**Università degli studi di Salerno**

**Facoltà di Informatica**



**Dipartimento di Informatica**

**Corso di laurea magistrale in Informatica**

**Data science e machine learning**

**Studio di usabilità degli operatori iconici di CoDIT**

**Docenti Candidati**

Francesca Tassatone 0522500568

Roberta Gesumaria 0522500569

Daniele Lupo

0522500579

Prof. Giuseppe Polese

Dott.ssa Loredana Caruccio

Dott. Bernardo Breve

**Anno accademico 2018-2019**

Sommario

[Introduzione 4](#_Toc21525060)

[iBench 5](#_Toc21525061)

[Installazione 5](#_Toc21525062)

[Problemi riscontrati 5](#_Toc21525063)

[Utilizzo 5](#_Toc21525064)

[File di configurazione 5](#_Toc21525065)

[Problemi riscontrati 8](#_Toc21525066)

[CoDIT 11](#_Toc21525067)

[Installazione 11](#_Toc21525068)

[Problemi riscontrati durante l’installazione 11](#_Toc21525069)

[Utilizzo 11](#_Toc21525070)

[Struttura XMI riconosciuta da CoDIT 12](#_Toc21525071)

[Problemi riscontrati durante l’utilizzo 14](#_Toc21525072)

[Connessione a MySQL e creazione database 17](#_Toc21525073)

[creaConnessione 17](#_Toc21525074)

[creaDb 18](#_Toc21525075)

[caricaDati 19](#_Toc21525076)

[Utilizzo 20](#_Toc21525077)

[Parser 21](#_Toc21525078)

[Implementazione 21](#_Toc21525079)

[Sezione iniziale 21](#_Toc21525080)

[Sezione entità e attributi 22](#_Toc21525081)

[Sezione relazioni 22](#_Toc21525082)

[Sezione finale 23](#_Toc21525083)

[Utilizzo 23](#_Toc21525084)

[Riscontro sull’utilizzo di CoDIT con schemi di grandi dimensioni 24](#_Toc21525085)

[Fasi e macchine utilizzate 24](#_Toc21525086)

[Generazione schemi con iBench 24](#_Toc21525087)

[Risultati 24](#_Toc21525088)

[*PC1* 26](#_Toc21525089)

[*PC2* 26](#_Toc21525090)

[*PC3* 26](#_Toc21525091)

[Tempistiche di esecuzione delle fasi preliminari 28](#_Toc21525092)

[*PC1* 28](#_Toc21525093)

[*PC2* 28](#_Toc21525094)

[*PC3* 29](#_Toc21525095)

[Studio di usabilità 30](#_Toc21525096)

[Indice delle tabelle 31](#_Toc21525097)

# Introduzione

Il progetto che sarà svolto riguarderà lo studio di usabilità sull’espressività e la complessità percepita degli operatori iconici di CoDIT utilizzando scenari che coinvolgono schemi di grosse dimensioni. Per generare tali schemi ci è stato indicato di utilizzare il tool iBench.

Vediamo nel dettaglio il lavoro da svolgere.

Utilizzo di CoDIT, un tool per l’integrazione di schemi. Gli schemi su cui lavora sono schemi concettuali sui quali è possibile applicare operatori iconici per l’integrazione. I formati richiesti da CoDIT sono XMI o SVG.

Per lo studio richiesto, avremo però bisogno di schemi più grandi di quelli che si hanno a disposizione, per ottenere ciò verrà utilizzato IBench.

IBench genera schemi logici in formato XML basandosi su un file di configurazione.

Le prime differenze da notare sono le tipologie di file e di schemi differenti che si ottengono da un tool e che vengono elaborati dall’altro. Inoltre, i file XMI riconosciuti da CoDIT hanno una struttura non generica, quindi sarà necessario estrarre dallo schema logico le informazioni utili per ottenere lo schema concettuale, ottenendo, quindi, una struttura adeguata comprensibile a CoDIT.

Successivamente, verrà effettuato uno studio di usabilità sull’espressività e la complessità percepita degli operatori iconici di CoDIT, sia dai membri del gruppo che da altri utenti, sempre con competenze informatiche.

# iBench

iBench è un generatore di metadati per la creazione di mappature, schemi e vincoli di schemi arbitrariamente grandi e complessi. Può essere utilizzato con un generatore di dati per generare efficacemente scenari realistici di integrazione dei dati con diversi gradi di dimensione e complessità. Inoltre, permette di creare benchmark per diverse attività di integrazione dei dati (virtuali), integrazione dei dati, scambio di dati, evoluzione degli schemi, mappatura degli operatori come composizione e inversione e corrispondenza degli schemi.

Modificando il file di configurazione predefinito, l'utente specifica il tipo di metadati, i dati da generare e le proprietà dello scenario generato. iBench produce un file XML che memorizza i metadati generati in base alla configurazione dell'utente. Se richiesto, genera anche i dati di istanza per gli schemi generati.

## Installazione

iBench è scritto in Java, per cui c’è bisogno di utilizzare ant per il building e ivy per scaricare le dipendenze.

La prima operazione svolta è stata quella di scaricare e installare ant.

Eseguendo il comando ant nella cartella principale, è stato scaricato ivy ed utilizzato automaticamente per scaricare il jar delle dipendenze. Successivamente, è stata generata una cartella build in cui sono presenti vari script di shell (.sh per Linux e Mac OS e .bat per Windows).

Il tutto è riportato in modo semplice nel file README messo a disposizione.

### Problemi riscontrati

Nell’utilizzo del file .bat per l’esecuzione di iBench è stata riscontrata la mancanza, nel codice, dell’indicazione dell’acquisizione dell’argomento dato da linea di comando. Successivamente all’aggiunta di tale sezione di codice, l’esecuzione è terminata senza alcun problema.

Codice iBench.bat non funzionante:

*java -Xmx4096m -jar ibench-fat.jar*

Codice iBench.bat funzionante:

*java.exe -jar ibench-fat.jar -c %2*

Inoltre, il parametro -*Xmx4096m* limitava l’utilizzo della RAM a 4GB.

## Utilizzo

iBench permette di generare source schema e target schema.

Per poter generare questi schemi viene messo a disposizione un file di configurazione. Attraverso quest’ultimo è possibile settare le dimensioni e la struttura che si vuole ottenere per gli schemi risultanti.

I nomi dell’intera struttura e i dati vengono generati in modo casuale.

### File di configurazione

Vediamo nel dettaglio le varie sezioni del file di configurazione che abbiamo utilizzato per la generazione degli schemi.

La prima parte è relativa alla definizione della *path* in cui devono essere salvati gli schemi e i dati generati, poi viene indicato il nome del file che conterrà i metadati relativi alla struttura dello schema generato.



Successivamente troviamo la parte per la configurazione sul numero di istanze, ossia entità per ogni schema, il numero di attributi per ogni entità, ed altre configurazioni relative ai dati generati.



Poiché i dati sono generati in modo casuale è possibile indicare il seme per il generatore casuale. Inoltre, per introdurre variabilità nella costruzione delle relazioni, è possibile indicare la deviazione relativa al numero di attributi per relazione.



La definizione delle dipendenze viene stabilita tramite percentuali. La percentuale delle dipendenze viene applicata al totale degli schemi generati, considerando sia source che target; quella delle chiavi esterne viene applicata sul numero di dipendenze generate.

Per esempio, utilizzando una configurazione con i seguenti parametri: 100 entità, 50% di dipendenze e 50% di chiavi esterne avremo 50 dipendenze con 25 chiavi esterne.



Per la generazione dei dati viene messa a disposizione una sezione in cui è possibile definire tipologie specifiche (dati UDT - User Definition Data Type), come e-mail, numero di telefono ed altri che possono essere scelti tra le classi che si trovano al seguente link <https://www.cs.toronto.edu/tox/toxgene/javadoc/allclasses-noframe.html>. In questa sezione bisogna indicare il numero di UDT che saranno definiti, poi per ognuno di esso bisognerà specificare un nome univoco, la classPath relativa alla classe della tipologia del dato che si vuole generare, la percentuale di generazione ed il tipo in cui verrà codificato il campo in un database SQL.



Inoltre, è possibile scegliere i dati, sempre in modo casuale, ma da un set fornito tramite file csv. Per configurare questa parte deve essere indicato il file csv, il nome della colonna all’interno del file, la percentuale di generazione ed infine il tipo in cui verrà convertito il campo in un database SQL.



Tramite i seguenti parametri è possibile gestire alcune differenze tra lo schema source e target.  
Utilizzandoli verrà effettuato il partizionamento (orizzontale, verticale), verranno aggiunti o rimossi alcuni attributi oppure verrà effettuato il merge di alcuni attributi.



Infine, sono presenti due sezioni relative ad opzioni sui dati di output che verranno generati.





### Problemi riscontrati

I problemi riscontrati si possono dividere in base all’obiettivo della configurazione, cioè due:

* Generazione dati per lo schema target;
* Utilizzo di csv per la generazione dei dati.

#### Generazione dati per lo schema target

Avendo bisogno dei dati relativi agli schemi generati, abbiamo configurato iBench in modo da abilitare la creazione dei file csv relativi alla generazione di tali dati. Per la creazione dei dati relativi allo schema target abbiamo avuto dei problemi in caso di particolari configurazioni.



Figura : proprietà per l'abilitazione della generazione dei csv del target schema e il numero di tuple da creare.

Le configurazioni che non sono risultate adatte alla generazione dei dati sono state:

* Abilitazione delle proprietà relative alle percentuali sulle dipendenze e chiavi esterne;
* Partizionamento orizzontale e verticale.

Vediamo nel dettaglio gli errori per entrambi i casi.

##### Abilitazione delle proprietà relative alle percentuali sulle dipendenze e chiavi esterne

La generazione dei file csv è stata possibile solo se le proprietà sulle percentuali relative alle dipendenze e chiavi esterne non erano state abilitate. Infatti, nel caso di abilitazione di tali proprietà, non venivano generati i dati relativi allo schema target, restituendo il seguente errore:



Seguendo lo stack di eccezioni giungiamo al seguente snipet di codice:



Figura : Snipet di codice presente nella classe MappingScenario.

Sì, probabilmente è questo il problema!

##### Partizionamento orizzontale e verticale

Nel generare schemi con l’abilitazione delle proprietà per il partizionamento orizzontale o verticale, i nomi delle entità che vengono generati nel target vengono creati con una sezione in maiuscolo, ad esempio “HP0FR0”. Con l’abilitazione delle proprietà per la generazione dei dati per il target, l’esecuzione restituisce il seguente errore:



Analizzando il codice, presumiamo che l’errore sia dovuto ad un’acquisizione con una forzatura al minuscolo del nome delle entità relative agli schemi.



#### Utilizzo di csv per la generazione dei dati

Nella sezione per UDT si possono specificare i file da cui prendere i valori per generare i dati.  
Provando con un solo file csv venivano generate righe contenenti più tuple. Inoltre, le percentuali delle chiavi esterne e delle dipendenze non erano coerenti alla configurazione specificata.

# CoDIT

CoDIT (Conceptual Data Integration Tool), supporta il processo di integrazione dei dati a livello concettuale, permette la costruzione di uno schema globale correlando sottoschemi delle sorgenti e, in questo modo, il DBA potrà specificare in maniera visuale come combinare le sorgenti di dati. Lo schema globale potrà essere specificato allineando i sottoschemi delle sorgenti oppure specificando nuovi costrutti e mappando i sottoschemi delle sorgenti ad essi.

A livello metodologico CoDIT si basa su un linguaggio visuale, CoDIL (Conceptual Data Integration Language).

CoDIL permette di gestire l’eterogeneità degli schemi, per far fronte alla necessità di risolvere i conflitti dovuti all’utilizzo di diversi costrutti e/o nomi per modellare la stessa realtà.

Attraverso CoDIL, il DBA potrà selezionare costrutti degli schemi sorgenti (uno per ogni schema) in relazione ed associare un operatore iconico per specificare come risolvere il conflitto.

Il linguaggio CoDIL fornisce un insieme di operatori iconici che possono essere applicati a coppie di sottoschemi concettuali. Attraverso l’utilizzo di tali operatori è possibile specificare in che modo i sottoschemi selezionati sono correlati e scegliere come questi devono essere mappati nello schema riconciliato.

## Installazione

Per l’installazione del tool è stato utilizzato il file README messo a disposizione. Il file è molto dettagliato sia per l’istallazione su Mac OS che per Windows. Inoltre, la guida per il sistema Mac OS può essere eseguita anche per Windows.

### Problemi riscontrati durante l’installazione

Nel file README viene menzionata una guida per modificare il file *web.xml* della configurazione di Catalina. Purtroppo, all’interno della guida sono presenti alcuni errori, come ad esempio dei tag di apertura o chiusura mancanti. Una volta risolti i vari errori, si ottiene il file *web.xml* allegato.

## Utilizzo

Le funzionalità offerte da CoDIT sono:

1. Disegna schema;
2. Caricamento dinamico;
3. Create target scheme.



Figura : Home tool CoDIT.

L’operazione sulla quale ci siamo focalizzati è il caricamento dinamico. I formati riconosciuti da CoDIT sono XMI e SVG. Il file XMI, però, ha una struttura specifica che riporteremo di seguito.

### Struttura XMI riconosciuta da CoDIT

La struttura del file XMI può essere suddivisa in quattro parti:

* Una sezione iniziale;
* Una sezione in cui vengono indicate tutte le entità con i relativi attributi;
* Una sezione in cui vengono indicate tutte le relazioni;
* Una sezione finale.

#### Sezione iniziale

La seguente sezione risulta essere identica in tutti i file forniti come esempi per un primo utilizzo di CoDIT.



Figura : Struttura XML CoDIT, sezione iniziale.

#### Sezione entità e attributi

Per indicare l’entità viene utilizzato il tag "packagedElement", con i seguenti attributi:

* xmi:type = uml:Class, per indicare che si tratta di un’entità;
* xmi:id, per indicare l’id;
* name, per indicare il nome dell’entità;
* type, per indicare il tipo di entità. Non sempre viene specificato, di solito è usato per indicare le entità deboli (WeakEntity).



Figura : Struttura XML CoDIT, entità.

Per ogni entità, come figlio dell’elemento "packagedElement", viene indicato l’attributo “xmi:Extension” per specificare che si tratta di un’entità.



Figura : Struttura XML CoDIT, estensione identificativa dell'entità.

Per indicare gli attributi relativi all’entità viene utilizzato il tag "ownedAttribute” come figlio dell’elemento "packagedElement".

Per ogni attributo vengono definiti:

* xmi:type=”uml: Property, per indicare che è un attributo;
* xmi:id, per indicare l’id;
* name, per indicare il nome dell’attributo;
* visibility, per indicare la visibilità dell’attributo.



Figura : Struttura XML CoDIT, attributo.

Nel caso in cui l’attributo sia chiave primaria, viene utilizzato il tag “xmi:Extension”, come figlio di “ownedAttribute”, per indicare che si tratta di una chiave primaria.



Figura : Struttura XML CoDIT, attibuto chiave primaria.

#### Sezione relazioni

In questa sezione vengono definite le relazioni tra le entità sopra definite.

Viene usato sempre il tag "packagedElement", questa volta però è utilizzato con lo scopo di definire le associazioni, e viene specificato nell’attributo “xmi: type= uml:Association”.

La relazione tra le entità è indicata attraverso l’attributo “xmi:id” in cui è specificato il nome delle due entità relazionate, usando il seguente formato E1\_\_E2\_\_id.

Inoltre, se si deve definire una specializzazione, oltre ad indicare un id in cui ci sono i nomi delle entità, viene inserito anche l’attributo “type=ISA”.



Figura : Struttura XML CoDIT, sezione relazioni tra le entità.

#### Sezione finale

La seguente sezione risulta essere identica in tutti i file forniti come esempi per un primo utilizzo di CoDIT.



Figura : Struttura XML CoDIT, sezione finale.

#### Coordinate elementi

Un’ulteriore funzionalità offerta da CoDIT è quella di salvare la posizione degli elementi, memorizzando le relative coordinate all’interno del file XMI. Quindi, per ogni elemento identificativo di una entità, attributo o relazione, sono presenti gli attributi x e y.



Figura : Struttura XML CoDIT, entità e attributi con coordinate.

### Problemi riscontrati durante l’utilizzo

Durante l’utilizzo preliminare di CoDIT, ancora non incentrato sullo studio di usabilità richiesto, sono stati riscontrati vari problemi, alcuni dei quali sono stati compresi e risolti per poter utilizzare schemi di grandi dimensioni.

#### Limitazione della visualizzazione degli schemi

Il primo problema riscontrato era legato alla visualizzazione degli schemi dalle dimensioni sotto indicate, generati con il tool iBench e successivamente convertiti nel formato xmi leggibile dal tool CoDIT.

Gli schemi utilizzati erano di tali dimensioni:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** **schema** | #**Entità** | **#Attributi per entità1** | **#Tuple** | **Percentuale dipendenze** | **Percentuale chiavi esterne** |
| Schema1 | 50 | 10 | 100 | 50 | 50 |
| Schema2 | 80 | 10 | 100 | 50 | 50 |

Tabella : Tabella relativa agli schemi nella fase iniziale dell'utilizzo di CoDIT.

Nel caricamento di questi schemi il risultato ottenuto era di due tipi:

1. per lo Schema1 la sezione atta alla visualizzazione dello schema appariva bianca. Gli errori che venivano riscontrati sulla console del browser (Chrome) e su Eclipse erano i seguenti:



Figura : Console browser





Figura : Console Eclipse

1. per lo Schema2, la schermata atta alla visualizzazione degli schemi restava bloccata sul caricamento indicando la dimensione del file che si voleva caricare.

L’errore veniva riportato solo nelle console del browser:



Discutendone con il dottorando Bernardo Breve, siamo riusciti a risolvere il problema relativo allo Schema1, modificando parte dell’algoritmo usato per visualizzare gli elementi dello schema (parser1.js). Per quanto riguarda il problema relativo allo Schema2, invece, non è stato ancora risolto poiché viene ancora riscontrato nella sezione di destra ogni qualvolta dovrebbe essere visualizzato un modal di errore. Quindi, supponiamo sia legato ad un errore relativo alla creazione del modal.

#### Limitazione sulla dimensione massima dei file caricati

Dopo aver risolto il problema precedente legato alla mancata visualizzazione di alcuni schemi, abbiamo provato a visualizzare schemi di grosse dimensioni. Non è stato possibile a causa della dimensione del file che eccedeva quella di caricamento massima.

Anche per questo errore, trovato il punto in cui veniva definita la dimensione massima (upload\_file.php, riga 9) è stato risolto il problema portando la dimensione ad un massimo di 1Gb.

#### Limitazione della sezione di visualizzazione degli schemi

Riuscendo a caricare e visualizzare schemi di grosse dimensioni ci siamo resi conto che la sezione atta alla visualizzazione è limitata dato che, pur permettendo lo scorrimento, essa non risulta infinita. In conclusione, non è possibile visualizzare nel complesso schemi eccessivamente grandi.

# Connessione a MySQL e creazione database

Per alcune funzionalità messe a disposizione da CoDIT è necessario che, oltre ad essere in possesso di uno schema concettuale, ci sia anche un database popolato. Abbiamo, quindi, creato uno script in Python, che permettesse di connettersi a MySQL e ci permettesse di creare e popolare nuovi database.



Di seguito verranno illustrati nel dettaglio i vari metodi implementati.

## creaConnessione



La libreria mysql.connector mette a disposizione un oggetto dove vengono salvate tutte le informazioni necessarie per la connessione. Dato che serve una connessione per ogni database, è stato implementato un metodo che si occupa di configurare la connessione al database, specificato come parametro.

## creaDb

Immagine che contiene screenshot

Descrizione generata automaticamenteImmagine che contiene screenshot, testo

Descrizione generata automaticamente

Immagine che contiene screenshot

Descrizione generata automaticamente

Questo metodo si occupa di leggere le informazioni dal file *xml* contenente la struttura degli schemi.  
Una volta eliminati i database, se già esistenti, crea i nuovi database che conterranno le relazioni degli schemi *Source* e *Target*.

Leggendo la struttura delle varie relazioni, si occuperà di effettuare le *query* di creazione delle varie tabelle con i relativi attributi.

## caricaDati



Una volta creata la singola tabella, verranno caricati i dati inerenti letti dal csv denominato come la tabella.  
Leggendo il csv, verrà costruita la query di inserimento utilizzando i dati divisi per riga.  
Dato che iBench, con alcune configurazioni, genera più valori rispetto al numero di campi, è stato introdotto un controllo sul numero di valori utilizzati.

## Utilizzo

Per eseguire lo script Python bisogna utilizzare il seguente comando:

*python3 ConnessioneDb.py pathFile.xml nomeDb*

Dove:

* *ConnessioneDb*.py, è il nome dello script che si vuole eseguire;
* pathFile.xml, è il nome del file dal quale si vuole acquisire la struttura del database;
* nomeDb, è il suffisso del nome che verrà dato ai database che saranno creati per il source e il target schema.

# Parser

Come specificato nel paragrafo introduttivo, iBench genera uno schema logico su file XML, mentre CoDIT lavora su schemi concettuali su file XMI con una struttura mostrata nel paragrafo *Struttura XMI riconosciuta da CoDIT*. Per risolvere questo problema è stato creato un parser, che letto il file XML genera un nuovo file XMI con la struttura richiesta.

Per effettuare questa traduzione da logico a concettuale non è stato possibile effettuare un reengineering poiché lo schema logico generato da iBench viene costruito con dei nomi casuali, quindi la struttura generata non ha una semantica utile per poter effettuare un vero e proprio reengineering semi-automatico o manuale. Ciò che è abbiamo potuto fare è stato semplicemente mappare ogni relazione, nello schema generato da iBench, in una entità nello schema concettuale risultante, ricostruendo attributi, chiavi primarie ed esterne dalle informazioni relative ad ogni relazione specificata nella struttura generata da iBench.

## Implementazione

Per implementare il parser è stato utilizzato il linguaggio di programmazione Python, il quale offre una soluzione finalizzata proprio all’interfacciamento con i file XML. Infatti, grazie all’utilizzo del modulo xml.etree.ElementTree è stato possibile leggere il contenuto del file e creare una struttura per il nuovo file XMI da generare.

Per individuare i nomi degli elementi all’interno del file XML è stato utilizzato XPath, un linguaggio di interrogazione, parte della famiglia XML, che permette di individuare i nodi all’interno di un documento.

Di seguito mostreremo le varie parti del parser in relazione alla struttura del file XML mostrata nel paragrafo *Struttura XMI riconosciuta da CoDIT*.

### Sezione iniziale



### Sezione entità e attributi





### Sezione relazioni



### Sezione finale



## Utilizzo

Per eseguire lo script Python bisogna utilizzare il seguente comando:

*python3 Parser.py pathFile.xml nomeFile*

Dove:

* Parser.py, è il nome dello script che si vuole eseguire;
* pathFile.xml, è il nome del file xml che si vuole convertire;
* nomeFile, è il suffisso del nome che verrà dato ai file finali xmi per il source e il target schema.

Cosa importante da ricordare è che, per non avere problemi con l’utilizzo dei file con il tool CoDIT, in nomeFile non utilizzare il segno di punteggiatura “punto”.

# Riscontro sull’utilizzo di CoDIT con schemi di grandi dimensioni

Di seguito saranno riportati i test svolti durante le fasi preliminari dello studio di usabilità degli operatori iconici del tool CoDIT per verificarne il funzionamento con schemi di grandi dimensioni.

## Fasi e macchine utilizzate

Fasi per la generazione di schemi utilizzabili con CoDIT:

* Generazione schema iBench, fase in cui vengono generati gli schemi logici (source e target) in formato xml e i dati relativi solo allo schema source, attraverso il file di configurazione suddetto (vedi paragrafo Generazione dati per lo schema target);
* Traduzione xml to xmi con parser, fase in cui viene tradotta la struttura dello schema logico generato da iBench in xml in schema concettuale compreso da CoDIT in xmi;
* Creazione database e caricamento dati su MySQL, fase in cui, basandoci sulla struttura dello schema generato da iBench, creiamo un database per ogni schema (sourceSchema e targetSchema). Ogni tabella relativa al sourceSchema verrà popolata in modo automatico utilizzando i file csv generati da iBench.

Per testare le tempistiche delle suddette fasi al fine di ottenere materiale atto allo studio di usabilità sono state utilizzate tre macchine con le seguenti specifiche:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **PC** | **CPU** | **RAM** | **Tipo memoria** | **Sistema operativo** |
| 1 | I7-5500U 2.40GHz | 8GB | HDD | Windows 10 64bit |
| 2 | Rayzen 5 2500U 2GHz | 8GB | SSD | Windows 10 64bit |
| 3 | I5 2.7GHz | 8GB | SSD | Mac OS Mojave |

Tabella : Specifiche dei pc.

## Generazione schemi con iBench

Uno degli obiettivi principali dello studio che ci è stato affidato era quello di poter generare schemi “abbastanza grandi” che poi sarebbero stati utilizzati per un successivo studio di usabilità di CoDIT.

### Risultati

Per comprendere le dimensioni degli schemi adatte per un buon utilizzo del tool CoDIT, abbiamo svolto le 3 fasi o parte di esse su schemi di varie dimensioni. Di seguito verranno mostrati i diversi casi valutati.

Inizialmente, con il tool iBench sono stati generati schemi di dimensioni relativamente grandi. Di seguito sono descritte le caratteristiche di ognuno di essi, considerando che sono gli schemi più grandi che siamo riusciti a generare con le macchine a disposizione.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** **schema** | #**Entità** | **#Attributi per entità[[1]](#footnote-1)** | **#Tuple** | **Percentuale dipendenze** | **Percentuale chiavi esterne** |
| SC0 | 200 | 5 | 100 | 100 | 50 |
| SC1 | 400 | 7 | 200 | 50 | 50 |
| SC2 | 470 | 9 | 150 | 80 | 50 |
| SC3 | 490 | 10 | 140 | 50 | 50 |

Tabella : Schemi che, convertiti in xmi, sono risultati di dimensioni superiori al massimo consentito.

Nella tabella seguente vengono riportati i tempi impiegati per la generazione degli schemi:

|  |  |  |  |
| --- | --- | --- | --- |
| **ID schema** | **PC1** | **PC2** | **PC3** |
| SC0 | 37s | 24s | 37s |
| SC1 | 378s | 319s | 308s |
| SC2 | 546s | 468s | 443s |
| SC3 | 700s | err[[2]](#footnote-2) | 490s |

Tabella : Tempi relativi alla generazione di schemi su iBench.

I suddetti schemi, successivamente alla conversione in xmi, non sono risultati adatti alla visualizzazione su CoDIT, poiché le dimensioni dei file risultavano superiori alla dimensione massima (vedi paragrafo *Limitazione sulla dimensione massima dei file caricati).*

Analizzando il codice, siamo riusciti ad aumentare la dimensione massima dei possibili file caricati e abbiamo provato a caricare lo schema più piccolo, cioè SC0. Il risultato non è stato dei migliori, in quanto il tempo impiegato risulta essere eccessivo per l’attesa (circa 15 minuti per il PC2, 23 minuti per il PC1). Inoltre, la visualizzazione risultava essere poco comprensibile e non totale dello schema, poiché la sezione per la visualizzazione è limitata. In conclusione, riteniamo sia poco utile generare grandi schemi per l’utilizzo di CoDIT, sia per i tempi di attesa che per la visualizzazione finale (vedi paragrafo *Limitazione della sezione di visualizzazione degli schemi*).

Abbiamo, quindi, generato altri tre schemi di dimensioni inferiori, che abbiamo ritenuto più consoni per un utilizzo con il tool CoDIT. Per ottenere schemi strutturati in modo differente, abbiamo abilitato alcune proprietà nel file di configurazione per lo schema SC5 e SC6.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID** **schema** | #**Entità** | **#Attributi per entità1** | **#Tuple** | **Percentuale dipendenze** | **Percentuale chiavi esterne** | **Proprietà**  **attivate** |
| SC4 | 50 | 5 | 50 | 100 | 50 | - |
| SC5 | 100 | 7 | 100 | 85 | 50 | Partizionamento verticale 10 |
| SC6 | 70 | 10 | 200 | 50 | 50 | AddDelAttribute 5 |

Tabella : Schemi utilizzati per lo studio di usabilità.

Per ogni macchina utilizzata è riportata una tabella con i tempi di esecuzione per ogni fase. I tempi che sono stati indicati sono il risultato di una media su tre esecuzioni.

### PC1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC4 | 10s | 0.3s | 39s | | 38s |
| SC5 | 39s | 0.5s | 274s | | 312s |
| SC6 | 20s | 0.5s | 1838s | | 1825s |

Tabella : Tempi ottenuti con il PC1.

### PC2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC4 | 7s | 1s | 60s | | 69s |
| SC5 | 18s | 1s | 327s | | 317s |
| SC6 | 24s | 1s | 320s | | 278s |

Tabella : Tempi ottenuti con il PC2.

### PC3

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC4 | 5s | 0.08s | 48s | | 48s |
| SC5 | 15s | 0.15s | 113s | | 46s |
| SC6 | 106s | 0.13s | 340s | | 330s |

Tabella : Tempi ottenuti con il PC3.

Anche questi schemi, però, non sono risultati adatti. I motivi legati all’inadeguatezza degli schemi sono principalmente dovuti alla struttura dell’interfaccia grafica di CoDIT. Dopo un primo utilizzo svolto da noi tre, per la riorganizzazione degli schemi da poter usufruire per effettuare lo studio di usabilità sugli operatori iconici, abbiamo ritenuto opportuno avere un riscontro anche da altri colleghi in modo da confermare la nostra teoria. Abbiamo, quindi, sottoposto ad un gruppo di colleghi della laurea di Informatica un test sulla visualizzazione degli schemi su descritti, con la possibilità di poterli riorganizzare in modo da renderli comprensibili. I risultati non sono stati molto diversi da ciò che avevamo ipotizzato.

Il gruppo di ragazzi può essere diviso in due categorie, quelli che si arrendevano dopo pochi minuti, ritenendo la situazione troppo stressante e quelli che cercavano di completare il task assegnatogli fino alla fine, nonostante lo stress dovuto alla poca usabilità del tool.

I riscontri finali ottenuti sull’utilizzo di CoDIT sull’usabilità relativa al task assegnato si possono riassumere nei seguenti punti:

* Sarebbe utile poter effettuare zoom in e zoom out sulle singole sezioni di visualizzazione degli schemi in modo da poter avere un quadro generale dello schema oppure poter visualizzare a schermo intero le due sezioni in modo da poterle utilizzare singolarmente;
* Data la possibilità di nascondere gli attributi relativi ad un’entità con il doppio click su di essa, sarebbe utile poter spostare l'entità facendo sì che alla ricomparsa degli attributi, questi siano riposizionati nelle vicinanze dell'entità stessa;
* Una migliore gestione sul rilascio della linea di collegamento tra elementi nel caso in cui la si clicchi per sbaglio;
* La possibilità di un pulsante che permetta di annullare le operazioni effettuate involontariamente;
* Consentire la creazione di linee di collegamento nel caso in cui vengano eliminate erroneamente.

In conclusione, riteniamo che data la poca usabilità dell’interfaccia grafica di CoDIT, non permetta l’utilizzo di schemi con eccessive entità e relazioni tra esse.

# Tempistiche di esecuzione delle fasi preliminari

Successivamente ai problemi riscontrati, abbiamo definito schemi di piccole dimensioni (dalle 6-10 entità, con circa 5 attributi a testa) che sono risultati adatti allo studio di usabilità degli operatori iconici di CoDIT sia per tempi di esecuzione che per comprensibilità.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID** **schema** | #**Entità** | **#Attributi per entità[[3]](#footnote-3)** | **#Tuple** | **Percentuale dipendenze** | **Percentuale chiavi esterne** | **Proprietà**  **attivate** |
| SC7 | 2 | 5 | 100 | 70 | 100 | VertPartition = 4 |
| SC8 | 4 | 5 | 100 | 70 | 100 | Merging = 2 |
| SC9 | 2 | 5 | 100 | 70 | 100 | VertPartition = 2 and  Merging = 2 |

Per ogni macchina a disposizione sono state svolte le tre fasi preliminari (vedi paragrafo *Fasi e macchine utilizzate*) con successiva visualizzazione sul tool. Per ogni fase sono stati riportati i tempi di esecuzione, determinati da una media su tre esecuzioni. Successivamente alla visualizzazione degli schemi, questi ultimi sono stati riorganizzati in modo da renderli comprensibili, salvando gli schemi con le relative posizioni. I file risultanti saranno allegati alla documentazione. La riorganizzazione di ogni singolo schema è stata effettuata all’incirca in cinque minuti.

### PC1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Creazione e popolamento del database su MySQL** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC7 | 1.93s | 0.56s | 10.4s | 1.14s | | 1.80s |
| SC8 | 2.20s | 0.12s | 9.71s | 1.61s | | 1.19s |
| SC9 | 2.30s | 0.13s | 16.10S | 1.66s | | 1.56s |

Tabella : Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC1.

### PC2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Creazione e popolamento del database su MySQL** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC7 | 2s | 0.27s | 1.67s | 1.23s | | 1.53s |
| SC8 | 1.84s | 0.14s | 1.99s | 1.52s | | 1.31s |
| SC9 | 1.9s | 0.14s | 2.49s | 1.24s | | 1.39s |

Tabella : Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC2.

### PC3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Creazione e popolamento del database su MySQL** | **Visualizzazione schema sul tool CoDIT** | | |
| ***SourceSchema*** | ***TargetSchema*** | |
| SC7 | 2s | 0.08s | 0.88s | 0.01s | | 0.02s |
| SC8 | 2.18s | 0.05s | 0.7s | 0.02s | | 0.02s |
| SC9 | 2.30s | 0.05s | 0.78s | 0.02s | | 0.02s |

Tabella : Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC3.

# Studio di usabilità

## Metriche utilizzate

Al fine di una maggiore comprensione ed un’individuazione delle problematiche relative all’usabilità di CoDIT sono state utilizzate le seguenti metriche:

* *Tempo d’azione*: è il tempo totale che serve all’utente per completare l’attività. Il cronometro viene fatto scattare quando l’utente ha compreso il compito da portare a termine e viene stoppato nel momento in cui tutte le azioni sono terminate;
* *Livello di soddisfazione circa il compito e il test*: si tratta di una metrica qualitativa che consiste nel chiedere all’utente di esprimere un parere circa la difficoltà incontrata nel portare a termine il compito assegnato e nell’effettuare il test. Tale metrica è stata ricavata attraverso il questionario post esperimento;
* *Errori*: durante lo svolgimento dei task sono state contate le azioni compiute erroneamente durante l’esecuzione dell’operazione, come selezione errata di un’icona, sezioni non cliccabili, click successivi nel tentativo di attivare un’azione, trascinamento di elementi in sezioni errate, etc.

## Le fasi del test

Lo studio dell’usabilità è stato condotto utilizzando le macchine sopra descritte (paragrafo Fasi e macchine utilizzate). Il test è composto dalle seguenti fasi:

1. Preparazione;
2. Esecuzione;
3. Analisi dei risultati.

### Preparazione

Questa fase prevede i seguenti aspetti:

* analisi preliminare del tool;
* quali operatori iconici utilizzare in base agli schemi generati;
* quanti utenti selezionare e quali tipologie di utenti scegliere;
* quali e quanti task preparare;
* come preparare i moduli per la raccolta dati;
* cosa fare prima dell’osservazione: il test pilota.

#### Analisi preliminare del tool

Essendo un tool a noi sconosciuto per avere una conoscenza sufficiente per poter illustrare al meglio le funzionalità offerte, abbiamo svolto un utilizzo preliminare. Inoltre, essendo sperimentale, abbiamo testato quali fossero le operazioni realmente funzionanti.

#### Quali operatori iconici utilizzare in base agli schemi generati

Valutando gli schemi generati con il tool *iBench*, abbiamo selezionato gli operatori iconici utilizzabili durante il test di usabilità.

Gli operatori iconici selezionati sono i seguenti con i rispettivi inversi che non sono riportati.

##### Mapping di attributi



Da utilizzare nel caso in cui lo stesso concetto sia stato modellato da una parte con un attributo semplice e dall’altro con uno derivabile.



Da utilizzare nel caso in cui lo stesso concetto sia stato modellato da una parte con un attributo semplice e dall’altro da uno composto.



Da utilizzare nel caso in cui lo stesso concetto sia stato modellato da una parte con un attributo semplice e dall’altro con una relazione tra due entità.

##### Partizionamento



Da utilizzare nel caso in cui la relazione dello schema di destra rappresenta una decomposizione verticale dell’entità dello schema di sinistra.

#### Quanti utenti selezionare e quali tipologie di utenti scegliere

Essendo un tool relativo ai database e all’integrazione degli stessi abbiamo ritenuto opportuno selezionare utenti del corso di laurea di Informatica, sia della triennale che della magistrale. Il numero di utenti selezionati è stato 11 per avere un campione variegato per poter avere abbastanza dai da poter analizzare.

#### Quali e quanti task preparare

I task da far svolgere durante il test sono stati preparati basandoci sugli operatori iconici scelti nella fase precedente (paragrafo Quali operatori iconici utilizzare in base agli schemi generati). Quindi, abbiamo preparato quattro task, uno per ogni operatore iconico.

##### Task 1

*Situazione*: ti trovi in una situazione in cui in uno schema hai un attributo semplice e nell'altro un attributo derivato. Quale operatore iconico utilizzeresti per mantenere una configurazione piuttosto che un'altra?

*Schema da utilizzare*: SC7.

*Informazioni preliminari*: l’entità coinvolta è cheese\_cp\_1\_nl0\_ce0.

*Elementi da selezionare*: Nello schema source, nell'entità cheese\_cp\_1\_nl0\_ce0 hai l'attributo ranch\_cp\_1\_nl0\_ae2 come attributo semplice. Nello schema target, medesima entità, hai l'attributo compare\_cp\_1\_nl0\_ae1 derivabile dall'attributo branch\_cp\_1\_nl0\_ae2.

*Task 1.1*: Mantenere la configurazione con l'attributo semplice.

*Task 1.2*: Mantenere la configurazione con l'attributo derivabile.

##### Task 2

*Situazione*: ti trovi in una situazione in cui in uno schema hai un attributo semplice e nell'altro un attributo composto. Quale operatore iconico utilizzeresti per mantenere una configurazione piuttosto che un'altra?

*Schema da utilizzare*: SC7.

*Informazioni preliminari*: l’entità coinvolta è test\_cp\_0\_nl0\_ce0.

*Elementi da selezionare*: Nello schema source, nell'entità test\_cp\_0\_nl0\_ce0 hai gli attributi slope\_cp\_0\_nl0\_ae2, touch\_cp\_0\_nl0\_ae4 e measure\_cp\_0\_nl0\_ae3 che rappresentano parti dell'attributo measure\_cp\_0\_nl0\_ae3 dello schema source. Nello schema target, medesima entità, hai l'attributo measure\_cp\_0\_nl0\_ae3 come attributo semplice.

*Task 2.1*: Mantenere la configurazione con l'attributo semplice.

*Task 2.2*: Mantenere la configurazione con l'attributo composto.

##### Task 3

*Situazione*: ti trovi in una situazione in cui in uno schema hai un attributo semplice e nell'altro è rappresentato da due entità relazionate tra loro. Quale operatore iconico utilizzeresti per mantenere una delle due configurazioni?

*Schema da utilizzare*: SC7.

*Informazioni preliminari*: le entità coinvolte sono red\_vp\_1\_nl0\_ce0, tail\_vp\_0\_nl0\_ce1 e warn\_vp\_0\_nl0\_ce0.

*Elementi da selezionare*: Nello schema source, nell'entità red\_vp\_1\_nl0\_ce0 hai l'attributo describe\_vp\_1\_nl0\_ae5 come attributo semplice. Nello schema target invece, l'attributo describe\_vp\_1\_nl0\_ae5 suddetto è stato modellato sulle entità tail\_vp\_0\_nl0\_ce1 e warn\_vp\_0\_nl0\_ce0, relazionate attraverso tail\_vp\_0\_nl0\_ce1/warn\_vp\_0\_nl0\_ce0.

*Task 3.1*: Far sì che anche nel target sia mantenuto l'attributo semplice.

*Task 3.2*: Far sì che anche nel source l'attributo semplice sia scomposto in due entità come accade nel target.

##### Task 4

*Situazione* : ti trovi in una situazione in cui in uno schema hai un'unica entità e nell'altro è rappresentata da due entità relazionate tra loro. Quale operatore iconico utilizzeresti per mantenere una delle due configurazioni?

*Schema da utilizzare*: SC9.

*Informazioni preliminari*: le entità coinvolte sono cheese\_cp\_1\_nl0\_ce0 e indicate\_vp\_0\_nl0\_ce1.

*Elementi da selezionare*: Nello schema target, l'entità cheese\_cp\_1\_nl0\_ce0 dello schema source, è descritta dalle due entità cheese\_cp\_1\_nl0\_ce0 e indicate\_vp\_0\_nl0\_ce1 relazionate da cheese\_cp\_1\_nl0\_ce0/indicate\_vp\_0\_nl0\_ce1.

*Task 3.1:* Far sì che anche nel target sia mantenuto l'attributo semplice.

*Task 3.2*: Far sì che anche nel source l'attributo semplice sia scomposto in due entità come accade nel target.

#### Come preparare i moduli per la raccolta dati

#### Cosa fare prima dell’osservazione: il test pilota

1. Compilazione questionario pre-esperimento, per la raccolta dei dati statistici come le conoscenze preliminari;
2. Lettura e comprensione delle slide che racchiudono le conoscenze preliminari alla comprensione degli operatori;
3. Spiegazione dello svolgimento delle successive fasi al candidato;

Le successive fasi verranno svolte unicamente dal candidato, senza nessun supporto dove possibile.

1. Selezione di due elementi casuali in modo da visualizzare la lista degli operatori iconici;
2. Individuazione dell’operatore iconico appropriato in base alla situazione spiegata;
3. Selezione degli elementi coinvolti indicati dai facilitatori;
4. Selezione dell’operatore appropriato;
5. Selezione della configurazione giusta in base all’operazione indicata dai facilitatori;
6. Completamento procedura aggiuntiva, ove prevista;
7. Utilizzo dei tasti relativi all’eliminazione dell’operatore, visualizzazione degli elementi coinvolti e modifica della configurazione, dove richiesto.

# Indice delle tabelle

[Tabella 1: Tabella relativa agli schemi nella fase iniziale dell'utilizzo di CoDIT. 14](#_Toc21083382)

[Tabella 2: Specifiche dei pc. 23](#_Toc21083383)

[Tabella 3: Schemi che, convertiti in xmi, sono risultati di dimensioni superiori al massimo consentito. 24](#_Toc21083384)

[Tabella 4: Tempi relativi alla generazione di schemi su iBench. 24](#_Toc21083385)

[Tabella 5: Schemi utilizzati per lo studio di usabilità. 24](#_Toc21083386)

[Tabella 6: Tempi ottenuti con il PC1. 25](#_Toc21083387)

[Tabella 7: Tempi ottenuti con il PC2. 25](#_Toc21083388)

[Tabella 8: Tempi ottenuti con il PC3. 25](#_Toc21083389)

[Tabella 9: Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC1. 26](#_Toc21083390)

[Tabella 10: Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC2. 27](#_Toc21083391)

[Tabella 11: Tempi relativi alle tre fasi e successiva visualizzazione su CoDIT ottenuti con il PC3. 27](#_Toc21083392)

1. Per introdurre variabilità nella costruzione delle relazioni, è stata introdotta la variabilità del numero di attributi. Durante i test è stata utilizzata una variabilità di 3 attributi. [↑](#footnote-ref-1)
2. La generazione dello schema SC3 non è stata possibile per il PC2 poiché durante la generazione dei file csv veniva saturata la RAM. [↑](#footnote-ref-2)
3. Per introdurre variabilità nella costruzione delle relazioni, è stata introdotta la variabilità del numero di attributi. Durante i test è stata utilizzata una variabilità di 2 attributi. [↑](#footnote-ref-3)