

UNIVERSITÀ DEGLI STUDI DI MILANO
Facoltà di Scienze e Tecnologie
Corso di Laurea in Informatica

SPAZI-UNIMI:
PROGETTAZIONE E IMPLEMENTAZIONE
DELL'INTEGRAZIONE E VALIDAZIONE DELLE
DIVERSE FONTI DI DATI EDILIZI

Relatore: Prof. Carlo Bellettini

Correlatore: Dr. Matteo Camilli

Tesi di:
Paolo Venturi
Matricola: 775021

Anno Accademico 2013-2014

dedicato a ...

Prefazione

hkjafgyruet.

Organizzazione della tesi

La tesi è organizzata come segue:

- nel Capitolo 1

Ringraziamenti

asdjhgtry.

Indice

	ii
Prefazione	iii
Ringraziamenti	iv
1 Il progetto Spazi-Unimi	1
1.1 Introduzione al progetto	1
1.2 Il problema e i dati forniti per risolverlo	2
1.3 Scelta di tool, tecnologie e tecniche di sviluppo	3
1.4 Estrazione dei dati dai file DXF	3
1.5 Estrazione dei dati dai file CSV	3
1.6 Salvataggio dei dati in MongoDB	3
2 Integrazione e validazione delle diverse fonti di dati edilizi	4
2.1 Data Integration: teoria e possibili utilizzi	4
2.2 Analisi dei dati (da DXF e CSV)	4
2.3 Definizione di un sistema non dipendente dall'ordine di esecuzione . .	4
2.4 Strategie di merging dei dati adottate	4
2.5 Possibili miglioramenti	4
2.6 Reporting degli errori e delle criticità	4
2.7 Indirizzo ben formato: teoria e possibile utilizzo	4
2.8 Definizione di un DBAnalysis per avere statistiche specifiche sul Merge	4
3 Considerazioni finali e presentazione dei risultati	5
3.1 Definizione di API e prototipo App	5
3.2 Statistiche sui tempi di calcolo e di risposta del DB	5
3.3 Considerazioni sullo sviluppo del progetto	5
3.4 Miglioramenti/crescita dal lato personale	5

Capitolo 1

Il progetto Spazi-Unimi

1.1 Introduzione al progetto

Il progetto Spazi-Unimi nasce dall'esigenza degli utenti (studenti, professori, etc.) dell'Università degli Studi di Milano di cercare in modo facile e veloce la posizione delle aule di loro interesse. Vista la dislocazione delle sedi universitarie in varie aree della città (e della regione) un nuovo studente o un visitatore può avere serie difficoltà nell'orientarsi: da qui l'idea di creare una App che semplifichi la ricerca degli edifici universitari e delle loro stanze.

Spazi-Unimi è stato ideato nell'ambito del progetto Campus Sostenibile, una collaborazione tra il Politecnico di Milano e l'Università degli Studi di Milano, che si propone di trasformare il quartiere Città Studi in un modello per quanto riguarda la qualità della vita e la sostenibilità. Dalla proposta del Prof. Carlo Bellettini è quindi partito lo sviluppo del progetto che è stato portato avanti con altri due studenti del dipartimento di Informatica: Samuel Brandao Gomes e Diego Costantino.

I file da cui estrarre le informazioni utili alla creazione dell'applicazione sono stati forniti da due diverse fonti:

- la Divisione Manutenzione edilizia e impiantistica che ha concesso le piantine delle sedi universitarie e le informazioni sulle aule didattiche;
- la Divisione sistemi informativi che ha concesso le informazioni sulle aule presenti sul sistema EasyRoom.

Durante le 18 settimane di stage interno il lavoro effettuato ha riguardato principalmente lo sviluppo della parte back-end che si propone di fornire agli addetti delle diverse fonti di dati un modo semplice e immediato per aggiornare le informazioni. La parte su cui più si è incentrato il mio lavoro è stata l'unione dei dati provenienti dalle diverse fonti cercando di rendere disponibili all'utente finale le migliori informazioni per quanto riguarda completezza e qualità.

Nell'ultima parte dello stage invece ci si è concentrati sulla definizione di un'interfaccia REST API utile alle necessità della futura applicazione multiplatforma scaricabile dagli utenti dell'università.

1.2 Il problema e i dati forniti per risolverlo

La prima attività svolta è stato uno studio di fattibilità: sapendo che sul mercato non era presente nessuna applicazione/tool simile a quella che si voleva sviluppare ci si è concentrati più sulla ricerca di possibili librerie utili all'analisi dei file forniti dalle varie fonti. Sia l'edilizia che i sistemi informativi hanno fornito dati sulle aule organizzati in fogli elettronici e scritti in formato XLS, per quanto riguarda le mappe messe a disposizione dall'edilizia invece le informazioni sono su file di tipo AutoCAD DWG.

I file XLS essendo per loro natura in formato tabellare risultano di non difficile lettura ma ancora più semplice risulta quella dei file CSV (Comma-separated values) un formato basato su file di testo ricavabile senza sforzo da fogli elettronici o da database.

Il formato AutoCAD DWG invece è risultato molto più complicato da analizzare in quanto risulta essere un file binario diviso in diverse sezioni la cui codifica è molto complessa. Vista l'impossibilità di ottenere dati in modo semplice si è cercato un formato più adatto ai nostri scopi in cui esportare la collezione di 606 file DWG forniti. La scelta è ricaduta sull'altro formato AutoCAD cioè il DXF: questo standard utilizza un file ASCII diviso in sezioni (HEADER, CLASSES, TABLES, ENTITIES, OBJECTS, THUMBNAILEDIMAGE ed END OF FILE) risultando abbastanza leggibile a chiunque. La sezione di maggior interesse per i nostri scopi è risultata ENTITIES che contiene tutti gli oggetti disegnati nel file con le loro caratteristiche.

Dimostrata la fattibilità del progetto partendo da questi formati di dati ci si è concentrati sulla ricerca degli scenari d'uso per l'applicazione:

- trovare le sedi universitarie vicine alla propria posizione;
- trovare le stanze di una certa categoria (biblioteche, aule, punti ristoro, etc...) più vicine;
- cercare le stanze per nome mostrando una lista in caso di ambiguità;
- mostrare le mappe interne degli edifici rendendole interattive;
- segnalare errori e problematiche con un apposito form in modo da rendere le informazioni disponibili sempre più corrette e affidabili.

1.3 Scelta di tool, tecnologie e tecniche di sviluppo

1.4 Estrazione dei dati dai file DXF

1.5 Estrazione dei dati dai file CSV

1.6 Salvataggio dei dati in MongoDB

Capitolo 2

Integrazione e validazione delle diverse fonti di dati edilizi

- 2.1 Data Integration: teoria e possibili utilizzi
- 2.2 Analisi dei dati (da DXF e CSV)
- 2.3 Definizione di un sistema non dipendente dall'ordine di esecuzione
- 2.4 Strategie di merging dei dati adottate
- 2.5 Possibili miglioramenti
- 2.6 Reporting degli errori e delle criticità
- 2.7 Indirizzo ben formato: teoria e possibile utilizzo
- 2.8 Definizione di un DBAnalysis per avere statistiche specifiche sul Merge

Capitolo 3

Considerazioni finali e presentazione dei risultati

- 3.1 Definizione di API
- 3.2 Statistiche sui tempi di calcolo e di risposta del DB
- 3.3 Considerazioni sullo sviluppo del progetto
- 3.4 Miglioramenti/crescita dal lato personale

Bibliografia

- [1] M. Gotti, I linguaggi specialistici, Firenze, La Nuova Italia, 1991.
- [2] R. Wellek, A. Warren, Theory of Literature , 3rd edition, New York, Harcourt, 1962.
- [3] A. Canziani et al., Come comunica il teatro: dal testo alla scena. Milano, Il Formichiere, 1978.
- [4] Ministry of Defence, Great Britain, Author and Subject Catalogues of the Naval Library, London, Ministry of Defence, HMSO, 1967.
- [5] H. Heine, Pensieri e ghiribizzi. A cura di A. Meozzi. Lanciano, Carabba, 1923.
- [6] L. Basso, “Capitalismo monopolistico e strategia operaia”, Problemi del socialismo, vol. 8, n. 5, pp. 585-612, 1962.
- [7] L. Avirovic, J. Dodds (a cura di), Atti del Convegno internazionale Umberto Eco, Claudio Magris. Autori e traduttori a confronto (Trieste, 27-28 novembre 1989), Udine, Campanotto, 1993.
- [8] E.L. Gans, The Discovery of Illusion: Flaubert’s Early Works, 1835-1837, unpublished Ph.D. Dissertation, Johns Hopkins University, 1967.
- [9] R. Harrison, Bibliography of planned languages (excluding Esperanto). <http://www.vor.nu/langlab/bibliog.html>, 1992, agg. 1997.