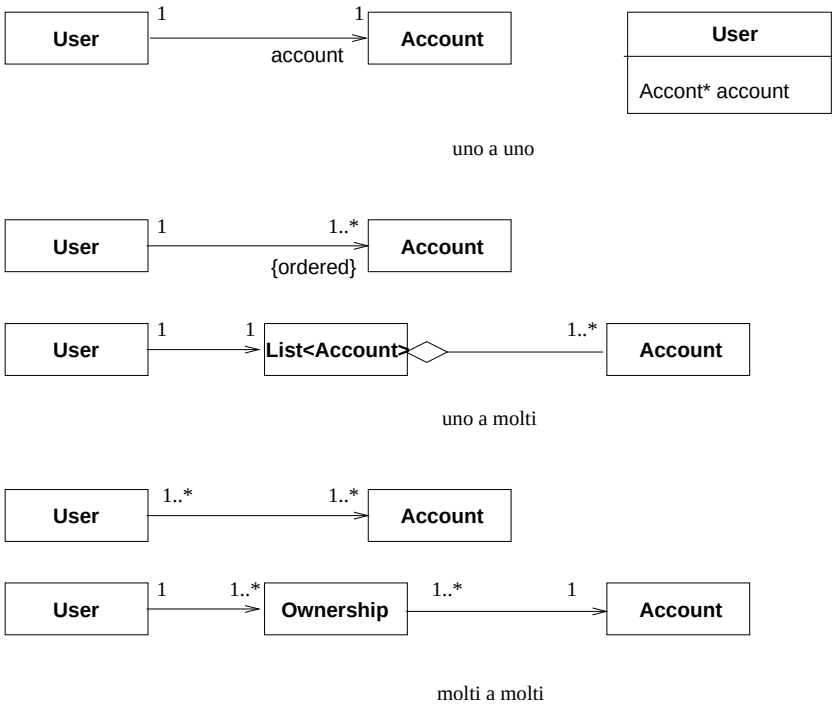


Figura 5.26: Realizzazione delle associazioni.

in particolare quelle di *unicità* (se ogni elemento appare una sola volta o no) e dell'*ordinamento*. Queste proprietà si possono rappresentare con le propretà UML unique, nonunique, ordered, unordered. Per modellare le associazioni si può scegliere di lasciarle come sono, con l'indicazione di queste proprietà che serve da guida per il programmatore nella scelta della classe contenitore, oppure rappresentare questa classe esplicitamente.

### Associazioni da molti a molti

Questo tipo di associazione si realizza introducendo una classe intermedia, come in Fig. 5.26, e risolvendo le asociazioni risultanti come nei casi precedenti.

# Capitolo 6

# Convalida e verifica

Ogni attività industriale richiede che vengano controllate la correttezza, la funzionalità e la qualità dei prodotti finiti, dei prodotti intermedi e dello stesso processo di produzione. Nell'ingegneria del software si parla di *convalida* e *verifica*: la differenza sta nei termini di riferimento usati per il controllo, in quanto nella convalida i prodotti vengono confrontati con i requisiti dell'utente, mentre nella verifica il confronto avviene con le specifiche. Ricordiamo che i requisiti sono generalmente imprecisi e necessariamente informali, mentre le specifiche (risultanti dalla fase di analisi dei requisiti) sono precise e possono essere formali. La verifica può quindi contare su procedimenti piú metodici e rigorosi, però non è sufficiente ad assicurare la bontà del prodotto, poiché le specifiche stesse, che ne sono il punto di riferimento, possono non essere corrette rispetto ai requisiti e devono essere a loro volta convalidate. Verifica e convalida devono quindi completarsi a vicenda. Aforisticamente, si dice che con la verifica ci accertiamo che il prodotto sia fatto bene (*building the product right*) e con la convalida ci accertiamo che il prodotto sia quello giusto (*building the right product*).

Nel seguito, faremo riferimento solo alla verifica (e marginalmente alla convalida) del codice, pur tenendo presente che, come già accennato, i prodotti di ogni fase del processo di sviluppo devono essere verificati o convalidati.

## 6.1 Concetti fondamentali

Introduciamo qui alcuni concetti fondamentali nel campo della convalida e della verifica, cominciando dai termini *errore*, *anomalia*, e *guasto*:

**Guasto** (*malfunzionamento*, *failure*) un comportamento del programma non corretto rispetto alle specifiche.
**Anomalia** (*difetto*, *fault*, *bug*) un'aspetto del codice sorgente che provoca dei guasti.
**Errore** errore concettuale o materiale che causa anomalie.

Gli errori sono quindi la causa delle anomalie, che a loro volta sono la causa dei malfunzionamenti. Osserviamo però che non c'è una corrispondenza diretta e biunivoca fra anomalie e guasti. Un particolare guasto può essere provocato da una o piú anomalie, e l'effetto di una anomalia può essere bilanciato, e quindi nascosto, da quello di un'altra. Alcune anomalie possono non provocare alcun guasto. Ma soprattutto, il problema fondamentale è che in genere un malfunzionamento avviene in presenza di alcuni dati di ingresso e non di altri.

Le attività rivolte alla ricerca e all'eliminazione delle anomalie si possono classificare come segue:

**Analisi statica** esame del codice sorgente.
**Analisi dinamica** esecuzione del codice.
**Debugging** ricerca delle anomalie a partire dai guasti, e loro eliminazione.

Conviene inoltre distinguere fra convalida o verifica *in piccolo* e *in grande*, cioè a livello di singolo modulo (o *unità*) o di sistema.

## 6.2 Analisi statica

L'analisi statica consiste nell'esame del codice sorgente. Fra i vari tipi di analisi statica citiamo i seguenti:

- *walk-through*: un gruppo di collaudatori e progettisti analizza il codice e ne simula l'esecuzione.
- *ispezione*: simile al walk-through, ma finalizzata alla scoperta di specifici errori.

- *analisi automatica*: compilazione, lint (ricerca di costrutti sintatticamente corretti ma potenzialmente errati), analisi di flusso dei dati.
- *prova formale*: dimostrazione formale (usando la logica) che non ci sono anomalie.

In questo corso tratteremo solo l'analisi di flusso dei dati. Citiamo però, a proposito dell'analisi statica automatica, lo strumento `gcov` (`http://gcc.gnu.org/onlinedocs/gcc-3.0/gcc_8.html`), uno strumento che permette, fra l'altro, di scoprire le parti di codice che non vengono eseguite.

### 6.2.1 Analisi di flusso dei dati

Per *flusso dei dati* s'intende la successione delle operazioni che cambiano il valore di una variabile. Le operazioni possibili sono la *definizione*, l'*uso*, e l'*annullamento*. Nel contesto dell'analisi di flusso dei dati, si dice che una variabile viene definita ogni volta che le viene assegnato un valore. Una variabile viene usata (ovviamente) quando viene letto il suo valore, e viene annullata quando il suo valore diviene indefinito (come avviene, per esempio, alle variabili locali di un sottoprogramma al termine dello stesso). Nella verifica del software, l'analisi di flusso dei dati serve a scoprire se nell'esecuzione di un programma si possono produrre sequenze di operazioni anomale per qualche variabile. Per esempio, se scopriamo che una variabile viene usata prima di essere definita, molto probabilmente tale sequenza è causata da qualche anomalia del software (in qualche raro caso il programma potrebbe essere corretto, per esempio se la variabile fosse condivisa con altri processi di cui l'analisi non tiene conto). Altre sequenze che possono essere sintomi di anomalie sono quelle in cui una definizione viene seguita immediatamente da un'altra definizione o da un annullamento.

Una sequenza di operazioni su di una variabile può essere descritta da una stringa di simboli appartenenti all'alfabeto $\{d, u, a\}$ (per *definizione*, *uso*, e *annullamento*, rispettivamente). Si può quindi usare il formalismo delle espressioni regolari per ottenere una rappresentazione finita delle sequenze (potenzialmente infinite) che possono essere generate da un segmento di programma.

Consideriamo, per esempio, il flusso di dati relativo alla variabile `a` nelle istruzioni successive all'istruzione 5 del seguente programma:

```
main()
{
    int a, b, t, x, y;     // 1   a
    cin >> x >> y;         // 2
    a = x;                 // 3   d
    a = y;                 // 4   d
    while (a != b)         // 5   u
        if (a > b) {       // 6   u
            t = a - b;     // 7   u
            a = t;         // 8   d
        } else {           // 9
            t = b - a;     // 10  u
            b = t;         // 11
        }                  // 12
    cout << a;             // 13  u
}
```

L'istruzione 6 è un uso di `a`, quindi l'espressione regolare cercata inizia col simbolo `u`. Successivamente possono essere eseguite le istruzioni 7 e 8, che generano la sequenza `ud`, *oppure* le istruzioni 10 e 11, che generano la sequenza `u`. Dopo l'esecuzione dell'istruzione composta `if` viene nuovamente eseguito il test all'istruzione 5 (operazione `u`), completando l'esecuzione del corpo del `while`. L'espressione relativa ad una sola esecuzione del corpo del `while` è quindi `u(ud|u)u`. Poiché il corpo può essere eseguito zero o piú volte, l'espressione regolare che descrive un numero arbitrario di iterazioni è `(u(ud|u)u)*`. Infine, all'uscita dal ciclo la variabile `a` viene usata nell'istruzione 13, per cui l'espressione ricercata è `(u(ud|u)u)*u`.

L'analisi delle espressioni regolari cosí ottenute permette di vedere se sono possibili sequenze di operazioni insicure. Le proprietà algebriche delle espressioni regolari permettono di verificare l'equivalenza fra espressioni.

## 6.3 Analisi dinamica

Nell'analisi dinamica (o *testing*) si esegue il programma da verificare o convalidare per osservare se i risultati corrispondono a quelli previsti dalle specifiche o dai requisiti. Generalmente è impossibile eseguire il programma per tutti i valori possibili dei dati di ingresso, poiché questo insieme può essere infinito o comunque troppo grande. Il primo problema da affrontare nell'attività di testing è quindi quello di selezionare un insieme di dati di ingresso che sia abbastanza piccolo da permettere l'esecuzione dei test entro i limiti di tempo e di risorse disponibili, ma sia anche abbastanza grande, e soprattutto significativo, da fornire risultati sufficientemente affidabili. Il secondo

problema è quello di stabilire i criterio di successo o fallimento del test, cioè capire quale deve essere il comportamento corretto del programma. Infine, bisogna stabilire il criterio di terminazione, per decidere quando interrompere l'attività di testing.

Riassumiamo qui, schematicamente, i problemi appena citati:

**Selezione dei dati** può essere guidata:

- dalle specifiche (test funzionale);
- dalla struttura del programma (test strutturale);

nessuno dei due criteri da solo è sufficiente, poiché ciascuno di essi permette di scoprire diversi tipi di guasti.

**Correttezza dei risultati** può essere decisa:

- dall'utente (convalida);
- dal confronto con le specifiche (verifica); in questo caso, le specifiche si rivelano tanto piú utili quanto sono formali; se le specifiche sono eseguibili, il confronto può avvenire con i risultati di un prototipo ottenuto direttamente dalle specifiche;
- dal confronto con versioni precedenti; avviene nei test di *regressione*, in cui si verifica una nuova versione del software;

**Terminazione** dovrebbe essere decisa in base a modelli statistici che permettano di stimare il numero di anomalie sopravvissute al test, ma spesso la fase di test termina allo scadere del periodo di tempo ad essa destinato (generalmente troppo presto).

Una forma particolare di analisi dinamica è l'*esecuzione simbolica*, che consiste nell'eseguire il programma da verificare per mezzo di un'interprete capace di rappresentare simbolicamente le variabili e di fornire i risultati sotto forma di relazioni algebriche (equazioni e disequazioni) fra le variabili di ingresso.

### 6.3.1 Definizioni e risultati teorici

Diamo di seguito le definizioni di alcuni concetti usati nell'analisi dinamica:

- il programma $P$ da testare viene rappresentato da una funzione $P : D \rightarrow R$, dove il dominio $D$ è l'insieme dei dati di ingresso ed il codominio $R$ è l'insieme dei risultati;

- un programma $P$ si dice *corretto per un dato* $d$ (e si scrive $\mathrm{ok}(P, d)$) se e soltanto se il risultato dell'esecuzione del programma col dato di ingresso $d \in D$ corrisponde a quello richiesto dalle specifiche;
- un programma $P$ si dice *corretto* (e si scrive $\mathrm{ok}(P)$) se e soltanto se il risultato dell'esecuzione del programma per ogni dato di ingresso $d \in D$ corrisponde a quello richiesto dalle specifiche: $\mathrm{ok}(P) \Leftrightarrow \forall_{d \in D} \mathrm{ok}(P, d)$
- un *test* $T$ è un sottoinsieme del dominio del programma ($T \subset D$), ed un *dato di test* $t$ è un elemento di tale sottoinsieme ($t \in T$); in generale, un dato di test è una $n$-upla di valori;
- un programma $P$ si dice *corretto per un test* $T$ (e si scrive $\mathrm{ok}(P, T)$) se e soltanto se il risultato dell'esecuzione del programma per ogni dato di test $t \in T$ corrisponde a quello richiesto dalle specifiche: $\mathrm{ok}(P, T) \Leftrightarrow \forall_{t \in T} \mathrm{ok}(P, t)$
- un test $T$ si dice *ideale* se la correttezza del programma per $T$ implica la correttezza (*tout-court*) del programma: $\mathrm{ideale}(T, P) \Leftrightarrow (\mathrm{ok}(P, T) \Rightarrow \mathrm{ok}(P))$

La teoria dei test ha prodotto vari risultati fra i quali riportiamo i seguenti:

**Teorema di Howden** Non esiste un algoritmo che, dato un programma arbitrario $P$, generi un test ideale per tale programma.

**Teorema di equivalenza dei programmi** (Brainerd e Landweber, 1974). Non è possibile stabilire se due generici programmi càlcolino la stessa funzione o no.

**Teorema di equivalenza dei cammini** Non è possibile stabilire se due generici cammini del grafo di flusso (v. oltre) di un programma càlcolino la stessa funzione o no.

**Teorema di Weyuker** I seguenti problemi sono indecidibili:

- esistenza di un dato di ingresso che causi l'esecuzione di una particolare istruzione.
- esistenza di un dato di ingresso che causi l'esecuzione di una particolare decisione.
- esistenza di un dato di ingresso che causi l'esecuzione di un particolare cammino.
- esistenza di un dato di ingresso che causi l'esecuzione di ogni istruzione.
- esistenza di un dato di ingresso che causi l'esecuzione di ogni decisione.
- esistenza di un dato di ingresso che causi l'esecuzione di ogni decisione.

Tutti questi risultati sono negativi in quanto sono risultati di indecidibilità, e la loro importanza sta nel fatto che esprimono i limiti teorici dell'attività di testing. Comunque bisogna tener presente che i teoremi di indecidibilità si riferiscono a programmi arbitrari, e non escludono che per classi particolari di programmi si possano ottenere delle risposte ai problemi dimostrati indecidibili. Inoltre, l'indecidibilità di un problema significa che una soluzione al problema non può essere costruita meccanicamente (o meglio, algoritmicamente), ma non significa che la soluzione non esista.

È fondamentale, però, tener presente il limite espresso dalla

**Tesi di Dijkstra** Un test può rilevare la presenza di malfunzionamenti, ma non dimostrarne l'assenza.

### 6.3.2 Test strutturale

Nel test strutturale la selezione dei dati di test viene guidata dalle informazioni che abbiamo sul codice da testare, e presuppone quindi che il codice sorgente sia disponibile. L'idea generale è di trovare dei dati di ingresso che causino l'esecuzione di tutte le operazioni previste dal programma, nelle varie sequenze possibili. Il problema fondamentale sta nel fatto che il numero delle possibili sequenze di operazioni in generale non è limitato: si pensi ai programmi contenenti cicli, che possono essere iterati per un numero di volte determinabile solo a tempo di esecuzione.

I dati di test vengono scelti in base a dei *criteri di copertura*, che definiscono l'insieme di sequenze di operazioni che devono essere eseguite nel corso del test (si dice che il test *esercita* determinate operazioni o sequenze).

#### Il grafo di controllo

Nel test strutturale ci si riferisce ad una rappresentazione astratta del programma, detta *grafo di controllo*. Nella sua forma piú semplice, il grafo di controllo viene costruito associando un nodo a ciascuna istruzione o condizione di istruzioni condizionali o iterative, e collegando i nodi con archi in modo tale che i cammini del grafo rappresentino le possibili sequenze di esecuzione. A seconda del grado di dettaglio desiderato, piú istruzioni possono essere rappresentate da un solo nodo, oppure una istruzione può essere scomposta in operazioni elementari, ciascuna rappresentata da un nodo.

### Criterio di copertura dei comandi

Questo criterio richiede che ogni istruzione (eseguibile) del programma sia eseguita per almeno un dato appartenente al test. Per esempio, consideriamo il seguente programma:

```
main()
{
    int x, y;
    cin >> x >> y;
    if (x != 0)
        x = x + 10;
    y = y / x;
    cout << x << endl << y;
}
```

Il dominio del programma è l'insieme delle coppie ordinate di interi. Qualsiasi test contenente coppie $(x, y)$ con $x$ diverso da zero soddisfa il criterio di copertura. Osserviamo però che questo criterio esclude i test contenenti coppie con $x$ nullo, in corrispondenza delle quali si verifica un malfunzionamento.

### Criterio di copertura delle decisioni

Il criterio di copertura delle decisioni richiede che ogni arco del grafo di controllo venga percorso almeno una volta. Per il programma visto precedentemente, il criterio di copertura delle decisioni richiede che il test contenga coppie sia con $x$ diverso da zero che con $x$ uguale a zero, e in questo modo si trova il malfunzionamento dovuto alla divisione per zero. Ma per altri programmi il questo criterio non basta. Consideriamo quest'altro programma:

```
main()
{
    int x, y;
    cin >> x >> y;
    if (x == 0 || y > 0)
        y = y / x;
    else
        x = y + 2 / x;
    cout << x << endl << y;
}
```

Per il criterio di copertura delle decisioni, ogni test deve contenere sia coppie tali che la condizione `(x = 0 || y > 0)` sia vera, sia coppie tali che la

condizione sia falsa. Un test accettabile è {`(x = 5, y = 5), (x = 5, y = -5)`}. Anche questo test è incapace di rilevare il malfunzionamento dovuto alla divisione per zero.

### Criterio di copertura delle condizioni

Il problema dell'esempio precedente sta nel fatto che nel programma la decisione se eseguire il ramo `then` o il ramo `else` dipende da una condizione logica complessa. Il criterio di copertura delle condizioni è piú fine del criterio di copertura delle decisioni, poiché richiede che ciascuna condizione elementare venga resa sia vera che falsa dai diversi dati appartenenti al test. Nell'esempio considerato, questo criterio viene soddisfatto, per esempio, dal test {`(x = 0, y = -5), (x = 5, y = 5)`}.

Questo test, però, non soddisfa il criterio di copertura delle decisioni (si può verificare che per il test visto la condizione complessiva è sempre vera), per cui non riesce a rilevare la divisione per zero nel ramo `else`.

### Criterio di copertura delle decisioni e delle condizioni

Questo criterio combina i due precedenti, richiedendo che sia le condizioni elementari sia le condizioni complessive siano vere e false per diversi dati di test. Nell'esempio considerato, questo criterio viene soddisfatto dal test {`(x = 0, y = -5), (x = 5, y = 5), (x = 5, y = -5)`}; osserviamo però che in questo caso non si rileva la divisione per zero nel ramo `then`.

### Criterio di $n$-copertura dei cicli

I criteri d copertura visti finora non dànno indicazioni per la scelta dei dati di test quando il programma contiene dei cicli, e quindi non aiutano a trovare i (numerosi) malfunzionamenti che si verificano dopo un certo numero di iterazioni. Il criterio di $n$-copertura dei cicli richiede che ogni cammino contenente al piú $n$ iterazioni di ogni ciclo venga eseguito per almeno un dato di test.

Che esegue i test deve scegliere dei valori appropriati di $n$. È comune scegliere $n$ uguale a zero (il ciclo non viene eseguito), uguale ad uno e a due, ed uguale al numero massimo prevedibile di iterazioni, al massimo meno uno ed al massimo piú uno. Bisogna però osservare che in molti casi il numero massimo prevedibile di iterazioni è tanto grande da rendere inapplicabile

il criterio di $n$-copertura corrispondente. Inoltre, certi malfunzionamenti si verificano dopo un numero ben definito di iterazioni, ed eseguire il ciclo ulteriormente non aggiunge alcuna informazione.

### Criteri di copertura Data Flow

Questi criteri permettono di trovare il numero ottimo di iterazioni di ciascun ciclo che devono essere eseguite in un test, basandosi sui flussi di dati relativi alle variabili che appaiono nel ciclo. L'idea alla base di questi criteri è di scegliere una particolare categoria di operazioni sulle variabili ed eseguire ciascun ciclo un numero di volte sufficiente ad esercitare tutte le possibili sequenze delle operazioni prescelte.

Premettiamo le seguenti definizioni:

- insieme *def* di un nodo: insieme delle variabili che vengono definite nell'istruzione associata al nodo.
- insieme *use* di un nodo: insieme delle variabili che vengono usate nell'istruzione associata al nodo.
- cammino *libero da definizioni* per una variabile $x$ dal nodo $i$ al nodo $j$: un cammino $(i, n_1, \ldots, n_m, j)$ tale che $x$ non appartenga ad alcuno degli insiemi *def* dei nodi sul cammino.
- insieme *du* di un nodo $i$ e di una variabile $x$: $\mathrm{du}(x, i)$ è l'insieme di nodi tale che $x \in \mathrm{def}(i)$, ogni nodo $j \in \mathrm{du}(x, i)$ sia raggiungibile da $i$ attraverso almeno un cammino libero da definizioni per $x$, e $x \in \mathrm{use}(j)$.

I criteri di copertura considerati sono:

**copertura delle definizioni** per ogni nodo $i$ ed ogni variabile $x \in \mathrm{def}(i)$, viene eseguito un cammino libero da definizioni da $i$ ad almeno un elemento di $\mathrm{du}(x, i)$.

**copertura di tutti gli usi** per ogni nodo $i$ ed ogni variabile $x \in \mathrm{def}(i)$, viene eseguito un cammino libero da definizioni da $i$ a ciascun elemento di $\mathrm{du}(x, i)$.

**copertura di tutti i cammini du** per ogni nodo $i$ ed ogni variabile $x \in \mathrm{def}(i)$, vengono eseguiti tutti i cammini liberi da definizioni da $i$ a ciascun elemento di $\mathrm{du}(x, i)$.

Consideriamo il seguente esempio:

```
main()
{
    int a, b, t, x, y;    // 1
    cin >> x >> y;        // 2
    a = x;                // 3   d
    b = y;                // 4
    while (a != b)        // 5   u
        if (a > b) {      // 6   u
            t = a - b;    // 7   u
            a = t;        // 8   d
        } else {          // 9
            t = b - a;    // 10  u
            b = t;        // 11
        }                 // 12
    cout << a;            // 13  u
}
```

Limitiamoci ad esaminare la variabile `a`, i cui insiemi *du* per i nodi 3 e 8 sono, rispettivamente $\text{du}(3, a) = \{5, 6, 7, 10, 13\}$ e $\text{du}(8, a) = \{5, 6, 7, 10, 13\}$ (i due insiemi sono uguali). Applicando, per esempio, il criterio di copertura dei cammini *du*, dobbiamo selezionare dei test che facciano eseguire dei cammini contenenti i nodi $(12, 5, 6, 7)$, raggiungibili con una iterazione, ed i nodi $(12, 5, 6, 10)$, raggiungibili con un'altra iterazione. Ulteriori iterazioni non esercitano altri cammini *du*, per cui due iterazioni bastano a soddisfare il criterio. Un test possibile è $\{(x = 9, y = 6)\}$.

### 6.3.3 Testing funzionale

Nel testing funzionale bisogna prima di tutto analizzare un modello concettuale del sistema da collaudare, con l'obiettivo di trovare dei sottoinsiemi del dominio di ingresso da cui estrarre i dati di test.

Il modello si basa sui requisiti, o piú precisamente sulle specifiche. Se queste ultime sono abbastanza precise, esse stesse costituiscono un modello da cui estrarre le informazioni necessarie per il test. Se le specifiche sono informali, devono essere analizzate per costruire il modello.

Un approccio generale al testing strutturale si basa sulla costruzione di un grafo, al quale si possono applicare i criteri di copertura visti nel testing strutturale (ove il grafo considerato è il grafo di controllo del *codice*). Questo grafo si riferisce, a seconda dell'applicazione e/o del metodo di specifica, a vari aspetti *dell'applicazione* (e non dell'implementazione), quali il flusso del controllo, il flusso dei dati, l'evoluzione dello stato, ed altri ancora.

### Tabelle di decisione e grafi causa/effetto

Le condizioni sui dati di ingresso e su quelli di uscita vengono rappresentate da variabili booleane. Le dipendenze logiche (AND, OR, NOT, esclusione) possono essere rappresentate in forma tabulare o grafica, ma anche quando si usa la forma grafica, generalmente questa viene ricondotta ad una forma tabulare, piú adatta ad un'analisi sistematica e piú facilmente automatizzabile.

In forma tabulare, si usano *tabelle di decisione*. In queste tabelle le righe rappresentano i possibili valori booleani delle condizioni sui dati di ingresso e di uscita, e le colonne rappresentano le possibili combinazioni delle condizioni. I test vengono scelti in modo tale da esercitare le varie combinazioni degli ingressi.

Consideriamo, per esempio, un'interfaccia per un database, che accetti dei comandi per mostrare il contenuto di un indice, diviso in dieci sezioni. Per chiedere la visualizzazione di una sezione, l'utente deve dare un comando formato da due caratteri, dei quali il primo è `D` oppure (indifferentemente) `L`, ed il secondo è la cifra corrispondente alla sezione richiesta (da 0 a 9). Se il comando è formato correttamente il sistema risponde visualizzando la sezione, se il primo carattere non è corretto il sistema stampa un messaggio di errore (messaggio $A$), e se il secondo carattere non è corretto viene stampato un altro messaggio di errore (messaggio $B$). Per scrivere la tabella di decisione, dobbiamo individuare le condizioni booleane sugli ingressi (cause) e quelle sulle uscite (effetti), e numerarle per comodità di riferimento (i numeri sono stati scelti in modo da distinguere condizioni sugli ingressi, sulle uscite, e intermedie):

| Cause | |
|---|---|
| 1 | il primo carattere è `D` |
| 2 | il primo carattere è `L` |
| 3 | il secondo carattere è una cifra |
| 20 | il primo carattere è `D` oppure `L` |

| Effetti | |
|---|---|
| 50 | visualizzazione della sezione richiesta |
| 51 | messaggio $A$ |
| 52 | messaggio $B$ |

La causa 20 è intermedia fra le condizioni sugli ingressi e quelle sulle uscite. La tabella di decisione è la seguente:

| | | | | | | |
|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 1 | 1 |
| 3 | 0 | 1 | 0 | 1 | 0 | 1 |
| 20 | 1 | 1 | 0 | 0 | 1 | 1 |
| 50 | 0 | 1 | 0 | 0 | 0 | 1 |
| 51 | 0 | 0 | 1 | 1 | 0 | 0 |
| 52 | 1 | 0 | 1 | 0 | 1 | 0 |

La prima colonna, per esempio, rappresenta la situazione in cui la condizione 1 è vera, le condizioni 2 e 3 sono false, e conseguentemente le condizioni 50 e 51 sono false e la condizione 52 è vera.

In forma grafica, si usano *grafi causa/effetto* i cui nodi terminali sono condizioni sui dati di ingresso (cause) e sui dati di uscita (effetti), ed i nodi interni sono condizioni intermedie, ciascuna espressa come combinazione logica semplice (AND, OR, NOT) delle rispettive cause. La Fig. 6.1 mostra il grafo corrispondente all'esempio già esaminato.

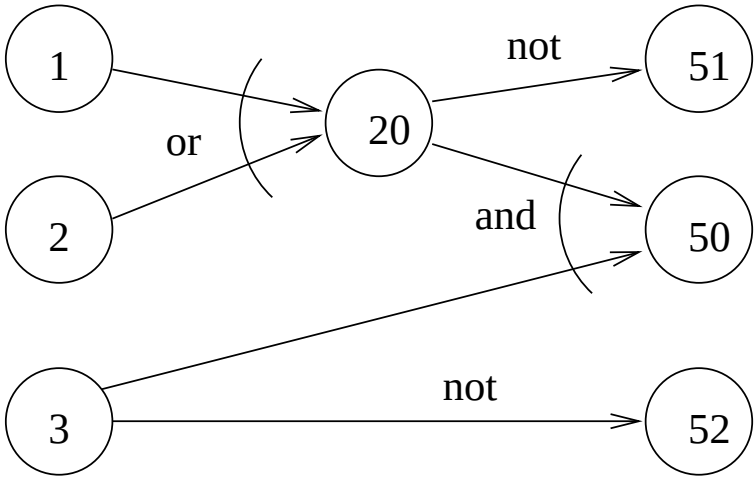


Figura 6.1: Grafo causa/effetto

### Testing sintattico

Spesso i dati di ingresso sono descritti formalmente da una grammatica. Questo vale non solo per i compilatori e gli interpreti di linguaggi di programmazione, ma anche per gli interpreti di comandi dei sistemi operativi e in generale per le applicazioni che richiedono informazioni in forma testuale, come, per esempio, file di configurazione, numeri telefonici, basi di dati.

I dati di test si ottengono applicando le regole di produzione della grammatica per generare sequenze corrette di dati di ingresso (*clean test*). Inoltre si possono generare dati scorretti (*dirty test*) per verificare che vengano riconosciuti come tali e trattati adeguatamente.

Una grammatica può essere rappresentata anche sotto forma di grafo, per cui anche il testing sintattico può essere affrontato con i metodi generali basati su grafi.

#### Flussi di controllo

Il comportamento di un sistema può essere descritto da un algoritmo, a cui possiamo associare un grafo del flusso di controllo (ovviamente distinto e indipendente dal grafo del flusso di controllo dell'implementazione). A questo grafo si possono applicare i criteri di copertura ed i metodi di analisi visti nel testing strutturale.

#### Flussi di dati

La descrizione di un sistema in termini di flussi di dati si presta all'uso di tecniche analoghe a quelle basate sul flusso dei dati impiegate nel testing strutturale.

#### Macchine a stati finiti

Se il sistema viene descritto da un automa a stati finiti, il grafo usato nel testing è ovviamente il diagramma di stato.

## 6.4 CppUnit e Mockpp

I framework CppUnit servono a facilitare la scrittura di programmi di test, sia nel test di unità che nel test di integrazione. La Fig. 6.2 mostra l'architettura di un generico sistema in cui mettiamo in evidenza un componente da collaudare. Supponendo che prima di **UnderTest** sia stato implementato e collaudato solo il componente Provider1, bisogna realizzare l'infrastruttura di test (*test harness*) mostrata in Fig. 6.3. I framework CppUnit e Mockpp servono a realizzare, ripettivamente, i driver e gli stub.

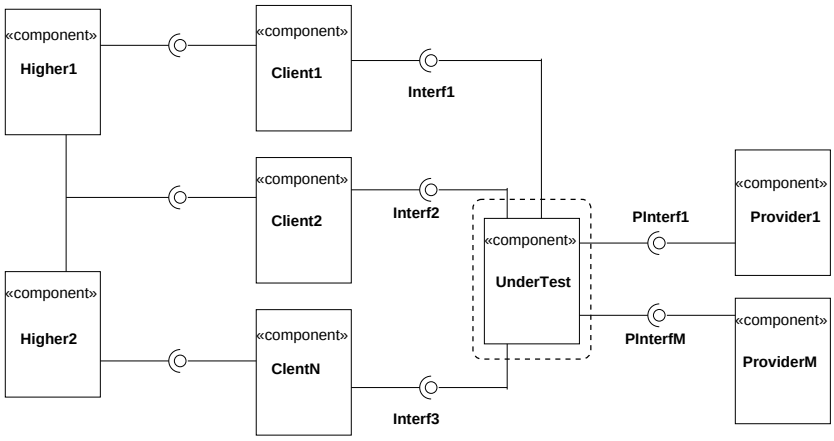


Figura 6.2: Integrazione di un componente.

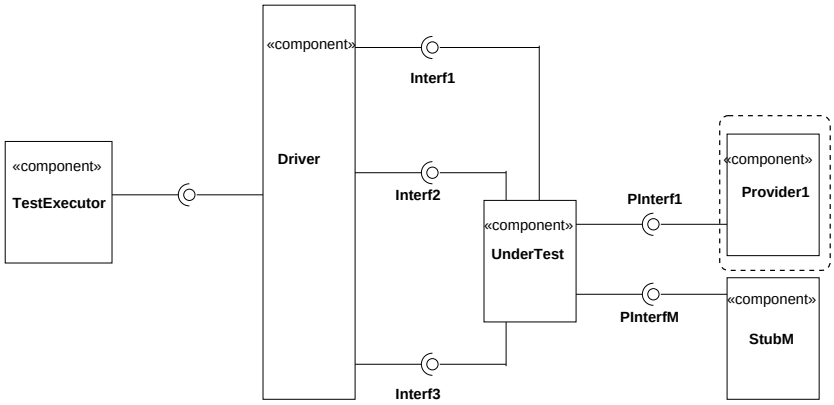


Figura 6.3: Infrastruttura di test.

### 6.4.1 Il framework CppUnit

Per eseguire un test con questo framework, il collaudatore deve scrivere una classe driver. Ciascun metodo di questa classe esegue un caso di test, e il framework fornisce un programma che esegue i test e raccoglie i risultati.

Il framework CppUnit si può descrivere, in modo semplificato, come in Fig. 6.4, dove **UnitA** è la classe da collaudare, **UnitB** è una classe che collabora con la precedente, e **UnitATest** è il driver. La classe **TestFixture** crea ed inizializza le istanze della classe sotto test e le altre classi, e poi le ddistrugge alla fine del test. **TestCaseAaCaller** esegue il caso di test *a*, cioè il metodo testCaseAa() della classe **UnitATest**, e **TestRunner** è la classe di piú

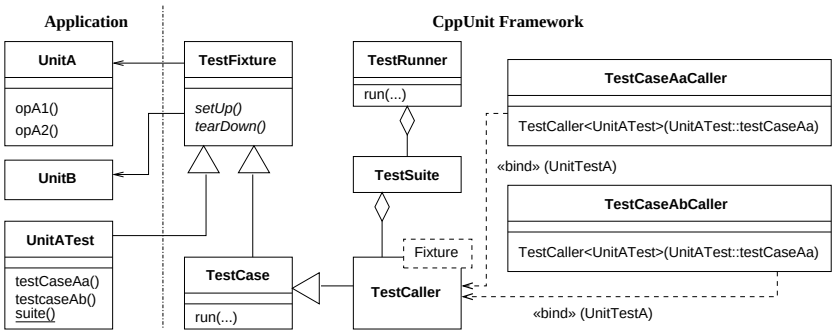


Figura 6.4: Infrastruttura CppUnit.

alto livello nel programma di test, quella che coordina le altre. La Fig. 6.5 l'applicazione del framework al test di una classe **Complex**.

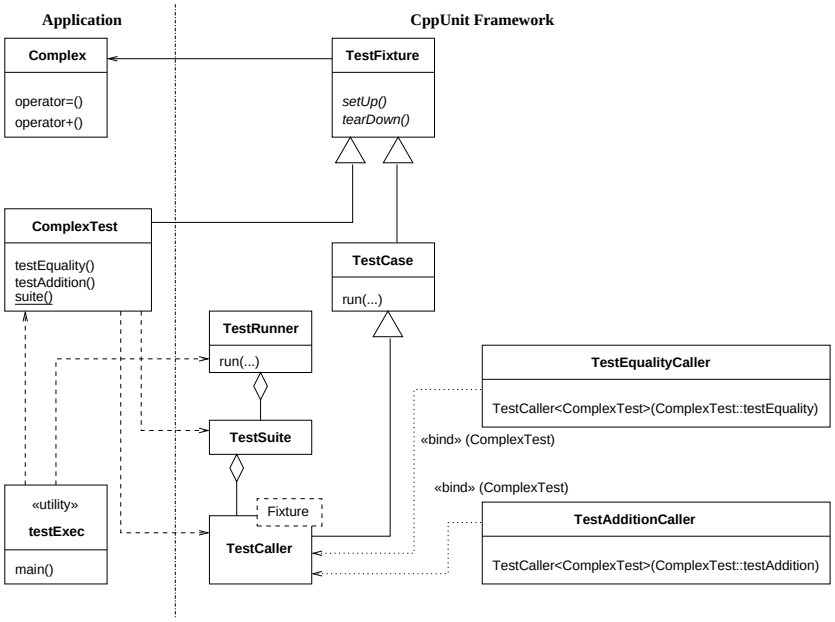


Figura 6.5: Infrastruttura CppUnit.

Supponiamo che la classe `Complex` sia cosí definita:

```
class Complex {
    double real;
    double imaginary;
public:
```

```
    Complex(double r, double i = 0) : real(r), imaginary(i) {};
    friend bool operator==(const Complex& a, const Complex& b);
    friend Complex operator+(const Complex& a, const Complex& b);
};

bool
operator==(const Complex &a, const Complex &b)
{
    return a.real == b.real && a.imaginary == b.imaginary;
}

Complex
operator+(const Complex &a, const Complex &b)
{
    return Complex(a.real, a.imaginary + b.imaginary);
}
```

La classe driver è `ComplexTest`:

```
class ComplexTest : public CppUnit::TestFixture  {
private:
    Complex* m_10_1;
    Complex* m_1_1;
    Complex* m_11_2;
public:
    static CppUnit::Test *suite();
    void setUp();
    void tearDown() ;
    void testEquality();
    void testAddition();
};

CppUnit::Test*
ComplexTest::
suite()
{
    CppUnit::TestSuite* suiteOfTests = new CppUnit::TestSuite("ComplexTest");
    suiteOfTests->addTest(new CppUnit::TestCaller<ComplexTest>(
                                    "testEquality",
                                    &ComplexTest::testEquality));
    suiteOfTests->addTest(new CppUnit::TestCaller<ComplexTest>(
                                    "testAddition",
                                    &ComplexTest::testAddition));
    return suiteOfTests;
}
void
ComplexTest::
setUp()
{
    m_10_1 = new Complex(10, 1);
    m_1_1 = new Complex(1, 1);
    m_11_2 = new Complex(11, 2);
```

```
}

void
ComplexTest::
tearDown()
{
    delete m_10_1;
    delete m_1_1;
    delete m_11_2;
}

void
ComplexTest::
testEquality()
{
    CPPUNIT_ASSERT(*m_10_1 == *m_10_1);
    CPPUNIT_ASSERT(!(*m_10_1 == *m_11_2));
}

void
ComplexTest::
testAddition()
{
    CPPUNIT_ASSERT(*m_10_1 + *m_1_1 == *m_11_2);
}
```

dove `testEquality()` e `testAddition()` sono i casi di test. Gli altri metodi creano la sequenza di test da eseguire (`suite()`), creano le istanze di `Complex` da usare nei test (`setup()`), e le distruggono (`tearDown()`).

Il programma principale crea un istanza di (una sottoclasse di) `TestRunner`, la inizializza e la esegue:

```
int
main()
{
    CppUnit::TextTestRunner runner;
    runner.addTest(ComplexTest::suite());
    runner.run();
    return 0;
}
```

Le macro `CPPUNIT_ASSERT` nei casi di test verificano le condizioni specificate, e se non risultano vere scrivono dei messaggi di errore.

### 6.4.2 Il framework Mockpp

Il framework Mockpp si può descrivere come in Fig. 6.6, dove **UnitA** è la classe da collaudare, **Provider1** è la classe richiesta da **UnitA** che dobbiamo simulare con uno stub, **IProvider1** è la sua interfaccia, e **MockProvider1** è lo stub. La classe **MockObject** è la base degli stub, e contiene un insieme di istanze di classi derivate da **Expectation**; queste ultime descrivono i valori che ci si aspetta vengano passati ai metodi di **Provider1** nell'esecuzione dei metodi di **UnitA**.

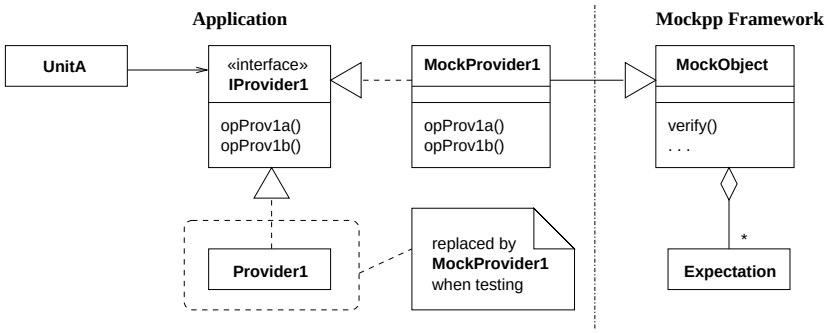


Figura 6.6: Infrastruttura Mockpp.

Le *expectation* definiscono il comportamento dello stub. Sono dei contenitori che, all'atto della creazione dello stub, vengono riempiti con valori o vincoli definiti dal collaudatore. Durante l'esecuzione del test le istanze dell'unità sotto test invocano i metodi dello stub, che passano i loro argomenti alle *expectation*. Queste ultime verificano se i valori degli argomenti corrispondono ai valori previsti o rispettano i vincoli specificati, sollevando un'eccezione se ci sono delle discrepanze. Le *expectation* possono eseguire diversi tipi di verifiche, come confrontare valori e controllare il loro ordinamento temporale.

La Fig. 6.7 mostra un'applicazione di questo framework al collaudo di una classe **Consumer**, che richiede l'interfaccia **IFile**.

Un'istanza di `Consumer`, con l'operazione `load()` apre un file, legge tre righe e chiude il file, poi compie delle elaborazioni (`process()`) che non ci interessano:

```
class IFile;

class Consumer {
```

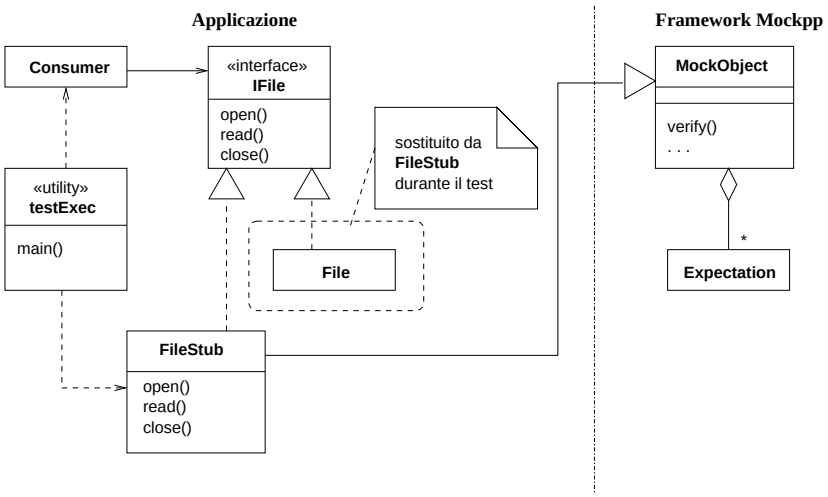


Figura 6.7: Infrastruttura Mockpp.

```
    IFile   *configfile;
    std::string  config1;
    std::string  config2;
    std::string  config3;
public:
    Consumer(IFile *intf) : configfile(intf) {};
    void load();
    void process();
};

void
Consumer::
load()
{
    configfile->open("file1.lst");
    config1 = configfile->read();
    config2 = configfile->read();
    config3 = configfile->read();
    configfile->close();
}
```

Nella classe `FileStub` definiamo due *expectation*: `open_name` deve verificare che il file venga aperto passando il nome previsto, `close_counter` deve verificare che il file venga chiuso. L'oggetto `read_data` contiene i valori che devono essere restituiti alle chiamate `read()`. I metodi della sono implementati per mezzo di operazioni su questi oggetti:

```
class FileStub : public IFile, public mockpp::MockObject {
```

```
public:
    mockpp::ExpectationList<std::string>  open_name;
    mockpp::ExpectationCounter            close_counter;
    mockpp::ReturnObjectList<std::string> read_data;
    FileStub() { /* inizializzazioni */ };
    virtual void open(const std::string &name);
    virtual std::string read();
    virtual void close();
};

void
FileStub::
open(const std::string &name)
{
    open_name.addActual(name);
}

std::string
FileStub::
read()
{
    return read_data.nextReturnObject();
}

void
FileStub::
close()
{
    close_counter.inc();
}
```

Il programma principale esegue il test creando un'istanza di `FileStub` e di `Consumer`, inizializzando lo stub e usando l'unità sotto test come previsto dal piano di test:

```
int
main(int argc, char **argv)
{
    try {
        FileStub mock;

        // nome passato a open()
        mock.open_name.addExpected("file1.lst");

        // numero di operazioni close()
        mock.close_counter.setExpected(1);

        // righe lette da Consumer
        mock.read_data.addObjectToReturn("record-1");
        mock.read_data.addObjectToReturn("record-2");
        mock.read_data.addObjectToReturn("record-3");
```

```
        // esecuzione
        Consumer consumer(&mock);
        consumer.load();
        consumer.process();

        // controllo
        mock.verify();
        std::cout << "Test eseguito con successo"
                  << std::endl;
    } catch(std::exception &ex) {
        std::cout << std::endl
            << "Errori.\n" << ex.what() << std::endl
        return 1;
    }
    return 0;
}
```

## 6.5 Test in grande

### 6.5.1 Test di integrazione

Nel test di unità, il test di ciascun modulo richiede l'uso di moduli ausiliari che simulano i moduli che, nel sistema reale, interagiscono col modulo in esame. I moduli ausiliari che simulano i moduli cliente sono detti *driver*, quelli che simulano i moduli servente si dicono *stub*. I driver e gli stub devono simulare degli interi sottosistemi, ma ovviamente devono essere realizzati in modo semplice. Per esempio, invece di eseguire i calcoli richiesti, possono fornire al modulo sotto test i valori particolari richiesti dal test, ricavandoli da tabelle o dall'interazione con il collaudatore.

Nel test di integrazione si vuole testare l'interfacciamento fra moduli il cui funzionamento è già stato testato individualmente. È possibile eseguire il test di integrazione dopo che tutti i moduli sono stati testati individualmente, assemblandoli e provando il sistema completo (*big-bang test*). In genere si preferisce una strategia incrementale del test di integrazione, che permette di integrare i moduli man mano che vengono completati e che superano il test di unità. In questo modo gli errori di interfacciamento vengono scoperti prima ed in modo piú localizzato, e si risparmia sul numero di driver e di stub che devono essere sviluppati.

La strategia del test di integrazione generalmente ricalca la strategia di sviluppo, e quindi può essere top-down o bottom-up, o una combinazione

delle due. Con la strategia top down non c'è bisogno di driver, poiché i moduli sviluppati e testati in precedenza fanno da driver per i moduli integrati successivamente. Analogamente, con la strategia bottom-up non c'è bisogno di stub. La strategia top-down permette di avere presto dei prototipi, mentre la strategia bottom-up permette di testare subito i moduli terminali, che generalmente sono i piú critici.

### 6.5.2 Test di sistema

Il test di integrazione, sebbene nella sua ultima fase si applichi al sistema completo, è rivolto a verificare la correttezza dei singolo moduli rispetto all'interfacciamento col resto del sistema. Quindi il test di integrazione non può verificare le proprietà globali del sistema, che non si possono riferire a qualche particolare modulo o sottosistema. Alcune di queste proprietà sono la robustezza e la riservatezza (*security*). Fra i possibili test di sistema citiamo il *test di stress* ed il *test di robustezza*.

Il test di stress consiste nel sottoporre il sistema ad uno sforzo superiore a quello previsto dalle specifiche, per assicurarsi che il superamento dei limiti non porti a malfunzionamenti incontrollati, ma solo ad una degradazione delle prestazioni. La grandezza che definisce lo sforzo dipende dall'applicazione: per esempio, potrebbe essere il numero di utenti collegati contemporaneamente ad un sistema multiutente, od il numero di transazioni al minuto per un database.

Il test di robustezza consiste nell'inserire dati di ingresso scorretti. Anche in questo caso ci si aspetta che il sistema reagisca in modo controllato, per esempio stampando dei messaggi di errore e, nel caso di sistemi interattivi, rimettendosi in attesa di ulteriori comandi dall'utente.

### 6.5.3 Test di accettazione

Il test di accettazione è un test di sistema eseguito dal committente (invece che dal produttore), che in base al risultato decide se accettare o no il prodotto. Il test di accettazione ha quindi una notevole importanza legale ed economica, e la sua pianificazione può far parte del contratto.

Un tipo di test di accettazione è quello usato per prodotti destinati al mercato. In questo caso il test avviene in due fasi: nella prima fase ($\alpha$-*test*)

il prodotto viene usato all'interno dell'organizzazione produttrice, e nella seconda ($\beta$-*test*) viene distribuito in prova ad alcuni clienti selezionati.

### 6.5.4 Test di regressione

Il test di regressione serve a verificare che una nuova versione di un prodotto non produca nuovi errori rispetto alle versioni precedenti e sia compatibile con esse. Il test di regressione può trarre grande vantaggio da una buona gestione del testing, che permetta di riusare i dati ed i risultati dei test delle versioni precedenti.

## 6.6 Gestione del processo di testing

L'attività di testing, al pari delle altre fasi della produzione del software, deve essere pianificata e controllata. In particolare, deve essere documentata. Lo standard IEEE 829/1983 definisce i deliverable relativi alla fase di testing:

**Test Plan** descrive l'oggetto da testare, le caratteristiche da testare, il programma delle operazioni di testing, i criteri di accettazione e le risorse.

**Test Design Specification** specifica le tecniche di testing, i metodi di analisi dei risultati, la lista dei test da eseguire con le relative descrizioni generali e motivazioni.

**Test Case Specification** per ogni test (*test case*) individuato dalla TDS (v. sopra), vengono specificati i dati di test, i risultati attesi e le condizioni di esecuzione.

**Test Procedure Specification** istruzioni passo per passo sulla preparazione, l'esecuzione, e la valutazione dei risultati di ciascun test.

**Test Item Transmittal Report** documento che accompagna ogni modulo (o piú in generale ogni *test item*) consegnato ai collaudatori, contenente l'identificazione del modulo, la sua collocazione, ed altre informazioni.

**Test Log** il datbase della fase di testing, contenente la storia dettagliata di tutte le operazioni di test e dei rispettivi risultati.

**Test Incident Report** descrive tutti gli incidenti avvenuti nella fase di test, cioè tutti gli eventi notevoli che richiedono ulteriori analisi.

**Test Summary Report** rapporto conclusivo, contenente una valutazione critica.

## Letture

**Obbligatorie:** Cap. 6 Ghezzi, Jazayeri, Mandrioli, esclusa Sez. 6.9.

# Appendice A

# Metriche del software

Tutte le attività industriali richiedono che le proprietà dei prodotti e dei semilavorati, oltre che delle risorse e dei processi impiegati, vengano misurate. Per esempio, nell'industria meccanica possiamo considerare le dimensioni e le proprietà tecnologiche dei pezzi (prodotti e semilavorati), le dimensioni, il consumo energetico e la velocità delle macchine operatrici (risorse materiali), il numero e lo stipendio degli addetti (risorse umane), il tempo di produzione, i tempi morti e la produttività dei cicli di lavorazione (processi). Tutte queste grandezze, o *metriche*, sono necessarie per pianificare il lavoro e valutarne i risultati: è difficile immaginare un'attività ingegneristica che possa fare a meno di metriche.

Anche da questo punto di vista, l'industria del software è abbastanza anomala, in quanto l'uso di metriche è molto meno evoluto che negli altri settori. Questo dipende in parte dalla natura "immateriale" del software, che rende difficile la stessa definizione di grandezze misurabili. Questo non vuol dire che manchino delle metriche per il software: in effetti ne sono state proposte circa 500. Purtroppo non è chiara l'utilità effettiva di certe metriche, e spesso non si trova una giustificazione teorica per le metriche effettivamente usate in pratica. L'argomento delle metriche del software è quindi complesso ed in continua evoluzione.

Per definire una metrica, bisogna innanzitutto individuare le *entità* che ci interessano, e gli *attributi* di tali entità che vogliamo quantificare. Nell'ingegneria del software le entità possono essere *prodotti* (inclusi i vari semilavorati, o *artefatti*), *risorse*, e *processi*. Alcuni prodotti sono i documenti di specifica e di progetto, il codice sorgente, i dati ed i risultati di test. Fra le risorse ricordiamo il personale e gli strumenti hardware e software, come

esempi di processi citiamo le varie fasi del ciclo di vita.

Gli attributi possono essere *interni* o *esterni*. Gli attributi interni di un'entità possono essere valutati considerando l'entità isolatamente, mentre gli attributi esterni richiedono che l'entità venga osservata in relazione col proprio ambiente. Generalmente gli attributi esterni sono i piú interessanti dal punto di vista dell'utente e da quello del produttore, ma non sono misurabili direttamente e vengono valutati a partire dagli attributi interni. Per esempio, nei documenti di specifica e di progetto possiamo misurare gli attributi interni *dimensione* e *modularità*. Questi influenzano gli attributi esterni *comprensibilità* e *manutenibilità*: si suppone che un documento breve e ben strutturato sia piú comprensibile e facile da modificare di uno lungo e caotico.

Dei vari attributi ci può interessare una *valutazione* o una *predizione* (o *stima*). La stima di una metrica si ottiene a partire dai valori di altre metriche, applicando un *modello*, cioè una formula o algoritmo.

Nel seguito tratteremo alcune metriche, scelte fra le piú diffuse o piú citate in letteratura.

## A.1 Linee di codice

La metrica piú semplice e di piú diffusa applicazione è costituita dal numero di linee di codice (LOC) di un programma. Quando si usa questa metrica, bisogna specificare come vengono contate le linee di codice: per esempio, bisogna decidere se contare soltanto le istruzioni eseguibili o anche le dichiarazioni. Generalmente i commenti non vengono contati. Nei linguaggi che permettono di avere piú istruzioni su uno stesso rigo o di spezzare un'istruzione su piú righi, si può usare il numero di istruzioni invece del numero di linee di codice, ma anche il numero di istruzioni può essere difficile da definire.

Questa metrica è alla base di vari indici, quali la *produttività* (linee di codice prodotte per persona per unità di tempo), la *qualità* (numero di errori per linea di codice), il *costo unitario* (costo per linea di codice). Possiamo notare che l'uso di questi indici per valutare la qualità del software penalizza i programmi brevi.

## A.2 Software science

La teoria della *Software science* propone un insieme di metriche basate su quattro parametri del codice sorgente:

| | |
|---|---|
| $\eta_1$ | numero di operatori distinti |
| $\eta_2$ | numero di operandi distinti |
| $N_1$ | numero di occorrenze di operatori |
| $N_2$ | numero di occorrenze di operandi |

Gli operatori comprendono, oltre agli operatori aritmetici, le parole chiave per la costruzione di istruzioni composte e le chiamate di sottoprogramma. Da questi parametri si possono ottenere metriche sia effettive che stimate.

La *lunghezza* del programma è la somma dei numeri di occorrenze di operatori ed operandi:

$$N = N_1 + N_2$$

La lunghezza stimata si ricava invece dai numeri di simboli distinti:

$$\hat{N} = \eta_1 \log_2 \eta_1 + \eta_2 \log_2 \eta_2$$

(questa formula discende da certe ipotesi sulla struttura del programma).

Il *volume* del programma è il numero minimo di bit necessari a rappresentarlo:

$$V = N \log_2(\eta_1 + \eta_2)$$

Il *volume potenziale* del programma è il numero di bit necessari a rappresentare il programma se questo fosse ridotto ad un'unica chiamata di funzione. Questo dato si ottiene dalla formula:

$$V^* = (2 + \eta_2^*) \log_2(2 + \eta_2^*)$$

dove $\eta_2^*$ rappresenta il numero di argomenti della funzione equivalente al programma.

Il *livello* del programma, un indicatore del suo grado di astrazione, è

$$L = \frac{V^*}{V}$$

Il livello può essere stimato da:

$$\hat{L} = \frac{2\eta_2}{\eta_1 N_2}$$

Lo *sforzo* (*effort*) è una misura dell'impegno psichico richiesto dalla scrittura del programma. Questa misura si basa su un modello alquanto semplicistico dei processi mentali del programmatore, e viene cosí calcolata:

$$E = \frac{V}{L} = \frac{V^2}{V^*}$$

Dallo sforzo si ricaverebbe una stima del tempo richiesto per scrivere il programma:

$$\hat{T} = \frac{E}{S}$$

dove $S$ sarebbe il numero di decisioni elementari che un programmatore può compiere in un secondo, per cui si usa tipicamente il valore 18.

## A.3 Complessità

La complessità è un concetto molto importante per l'ingegneria del software, ma è difficile da definire e quindi da misurare. Le metriche di complessità sono quindi un argomento su cui la ricerca e il dibattito scientifico sono particolarmente vivi.

### A.3.1 Numero ciclomatico

Il *numero ciclomatico* è un concetto della teoria dei grafi che, applicato al grafo di controllo di un programma, dà una misura della sua complessità.

In un grafo contenente $n$ archi, un cammino può essere rappresentato da un vettore: se gli archi sono numerati da 1 a $n$, un cammino viene rappresentato da un vettore $V$ in cui $V_i$ è il numero di volte che l'$i$-esimo arco compare nel cammino. Un cammino è combinazione lineare di altri cammini se il suo vettore è combinazione lineare dei vettori associati agli altri cammini. Due o piú cammini si dicono linearmente indipendenti se nessuno di essi è combinazione lineare degli altri. Nel grafo della figura A.1, i cammini possibili sono rappresentati dalla seguente tabella:

| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| (a,b,c,d) | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 |
| (e,f,g,h) | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| (a,b,g,h) | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 |
| (e,f,c,d) | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 |

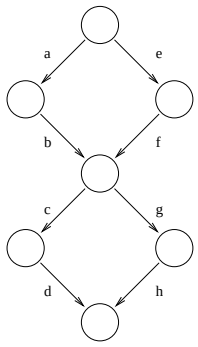


Figura A.1: Cammini linearmente indipendenti

Il cammino $(e, f, c, d) = (a, b, c, d)+(e, f, g, h)-(a, b, g, h)$ è una combinazione lineare degli altri tre, cioè non ne è linearmente indipendente.

Il numero ciclomatico di un programma è il numero di cammini completi (cioè cammini fra il nodo iniziale ed il nodo finale) linearmente indipendenti del grafo di controllo, modificato aggiungendo un arco fra il nodo terminale ed il nodo iniziale (in modo da rendere il grafo fortemente connesso). Se $e$ è il numero degli archi ed $n$ è il numero dei nodi, il numero ciclomatico è

$$v = e - n + 2$$

Se un programma è composto di $p$ sottoprogrammi, la formula diventa

$$v = e - n + 2p$$

Il numero ciclomatico può essere calcolato anche senza costruire il grafo di controllo. Se $d$ è il numero di decisioni (istruzioni condizionali) presenti nel programma, le formule per il numero ciclomatico, rispettivamente per un programma monolitico e per un programma composto di $p$ sottoprogrammi, diventano

$$v = d + 1$$

$$v = d + p$$

### A.3.2 I criteri di Weyuker

La ricercatrice Elaine Weyuker ha proposto una serie di criteri che dovrebbero essere soddisfatti dalle metriche di complessità. Sebbene questi criteri siano piú che altro di interesse teorico e siano oltretutto controversi, sono un

utile esempio del tipo di problemi che devono essere affrontati sia quando si inventa una metrica per qualche proprietà del software, sia quando vogliamo applicarne una già esistente, e dobbiamo quindi valutarne l'adeguatezza.

Nel seguito, $P$ e $Q$ denotano programmi o parti di programmi, $c()$ una misura di complessità, '$\equiv$' l'equivalenza funzionale, e ';' la concatenazione.

1. La misura di complessità non deve essere troppo grossolana (1).

   $$\exists P, Q \;\; c(P) \neq c(Q)$$

   (deve esistere almeno una coppia di programmi con complessità diverse).
2. La misura di complessità non deve essere troppo grossolana (2). Dato un numero non negativo $\bar{c}$, esiste solo un numero finito di programmi aventi complessità $\bar{c}$.
3. La misura di complessità non deve essere troppo fine.

   $$\exists P, Q \;\; c(P) = c(Q)$$

   (deve esistere almeno una coppia di programmi con la stessa complessità).
4. La complessità deve essere indipendente dalla funzionalità.

   $$\exists P, Q \;\; P \equiv Q \wedge c(P) \neq c(Q)$$

   (la misura deve poter discriminare due programmi che calcolano la stessa funzione).
5. La complessità deve essere monotòna rispetto alla concatenazione.

   $$\forall P, Q \;\; c(P) \leq c(P;Q) \wedge c(Q) \leq c(P;Q)$$

   (un programma è piú complesso di ciascuna sua parte).
6. Il contributo di una parte del programma alla complessità totale deve poter dipendere dal resto del programma.

   $$\exists P, Q, R \;\; c(P) = c(Q) \wedge c(P;R) \neq c(Q;R)$$

   $$\exists P, Q, R \;\; c(P) = c(Q) \wedge c(R;P) \neq c(R;Q)$$

   (due parti di programma di uguale complessità, concatenate ad una terza, possono dar luogo a programmi di complessità diversa).
7. La complessità deve dipendere dall'ordine in cui sono scritte le istruzioni.

   $$\exists P, Q \;\; \text{tali che } P \text{ sia una permutazione di } Q \text{ e } c(P) \neq c(Q)$$

8. La complessità non deve dipendere dalla scelta dei nomi di variabili e sottoprogrammi.

$$\forall P, Q \text{ tali che } P \text{ rinòmini } Q, c(P) = c(Q)$$

9. La complessità totale del programma può superare la somma delle complessità delle parti componenti.

$$\exists P, Q \;\; c(P) + c(Q) \leq c(P; Q)$$

La proprietà 6 fa sí che una misura di complessità tenga conto delle interazioni fra parti di programmi: due parti di programma di uguale complessità possono interagire in modo diverso col contesto in cui sono inseriti. Una misura che soddisfi questa proprietà, insieme alla proprietà 7, non è addititiva, cioè non è possibile ricavare la complessità di un programma sommando le complessità delle parti componenti. Alcuni ricercatori ritengono, al contrario, che le misure di complessità debbano essere additive.

La proprietà 7 permette di tener conto dell'ordinamento delle istruzioni di un programma, e in particolare di distinguere la complessità in base all'annidamento dei cicli. Il numero ciclomatico non rispetta questa proprietà.

La proprietà 8 rispecchia il fatto che si vuole considerare la complessità come un attributo strutturale ed interno, distinto dalla comprensibilità (attributo di carattere psicologico ed esterno). Evidentemente, un programma che usa identificatori significativi è piú comprensibile di un programma ottenuto sostituendo gli identificatori con sequenze casuali di lettere, ma la struttura dei due programmi è identica e quindi ugualmente complessa.

## A.4 Punti funzione

I *Punti Funzione* (*Function Points*) sono una metrica per i requisiti, e permettono, a partire dai requisiti utente, di ottenere una valutazione preliminare dell'impegno richiesto dal progetto. Piú precisamente, i FP sono un numero legato al tipo ed al numero di funzioni richieste al sistema, ed a varie caratteristiche del sistema stesso. Un numero di FP piú alto corrisponde ad un sistema piú complesso e costoso.

Il metodo di calcolo prevede che un sistema possa offrire cinque tipi di funzioni. Dall'analisi dei requisiti bisogna trovare quante funzioni di ciascun tipo sono richieste, e valutare la complessità delle operazioni per ciascun tipo

di funzione assegnandogli un peso numerico secondo la seguente tabella, che elenca i tipi di funzione previsti (contraddistinti da un indice $i$) con i pesi corrispondenti a tre livelli di complessità:

| | | Complessità | | |
|---|---|---|---|---|
| i | Tipi di funzione | bassa | media | alta |
| 1 | File logici interni | 7 | 10 | 15 |
| 2 | File di interfaccia esterni | 5 | 7 | 10 |
| 3 | Ingressi esterni | 3 | 4 | 6 |
| 4 | Uscite esterne | 4 | 5 | 7 |
| 5 | Richieste esterne | 3 | 4 | 6 |

I *file logici* sono gli insiemi di dati che devono essere mantenuti internamente dal sistema (ciascuno dei quali può corrispondere ad uno o piú file fisici, a seconda delle successive scelte di progetto che ne definiscono l'implementazione). I *file di interfaccia* sono i file scambiati con altri programmi, gli ingressi e le uscite sono le informazioni distinte fornite e, rispettivamente, ottenute dall'utente, le *richieste* sono le interrogazioni in linea che richiedono una risposta immediata dal sistema.

Alle caratteristiche del sistema (*General System Characteristics*), quali la riusabilità o la presenza di elaborazioni distribuite, che influenzano la complessità del progetto, viene assegnato un grado di influenza, cioè una valutazione dell'importanza di ciascuna caratteristica su una scala da zero a cinque. Le caratteristiche prese in considerazione dal metodo sono le quattordici mostrate nella tabella seguente, contraddistinte dall'indice $j$:

| j | Caratteristiche del sistema |
|---|---|
| 1 | Riusabilità |
| 2 | Trasmissione dati |
| 3 | Elaborazioni distribuite |
| 4 | Prestazioni |
| 5 | Ambiente usato pesantemente |
| 6 | Ingresso dati in linea |
| 7 | Ingresso dati interattivo |
| 8 | Aggiornamento in tempo reale dei file principali |
| 9 | Funzioni complesse |
| 10 | Elaborazioni interne complesse |
| 11 | Facilità d'istallazione |
| 12 | Facilità di operazione |
| 13 | Siti molteplici |
| 14 | Modificabilità |

La valutazione del grado d'influenza delle caratteristiche avviene sulla seguente scala:

| Grado d'influenza | |
|---|---|
| 0 | Nullo |
| 1 | Scarso |
| 2 | Moderato |
| 3 | Medio |
| 4 | Significativo |
| 5 | Essenziale |

Il numero di FP si calcola dalla seguente formula:

$$FP = (\sum_{i=1}^{5} N_i W_i)(0.65 + 0.01 \sum_{j=1}^{14} D_j)$$

dove $N_i$ è il numero di funzioni di tipo $i$, $W_i$ è il rispettivo peso, e $D_j$ è il grado d'influenza della $j$-esima caratteristica.

## A.5 Stima dei costi

Il costo di sviluppo di un sistema è ovviamente un parametro fondamentale per la pianificazione economica del lavoro. È importante poter prevedere i

costi con ragionevole accuratezza, e a questo scopo sono stati proposti vari modelli che calcolano una stima dei costi a partire da alcuni parametri relativi al sistema da realizzare ed all'organizzazione che lo deve sviluppare.

Il costo di un sistema viene generalmente riassunto in due fattori: la quantità di lavoro, cioè l'impegno di risorse umane, ed il tempo di sviluppo. L'impegno di risorse umane viene di solito quantificato in termini di *mesi-uomo*, cioè del tempo necessario ad una persona per eseguire tutto il lavoro, ovvero il numero di persone necessarie a fare tutto il lavoro nell'unità di tempo. È bene ricordare che questa misura è soltanto convenzionale, ed evitare l'errore di credere che tempo di sviluppo e personale siano grandezze interscambiabili. Non è possibile dimezzare il tempo di sviluppo raddoppiando l'organico, poiché nel tempo di sviluppo entrano dei fattori, quali il tempo di apprendimento ed i tempi di comunicazione all'interno del gruppo di lavoro, che crescono all'aumentare del numero di partecipanti. Il tempo di sviluppo, quindi, viene di solito calcolato separatamente dalla quantità di lavoro. Fissati questi due fattori, si può quindi valutare il numero di persone da impiegare (*staffing*).

### A.5.1 Il modello COCOMO

Il *COCOMO* (*COnstructive COst MOdel*) è un metodo di stima dei costi basato su rilevazioni statistiche eseguite su un certo numero di progetti di grandi dimensioni sviluppati presso una grande azienda produttrice di software. Esistono tre versioni del modello (*base*, *intermedio* e *avanzato*) di diversa accuratezza.

Ciascune delle tre versioni prevede che le applicazioni vengano classificate in tre categorie:

**Organic** Applicazioni semplici, di dimensioni limitate, di tipo tradizionale e ben conosciuto.

**Semi-detached** Applicazioni di complessità media (software di base).

**Embedded** Applicazioni complesse con requisiti rigidi di qualità ed affidabilità (sistemi in tempo reale).

Il principale dato d'ingresso del metodo è la dimensione del codice espressa il migliaia di linee di codice (KDSI), che deve essere stimata indipendentemente. Altre informazioni vengono usate per raffinare i risultati del calcolo, come vedremo fra poco.

### Il modello base

Se $S$ è la dimensione del codice in KDSI, la quantità di lavoro $M$ in mesi-uomo ed il tempo di sviluppo $T$ in mesi sono dati dalle formule:

$$M = a_b S^{b_b}$$

$$T = c_b M^{d_b}$$

dove i coefficienti $a_b$, $b_b$, $c_b$, e $d_b$ si ricavano dalla seguente tabella:

| Tipo dell'applicazione | $a_b$ | $b_b$ | $c_b$ | $d_b$ |
|---|---|---|---|---|
| Organic | 2.4 | 1.05 | 2.5 | 0.38 |
| Semi-detached | 3.0 | 1.12 | 2.5 | 0.35 |
| Embedded | 3.6 | 1.20 | 2.5 | 0.32 |

Il metodo permette di calcolare la ripartizione della quantità di lavoro e del tempo di sviluppo fra le diverse fasi del ciclo di vita. Il COCOMO presuppone un ciclo a cascata articolato nelle quattro fasi di *pianificazione ed analisi*, *progetto*, *sviluppo*, ed *integrazione e test*. I costi si riferiscono alle ultime tre fasi, e la loro ripartizione fra queste fasi si ottiene da due tabelle (una per la quantità di lavoro ed una per il tempo di sviluppo) che dànno la rispettiva percentuale secondo il tipo e la dimensione dell'applicazione. Le stesse tabelle forniscono anche la percentuale relativa al costo aggiuntivo della fase di pianificazione ed analisi.

La ripartizione percentuale della quantità di lavoro fra le fasi del ciclo di vita si ricava dalla tabella seguente (P & A: pianificazione ed analisi; I & T: integrazione e test):

| Tipo dell'applicazione | Fase | Dimensione (KDSI) 2 | 8 | 32 | 128 | 512 |
|---|---|---|---|---|---|---|
| Organic | P & A | 6 | 6 | 6 | 6 | |
| | Progetto | 16 | 16 | 16 | 16 | |
| | Sviluppo | 68 | 65 | 62 | 59 | |
| | I & T | 16 | 19 | 22 | 25 | |
| Semi-detached | P & A | 7 | 7 | 7 | 7 | 7 |
| | Progetto | 17 | 17 | 17 | 17 | 17 |
| | Sviluppo | 64 | 61 | 58 | 55 | 52 |
| | I & T | 19 | 22 | 25 | 28 | 31 |
| Embedded | P & A | 8 | 8 | 8 | 8 | 8 |
| | Progetto | 18 | 18 | 18 | 18 | 18 |
| | Sviluppo | 60 | 57 | 28 | 31 | 34 |
| | I & T | 22 | 25 | 28 | 31 | 34 |

La ripartizione del tempo di sviluppo si ricava dalla tabella seguente:

| Tipo dell'applicazione | Fase | Dimensione (KDSI) 2 | 8 | 32 | 128 | 512 |
|---|---|---|---|---|---|---|
| Organic | P & A | 10 | 11 | 12 | 13 | |
| | Progetto | 19 | 19 | 19 | 19 | |
| | Sviluppo | 63 | 59 | 55 | 51 | |
| | I & T | 18 | 22 | 26 | 30 | |
| Semi-detached | P & A | 16 | 18 | 20 | 22 | 24 |
| | Progetto | 24 | 25 | 26 | 27 | 28 |
| | Sviluppo | 56 | 52 | 48 | 44 | 40 |
| | I & T | 20 | 23 | 26 | 29 | 32 |
| Embedded | P & A | 24 | 28 | 32 | 36 | 40 |
| | Progetto | 30 | 32 | 34 | 36 | 38 |
| | Sviluppo | 48 | 44 | 40 | 36 | 32 |
| | I & T | 22 | 24 | 26 | 28 | 30 |

### Il modello intermedio

Nel modello intermedio si calcola dapprima la quantità di lavoro, detta *nominale*, che viene poi corretta secondo una formula che tiene conto di quindici coefficienti relativi ad altrettanti attributi dell'applicazione, del sistema di calcolo, del personale e dell'organizzazione del progetto.

La formula della quantità nominale di lavoro è la seguente:

$$M_n = a_i S^{b_i}$$

dove $a_i$ e $b_i$ si ottengono da questa tabella:

| Tipo dell'applicazione | $a_i$ | $b_i$ |
|---|---|---|
| Organic | 3.2 | 1.05 |
| Semi-detached | 3.0 | 1.12 |
| Embedded | 2.8 | 1.20 |

I coefficienti di correzione $c_j$ vengono ottenuti dalla tabella seguente:

| j | Attributo | Molto basso | Basso | Normale | Alto | Molto alto | Extra alto |
|---|---|---|---|---|---|---|---|
| 1 | RELY affidabilità | 0.75 | 0.88 | 1.0 | 1.15 | 1.40 | |
| 2 | DATA dimens. database | | 0.94 | 1.0 | 1.08 | 1.16 | |
| 3 | CPLX complessità | 0.70 | 0.85 | 1.0 | 1.15 | 1.30 | 1.65 |
| 4 | TIME req. di efficienza | | 1.0 | 1.11 | 1.30 | 1.66 | |
| 5 | STOR req. di memoria | | 1.0 | 1.06 | 1.21 | 1.56 | |
| 6 | VIRT variabil. ambiente | | 0.87 | 1.0 | 1.15 | 1.30 | |
| 7 | TURN tempo di risposta | 0.87 | 1.0 | 1.07 | 1.15 | | |
| 8 | ACAP esper. analisti | 1.40 | 1.19 | 1.0 | 0.86 | 0.71 | |
| 9 | AEXP esper. applicazione | 1.29 | 1.13 | 1.0 | 0.91 | 0.82 | |
| 10 | PCAP esper. programmatori | 1.42 | 1.17 | 1.0 | 0.86 | 0.70 | |
| 11 | VEXP conosc. ambiente | 1.21 | 1.10 | 1.0 | 0.90 | | |
| 12 | LEXP conosc. linguaggio | 1.21 | 1.10 | 1.0 | 0.90 | | |
| 13 | MODP tecniche moderne | 1.24 | 1.10 | 1.0 | 0.91 | 0.82 | |
| 14 | TOOL strumenti software | 1.24 | 1.10 | 1.0 | 0.91 | 0.83 | |
| 15 | SCED scadenze per lo svil. | 1.24 | 1.10 | 1.0 | 0.91 | 0.83 | |

La quantità di lavoro corretta $M$ viene quindi calcolata dalla formula:

$$M = M_n \prod_{j=1}^{15} c_j$$

Il tempo di sviluppo si ottiene come nel modello base.

Il modello intermedio comprende una procedura per ripartire la quantità di lavoro fra i componenti dell'applicazione, cosí riassumibile:

1. Si stimano le dimensioni $S^{(k)}$ di ciascun componente (che identifichiamo con un indice $k$), da cui si ottiene la dimensione complessiva $S = \sum_k S^{(k)}$.
2. Si calcola la quantità nominale di lavoro complessiva $M_n$ secondo la procedura già vista.
3. Si calcola la produttività nominale $p = S/M_n$, cioè il numero di linee di codice che dovrà essere prodotto per persona e per unità di tempo.
4. Si calcola la quantità nominale di lavoro $M_n^{(k)}$ di ciascun componente: $M_n^{(k)} = S^{(k)}/p$.
5. Si calcola la quantità corretta di lavoro $M^{(k)}$ di ciascun componente usando i rispettivi coefficienti correttivi: $M^{(k)} = M_n^{(k)} \prod_j c_j^{(k)}$.

### Il modello avanzato

Il modello COCOMO avanzato introduce dei coefficienti correttivi anche nel calcolo della ripartizione dei costi fra le fasi di sviluppo. Inoltre, le stime si basano su un modello dell'applicazione piú articolato, che prevede un'architettura strutturata gerarchicamente in sottosistemi e moduli. Per ulteriori informazioni sul modello avanzato si faccia riferimento alla letteratura relativa.

### Valutazione del metodo

Da alcune esperienze di applicazione del metodo COCOMO, risulta che è in grado di stimare i costi con un errore inferiore al 20% nel 25% dei casi per il modello base, nel 68% dei casi per il modello intermedio, e nel 70% dei casi per il modello avanzato.

Nel valutare l'applicabilità del metodo, bisogna tener presente che le formule e le tabelle sono state ottenute per mezzo di analisi statistiche su progetti con queste caratteristiche:

1. applicazioni generalmente complesse e critiche;
2. ciclo di vita a cascata;
3. requisiti stabili;
4. attività di sviluppo svolte in parallelo da piú gruppi di programmatori;
5. elevata maturità dell'organizzazione produttrice e competenza dei progettisti.

Un'organizzazione che voglia applicare questo metodo di stima deve quindi interpretare i risultati tenendo conto delle differenze fra la propria situazione effettiva (dal punto di vista dell'applicazione specifica e dell'ambiente di sviluppo) e quella presupposta dal modello. In particolare, bisogna tener conto degli aggiornamenti delle tabelle che vengono eseguiti e pubblicati periodicamente in base alle esperienze piú recenti. Inoltre è opportuno che le organizzazioni produttrici raccolgano ed elaborino le informazioni relative ai propri progetti, in modo da poter adattare il metodo alla propria situazione.

### A.5.2 Il modello di Putnam

Nel modello di Putnam la dimensione $L$ del programma (in LOC), la quantità di lavoro $K$ (in anni–uomo), ed il tempo di sviluppo $T_d$ (in anni) sono legate dalla formula

$$K = \frac{L^3}{C_k^3 t_d^4}$$

dove la *costante tecnologica* $C_k$ è una caratteristica dell'organizzazione produttrice, valutata in base all'analisi dei suoi dati storici, che tiene conto della qualità delle metodologie applicate.

# Appendice B

# Gestione del processo di sviluppo

La gestione del processo di sviluppo è un argomento ampio che richiede di essere trattato da vari punti di vista, quali l'organizzazione aziendale, l'economia e la sociologia. Rinunciando ad affrontare tutte queste problematiche, alcune delle quali sono oggetto di altri corsi, nel seguito ci limiteremo ad accennare ad alcuni strumenti di uso comune, ed infine introdurremo il problema della gestione delle configurazioni.

## B.1 Diagrammi WBS

I diagrammi di Work Breakdown Structure (struttura della suddivisione del lavoro) descrivono la scomposizione del progetto in base agli elementi funzionali del prodotto (WBS funzionale, Fig. B.1) oppure in base alle attività costituenti lo sviluppo (WBS operativa, Fig. B.2). Le unità di lavoro definite dalla WBS, dette task, sono raggruppate in un insieme di *work packages*. Ciascun work package definisce un insieme di task in modo abbastanza dettagliato da permettere ad un gruppo di persone di lavorarci indipendentemente dal resto del progetto. Per ogni work package bisogna stabilire le date di inizio e fine ed una stima del costo (*effort*) in termini di lavoro (mesi-persona) ed eventualmente di denaro (*budget*).

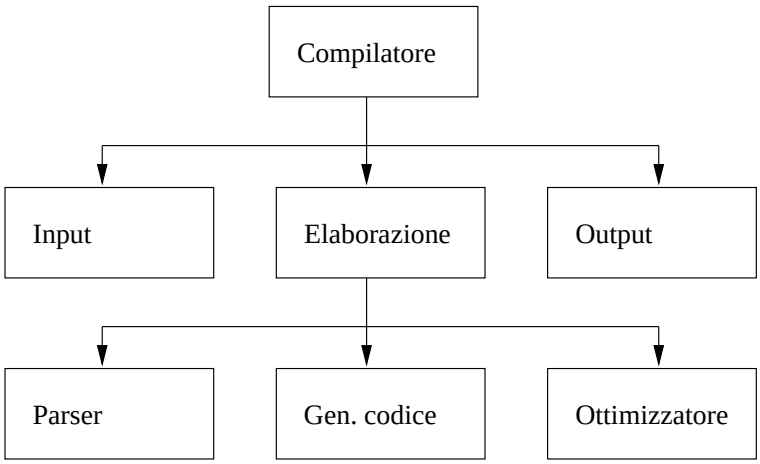


Figura B.1: WBS funzionale

## B.2 Diagrammi PERT

I diagrammi PERT (*Program Evaluation and Review Technique*) rappresentano i legami di precedenza fra le diverse attività del progetto. La dipendenza piú comune è quella di tipo "finish-to-start", in cui una attività può terminare solo dopo che un'attività precedente è terminata. Altre relazioni sono "start-to-start", in cui un'attività può iniziare insieme ad un'altra o dopo, e "finish-to-finish", in cui un'attività può terminare insieme ad un'altra o dopo. La figura B.3 mostra un ipotetico diagramma PERT con relazioni "finish-to-start", dove si suppone che il work package relativo al generatore di codice comprenda fra i propri task la definizione dei formati intermedi fra il parser ed il generatore e fra il generatore e l'ottimizzatore, per cui il completamento del generatore di codice deve precedere le attività di sviluppo del parser e dell'ottimizzatore.

A seconda dei legami di precedenza, può accadere che alcune attività possano essere ritardate entro un certo limite senza che il loro ritardo si ripercuota sul tempo complessivo di sviluppo. Questo margine di tempo si chiama *slack time*. Se nel grafo esiste una sequenza di attività senza slack time, cioè tali che un ritardo di ciascuna di esse porti ad un ritardo di tutto il progetto, la sequenza è un *cammino critico*.

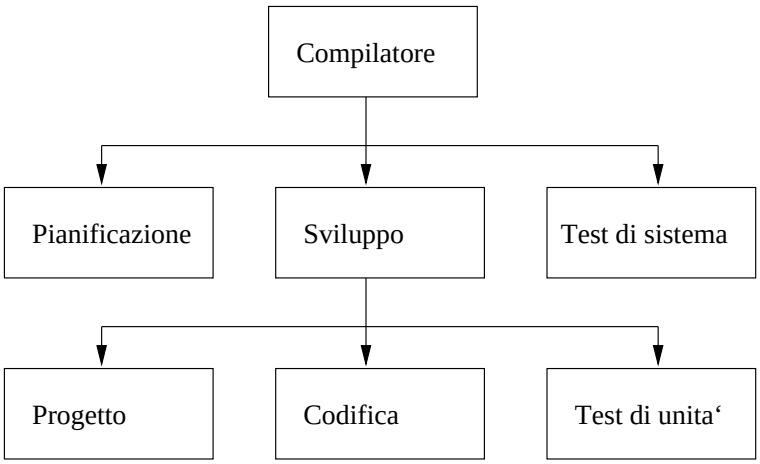


Figura B.2: WBS operativa

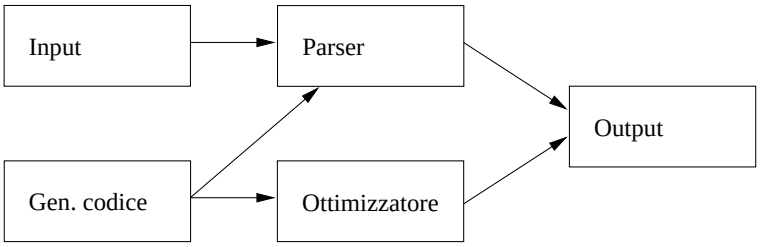


Figura B.3: Diagramma PERT

## B.3 Diagrammi di Gantt

I diagrammi di Gantt rappresentano ciascuna attività con una linea o una barra orizzontale parallela ad un asse dei tempi, e permette quindi di osservare immediatamente le date di inizio e fine, le durate, ed il parallelismo fra le varie attività (Fig. B.4). Possono essere integrati dalle relazioni di precedenza viste nei diagrammi PERT. Nei diagrammi vengono messi in evidenza i *milestone*, cioè alcuni eventi particolarmente importanti, quali revisioni del lavoro e produzione di deliverable.

## B.4 Esempio di documento di pianificazione

Riportiamo un breve estratto da un documento reale di pianificazione. Si tratta di un sistema embedded, e in particolare di un simulatore di veicoli capace di riprodurre gli stimoli meccanici, visivi ed acustici provati dal

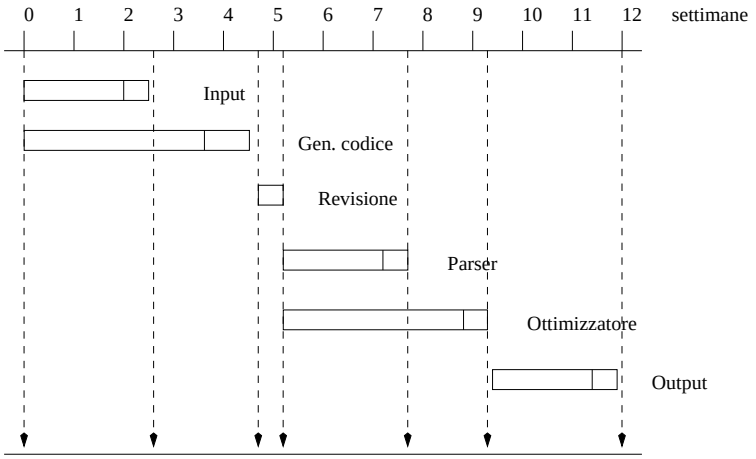


Figura B.4: Diagramma di Gantt

guidatore; il documento di pianificazione si riferisce al sistema completo, e non solo alla componente software. Il progetto viene svolto da un consorzio di quattro organizzazioni, che chiameremo AAA, BBB, etc., diretto da un comitato di gestione (*Steering Committee*). Il progetto è finanziato in parte dalla Comunità Europea (EC).

Il documento di pianificazione comprende un capitolo di introduzione ed analisi, che descrive la struttura del progetto in termini di workpackages e di fasi di sviluppo, e quindi descrive analiticamente i singoli workpackages ed i rispettivi task. L'estratto che riproduciamo viene da questo capitolo, e comprende l'introduzione, la descrizione del primo workpackage, e la descrizione del primo dei suoi task.

Il documento originale comprende anche tabelle che riassumono la ripartizione del lavoro fra i partner, diagrammi PERT e di Gantt, e la definizione delle politiche relative alla conformità (*compliance*) agli standard internazionali.

---

## Workplan, deliverables

### Introduction and analysis

The duration of the project is 36 months.

The project is organized in 10 workpackages and it is divided in 3 sequential phases:

1. user requirements and functional specification of the simulator;
2. design, implementation and testing of the main subsystems;

3. integration, final testing and user evaluation.

The first phase is composed by three work packages: WP1, WP2 and WP3, and its duration is 9 months. During this phase, all the high level requirements for the simulator and the virtual scenario are defined. The mathematical models of the vehicle and of the acoustic rendering are chosen. Finally, the global architecture of the simulator is designed and the functional requirements of the main components are specified.

The second phase is composed by five workpackages: WP4, WP5, WP6, WP7 and WP8, and its duration is 19 months. Each work package of this phase is devoted to one of the main subsystems of the simulator: the mechanical frame, the computer system for the mathematical model, the real-time system for the simulator control, the graphical and the acoustic subsystems. During this phase, each subsystem is designed, implemented and functionally tested.

The final phase is composed by the work package WP9 and its duration is 8 months. During this phase, all the subsystems are integrated in order to build the simulator prototype. The integration and the field tests are executed and the prototype is evaluated.

During the progress of the project it might be necessary to modify the workplan, both for technical reasons and for taking into account the feedback coming from market analysis. Any change in the workplan will be first agreed within the Steering Committee, and afterwards presented to the EC for approval. Each partner should update its activities in agreement with the consortium.

The activity of project management has to be performed during all phases.

The information dissemination will be started after the first phase.

These activities together are included in work package WP10.

There are 4 planned milestones during the project:

**M0** effective commencement date T0.
**M1** after 9 months. End of the specification phase and start of the implementation. All the modules are defined and their requirements specified.
**M2** after 19 months. End of detailed design and beginning of the implementation phase. Checkpoint for the task of design and implementation of the five main subsystems.
**M3** after 28 months. End of the implementation phase and start of the system integration. All the main subsystems are implemented and tested.
**M4** end of the project. The simulator prototype is operational.

One critical path has raised up during the planning in correspondence with the last project activity. This is due to the low flexibility of the activity of field testing .... The easiest alternative action consists in increasing the project duration of at least one month .... Another alternative action is the overlapping of the “Integration test” task with the integration activity itself in such a way to provide an easy management of possible feedback from testing activities.

| WP ID | WORKPACKAGE | PARTNER |
|---|---|---|
| WP1 | Simulator User Requirements Definition | AAA |
| WP2 | Mathematical Model Definition | AAA |
| WP3 | Global architecture and Subsystem Specification | BBB |
| WP4 | Mechanical Subsystem: Design, Implem.n and Test | BBB |
| WP5 | Computational Subsystem: Design, Implem.n and Test | CCC |
| WP6 | Real-time Subsystem: Design, Implem.n and Test | BBB |
| WP7 | Graphical Subsystem: Design, Implem.n and Test | DDD |
| WP8 | Acoustic Subsystem: Design, Implem.n and Test | DDD |
| WP9 | Simulator Integration and Test | BBB |
| WP10 | Project Management | AAA |

**WP1 Simulator User Requirements Definition [AAA]**

Objectives: To define the specifications of the simulator from the user's point of view. This will be done first by a state-of-the art-analysis of existing simulators ... (task T1.1). A simultaneous market analysis will provide the feedback necessary to define a product suitable for a commercial exploitation. After that, the simulator functionality will be defined (task T1.2), and the virtual scenario will be outlined (task T1.3). Finally, specifications will be given (task T1.4), and the field test to be carried out at the end ... will be planned (Task T1.5).

**Task T1.1 State-of-the-art of simulator systems [AAA]**

Baselines: RP 1.1 State-of-the-art of simulator systems report
Duration: 1 month
Effort: AAA: 2.0, BBB: 1.0
Links: output to T1.2, T1.3, T1.6
Objectives: To gain a deeper knowledge about the state-of-the-art in simulator systems, both from the technical and the commercial point of view.
Approach: The various types of simulators ... will be analysed. Existing patents will be researched, as well as any other documentation available. If necessary, the manufacturers will be interviewed. The potential market will be estimated, and the potential customers will be identified.
Activities: Market analysis [AAA].
Overview on the existing simulators, with special attention on the control strategies related to inertial feedback [AAA].
Analysis of mechanical actuation solutions and computer architectures of existing simulator systems [BBB].
Risks: The available literature may not be sufficient for the proposed objective, and the interviewed manufacturers may not be well disposed to give the requested information, since this might be regarded as industrial know-how. The partners could propose to sign an agreement whereby they

> commit themselves to use the information only for the project, and not to transfer them to any third party. ...

---

È utile osservare che nell'introduzione c'è un paragrafo riguardante la possibilità di modificare il piano di lavoro, ed accenna implicitamente (senza specificarle) a delle procedure di modifica, indicando le strutture ed orgnanizzazioni competenti (comitato di gestione e Comunità Europea). Viene inoltre messo in evidenza un cammino critico e vengono proposte delle azioni da prendere nel caso che il progetto subisca ritardi.

Ogni workpackage è affidato ad un partner, come risulta dalla tabella, che ha il ruolo di coordinatore per tale workpackage, a cui contribuiscono però anche gli altri partner. Quindi la descrizione di ogni workpackage, di ogni task e di ogni attività entro i task specifica il partner responsabile. In particolare, La descrizione di ciascun task ha una voce (*Effort*) che specifica il contributo in mesi-uomo di ciascuna organizzazione.

Nella descrizione dei singoli task, la prima voce (*Baselines*) definisce il "semilavorato" che deve essere prodotto dal task. Nell'esempio riportato, si tratta di un rapporto, identificato dalla sigla RP 1.1 e dal titolo "State-of-the-art ...". La voce *Links* índica i rapporti di dipendenza fra i vari task, fornendo quindi le informazioni caratteristiche dei diagrammi PERT. Nell'esempio, vediamo che i task T1.1, T1.3 e T1.6 dipendono dal task T1.1. Infine, osserviamo che per ogni task vengono individuati i rischi associati e proposte eventuali soluzioni.

## B.5 Gestione delle configurazioni

La gestione delle configurazioni si riferisce all'identificazione, l'organizzazione ed il controllo delle diverse versioni del prodotto, dei suoi componenti, della documentazione, e di tutte le informazioni relative allo sviluppo (per esempio, opzioni di compilazione, dati e risultati dei test, ... ). Un insieme di elementi considerati come un'unità dal punto di vista della gestione delle configurazioni è un *configuration item*.

Piú precisamente, ogni versione del prodotto è individuata da un insieme di unità di configurazione (o moduli), cioè da una *configurazione*. Organizzare e controllare le versioni del prodotto significa controllare la distribuzione e la modifica degli elementi di configurazione, registrare e riferire lo stato

degli elementi di configurazione e le relative richieste di modifica, e verificare la correttezza e completezza degli elementi di configurazione (ANSI/IEEE Std. 729-1983). In pratica, si tratta di far sí che ogni persona che debba lavorare sul prodotto, su un suo componente o su qualunque altro artefatto del processo di sviluppo, in qualsiasi fase del ciclo di vita, abbia a disposizione la versione corrente, e che non si creino inconsistenze dovute all'esistenza contemporanea di versioni diverse.

La pianificazione del progetto richiede anche la produzione di un piano di gestione delle configurazioni.

Una *versione* è uno stadio dell'evoluzione di un elemento di configurazione. Diverse versioni di un elemento di configurazione possono differire perché rispondono a requisiti diversi (offrono diverse funzioni e/o hanno diverse interfacce) o perché hanno diverse implementazioni degli stessi requisiti. Ciascuna versione può appartenere a diverse configurazioni. Ciascun elemento di configurazione e ciascuna delle sue versioni deve avere un identificatore. I numeri di versione sono strutturati in modo da distinguere i cambiamenti piú importanti (versione/edizioni) da quelli di minore portata (revisioni). I cambiamenti piú importanti naturalmente richiedono procedure di approvazione.

Una *baseline* è un documento o prodotto, costituito da un insieme di elementi di configurazione, che è stato revisionato ed accettato formalmente. Una baseline è una base per gli sviluppi ulteriori.

Per permettere la gestione delle configurazioni, ogni modulo o documento deve contenere le informazioni necessarie, quali identificatore, autore, data, e storia dei cambiamenti. È fondamentale l'uso di strumenti software, come SCCS o RCS.

# Appendice C

# Qualità

La qualità è "*l'insieme delle caratteristiche di un'entità che conferiscono ad essa la capacità di soddisfare esigenze espresse ed implicite*", dove un'entità può essere, fra l'altro, un prodotto, un processo, un servizio, o un'organizzazione (ISO 8402). Alcuni aspetti (parziali) della qualità sono *idoneità all'uso*, *soddisfazione del cliente*, o *conformità ai requisiti*. Osserviamo che la conformità ai requisiti è solo un aspetto necessario, ma non sufficiente, della qualità: in un certo senso, la qualità è ciò che fa preferire un prodotto ad un altro che pure soddisfa gli stessi requisiti.

La *gestione per la qualità* è l'insieme delle attività di gestione che traducono in pratica la *politica per la qualità* di un'organizzazione, cioè gli obiettivi e gli indirizzi generali riguardanti la qualità.

Il *sistema qualità* è costituito dalla struttura organizzativa, le procedure, i processi, e le risorse necessarie ad attuare la gestione per la qualità.

## C.1 Le norme ISO 9000

Le norme ISO 9000 definiscono un insieme di requisiti necessari ad assicurare che un processo di sviluppo fornisca prodotti della qualità richiesta in modo *consistente*, cioè predicibile e ripetibile.

Gli standard principali dal punto di vista dell'industria del software sono:

**ISO 8402** Gestione per la qualità e di assicurazione della qualità. Terminologia.

**ISO 9000-1** Regole riguardanti la conduzione aziendale per la qualità e l'assicurazione della qualità. Guida per la scelta e l'utilizzazione.

**ISO 9000-3** Regole riguardanti la conduzione aziendale per la qualità e l'assicurazione della qualità. Guida per l'applicazione della ISO 9001 allo sviluppo, alla fornitura e alla manutenzione del software.

**ISO 9001** Sistemi qualità. Modello per l'assicurazione della qualità nella progettazione, sviluppo, fabbricazione, installazione ed assistenza.

**ISO 9004-1** Gestione per la qualità e del sistema qualità. Guida generale.

**ISO 9004-2** Elementi di gestione per la qualità e del sistema qualità. Guida per i servizi.

Il "cuore" di questo sistema di norme è la ISO 9001, che dà le direttive per un sistema di qualità che copre tutto il ciclo di vita di qualsiasi tipo di prodotto industriale. Affiancate alla ISO 9001 si trovano le norme ISO 9002 e ISO 9003 che si riferiscono a sottoinsiemi del ciclo di vita.

La norma ISO 9000 propriamente detta, in tre parti, è una guida che dà indicazioni per la scelta fra le norme (ISO 9001, 9002 o 9003) da applicare, e sulla loro applicazione. In particolare, la 9000-1 illustra i concetti principali e introduce l'insieme delle norme. La ISO 9000-3 è una guida all'applicazione della 9001 nelle aziende informatiche. La necessità di una guida rivolta specificamente all'industria del software discende dal fatto che la 9001, pur essendo proposta come riferimento del tutto generale, è stata concepita inizialmente per l'industria manifatturiera, per cui molte delle sue indicazioni non sono applicabili direttamente all'industria informatica. La ISO 9000-3 fornisce un'interpretazione della 9001 adatta alle condizioni proprie dell'industria informatica.

La ISO 9004 è una specie di "manuale utente" per le 9001, 9002 e 9003, dà una descrizione piú ampia dei rispettivi requisiti e fornisce indicazioni riguardo alle possibili scelte delle aziende per soddisfare tali requisiti. In particolare, la ISO 9004-2 si riferisce alle aziende fornitrici di servizi. Questa norma è rilevante per l'industria del software, poiché le aziende informatiche offrono dei servizi oltre che dei prodotti.

È importante rilevare che le norme contengono soltanto dei *requisiti* per il sistema qualità, e non prescrivono i metodi da applicare per soddisfarli: tocca alle singole organizzazioni fare queste scelte.

### C.1.1 La qualità nelle norme ISO 9000

La norma ISO 9000-3 tratta i seguenti aspetti della gestione per la qualità:

1. Responsabilità della direzione;
2. sistema qualità;
3. verifiche ispettive interne del sistema qualità;
4. azioni correttive;
5. riesame del contratto;
6. specifica dei requisiti del committente;
7. pianificazione dello sviluppo;
8. pianificazione della qualità;
9. progettazione e realizzazione;
10. prova e validazione;
11. accettazione;
12. duplicazione, consegna ed installazione;
13. manutenzione;
14. gestione della configurazione;
15. controllo della documentazione;
16. documenti di registrazione della qualità;
17. misure;
18. regole, pratiche e convenzioni;
19. strumenti e tecniche;
20. approvvigionamenti;
21. prodotti software inclusi;
22. addestramento;

Quasi tutti i requisiti prevedono che le relative attività si svolgano secondo procedure pianificate e documentate, e che i risultati conseguenti vengano registrati, analizzati e conservati. Un testo sulle norme ISO 9000 sintetizza questo approccio con la frase "*dire cosa si fa, fare quel che si dice, e documentare*".

I primi tre requisiti si riferiscono al sistema di qualità in sé, e stabiliscono che la direzione predisponga una struttura ad esso dedicata, fornita delle risorse e dell'autorità necessarie ad applicare la politica della qualità. Il sistema qualità deve essere documentato, e in particolare può essere prodotto e messo a disposizione del personale un *manuale della qualità* contenente le regole e procedure aziendali. Le attività relative all'assicurazione della qualità devono essere pianificate.

Altri requisiti si riferiscono alle varie fasi del ciclo di vita. Come già osservato, le norme non prescrivono determinati modelli di sviluppo o metodologie, ma chiedono che vengano scelti ed applicati quelli che l'organizzazione produttrice ritiene adeguati. L'importante è che tutto sia documentato e verificato.

L'ultimo gruppo di requisiti (da *gestione della configurazione* in poi) si riferisce alle attività di supporto.

## C.2 Certificazione ISO 9000

Un'organizzazione che ha adeguato il proprio sistema qualità allo standard ISO 9000 può ottenere una *certificazione* in merito. In alcuni casi la certificazione ISO 9000 è indispensabile per ottenere certi contratti, particolarmente con enti pubblici. La certificazione è importante anche nei rapporti fra aziende private, anche perché le norme stesse richiedono che il fornitore si accerti della qualità dei prodotti ottenuti dai subfornitori. Quindi ogni azienda che si impegna ad adottare lo standard ISO 9000 è motivata a rivolgersi a subfornitori che abbiano anch'essi adottato lo standard.

La certificazione viene rilasciata da organismi di certificazione, pubblici o privati, che a loro volta devono essere *accreditati*, cioè autorizzati a certificare, da enti di accreditamento. L'organismo di accreditamento italiano è il SINCERT (SIstema Nazionale di accreditamento per gli organismi di CERTificazione). È possibile ottenere la certificazione da piú di un organismo di certificazione, ed anche da organismi stranieri.

Il certificatore prima di tutto analizza la documentazione fornita dall'organizzazione richiedente relativa al sistema qualità, e in seguito compie una *verifica ispettiva* (*audit*) inviando sul posto degli ispettori, che visitano i vari reparti dell'organizzazione ed intervistano il personale. Se il risultato dell'ispezione è positivo, viene emesso il certificato, che è valido per tre anni. In questi tre anni devono essere svolte altre verifiche ispettive, meno estese della prima, per accertare che la conformità allo standard venga mantenuta e migliorata. Il numero di verifiche svolte dopo la certificazione dipende dal certificatore (può quindi essere un fattore nella scelta del certificatore da parte del richiedente), e va da una a quattro all'anno. Dopo tre anni è necessario un nuovo processo di certificazione.

Nel corso dell'ispezione si possono rilevare delle non conformità. In questo caso vengono stabilite delle azioni correttive che saranno verificate in seguito.

È possibile, su iniziativa del richiedente, svolgere un'ispezione di prova prima della verifica ispettiva per la certificazione.

Le verifiche ispettive seguono la norma ISO 10011.

## C.3 Il Capability Maturity Model

Il *Capability Maturity Model* (*CMM*) è un modello, sviluppato dal Software Engineering Institute della Carnegie Mellon University, per valutare e quindi migliorare la qualità dei processi di sviluppo del software.

Il modello permette di valutare la *maturità* di un'organizzazione produttrice di software, cioè la sua capacità di gestire i processi di sviluppo. Sono previsti cinque livelli di maturità:

1. **Iniziale** (*Initial*) Ogni nuovo progetto viene gestito *ad hoc*, con poca o punta pianificazione[1].
2. **Ripetibile** (*Repeatable*) Esiste una gestione del progetto, capace di applicare le esperienze dei progetti passati a nuovi progetti riguardanti applicazioni simili.
3. **Definito** (*Defined*) Vengono applicate metodologie e tecniche ben definite sia per il processo di sviluppo che per la sua gestione. Tutti i progetti usano versioni particolari di un unico processo standardizzato nell'àmbito dell'organizzazione.
4. **Gestito** (*Managed*) Viene applicato un piano di misura per il controllo della qualità del prodotto e del processo.
5. **Ottimizzante** (*Optimizing*) I risultati delle misure vengono analizzati com metodi quantitativi ed usati per migliorare continuamente il processo di sviluppo. È prevista l'introduzione di nuove metodologie e tecniche.

Il CMM prevede delle strategie di miglioramento per le organizzazioni che vogliono passare ad un livello superiore.

**da 1 a 2** Introduzione di gestione di progetto, pianificazione, gestione della configurazione, assicurazione della qualità.

**da 2 a 3** Capacità di adattamento a nuovi problemi, di cambiare i processi.

[1] "*Few processes are defined, and success depends on individual effort and heroics*". SEI, `http://www.sei.cmu.edu/technology/cmm/cmm.sum.html`

**da 3 a 4** Raccolta di dati quantitativi per misurare le proprietà dei processi e dei prodotti.

**da 4 a 5** Aggiornamento dei processi in base ai risultati misurati.

Non si può stabilire una corrispondenza univoca fra la certificazione ISO 9000 ed il livello di maturità secondo il CMM, però un'organizzazione certificata ISO 9000 tipicamente ha una maturità corrispondente al terzo livello del CMM.

Il numero di organizzazioni che hanno conseguito una maturità corrispondente al massimo livello del CMM è tuttora molto ridotto.

# Appendice D

# Moduli in Ada

In Ada, i moduli sono rappresentati per mezzo dei *package*. Un package ha una specifica (interfaccia) ed un corpo (implementazione), che possono essere definiti in due file distinti e compilabili separatamente, come nel seguente esempio:

```
-- file StackADT_.ada
-- interfaccia

package StackADT is
    type Stack is private;
    procedure Push(S: in out Stack; E : in  Integer);
    procedure Pop (S: in out Stack; E : out Integer);
private
    type Table is array (1 .. 10) of integer;
    type Stack is
        record
            t : Table;
            index : integer := 0;
        end record;
end StackADT;

.............................
-- file StackADT.ada
-- implementazione

package body StackADT is
    procedure Push(S : in out Stack; E : in Integer) is
    begin
        S.index := S.index + 1;
        S.t(S.index) := E;
    end Push;

    procedure Pop(S : in out Stack; E : out Integer) is
```

```
    begin
        -- ...
    end Pop;
end StackADT;
```

In questo esempio si realizza un tipo di dato astratto che presenta la tipica interfaccia degli stack. L'implementazione si basa su due strutture dati: il vettore `Table` e il record (o struttura, nella terminologia del C++) `Stack`. Queste due strutture sono dichiarate nella parte privata della specifica. Il package `StackADT` può essere usato come segue:

```
-- file StackADTTest.ada

with StackADT; use StackADT;
procedure StackADTTest is
    i : INTEGER;
    s : Stack;
begin
    push(s, 1);
    push(s, 2);
    pop(s, i);
    -- ...
end StackADTTest;
```

La clausola `with` esprime la dipendenza di `StackADTTest` da `StackADT`. La clausola `use` rende visibili i nomi dichiarati nel package `StackADT`, permettendo di usarli senza qualificarli esplicitamente col nome del package.

## D.1 Moduli generici

La specifica di un package generico inizia con la parola chiave `generic`, seguita dai nomi e tipi dei parametri:

```
generic
    Size : Positive;
    type Item is private;
package Stack is
    procedure Push(E : in  Item);
    procedure Pop (E : out Item);
    Overflow, Empty : exception;
end Stack;

package body Stack is
```

```
    type Table is array (Positive range <>) of Item;
    Space : Table(1 .. Size);
    Index : Natural := 0;

    procedure Push(E : in Item) is
    begin
        end if;
        Index := Index + 1;
        Space(Index) := E;
    end Push;

    -- altre definizioni
    -- ...
end Stack;
```

Questo package può essere usato nel modo seguente:

```
package Stack_Int is new Stack(Size => 200, Item => Integer);

Stack_Int.Push(7);
```

Amche in Ada un modulo può essere parametrizzato rispetto a delle funzioni. I parametri funzione di un package generico vengono dichiarati con la parola chiave `with`:

```
generic
    type ITEM is private;
    type VECTOR is array(POSITIVE range <>) of ITEM;
    with function SUM(X, Y : ITEM) return ITEM;
        -- questa e' la somma generica fra elem. di VECTOR
package VECT_OP is
    function SUM(A, B : VECTOR) return VECTOR;
        -- questa e' la somma fra oggetti VECTOR
    function SIGMA(A : VECTOR) return ITEM;
        -- questa e' la somma degli elementi di un VECTOR
end;
```

Questo package può essere istanziato in questo modo:

```
package INT_VECT is new VECT_OP(INTEGER, TABLE, "+");
```

## D.2 Eccezioni

In Ada le eccezioni sollevate in un package possono essere dichiarate come nel seguente esempio:

```
package StackADT is
    type Stack is private;
    procedure Push(S: in out Stack; E : in  Integer);
    procedure Pop (S: in out Stack; E : out Integer);
    Full, Empty : exception;
private
    type Table is array (1 .. 10) of integer;
    type Stack is
        record
            t : Table;
            index : integer := 0;
        end record;
end StackADT;
```

L'istruzione `raise` viene usata per sollevare un'eccezione:

```
package body StackADT is
    procedure Push(S : in out Stack; E : in Integer) is
    begin
        if S.index >= 10 then
            raise Full;
        end if;
        -- ...
    end Push;

    procedure Pop(S : in out Stack; E : out Integer) is
    begin
        if S.index = 0 then
            raise Empty;
        end if;
        -- ...
    end Pop;
end StackADT;
```

Uno handler è una sequenza di istruzioni compresa fra la parola chiave `exception` e la fine del blocco cui appartiene lo handler. Generalmente uno handler è costituito da una sequenza di clausole `when`, che specificano le azioni da compiere quando viene sollevata un'eccezione. Nel seguente esempio, si suppone che la procedura `StackADTTest` possa sollevare l'eccezione `Full`:

```
procedure StackADTTest is
    i : INTEGER;
    s : Stack;
begin
    push(s, 1);
    push(s, 2);
    -- ...
exception
    when Full =>
        -- ...
end StackADTTest;
```

Il linguaggio fornisce delle eccezioni predefinite, fra cui, per esempio, quelle relative ad errori nell'uso di vettori, record, puntatori (`constraint_error`), e ad errori aritmetici (`numeric_error`). Le eccezioni predefinite vengono sollevate implicitamente da dei controlli (*check*) eseguiti dal supporto run-time del linguaggio, che possono essere disabilitati. Per esempio, i check associati alle eccezioni `constraint_error` sono `access_check` (uso di puntatori), `index_check` (indici di array), ed altri ancora. Per disabilitare un check, si usa la direttiva `pragma`:

```
pragma SUPPRESS(INDEX_CHECK);
```

## D.3 Eredità in Ada95

Accenniamo brevemente al modo in cui l'eredità è stata introdotta in Ada 95, la versione piú recente del linguaggio Ada.

I tipi che possono servire da base per l'eredità sono chiamati *tagged*. Il seguente esempio mostra la specifica di un package in cui viene definito un tipo base (tagged) `Figure` da cui vengono derivati i tipi `Circle`, `Rectangle` e `Square`. Quest'ultimo deriva da `Rectangle`.

```
package Figures is

    type Point is
        record
            X, Y: Float;
        end record;

    type Figure is tagged
        record
```

```
            Start : Point;
        end record;
    function Area (F: Figure) return Float;
    function Perimeter (F:Figure) return Float;
    procedure Draw (F: Figure);

    type Circle is new Figure with
        record
            Radius: Float;
        end record;
    function Area (C: Circle) return Float;
    function Perimeter (C: Circle) return Float;
    procedure Draw (C: Circle);

    type Rectangle is new Figure with
        record
            Width: Float;
            Height: Float;
        end record;
    function Area (R: Rectangle) return Float;
    function Perimeter (R: Rectangle) return Float;
    procedure Draw (R: Rectangle);

    type Square is new Rectangle with null record;

end Figures;
```

In questo esempio, i tipi `Circle` e `Rectangle` ereditano da `Figure` il campo `Start`, a cui aggiungono i campi propri dei tipi derivati. Le funzioni appartenenti all'interfaccia vengono ridefinite per ciascun tipo; osserviamo che l'oggetto su cui operano tali funzioni viene passato esplicitamente come parametro. Il tipo `Square` eredita da `Rectangle` senza aggiungere altri campi, come índica la clausola `null record`.

Il linguaggio Ada 95 permette anche di definire funzioni astratte, analogamente alle funzioni virtuali pure del C++.

# Bibliografia


[1] J. Arlow and I. Neustad. *UML 2 and the Unified Process, Second Edition*. Addison-Wesley, 2005.

[2] R. Fairley. *Software Engineering Concepts*. McGraw-Hill, 1985.

[3] E. Gamma, R. Helm, R. Johnson, and J. Vlissides. *Design Patterns – Elements of Reusable Object-Oriented Software*. Addison-Wesley, 1995.

[4] C. Ghezzi, A. Fuggetta, S. Morasca, A. Morzenti, and M. Pezzè. *Ingegneria del Software – Progettazione, sviluppo e verifica*. Mondadori, 1991.

[5] I. Jacobson, G. Booch, and J. Rumbaugh. *The Unified Software Development Process*. Addison-Wesley, 1999.

[6] F. Jahanian and A. K. Mok. Safety analysis of timing properties in real-time systems. *IEEE Transactions on Software Engineering*, SE-12(9):890–904, September 1986.

[7] J.R. Levine, T. Mason, and D. Brown. *lex & yacc*. O'Reilly & Associates, 1992.

[8] H. Lichter, M. Schneider-Hufschmidt, and H. Züllighoven. Prototyping in industrial software projects – bridging the gap between theory and practice. *IEEE Trans. on Software Engineering*, 20(11), November 1994.

[9] R.C. Linger. Cleanroom process model. *IEEE Software*, 11(2):50–58, 1994.

[10] E. Mendelson. *Introduction to Mathematical Logic*. 3rd Edition, Van Nostrand, 1987.

[11] P. Naur, B. Randell, and J. Buxton, editors. *Software Engineering: Concepts & Techniques*. Petrocelli/Charter, 1976.

[12] J. Rumbaugh, I. Jacobson, and G. Booch. *The Unified Modeling Language Reference Manual*. Addison-Wesley, 1999.

[13] I. Sommerville. *Software Engineering*. Addison-Wesley, 1995.

[14] J. Woodcock and J. Davies. *Using Z – Specification, Refinement, and Proof*. Prentice Hall, 1996.