Metriche Software

Riferimenti

- I. Sommerville – Ingegneria del Software – 8a edizione – Cap. 27
- R. Pressman – Principi di Ingegneria del Software – 5a edizione italiana - Cap. 17
- Ghezzi, Jazayeri, Mandrioli, Ingegneria del Software, 2a edizione, Capitolo 2

- N. Fenton, Software Metrics - A Rigorous and Practical Approach, 1997

Misurare il software: una premessa

- L'Ingegneria del Software non studia le leggi quantitative della fisica, ma processi e prodotti legati ad attività umane.
- Non è possibile effettuare misure assolute, ma è necessario ricorrere a misure indirette.

  – *Negli ultimi trent'anni molti studiosi hanno tentato di sviluppare una singola metrica che offrisse una misura globale della complessità del software*

    – Fenton paragona questi tentativi alla ricerca del "sacro Graal"

    …

Misurare il software

- La misurazione del software ha lo scopo di assegnare un valore ad un attributo caratterizzante un processo o prodotto software
  - consente una comparazione obiettiva tra prodotti/processi
  - rende misurabili processi, risorse, prodotti rilevanti della Ingegneria del Sw
- Sebbene molte aziende produttrici di software utilizzino procedure di misurazione, l'uso sistematico della misurazione del software non è ancora una pratica comune
  - Ancora pochi standard in questa area

Metrica

Metrica: misura quantitativa del grado di possesso di uno specifico attributo da parte di un sistema, un componente, o un processo [IEEE Std 610.12-1990: IEEE Standard Glossary of Software Engineering Terminology]

- Es: numero di linee di codice di un modulo, numero di casi d'uso di un'applicazione, ...

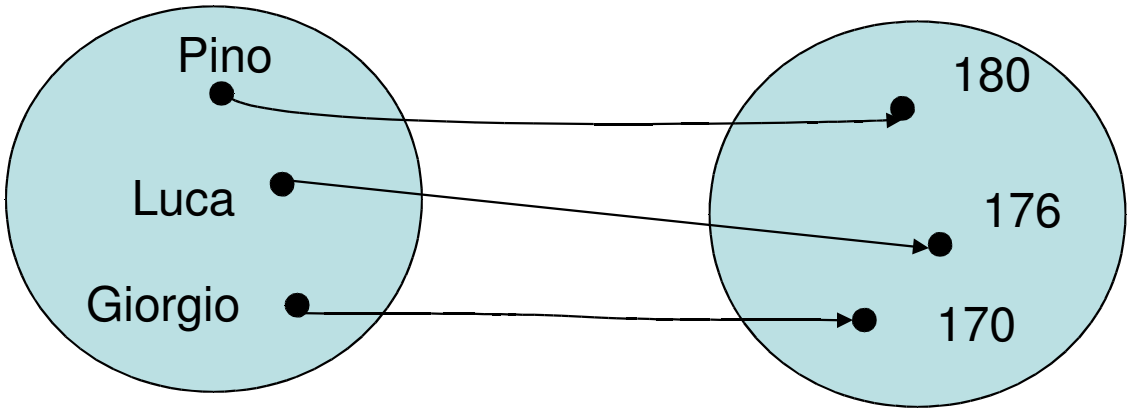
- Molto spesso, metrica è usato come sinonimo di misura o misurazione
- la definizione di metrica include procedure e modalità di misurazione.

Misurazione e Misura

- La **Misurazione** è il processo mediante il quale si assegnano numeri o simboli ad attributi di entità del mondo reale in modo tale da descriverle secondo regole chiaramente definite [Fenton]
  - **L'entità** rappresenta l'oggetto che si vuole sottoporre a misurazione.
  - **L'attributo** dell'entità, invece, è l'aspetto di tale oggetto che interessa descrivere o rappresentare.
  - L'insieme dei simboli o valori che si possono assegnare all'attributo costituisce la "forma" di rappresentazione dell'attributo, o "**scala dei valori**"
- Una **misura** è una funzione che mappa un insieme di oggetti in un altro insieme di oggetti (tipicamente numeri o insiemi di numeri, ma anche simboli)
  - La Misurazione è l'attività generale, una Misura è l'effettiva assegnazione di valori.

Un esempio di Misura di attributi

Entità= persone; Attributo= Altezza;



Misura dell'altezza di persone come funzione di mapping fra due insiemi

Misura e Metrica

- Spesso i termini metrica e misura (e anche misurazione) sono usati come sinonimi
  - Dai testi classici della misura
    - *Una misura è una empirica oggettiva assegnazione di un numero (o simbolo) ad una entità al fine di caratterizzarne uno specifico attributo*
  - Un tentativo di distinzione:
    - *Una metrica caratterizza con valori (numericamente o con simboli) attributi semplici*
    - *Una misura è una funzione di metriche che può essere usata per valutare o predire attributi più complessi*

Attributi

Attributi esterni:

- attributi di un'entità che sono visibili e di interesse per *l'utente* del prodotto software; descrivono l'aspetto esterno di un'entità e il loro rapporto con l'ambiente in cui vengono usate, indipendentemente dalla implementazione;
  - *ad esempio: facilità d'uso; portabilità, efficienza, affidabilità*

Attributi interni:

- attributi di un'entità che sono visibili e di interesse del *produttore*, i cui valori dipendono dalla implementazione;
  - *ad esempio: modularità, strutturazione, tracciabilità, testabilità, dimensione, complessità*

Misure e metriche

- Misura diretta: **la misura di un attributo che non dipende da quella di altri attributi**
- Misura indiretta: **la misura di un attributo che dipende da quella di almeno un altro attributo**

- Metriche: **tutto ciò per cui possiamo fare misure dirette**
  - Misure dirette ≡ Metriche

- Misure: **tutto ciò per cui dobbiamo fare misure indirette**
  - Misure indirette ≡ f(metriche)

- ad attributi semplici sono associate le metriche
- ad attributi complessi sono associate le misure

Entità

Entità di interesse per la misurazione nella IS sono:

- **Prodotti:**
  - tutto ciò che viene prodotto nel CVS (documentazione, software, test data, …);
- **Processi (attività)**
  - produzione specifiche, progettazione, codifica, testing, manutenzione, quality assurance, riuso, reengineering, …;
- **Risorse:**
  - hardware, software, documentazione (della risorsa), risorse umane, banche dati, conoscenza, ….

Esempi di Entità e attributi

	ENTITA	ATTRIBUTI INTERNI	ATTRIBUTI ESTERNI
Prodotti	Specifica	Dimensione, riuso, modularizzazione, funzionalità, correttezza sintattica	Comprensibilità, manutenibilità
	Progetto	Dimensione, riuso, accoppiamento, coesione, modularizzazione, funzionalità	Qualità, comprensibilità, manutenibilità
	Codice	Dimensione, riuso, accoppiamento, modulariz-zazione, funzionalità, complessità algoritmica	Affidabilità, manutenibilità, usabilità
	Dati test	Dimensione, copertura	Qualità, affidabilità
Processi	Specifica	Durata, sforzo, n. cambiamenti nei requisiti	Qualità, costo, stabilità
	Progettazione	Durata, sforzo, numero di errori individuati	Costo, stabilità
	Test	Durata, sforzo, numero di errori trovati	Costo, controllabilità
Risorse	Personale	Età,costo	Produttività, esperienza
	Team	Dimensione, livello di comunicazione, strutturazione	Produttività
	Software	Costo, dimensione	Affidabilità
	Hardware	Costo, prestazioni	Affidabilità

Scale di valori

Una Scala di valori è l'insieme:

- dei numeri / simboli da assegnare ad un attributo di un'entità
- delle relazioni tra tali numeri / simboli
  - Esistono 5 tipologie di scala:

1. Nominale

- La relazione tra i valori consente una semplice *classificazione* degli oggetti della misurazione, ma non ci indica nulla sulla relazione fra di essi
- Esempio:
  - *la classificazione degli errori di programmazione in lessicali, sintattici e semantici ci consente di suddividerli ma non di sapere se un errore sia più o meno grave di un altro;*
- in termini più analitici questo tipo di scala coincide con l'introduzione nell'insieme di partenza di classi di equivalenza

Scale di valori

2. Ordinale

- si introduce una relazione d'ordine fra gli oggetti, per cui si è in grado di determinare le posizioni relative degli oggetti (cioè dire se uno venga prima di un altro);
- non si è però in grado di quantificare la distanza fra gli oggetti
- Esempio:
  - *in una scala di valutazione (sufficiente, buono, ottimo) possiamo dire che buono è peggiore di ottimo ma non di sufficiente*

Scale di valori

3. **Intervalli**:
   - Si è in grado di misurare la distanza fra gli oggetti del dominio, e non solo posizionarli tra di loro come con le scale ordinali
   - Esempio:
     - *la classificazione del rendimento scolastico con "6", "8" e "10" in sostituzione di "sufficiente", "buono" e "ottimo", rispettivamente, consente non solo di dire che "6" precede "8", ma anche di quanto*

4. **Ratio**
   - introduce l'elemento zero che rappresenta l'assenza totale dell'attributo che si sta misurando nell'entità sottoposta a misurazione. L'introduzione dello zero, inoltre, conferisce al valore di un attributo il senso di positività, negatività o nullità.
   - Esempio:
     - *la misura di una temperatura, quantificata con un numero maggiore, minore, o uguale a zero.*

Scale di valori

5. Assoluta

- esiste una strategia di conteggio per la quale possiamo assegnare ad ogni oggetto un numero univocamente determinato.
- La scala assoluta è usata per quegli attributi di un oggetto che richiedono un semplice conteggio di elementi.
- Esempio:
  - *si consideri l'entità "studente" e si voglia quantificare il suo attributo "numero di esami superati". La misura in questo caso consiste nel semplice conteggio degli esami superati da uno studente; il totale conteggiato rappresenta lo studente in esame, visto attraverso l'attributo "numero di esami superati".*

Caratteristiche di metriche software efficaci

- Semplici e Calcolabili
- Convincenti a livello empirico ed intuitivo.
- Coerenti e obiettive.
- Coerenti nell'uso di unità e dimensioni.
- Indipendenti dal linguaggio di programmazione.
- Devono essere un meccanismo efficace per un feedback sulla qualità.
  - "L'esperienza insegna che la metrica di un prodotto viene utilizzata solo se è intuitiva e facile da calcolare. Se occorre svolgere decine di calcoli complessi, è improbabile che tale metrica venga largamente adottata" [Pressman]
  - Non tutte le metriche soddisfano tutte le caratteristiche di qualità indicate.

Metriche di sforzo

- Per sforzo (effort) si intende la misura del quantitativo di lavoro umano necessario per svolgere un lavoro relativo ad un processo di sviluppo software
- L'unità di misura più comunemente usata è il mese/uomo (man-month) e tutti i suoi multipli e sottomultipli
  - Un mese/uomo si definisce, teoricamente, come il quantitativo di lavoro che un singolo dipendente può svolgere in un mese lavorativo

The Mythical Man-Month

- Problemi della definizione del mese-uomo
  - Ogni persona ha una sua diversa produttività, che può variare nel tempo (con l'età e la conoscenza, ad esempio)
  - La produttività di una persona dipende dalla complessità del problema da risolvere
  - Aumentando la dimensione del gruppo di lavoro aumenta proporzionalmente la produttività?
    - *Se una persona può svolgere 1 man-month di lavoro in un mese, quanto lavoro potranno svolgere 10 persone nello stesso mese?*
- Tutti questi problemi furono espressi compiutamente per la prima volta da Brooks nel famoso libro The Mythical Man-Month: Essays on Software Engineering, nel 1975
- In conclusione: si tratta di una metrica di facile misura (ma solo a posteriori) ma di difficilissima stima

Brooks (1975). *The Mythical Man-Month*. Addison-Wesley.

Una Classificazione di Metriche software relative al prodotto

- **Metriche per il Modello di Analisi** :
  - Funzionalità fornita
  - Dimensioni del sistema (in termini di informazioni presenti sul modello di analisi)
  - Qualità delle specifiche
- **Metriche per il Modello di Design**:
  - Metriche dell'architettura
  - Metriche a livello dei componenti
  - Metriche della progettazione dell'interfaccia
  - Metriche specializzate relative al design object-oriented
- **Metriche relative al codice sorgente**
  - Metriche di Halstead
  - Metriche di complessità
  - Metriche di lunghezza
- **Metriche di testing**
  - Metriche sulle istruzioni e ramificazioni
  - Metriche relative ai difetti
  - Test dell'efficacia
  - Metriche sul processo

Metriche per il Modello di Analisi

- Che tipo di metriche si possono usare in fase di analisi?
  - Metriche che consentono al management di monitorare e controllare Costi, Scheduling, Qualità
  - Ad esempio: Metriche aventi lo scopo di poter "prevedere" quale sarà la "taglia" del processo software all'inizio del suo ciclo di vita, in modo da poter dimensionare le risorse umane e temporali da dedicare ad esso.

- Metriche basate sulle funzionalità
  - *Function Points (FP)*

- Metriche per la qualità delle specifiche

Function Point Analysis

- Tecnica proposta da **Albrecht** [*Measuring Application Development Productivity, 1979*]
- Approccio indipendente dal linguaggio di programmazione per misurare le funzionalità del sistema.
- Può essere usata per:
  - Stimare il costo richiesto per programmare, testare il software
  - Prevedere il numero di errori che si rileveranno durante il testing
  - Prevedere il numero di componenti o di LOC del sistema

Cosa sono i Function Point (FP)

- I FP sono una misura di funzionalità basata su entità logico-funzionali che l'utente facilmente comprende (es. Input, Output,etc.)
- I FP sono pertanto indipendenti dal linguaggio di programmazione, quindi la produttività può essere confrontata tra diversi linguaggi
- Un FP non è una singola caratteristica ma una combinazione di caratteristiche del sistema
  - *Destinati a supportare stime, pianificazioni relative a sistemi Sw*
  - *Nati per essere applicati a Sistemi Sw di tipo Business*
  - *Misura che può essere effettuata 'presto' nel CVS*
  - *Misura solo i requisiti funzionali*

Formulazione iniziale del metodo dei Function Point

- I FP vengono derivati usando una relazione empirica che si basa su:
  - Misure dirette (espresse in valori interi) del Dominio delle informazioni del software
  - Valutazione della complessità del software

- Le Misure del Dominio delle Informazioni sono catturate con riferimento al Confine (Boundary) del sistema software.

Boundary

- Secondo il concetto di Boundary, esiste una linea chiusa che definisce il confine del sistema software cui è rivolta la misura.
- Con riferimento al confine, sono misurate o valutate le seguenti caratteristiche del programma:
  - External input (EI)
  - External output (EO)
  - External inquiry (EQ)
  - Internal Logical File (ILF)
  - External Interface File (EIF)

External Input (EI)

- Un EI rappresenta un tipo di input unico (proveniente dall'esterno dell'applicazione) fornito o da un utente o da un altro sistema, che verrà elaborato dall'applicazione.
  - Un input è unico se il formato è unico, o la logica con cui viene elaborato è unica.
- Questi input tipicamente devono aggiungere, cambiare o cancellare dati in un File Logico Interno (ILF).
- I dati possono essere informazioni di controllo o informazioni del dominio applicativo (business information). Se i dati sono informazioni di controllo non devono aggiornare nessun ILF.

External Output (EO)

- Un EO è un tipo di dati unico che viene creato nell'applicazione ed è destinato all'utente, o ad altre applicazioni, attraversando il confine dell'applicazione.
  - Un output è unico se il formato o la logica elaborativa è unica.
- Un EO consiste in report, schermate, file, messaggi d'errore…
- Un EO può addizionalmente aggiornare un ILF.
- Questi reports o files sono creati a partire da uno o più ILF e EIF.

External Inquiry (EQ)

- Una interrogazione esterna EQ è definita come una coppia input/output unica, dove un input online produce la generazione di una risposta immediata del software, sotto forma di output online.
  - Un'interazione è unica se il formato o la logica di elaborazione è unica.
- Una EQ ricerca dati o informazioni di controllo (da uno o più ILF e/o EIF) per una risposta immediata.
- Il processo di input non deve aggiornare nessun ILF, mentre il processo di output non contiene dati "derivati"
- Se c'è un coinvolgimento di updating del file logico interno o una produzione di dati derivati o calcolati, deve essere considerato un input seguito da un output.

Internal Logical File

- Un ILF è un raggruppamento unico di dati logicamente correlati, o di informazioni di controllo, identificabile dall'utente, usato ed aggiornato dall'applicazione.

- I dati relativi non attraversano il confine, ma risiedono internamente al confine dell'applicazione e sono gestiti attraverso gli external input (EI).

External Interface File

- Una EIF è un unico gruppo di dati logicamente correlati, o di informazioni di controllo, identificabile dall'utente che viene usato dall'applicazione.

- I dati relativi attraversano il confine e risiedono interamente all'esterno del confine dell'applicazione e sono mantenuti da un'altra applicazione.

- Un EIF è un ILF di un'altra applicazione.

Calcolo dei Function Point

- Esistono diverse teorie, più o meno complesse per calcolare i FP in funzione del conteggio degli elementi prima descritti
  - Il livello di complessità di ciascun elemento individuato per ognuna delle 5 precedenti categorie viene classificato in:
    - *Basso (semplice)*
    - *Medio*
    - *Alto (complesso)*
  - Ad ognuno dei livelli è associato un peso che varia da 3 (più semplice) a 15 (più complesso).

Matrice dei pesi: un esempio

	Componenti	Livello di complessità		
		Basso	Medio	Alto
Transactional function types	Ext. Input (EI)	3	4	6
	Ext.Output (EO)	4	5	7
	Ext. Inquiry (EQ)	3	4	6
Data function types	Internal Logical files (ILF)	7	10	15
	External Interface files (EIF)	5	7	10

UFC e FP

Una prima stima dei FP è detta **UFC** (Unadjusted Function Point Count):

$$UFC = \sum(\text{number of elements of given type}) \times (\text{weight})$$

Lo UFC è poi corretto da un fattore moltiplicativo TFC compreso tra 0.65 e 1.35

DFP = UFC * TFC

(Delivered Function Point)

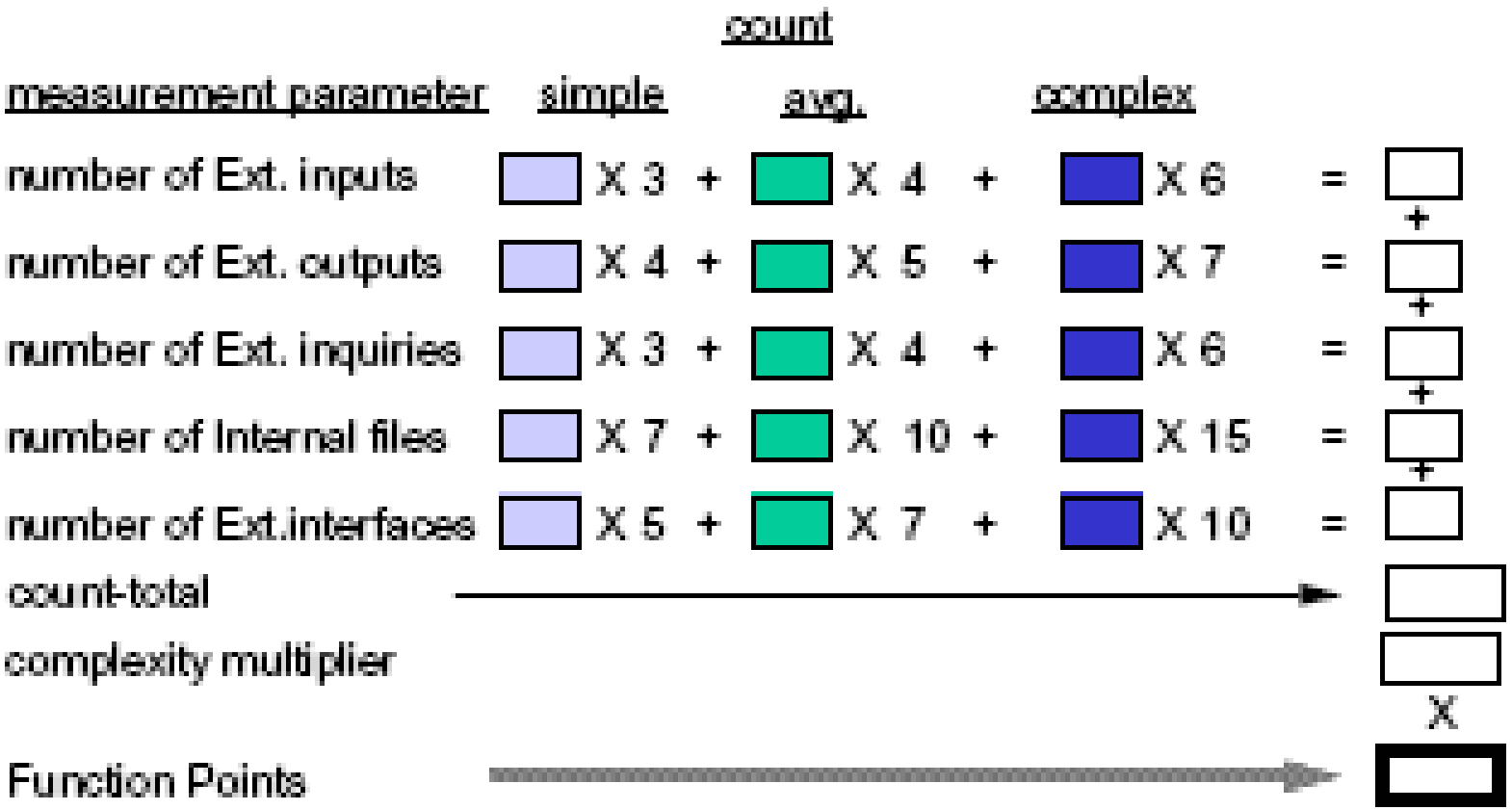
TFC- Fattore di Complessità Totale

- Il TFC è calcolato sulla base di 14 Fattori di complessità (FC), valutati su una scala da 0 (ininfluente) a 5 (essenziale):

1. Il sistema richiede backup e recovery affidabili?
2. Sono richieste comunicazioni specializzate per trasferire dati da/verso l'applicazione?
3. Vi sono funzionalità richiedenti elaborazioni distribuite?
4. Le prestazioni sono un elemento critico?
5. Il software funzionerà in un ambiente operativo esistente già pesantemente utilizzato?
6. Il sistema richiede inserimenti online di dati?
7. L'input online di dati richiede che la transazione relativa sia progettata su più schermate o più operazioni?
8. Il file principali IFL sono aggiornati online?
9. I dati di I/O, i file, le interrogazioni sono complesse?
10. L'elaborazione interna è complessa?
11. Il codice è progettato/scritto per essere riusabile?
12. Il progetto comprende anche l'installazione e la conversione?
13. Il software è stato progettato per essere installato presso diversi utenti?
14. Il software è stato progettato per facilitare le modifiche e per un semplice uso da parte dell'utente finale?

$$TFC = 0.65 + 0.01 * \sum_{i=1..14} FC_i$$

Schema di calcolo FP



Applicazione dei FP

- Dalla loro definizione, molti gruppi si sono internazionalmente occupati di poter fornire definizioni e metodologie di stima dei valori dei FP che fossero il più soddisfacenti possibili. Ad esempio:
  - International Function Point User Group, http://www.ifpug.org/
  - Gruppo Utenti Function Point Italia, http://www.gufpi-isma.org/
- Ciò nonostante, le metodologie che portano alla migliore stima sono quelle che sono state pensate, provate e migliorate con l'utilizzo nell'ambito di un gruppo di lavoro stabile e nell'utilizzo di cicli di vita e tecnologie ben stabilizzati

Utilizzi dei FP

- Per stimare la dimensione finale del codice:
  - Studi hanno cercato di rilevare una correlazione tra FP e numero di LOC, variabile a seconda del linguaggio di programmazione.
  - Es. COBOL: 1 FP -> 100 LOC
  - PL1 : 1FP -> 80 LOC
  - OO lang : 1 FP ->60 LOC
- Per stimare lo sforzo (in ore/uomo) di sviluppo:
  - Es. se 1 mese/uomo ->12 FP , allora …
- Per valutare la completezza del testing
  - studi hanno misurato la correlazione fra FP e numero di difetti scoperti durante il testing

Use Case Point

- Una proposta più recente sposta l'attenzione verso la misura degli use case (Karner, 1993)
- Nella tecnica degli use case point vengono tenuti in conto nella misura:
  - *Casi d'uso*
  - *Attori*
  - *Scenari*
- La tecnica di stima è simile a quella dei FP

Gautam Banerjee,Use case points, http://www.bfpug.com.br/Artigos/UCP/Banerjee-UCP_An_Estimation_Approach.pdf

Use Case Point

- Gli attori sono pesati in base alla loro complessità (altri programmi, utente esperto, utente inesperto)
- I casi d'uso sono pesati in base alla complessità delle interazioni (in funzione del numero di interazioni dei suoi scenari, ad esempio)
  - Nella complessità del caso d'uso si tiene conto anche della possibilità di riutilizzare software esistente
- Si aggiusta il conteggio con un fattore tecnico dipendente dai requisiti non funzionali richiesti e da condizioni ambientali legate al processo e agli strumenti di sviluppo
- Si calcola una stima generale dello sforzo in giorni/uomo

Vantaggi e svantaggi dei FP

Vantaggi:

- Indipendenti dal linguaggio
- Ottime indicazioni per applicazioni di elaborazione dati, che usano linguaggi convenzionali o non procedurali
- Basati su quei dati che hanno la maggior probabilità di essere noti all'inizio di un progetto

Svantaggi

- Soggettività nell'assegnazione dei pesi
- Dati sul dominio delle informazioni possono essere difficili da reperire
- Nessuna valutazione della complessità dell'algoritmo (a parità di pochi input e output potrebbe essere banale ed estremamente complicato)
- I FP non hanno un diretto significato fisico

Feedback e Tuning

- La stima dello sforzo con FP e UCP può diventare più affidabile con l'esperienza che si acquisisce nel loro utilizzo, specie in un ambiente di sviluppo invariato

- Al termine di ogni progetto è possibile cercare di ricavare dei feedback per il tuning dei parametri del metodo di calcolo dei FP

Metriche per la qualità delle specifiche

- Sono stati proposti alcuni **attributi di qualità** dello SRS (non ambiguità, completezza, correttezza, comprensibilità,verificabilità, coerenza interna ed esterna, realizzabilità, concisione, tracciabilità, modificabilità, precisione e riusabilità)
- Un metodo per valutare tali attributi:
  - *n_R requisiti, di cui n_f funzionali e n_{nf} non funzionali : $n_R=n_f+n_{nf}$*
  - *Una metrica proposta per ognuno degli attributi di qualità:*
    - Ad esempio, per l'assenza di ambiguità:

      $Q_1 = n_{ui}/n_R$

      Rapporto tra il numero di requisiti che non risultano ambigui (secondo i revisori dell'SRS) e il numero di requisiti funzionali
- Metriche analoghe sono proposte per tutti gli altri attributi di qualità dell'SRS

Metriche di progetto architetturale

- Si concentrano sulle caratteristiche dell'architettura del software, non tenendo conto della struttura interna dei singoli moduli.
- Si basano sull'analisi di modelli di progetto nei quali si evidenziano i moduli del sistema (qualora esso sia scomponibile) e i dati che vengono scambiati tra essi.
- Molto semplici da misurare, ma poco affidabili in quanto a legame con lo sforzo effettivo legato allo sviluppo del sistema.

Esempi di Metriche di Progetto Architetturale

- Architectural design metrics (*applicabili a moduli di architetture gerarchiche)*
  - **Structural complexity = f(fan-out)**
  - **Data complexity(mi) = f(input & output variables, fan-out)**
  - **System complexity = f(structural & data complexity)**
- Henry e Kafura (HK) metric**:** complessità architetturale vista come funzione del fan-in e fan-out dei moduli
- Morphology metrics (Fenton): una funzione del numero dei moduli e del numero delle interfacce tra i moduli

Metriche per il Design Object Oriented

- Peculiarità Sistemi O-O
  - Localizzazione: il modo con cui le informazioni e loro elaborazioni sono 'confinate' in un sistema
    - Nei sistemi Function oriented le informazioni sono localizzate intorno alle funzioni (moduli procedurali)
    - Nei sistemi Object Oriented sono concentrate incapsulando dati e procedure che li elaborano
  - Incapsulamento:
    - *una delle conseguenze sulle metriche è che l'unità da considerare è l'oggetto piuttosto che il sotto-programma*
  - Information hiding:
    - *influisce sull'accoppiamento fra oggetti*
  - Inheritance: *varie forme di influenza sul progetto*
  - Abstraction:
    - *possibilità di considerare gli oggetti da vari livelli di astrazione*

Metriche orientate alle classi

- Metriche di Chidamber and Kemerer **[A Metrics Suite for Object-Oriented Design, 1994]:**
  - weighted methods per class (WMC)
  - depth of the inheritance tree (DIT)
  - number of children (NOC)
  - coupling between object classes (CBO)
  - response for a class (RFC)
  - lack of cohesion in methods (LCOM)

Weighted Method per Class (WMC)

- **WMC è la somma pesata dei metodi di una classe**
  - Il peso di un metodo è dato da un fattore di complessità a scelta.
    - Il fattore di complessità dei metodi può essere una metrica di complessità proposta per le funzioni (ad esempio la complessità di Mc Cabe)
    - Al crescere di WMC aumenta la complessità della classe e diminuisce la speranza di poterla riusare.
    - Problema: come tener conto della complessità dei metodi ereditati?

Profondità dell'albero di ereditarietà di una classe- DIT

- **Distanza massima di un nodo (una classe) dalla radice dell'albero rappresentante la struttura ereditaria**
  - Maggiore è la profondità della classe nella gerarchia, maggiore è il numero di metodi che essa può ereditare, rendendo più complesso predire il suo comportamento.
  - Alberi di ereditarietà con elevata profondità possono aumentare la complessità del progetto (più classi e metodi sono coinvolti)
  - Maggiore è la profondità di una classe in una gerarchia, maggiore è il riuso potenziale dei metodi ereditati.
  - Ma i metodi della superclasse devono essere nuovamente testati per ciascuna sottoclasse.

Numero di figli - NOC

- **Numero di sottoclassi, figlie di una super-classe**
  - Al crescere del NOC, cresce il livello di riuso (NOC è dunque un indicatore di Riuso) …
  - ma tende ad 'evaporare' l'astrazione della Classe madre.
  - aumenta la possibilità che una Classe figlia non sia membro appropriato della madre
  - Al crescere del NOC, cresce la quantità di test case necessari a testare ogni figlia.

Risposte per classe RFC

- **Insieme dei metodi che possono essere potenzialmente eseguiti in risposta ad un messaggio ricevuto da un oggetto della Classe**
  - Include l'insieme dei metodi della classe e di tutte le classi a cui essa invia messaggi.
  - RFC è un indicatore del 'volume' di interazione fra classi
  - A valori elevati di RFC:
    - *aumenta la complessità progettuale della classe*
    - *cresce lo sforzo per il testing*

Accoppiamento tra le classi - CBO

- **Numero di collaborazioni di una classe, ovvero numero di altre classi cui essa è accoppiata**
  - L'accoppiamento può avvenire a seguito di lettura/modifica attributi, chiamata metodi, instanziazione oggetti
  - Eccessivo accoppiamento è negativo per la modularità ed il riuso;
  - Maggiormente indipendente è una classe più facilmente è riusabile;
  - Per migliorare la modularità, l'accoppiamento va tenuto al minimo; esso influisce sull'impatto di modifiche in altri moduli
  - Valori elevati del CBO complicano testing e modifiche.

Mancanza di coesione nei metodi - LCOM

- La coesione di un metodo esprime la proprietà di un metodo di accedere in maniera esclusiva ad attributi della classe.
- La Mancanza di Coesione deriva dalla presenza di più metodi che accedono ad attributi comuni.

- Se ogni attributo è acceduto da un solo metodo, allora LCOM=0 (massima coesione dei metodi).
- Altrimenti, LCOM >0 ed aumenta quanto più I metodi vanno ad operare sugli stessi attributi, rendendo quindi difficile il controllo di tutte le interazioni tra i metodi dovuti all'uso degli attributi comuni.

- Quindi se LCOM è elevato, I metodi potrebbero essere molto accoppiati tra loro attraverso gli attributi, ed aumenta la complessità del design della classe.

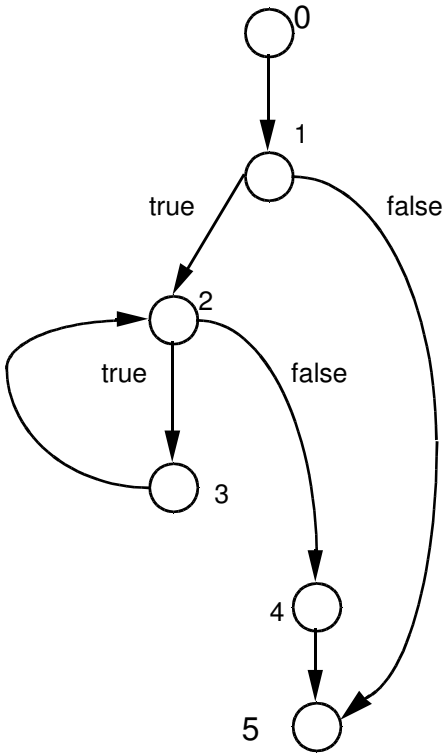
Metriche di design a livello dei componenti

- Metriche che valutano caratteristiche interne dei moduli, quali **coesione**, **coupling** e **complessità**. Ad esempio:
  - Le metriche di **coesione** di Bieman [Measuring Functional Cohesion, 1994] si basano sull'analisi delle funzioni e dei dati elaborati da un modulo, cercando di stabilire se il modulo svolge funzioni che operano sugli stessi dati (buona coesione), oppure su dati disgiunti (scarsa coesione)
  - La Metrica di **accoppiamento** proposta da Dhama [Quantitative Models of Cohesion and Coupling in Software, 1995] tiene conto del flusso di dati e di controllo fra i moduli, dell'accoppiamento a variabili globali, e dell'accoppiamento ambientale (mediante fan-in e fan-out) fra moduli.
  - Le metriche di **complessità** misurano in vario modo la complessità del CFG.

Metrica di McCabe

- Metrica strutturale relativa al Control Flow di un programma G
- E' detta anche numero ciclomatico V(G)
- Rappresenta la complessità logica del programma e quindi lo sforzo per realizzarlo (o per comprenderlo)
- per programmi strutturati V(G) è uguale al n.ro di predicati aumentato di uno:
  - V(G) = P + 1
    - *P: n.ro di predicati (semplici ed a due vie) in G*
    - *Predicati: decisioni - semplici, ovvero non composte con connettivi logici - in strutture di controllo (if-then-else, repeat-until, whiledo,…);*

Complessità ciclomatica di un programma



Complessità ciclomatica del programma è 3

Dato un grafo G di *n nodi* ed *e archi* il numero **ciclomatico** è dato da:

V(G) = *e- n+ 2*
oppure:
V(G)= *d + 1*
d= # nodi branch

V(G)= 3 =>3 cammini indipendenti

c1= 0-1-2-4-5
c2= 0-1-**2-3-2**-4-5
c3= 0-**1-5**

Cammini linearmente indipendenti

- V(G) rappresenta il n° di cammini linearmente indipendenti del Control Flow Graph nella teoria dei grafi V(G) rappresenta il numero di circuiti linearmente indipendenti che congiungono il nodo iniziale con quello finale
- Un cammino si dice indipendente se introduce almeno un nuovo insieme di istruzioni o una nuova decisione, rispetto ad un altro;
- In un CfG un cammino è indipendente se attraversa almeno un arco non ancora percorso dagli altri già considerati.
- L'insieme di tutti i cammini linearmente indipendenti di un programma formano i cammini di base;
- Tutti gli altri possibili cammini nel CfG sono generati da una combinazione lineare di quelli di base.
- Dato un programma, l'insieme dei cammini di base non è unico.

Metriche per il Codice Sorgente

- Linee di Codice
- Metriche di Halstead

Lines of Code (LOC)

- E' una metrica dimensionale
- Misura la lunghezza di un programma con il numero di linee di codice
- Esistono diverse possibili definizioni
  - qualsiasi linea di codice, inclusi i commenti;
  - tutte le linee di codice, esclusi i commenti (in questo caso si distinguono in Non Comment LOC (NCLOC) e Effective LOC (ELOC)
  - solo le istruzioni eseguibili (EXLOC), con anche le Data Declaration LOC (DDLOC), cioè le istruzioni dichiarative dei dati.

- La definizione più accettata è la seguente:
  - Una linea di codice è ogni linea di testo di un programma che non sia bianca o un commento, indipendentemente dal numero di istruzioni in essa inclusa. In particolare, tali linee di codice possono contenere intestazioni di unità di programma, dichiarazioni ed istruzioni esecutive

Osservazioni sulle LOC

- I commenti sono una parte importante del processo di sviluppo di un programma, per cui ignorarle, e quindi non invogliare il programmatore ad inserirle, può essere deleterio; ovviamente vanno prodotti in modo consistente (e non inseriti all'ultimo solo per motivi di tariffazione).
  - Le Comment LOC (CLOC) possono essere valutate in base alla loro importanza
    - *differenza fra commenti inseriti utilizzando metodi formali (per cui, utilizzando appositi tool, si possa effettuare un testing automatico del programma) e commenti inseriti all'ultimo momento prima della consegna.*
- Alcune misure derivate dalle LOC sono:
  - Densità di commento = CLOC/LOC
  - Effort = $5.2*KLOC^{0.91}$
    - *con Effort misurato in mesi/uomo*

Vantaggi e svantaggi

Vantaggi:

- Molto semplici da calcolare automaticamente
- Molto intuitive

Svantaggi

- Dipendenti dal linguaggio di programmazione
  - *Studi empirici sono stati svolti per valutare la dipendenza del numero di LOC dal linguaggio, a parità di algoritmo*
- Dipendenti dallo stile di programmazione
  - *Bisogna definire uno standard di formattazione del testo sorgente cui conformarsi, oppure uno strumento di formattazione automatica*
  - *Facilmente "falsificabili"!*
- Diventano una misura inconsistente in presenza di generatori automatici di codice

Altri usi delle LOC

- Le LOC sono utilizzate per misurare indirettamente altri attributi
  - Esempi:
    - *la probabilità di presenza di errori nel programma (cresce con la lunghezza)*
    - *il tempo per produrlo*
    - *la produttività (o il compenso) di un programmatore.*
    - *stima dei costi di un progetto SW*
- Nelle LOC bisognerebbe contare anche tutte le linee di codice non direttamente legate ai prodotti consegnati, ad esempio:
  - Codice per il testing
  - Codice per I prototipi

Metriche di Halstead (1977)

Halstead [Elements of Software Science, 1977] ha proposto un insieme di metriche per la lunghezza ed il volume di un programma. Basate su:

- n1 = numero di operatori distinti che compaiono in un programma
  - *Ad esempio costrutti di controllo, condizionali, assegnazioni, etc.*
- n2 = numero di operandi distinti che compaiono in un programma
  - *Ad esempio variabili e costanti di un programma*
- N1 = numero totale delle occorrenze di operatori
- N2 = numero totale delle occorrenze di operandi

- Si definiscono:
  - Vocabolario: $n=n1+n2$
  - Lunghezza: $N=N1+N2$ (è una lunghezza concettuale)
  - Volume: $V=N*\log_2 n$

Metriche di Halstead

Volume $V=N*\log_2 n$

- $\log_2 n$ è il numero di bit necessari per rappresentare il vocabolario
- V è il numero di bit necessari per rappresentare il programma nella sua forma minima, cioè indipendentemente dalla lunghezza dei nomi dei token;

- il concetto di Volume è legato quindi al contenuto di informazione del programma e dovrebbe dipendere unicamente dall'algoritmo scelto, non dall'espressività del linguaggio di programmazione.
- Il Volume è usato come una **metrica dimensionale** del software.
- Alcuni esperimenti hanno mostrato la correlazione fra il volume e le LOC, così come tra V e la difettosità dei moduli…

Volume Potenziale

Molte altre metriche sono state derivate da Halstead:

- **Volume potenziale**:
  in un caso ideale
  - *N1= n1 = 2 (solo due operatori: nome funzione e parentesi)*
  - *N2 = n2 (tutti operandi distinti)*

$$V^* = (2+n_2)\ log_2(2+n_2)$$

- É il volume del programma più sintetico nel quale può essere codificato un algoritmo.
- Viene usato per introdurre il concetto di **Livello di Implementazione.**

Livello di Implementazione, Difficoltà e Sforzo

Livello di implementazione

L=V* / V

- maggiore è il volume V minore è L; un algoritmo implementato con un linguaggio di programmazione con basso livello espressivo richiede un maggiore volume.
  - *È possibile approssimare L con L^ = (2 * n_2)/(n_1*N_2)*

Difficoltà

- D=1/L
  - *difficoltà di implementare un algoritmo in un certo linguaggio di programmazione*

Sforzo

- E = V*D = V/L = V^2/V*
  - *può essere pensato come lo sforzo richiesto per comprendere una implementazione piuttosto che quello per produrla*

Considerazioni

Nella storia della misurazione del software il lavoro di Halstead ha avuto dei meriti importanti:

- primo tentativo, anche se non perfetto, di derivare una misura software a partire dalla teoria
- prima misura software usata in un contesto industriale
- primo lavoro che ha evidenziato le potenzialità della misurazione del software nello sviluppo del software
- Le conclusioni che essa propone sono suscettibili di verifica sperimentale

Difetti

- Misure "generali" e non orientate ad un particolare compito
- Metrica basata sul codice e ideata quando (1970) i fondamenti dell'ingegneria del software non erano ancora pienamente apprezzati (in particolare l'importanza della progettazione del sistema)
- Regole di conteggio non precisamente definite
  - *Problemi a causa di linguaggi ambigui*

Metriche per il Testing

- Molte metriche proposte in letteratura si riferiscono al processo di testing (es. sforzo, durata, # difetti rilevati…) , non alle caratteristiche dei test in sé.

- Per progettare ed eseguire l'attività di testing, possono essere utili le metriche del livello analisi, design e codice.

Metriche per il testing

- Per prevedere l'impegno delle varie attività di testing si possono usare:
  - metriche per la misura della funzionalità, per pianificare il testing black-box.
  - metriche per il design architetturale, per stimare la complessità del testing di integrazione e la quantità di driver e stub da sviluppare.
  - La complessità ciclomatica può essere utile per avere una misura della complessità del testing di unità (white-box).
- É inoltre possibile relazionare sforzo e durata del testing, errori scoperti e numero di test case prodotti ai FP (per fare stime per analogia).

Metriche di Binder

- Binder (Object Oriented Software Testing, 1994) propone una varietà di metriche per valutare la testabilità di sistemi object oriented
  - Mancanza di coesione nei metodi (già vista)
  - Percentuale di attributi pubblici o protetti
    - *Creano elevate potenzialità di coupling*
  - Accesso pubblico ai dati (PAD)
    - *Elevati valori di PAD portano a side effect fra le classi.*
  - Numero di classi radice
    - *Più classi radice, più gerarchie da testare*
  - Fan-In (conta il numero di classi da cui una classe eredita)
  - Numero di figli e Profondità dell'albero di ereditarietà

Metriche per la manutenzione

- Le metriche per la manutenzione si occupano di prevedere la complessità del processo di manutenzione
  - Dipendono, quindi, dal processo di manutenzione che si va a istanziare

Ad esempio (Indice di Maturità del Software):

$SMI = [M_T - (F_a + F_C + F_d)] / M_T$

- M_T: numero di moduli nella versione corrente
- F_C: numero di moduli modificati
- F_a: numero di moduli aggiunti
- F_d: numero di moduli eliminati

- **Al tendere di SMI a 1 il prodotto tende a stabilizzarsi**
- **Il tempo medio per produrre una nuova versione del prodotto è correlato a SMI.**

Proporre delle metriche

- Per proporre delle metriche valide, bisogna trovare le relazioni tra entità e metriche
- Tali relazioni possono essere ottenute tramite studi empirici…
  - *si cerca di trovare una legge secondo la quale la misura di un'entità possa essere valutata come funzione delle metriche di attributi direttamente misurabili.*
- Oppure tramite la costruzione di modelli metrici arbitrari, nei quali le relazioni tra attributi interni ed esterni sono ipotizzati da esperti e successivamente validati.

- In [Kan, Metrics and Models in Software Quality Engineering, Second Edition, 2002] si possono trovare molte indicazioni su come effettuare una campagna di misure e su come validare empiricamente l'efficacia di metriche proposte nella descrizione di dati fattori di qualità
- Molto utile anche il libro di Fenton e Pfleeger [Software Metrics, Second Edition, 1997]

Goal-Question-Metric Paradigm

- Introdotto da Basili e Rombach a partire dal 1988
  - Basili, V. R., Caldiera, G. and Rombach, H. D., "The Goal Question Metric Approach", Encyclopedia of Software Engineering, Wiley&Sons Inc., 1994

- L'applicazione di un processo GQM permette di effettuare un programma di raccolta dati
  - Per effettuare l'assessment (valutazione) dell'efficacia dei processi aziendali
  - Per valutare la qualità di prodotti esistenti

GQM

- Basato sull'idea che per poter misurare con successo occorre :
  - Stabilire i Goals (gli Obiettivi) che si intendono perseguire con il programma di raccolta dati;
  - A partire dagli obiettivi, derivare i dati che permettono di valutare il raggiungimento degli obiettivi;
  - Definire uno schema di riferimento per interpretare i dati rispetto agli obiettivi fissati.
- Il risultato sarà un modello di misurazione a tre livelli (Concettuale, Operazionale, Quantitativo) corrispondenti a Goal/ Questions/ Metrics, ottenuto in tre macro-fasi.

Goal-Question-Metric Paradigm

Goals - Obiettivi

- Quali sono gli obiettivi (attributi esterni) che si vogliono perseguire?
  - *Un goal si definisce individuando **l'oggetto** dello studio, le **motivazioni** per la misurazione, un **modello di qualità** di riferimento, un **punto di vista** per l'osservazione, **l'ambiente** a cui l'oggetto appartiene.*
  - *Possibili oggetti: Prodotti/ Processi/ Risorse.*
- Esempio di Goal:
  - *Analizzare le attività di progettazione e verifica del processo di sviluppo, al fine di valutare l'efficacia del processo di rilevazione dei difetti presenti nel software, dal punto di vista del management e del team di sviluppo. Si vuole verificare se una tipologia di test ha un rapporto costo/benefici soddisfacente.*

Questions

Questions - Domande

- Quali domande ci permettono di valutare se e fino a che punto gli obiettivi vengono raggiunti?

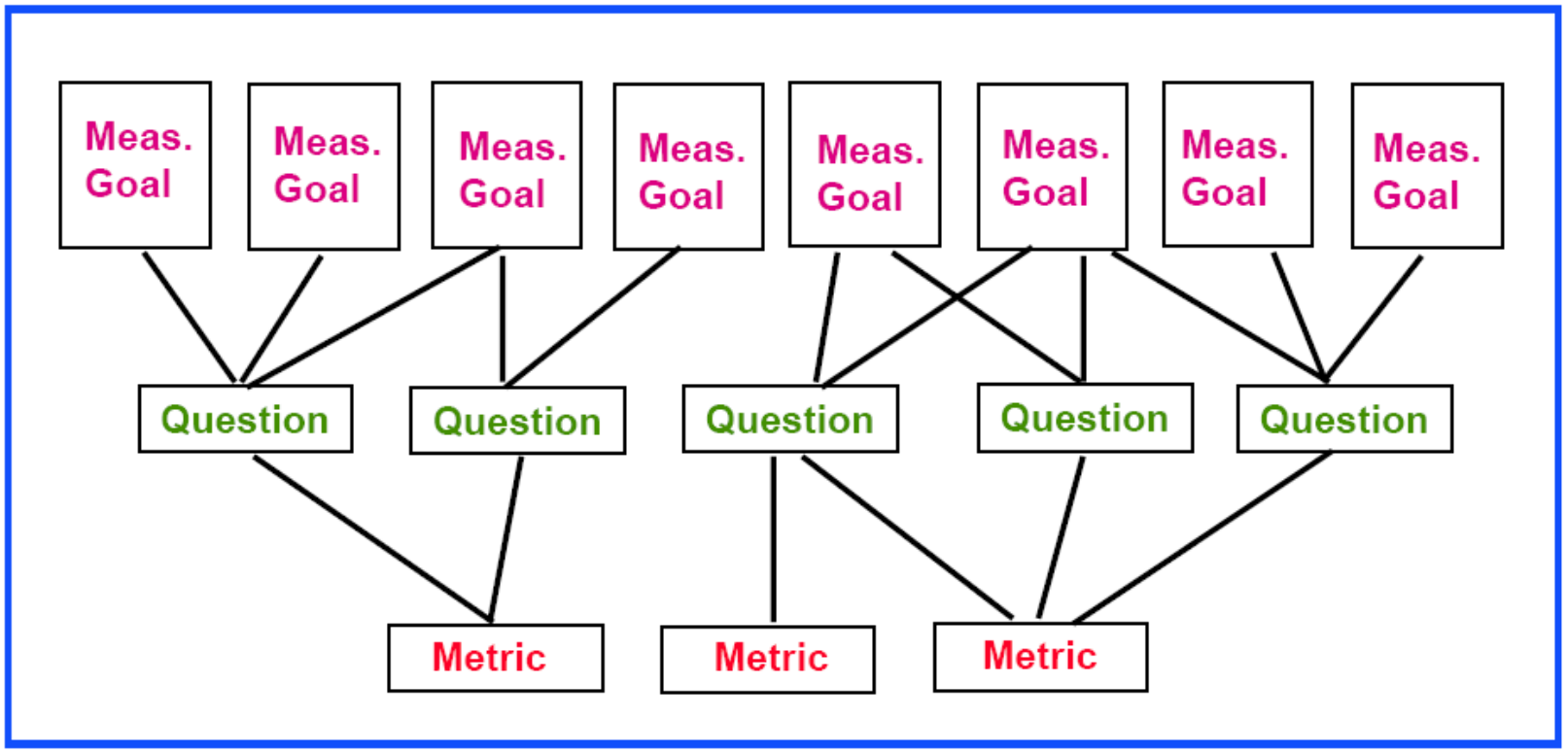
- Esempio di domande:
  - *Q1) Quale è il costo del test?*
  - *Q2) Quali sono I vantaggi?*

Metrics - Metriche

- Quali metriche consentono di dare risposta alle domande?

- Q1)-> M11: Giornate per singolo test
  - -> M12: Costo per figura professionale
  - -> M13: Numero di Persone coinvolte
- -Q2) -> M21: Numero di errori gravi rilevati
  - -> M22: Numero di errori lievi rilevati

Misurazione Goal-Based



Osservazioni

- I valori per le matrici QM e GQ sono spesso ottenuti da questionari compilati dagli esperti dei processi aziendali e dai responsabili della qualità che utilizzano un insieme discreto di valori possibili (poi normalizzati tra 0 e 1)
- La difficoltà nell'esecuzione di un GQM sta nella ricerca di matrici che portino a risultati soddisfacenti
- I valori possono essere aggiustati a seguito di osservazioni fatte a posteriori che vadano a correggere delle dipendenze tra question e goal che non siano soddisfacenti
  - Dei feedback possono essere ottenuti dall'applicazione dello stesso modello GQM a diversi progetti nell'ambito della stessa azienda e dal confronto dei risultati ottenuti