

Servizi innovativi *Web-GIS*: impiego di *FuLL (Fuzzy Logic and Language)* per l'accesso in linguaggio naturale ai *DB* geografici

Maurizio BOMBARA (*), Davide CALI' (*), Ivana CALI' (*), Giuseppe TROPEA (*)

(*) BC S.r.l. Software Company, Via Muscatello 19/21 – 95125 Catania, Tel. +39 0957286481, Fax +39 095444199,
E-mail info@bcsoftware.it

Riassunto

Nell'ambito dei sistemi avanzati di *e-government* ed *e-democracy*, una sfida tecnologica attuale è quella di rendere l'usabilità dell'interfaccia delle applicazioni *Web-GIS*, ed in genere dei *DBMS*, quanto più immediata e naturale possibile, senza dover richiedere all'utente cittadino/impresa competenze informatiche specifiche.

L'innovazione della tecnologia *FuLL (Fuzzy Logic and Language)* consiste nella concezione di un motore di ricerca intelligente che, impiegato come componente di interfaccia utente integrata nei motori *Web-GIS*, permetta di consultare in linguaggio naturale le banche dati geografiche, rendendo assolutamente *user-friendly* le diverse applicazioni. *FuLL* nasce da un progetto di ricerca e sviluppo della *Software Company BC S.r.l.*, finanziato dal Ministero delle Attività Produttive, nell'ambito del P.I.A. Innovazione; i partner scientifici sono gli istituti I.L.C. e I.S.T.I. del C.N.R. ed i dipartimenti di Linguistica e Informatica dell'università di Pisa.

I primi risultati di *FuLL* sono relativi alla sperimentazione di dimostratore SW prototipale, condotta in collaborazione con le Amministrazioni Provinciali di Bologna e di Catania.

Abstract

Within the advanced systems of e-government and e-democracy, a current technological challenge is that of making the usability of interfaces to Web-GIS applications, and to general-purpose DBMS also, as quick and natural as possible, so that no specific computer abilities from users, citizens or enterprises, are required.

The innovation of FuLL (Fuzzy Logic and Language) technology lies in an intelligent search engine that, integrated besides the usual Web-GIS applications' interface, allows users to query the geographical data banks in natural language, making the different applications completely user-friendly. FuLL has been started as a research and development project of BC S.r.l. Software Company, financed by the Ministry of Productive Activities, within the P.I.A. Innovation; project scientific partners are I.L.C. and I.S.T.I institutes of C.N.R. and Linguistics and Information Technology departments of Pisa University.

First results of FuLL have been acquired through a SW prototype tested in collaboration with Bologna and Catania Provincial Administrations.

Introduzione

La rapidità dell'evoluzione e della diffusione delle tecnologie, avvenuta in questi ultimi anni, ha mutato in modo significativo i processi di sviluppo e di governo del Paese, rendendo necessaria l'attuazione di sistemi avanzati di *e-government* e *e-democracy*, in cui il cittadino-utente, tramite la comunicazione *Web*, si avvicini sempre di più alle istituzioni in un rapporto di interazione e cooperazione. L'obiettivo da perseguire, affinché il sistema democratico funzioni, è quello di costruire una società dell'informazione fruibile e accessibile a tutti, in cui attraverso tecnologie innovative dell'informazione e della comunicazione la Pubblica Amministrazione possa offrire servizi sempre migliori, più efficaci ed efficienti, ai cittadini e alle imprese.

Nell'ambito dei servizi informativi territoriali (*SIT*), la tecnologia *Web-GIS* (*Geographical Information System*), in corso di maturazione in questi anni, rappresenta un valido strumento di supporto poiché permette ad un vasto pubblico di ottenere informazioni strutturate e chiare, accedendo via *Web* alle banche dati geografiche *SIT-GIS* degli enti gestori.

Una sfida tecnologica attuale è quella di rendere l'usabilità dell'interfaccia delle applicazioni *Web-GIS*, ed in genere dei *DBMS* (*Database Management System*), quanto più immediata e naturale possibile, senza dover richiedere all'utente competenze informatiche specifiche.

Lo strumento ideato è pienamente in linea con gli sforzi attuali della Commissione Europea di incentivare lo sviluppo di strumenti per la fruizione *cross-border* delle basi dati disponibili in area *GIS*. La concentrazione maggiore è stata finora nella realizzazione di metalinguaggi e/o *middleware* che consentano l'accesso al dato in maniera trasparente su strutture dati diverse tra loro. L'approccio *FuLL* si astrae dalla struttura dei dati ed offre interfaccia unica privilegiando in questo modo l'utente finale, il cittadino Europeo (*digital divide*, *e-democracy*).

L'innovazione tecnologica di FuLL

L'innovazione della tecnologia *FuLL* (*Fuzzy Logic and Language*) consiste nella concezione di un motore di ricerca intelligente che faciliti l'utente nell'acquisizione di informazioni contenute in banche dati strutturate, traducendo ed elaborando le domande poste dall'utente in linguaggio naturale tramite tecniche fuzzy ed un approccio orientato al dialogo.

L'impiego di *FuLL* come componente di interfaccia utente integrata nei motori *Web-GIS* permette di consultare in linguaggio naturale le banche dati geografiche, rendendo assolutamente user-friendly le diverse applicazioni.

FuLL è un progetto di ricerca e sviluppo della *Software Company BC S.r.l.*, finanziato dal Ministero delle Attività Produttive nell'ambito del P.I.A. Innovazione, a cui collaborano istituti del CNR (ILC, ISTI) e dipartimenti universitari di Pisa (Linguistica, Informatica), aventi competenze altamente specialistiche nel settore. La tecnologia alla base di *FuLL* mira ad essere in prima posizione nel mercato attuale per ciò che concerne i sistemi *NLIDB* (*Natural Language Interfaces to Databases*). Gli elementi innovativi e le caratteristiche tecniche principali di *FuLL* sono:

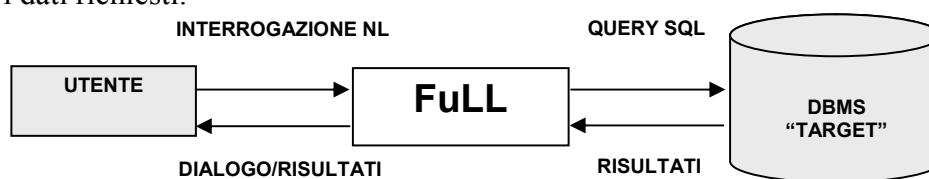
- 1. Interfaccia in linguaggio naturale:** per consentire all'utente l'utilizzo del linguaggio naturale senza dover conoscere linguaggi formali specifici per l'interrogazione di basi di dati. Il motore NLP (*Natural Language Processing*) di analisi e interpretazione semantica di testo si basa su di un *parser* flessibile e capace di sfruttare una conoscenza strutturata sul dominio di interesse per risolvere le ambiguità ed identificare i contenuti significativi della richiesta; è stata progettata una analisi sintattica e semantica sottospecificata, tale da gestire una comprensione parziale anche nel caso in cui alcuni elementi della frase non siano stati colti perfettamente. Gli stadi di analisi sintattica e semantica sono robusti, ad esempio non interamente tarati sulla comprensione di semplici forme interrogative, bensì supportati da dizionari, regole morfologiche e sintattiche flessibili, e da una rete semantica costruita sì sul dominio ma facilmente estendibile. Le strutture dati che intercettano la conoscenza del sistema sul mondo (ed anche le strutture sintattiche) si adattano ai differenti contesti senza essere riprogettate. *FuLL* è quindi in grado di agganciarsi a questa o quella base di dati ("*DB Target*") in maniera indolore quando queste ricadano nello stesso dominio linguistico/applicativo. Ciò garantisce la possibilità di personalizzare *FuLL* per contesti e domini applicativi anche diversi da quello *SIT-GIS*.
- 2. Tecniche Fuzzy per l'interpretazione linguistica:** per integrare e supportare l'analisi semantica delle richieste poste dagli utenti, garantendo una maggiore flessibilità di linguaggio e consentendo al sistema di rappresentare e comprendere aspetti soggettivi e qualitativi. L'impiego di tecniche basate sulla logica *fuzzy* facilita la corrispondenza fra richieste specificate in maniera vaga e imprecisa (caratteristica comune della maggior parte delle interrogazioni che si formulano usando il linguaggio naturale) e valori codificati in maniera numericamente precisa (caratteristica comune della maggior parte dei dati strutturati raccolti in un *DB* di interesse). Utilizzando gli insiemi *fuzzy* per definire l'appartenenza "sfumata" di un attributo ad un certo

intervallo di valori numerici, è possibile interpretare in maniera più vicina all'intuito umano alcune richieste numeriche; in tal modo è anche possibile combinare insieme più attributi in modo da permettere un ordinamento dei risultati ottenuti (*ranking*) che sia più rispondente a quanto chiesto. L'approccio di *FuLL* basato sulla *fuzzy logic* riguarda anche tecniche di valutazione fuzzy dei quantificatori linguistici, affiancate alla segmentazione della *query* in sub-condizioni e alla successiva aggregazione pesata delle sub-condizioni; tecniche di questo tipo consentono di tenere correttamente conto dell'effetto di connettivi logici più sfumati rispetto ai classici AND, OR, NOT booleani, superandone limitazioni ed aspetti contro-intuitivi (nell'utilizzo) per gli utenti umani.

3. **Interazione dialogica con l'utente:** per ridurre l'impersonalità dell'interfaccia utente e permettere al sistema di risolvere casi di ambiguità (tipici del linguaggio naturale) e di richieste incomplete, interagendo con l'utente a vari livelli.
4. **Dominio applicativo limitato e specifico:** per semplificare la risoluzione delle ambiguità e difficoltà di interpretazione legate al contesto (*word sense disambiguation*) eventualmente presenti nelle interrogazioni poste dall'utente. La gestione della conoscenza sul mondo nel dominio d'interesse, a supporto dell'analisi semantica, viene in *FuLL* strutturata tramite una architettura a più livelli, partendo da un livello più astratto sino ad uno schema logico a più basso livello, ovvero a ridosso dei dati stessi del database.
5. **Modellazione di entità e relazioni spaziali fra i dati:** per la possibilità di interrogare gli *Spatial Databases* dei motori *GIS*, in cui i dati sono relativi ad entità georeferenziate, e spesso sono organizzati in "strati informativi" (*layers*) e relazionati tramite operatori spaziali, come ad esempio *distance*, *overlay*, *buffering*, *intersect*, ed altri. Per estrarre i dati da questo tipo di *DB*, *FuLL* utilizza un formalismo di rappresentazione della conoscenza sul dominio ideato ad hoc per rappresentare e gestire correttamente le informazioni spaziali, sia a livello semantico che a livello di linguaggio di *query SQL* a basso livello, conformandosi poi agli standard di *SQL* spaziale (come *OpenGIS*) che si vanno affermando sul mercato.
6. **Architettura orientata all'integrazione in ambienti operativi eterogenei:** per avere la possibilità di includere il software sviluppato in architetture complesse, anche preesistenti. La progettazione modulare di *FuLL* permette di adattare la tecnologia a vari ambiti di interfaccia.

Interfaccia Utente-FuLL

Dal punto di vista tecnico *FuLL* è un software in grado di convertire una richiesta formulata in linguaggio naturale in una *query* formulata in *SQL*, da inviare ad un "DBMS Target" già esistente, per estrarre i dati richiesti.



Utente: Quali sono i centri commerciali molto grandi nei pressi della stazione di Bologna?

FuLL: Intendi la stazione ferroviaria o la stazione degli autobus?

Utente: La stazione ferroviaria

FuLL: I centri commerciali che meglio rispondono alla tua richiesta sono:

		<i>rank</i>
1	Centro commerciale "Il Mondo"	100%
2	Area commerciale Convenienza	75%
3	Pianeta Shopping	70%

Figura 1 – Esempio di interfaccia Utente-FuLL

Nell'esempio riportato è possibile notare che:

- espressioni vaghe e imprecise ("nei pressi di", "molto grandi") sono comprese dal sistema e interpretate in modo sfumato secondo la logica *fuzzy*;

- il termine “stazione” viene interpretato automaticamente come “stazione ferroviaria” o “stazione degli autobus” riconoscendo la presenza di un’ambiguità lessicale e risolvendola per mezzo del dialogo;
- i risultati vengono presentati all’utente interpretando la richiesta e ordinando i dati in base ad un *ranking* che con tecniche *fuzzy* tiene conto del grado (*rank*) di rispondenza dei risultati alla richiesta. Nell’esempio il centro commerciale “Il Mondo” risulta il più rispondente alla richiesta rispetto ai vincoli richiesti di vicinanza alla stazione e grande dimensione, combinati assieme secondo un AND *fuzzy*.

Architettura funzionale

Per ottenere il massimo dalle tecnologie e dalle ricerche consolidate disponibili in ogni campo di interesse del progetto, il sistema *FuLL* è costituito da una cascata di moduli che in sequenza operano su risultati intermedi dei moduli a monte e fanno uso di strutture dati sviluppate specificatamente per contenere le informazioni sintattiche e semantiche del dominio applicativo di riferimento e del database da interrogare, denominato “*DB Target*”. Nella figura seguente è schematizzata l’architettura funzionale di *FuLL*.

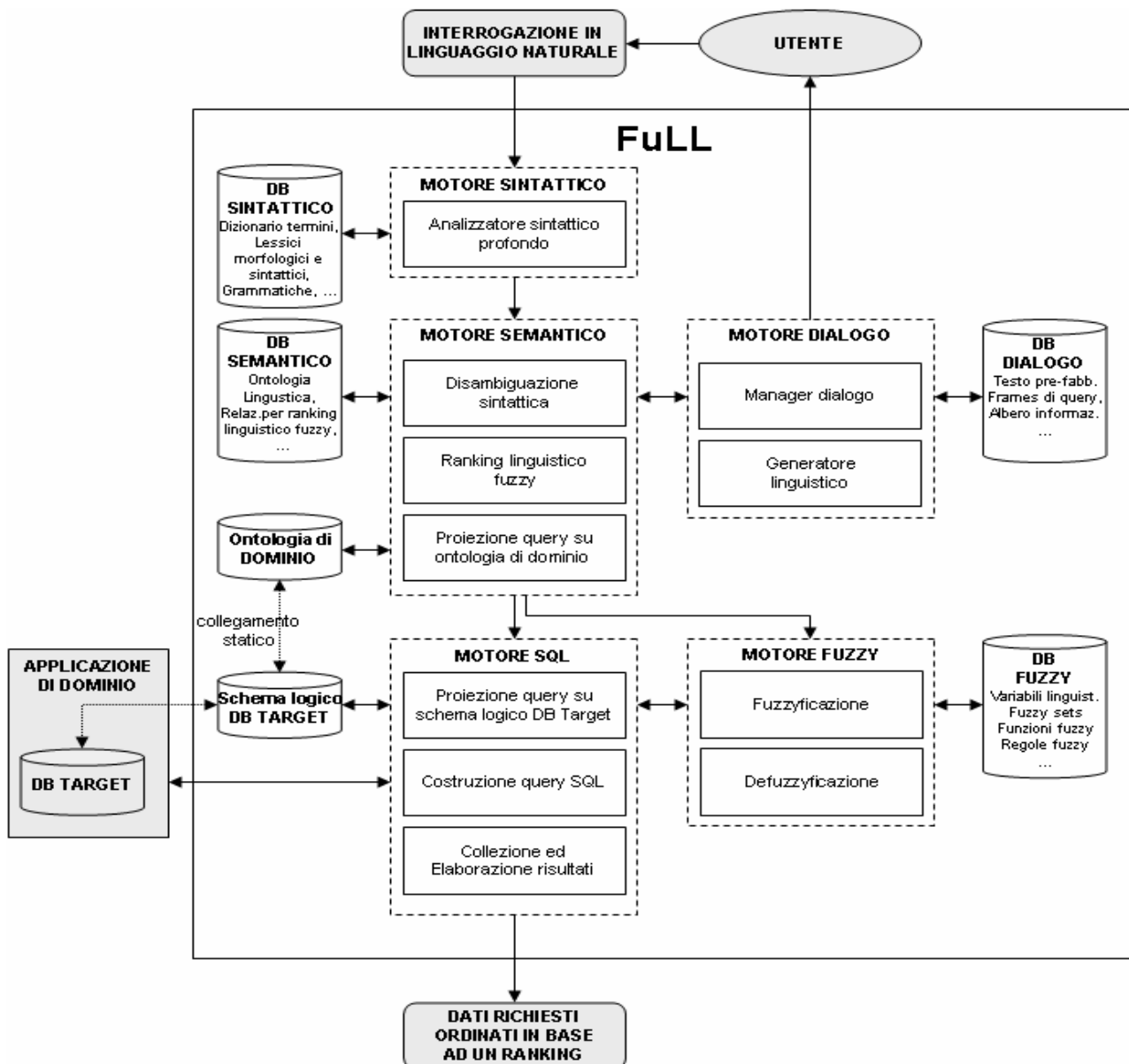


Figura 2 – Architettura funzionale di *FuLL*

Risultati

I primi risultati del progetto *FuLL* sono relativi alla sperimentazione di un dimostratore SW prototipale, specializzato per interrogazioni in linguaggio naturale verso banche dati geografiche contenenti informazioni sui temi “mobilità e trasporti” e “sistema insediativo e territoriale”.

Alla sperimentazione del prototipo partecipano le Amministrazioni Provinciali di Bologna e di Catania, interessati all'erogazione di servizi innovativi di consultazione di dati territoriali a cittadini e imprese a favore della fruizione e dello sviluppo del territorio, che collaborano in qualità di “esperti di dominio” e di “utenza target” e mettendo a disposizione database dei *SIT-GIS* di cui sono dotati.

Di seguito è riportato un esempio di processamento della *query*, in cui a partire dalla richiesta in linguaggio naturale posta dall'utente sono rappresentate la forma logica intermedia in uscita dal Motore Semantico e la forma finale in *SQL* con cui *FuLL* accede al *DB* per selezionare e presentare i risultati all'utente dopo che questi siano stati elaborati ed interpretati dal Motore *Fuzzy*.

La forma logica viene processata dai moduli di *FuLL* ed in particolare vengono estratte le informazioni intermedie *fuzzy* per il Motore *Fuzzy*.

Query in linguaggio naturale

Quali sono i centri commerciali molto grandi nei pressi della stazione ferroviaria di Bologna?

Forma logica intermedia

```
?(x1): ep1: centroCommerciale (X1) AND
      ep2: stazioneTreni (X2) AND
      ep3: comune (X3) AND
      ep4: nomeComune (X3, 'bologna') AND
      ep5: locStazioneComune (X2, X3) AND
      ep6: distanza (X1, X2, d) AND
      ep7: EQ (d, 2, default, default) AND
      ep8: dimensione (X1, c) AND
      ep9: EQ (c, 3, default, MUCH)
```

Informazioni intermedie Fuzzy

```
focusInfo = [{"x1", {"id_centroCommerciale"}}]
fuzzyInfo = FUZZY { EQ(d, 2, default, default), EQ(c, 3, default, MUCH) }
```

Forma SQL

```
SELECT Distance (R1.the_geom, R2.the_geom), Area (R1.area), fuzzy_rank, *
FROM FuzzyTable,
(SELECT * FROM CentroCommercialeView) AS R1,
(SELECT * FROM StazioneTreniView) AS R2,
(SELECT * FROM ComuneView) AS R3,
(SELECT * FROM NomeComuneView WHERE nomeComune = "bologna") AS R4,
(SELECT * FROM LocStazioneComuneView) AS R5
WHERE (R2.id_stazioneTreni = R5.id_locStazioneComune) AND
      (R3.id_comune = R4.id_comune) AND
      (fuzzy_rank > 70) AND
ORDER BY fuzzy_rank DESC
```

Figura 3 – Esempio di processamento della query

Nell'esempio si vede come, attraverso l'analisi semantica, si imponga che l'attributo “Distanza” assuma un valore 2 (in una scala di valori da 1 a 5), valore numerico che viene preventivamente fatto corrispondere con il valore linguistico “PRESSO” e tutti i suoi sinonimi. Allo stesso modo l'attributo “Area” deve assumere il valore “GRANDE” (che corrisponde a 3), al quale si applica in aggiunta il modificatore *MUCH*. E' il Motore *Fuzzy* che elabora queste informazioni e le combina assieme (facendo un *AND fuzzy* delle due condizioni) al fine di calcolare per ogni record il suo rank, ovvero il grado di rispondenza ai vincoli fuzzy imposti. Il Motore *Fuzzy* deposita i risultati nella

tabella *FuzzyTable* che si affianca alle altre, e l'insieme delle informazioni vengono poi composte nella *query SQL* finale e mostrate all'utente. Si noti come risultati che ottengono un voto (*rank*) inferiore al 70%, vengono scartati in quanto troppo lontani da quanto richiesto dall'utente.

In questa prima fase di prototipazione le strutture dati di *FuLL* relative al dominio applicativo in questione, anche se afferenti a due ambiti semantici distinti (“mobilità e trasporti” e “sistema insediativi territoriale”), sono state collegate a livello ontologico in strutture dati comuni, e ciò al fine di:

- tenere conto del fatto che i due ambiti scelti hanno a comune concetti e relazioni che un'utente “*Target*” può essere interessato a consultare contestualmente, come ad es. nei servizi *Web-GIS* erogati dalla pubblica amministrazione a favore dello sviluppo e della fruizione del territorio e rivolti a cittadini e imprese;
- verificare e sperimentare i limiti teorici delle dimensioni del dominio applicativo del prototipo in relazione a possibili ambiguità linguistiche e semantiche legate al contesto in cui l'utente esprime la richiesta.

Le successive sperimentazioni di *FuLL* riguarderanno l'estensione incrementale del dominio (al fine di valutarne i limiti teorici), la validazione dei modelli del prototipo su altri “*DB Target*” appartenenti allo stesso dominio, la specializzazione della tecnologia in altri possibili domini applicativi (logistica, ambiente, beni culturali, ...).

Partner e collaborazioni

I partner scientifici del progetto *FuLL* sono:

- Istituto di Linguistica Computazionale (I.L.C.) – C.N.R.
- Dipartimento di Linguistica - Università di Pisa
- Laboratorio KDD Lab (Knowledge Discovery and Delivery Lab): Istituto di Scienza e Tecnologie dell'Informazione (I.S.T.I.) - C.N.R. & Dipartimento di Informatica - Università di Pisa

La sperimentazione e la validazione del prototipo è condotta in collaborazione con:

- Area S.I.T. del Servizio Informatica e Sistemi Informativi (S.O.I.S.I.) - Provincia di Bologna
- 1° Servizio Pianificazione Territoriale e Trasporti del VII Dipartimento - Provincia di Catania

Riferimenti bibliografici

- [1] Androutsopoulos I., Ritchie G. D., Thanisch P. (1995), “Natural Language Interfaces to Databases - An Introduction”, *Journal of Natural Language Engineering*, vol.1, no.1, Cambridge University Press
- [2] Chu W., Meng F. (1999), “Database Query Formation from Natural Language using Semantic Modeling and Statistical Keyword Meaning Disambiguation”, *Technical Report 990003*, UCLA CS Dept., 16
- [3] Lenci A., Montemagni S., Pirrelli V., Soria C. (1999), “FAME: a Functional Annotation Meta-scheme for multi-modal and multi-lingual Parsing Evaluation”, *Proceedings of the ACL-IALL Workshop Computer-mediated Language Assessment and Evaluation in Natural Language Processing*, Maryland (USA)
- [4] Zadeh LA (1965), “Fuzzy sets”, *Information and Control*, Vol. 8, p. 338-353.
- [5] Burghard B. Rieger (1981), “Feasible Fuzzy Semantics. On some problems of how to handle word meaning empirically”, *Eikmeyer, H.J./Rieser, H. (eds.): Words, Worlds, and Contexts. New Approaches in Word Semantics (Research in Text Theory 6)*, Berlin/New York (W. de Gruyter), pp. 193-209.
- [6] Cappelli A., Giovannetti E., Michelassi P. (2004), “Ontological knowledge and language in modelling classical architectonic structures”, *Ontologies and Lexical Resources in Distributed Environment (Ontolex 2004, Lisboa, May 29th)*, LREC, pp. 85-89
- [7] Gruber, T. A (1992), “Translation Approach to Portable Ontology Specifications”, Knowledge Systems Laboratory, Stanford University, Stanford, CA