**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**

**Docente Candidati**

Francesca Tassatone 0522500568

Roberta Gesumaria 0522500569

Daniele Lupo

0522500579

Prof. Giuseppe Polese

Dott.ssa Loredana Caruccio

**Anno accademico 2018-2019**

Sommario

[Introduzione 3](#_Toc19287133)

[IBench 4](#_Toc19287134)

[Installazione 4](#_Toc19287135)

[Problemi riscontrati 4](#_Toc19287136)

[Utilizzo 4](#_Toc19287137)

[CoDIT 5](#_Toc19287138)

[Installazione 5](#_Toc19287139)

[Problemi riscontrati 6](#_Toc19287140)

[Struttura XMI riconosciuta da CoDIT 7](#_Toc19287141)

[Sezione iniziale 7](#_Toc19287142)

[Sezione entità e attributi 7](#_Toc19287143)

[Sezione relazioni 8](#_Toc19287144)

[Sezione finale 8](#_Toc19287145)

[Coordinate elementi 9](#_Toc19287146)

[Connessione a MySQL e creazione database 10](#_Toc19287147)

[Parser 11](#_Toc19287148)

[Sezione iniziale 11](#_Toc19287149)

[Sezione entità e attributi 11](#_Toc19287150)

[Sezione relazioni 12](#_Toc19287151)

[Sezione finale 13](#_Toc19287152)

[Studio di usabilità 14](#_Toc19287153)

# Introduzione

Il progetto svolto riguarda lo studio di usabilità sull’espressività e la complessità percepita degli operatori iconici di CoDIT utilizzando scenari che coinvolgano 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, abbiamo però bisogno di schemi più grandi di quelli che si hanno a disposizione, per ottenere ciò è stato 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 è stato necessario convertire lo schema, da logico a concettuale, e ottenere una strutturara adeguata in maniera da rendere il file comprensibile al tool.

Successivamente, è stato 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. Può essere utilizzato per 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.

L'utente fornisce un file di configurazione che specifica quali tipi di metadati e dati generare (es. schemi, dati, vincoli, mappature, ...) e 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 andata senza problemi.

Codice iBench.bat non funzionante:

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

Codice iBench.bat funzionante:

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

## 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. Per i dati, però, è messa a disposizione una sezione in cui è possibile definire alcune tipologie particolari (dati UDP), come e-mail, numero di telefono ed inoltre è possibile sceglierli da un set di dati fornito tramite file csv.

# 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, 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 corrispondenza 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.

Le funzionalità offerte da CoDIT sono:

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



Figura 1: 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.

## 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

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.

## 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 2: 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 3: 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 4: 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 5: 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 6: 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, ciò è 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 7: 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 8: 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 9: Struttura XML CoDIT, entità e attributi con coordinate.

# 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.



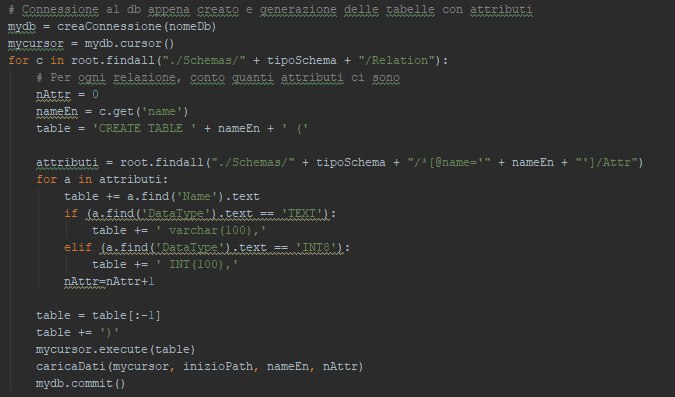
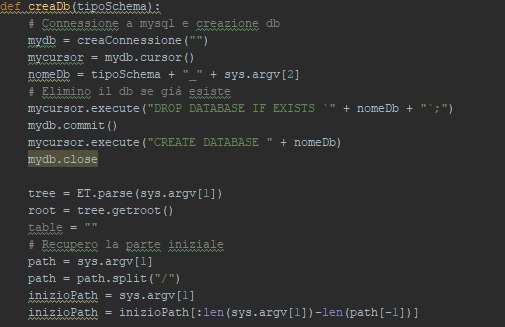
Di seguito verranno illustrati nel dettaglio i vari metodi implementati.

### 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.

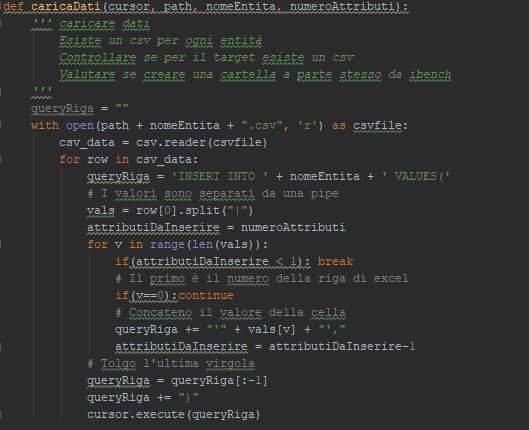
creaDb



Questo metodo di 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.

# 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 precedentemente. Per risolvere questo problema è stato creato un parser, che letto il file XML genera un nuovo file XMI con la struttura richiesta.

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.

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

## Sezione iniziale



## Sezione entità e attributi





## Sezione relazioni



## Sezione finale



# Statistiche di utilizzo

Per lo studio di usabilità sono stati generati diversi schemi con il tool *iBench*. Di seguito verranno descritte le caratteristiche di ognuno di essi.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **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 | 800 | 10 | 300 | 75 | 75 |

Per testare le tempistiche delle varie fasi per ottenere materiale atto allo studio di usabilità sono stati utilizzati diverse macchine, di seguito elencate.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **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 |

|  |  |  |  |
| --- | --- | --- | --- |
| **ID schema** | **Generazione schema iBench** | **Traduzione xml to xmi con parser** | **Creazione database e caricamento dati su MySQL** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

# Studio di usabilità

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)