

# TECNOLOGIE OPEN-SOURCE NEL PROGETTO FULL (FUZZY LOGIC AND LANGUAGE)

M. Bombara<sup>a</sup>, D. Cali<sup>a</sup>, I. Cali<sup>a</sup>, V. Cartelli<sup>a</sup>, E. Giovannetti<sup>b</sup>, M. Masserotti<sup>b</sup>, C. Renso<sup>b</sup>, L. Spinanti<sup>b</sup>, G. Tropea<sup>a</sup>

<sup>a</sup> BC S.r.l. Software Company – Via Caronda 136, 95128 Catania - [www.bcsoftware.it](http://www.bcsoftware.it) - e-mail: [info@bcsoftware.it](mailto:info@bcsoftware.it)

<sup>b</sup> KDDLAB ISTI CNR - Via Moruzzi, 1 - 56010 Pisa - e-mail: {emiliano.giovannetti, mavie.masserotti, chiara.renso, laura.spinanti}@isti.cnr.it

**KEY WORDS:** GIS, standardizzazione, interoperabilità, linguaggio naturale.

## RIASSUNTO

*FuLL (Fuzzy Logic and Language)* è un progetto di ricerca della Software Company BC, finanziato dal MAP e sviluppato in collaborazione con partner di ricerca del CNR e universitari, che aveva come obiettivo lo sviluppo di una tecnologia software che innova i sistemi di interrogazione dei database tramite l'utilizzo di un'interfaccia in linguaggio naturale. Nell'ambito del progetto, l'impiego di strumenti e tecnologie open-source ha consentito la sperimentazione del prototipo in tempi brevi e senza legare il sistema ad una specifica tecnologia proprietaria.

L'interfaccia in linguaggio naturale FuLL, alla base del progetto, è stata sperimentata collegandola a due distinti database GIS di due province italiane, Bologna e Catania. I dati originali sono stati importati all'interno di un DB PostgreSQL con estensioni PostGIS e collegati fra loro con l'impiego di funzioni spaziali e relazioni di natura semantica, per consentirne la consultazione in linguaggio naturale. Per visualizzare i dati geografici su web è stata integrata una web-application GIS, sviluppata con MapServer in tecnologia JSP e pienamente compatibile con PostGIS. Un risultato importante della scelta di tecnologie standard, come le estensioni OpenGIS spaziali di PostgreSQL, è la piena portabilità verso altri sistemi DBMS spaziali.

## ABSTRACT

*FuLL (Fuzzy Logic and Language)* is a research and development project of BC Software Company, founded by Italian MAP and developed together with University and CNR partners. The aim of the project is to innovate the software technology of NLIDB (Natural Language Interfaces to Databases). By using open-source tools and technologies, an easy and fast testing of the prototype has been possible, together with the added benefit of detaching the system from any proprietary solutions.

The natural language interface, which stands at the core of project FuLL, has been tested by linking it with two different GIS databases holding data from two Italian Provinces, Bologna and Catania. Original data has been imported inside a PostgreSQL database, augmented with PostGIS extensions, and the various layers of information were linked by means of spatial functions and semantic relationships between the entities, so that natural language querying of the data is enabled. On the other hand, to display query results and to navigate the data, a GIS web-application has been developed, by using JSP technology and a MapServer engine fully compatible with PostGIS. By choosing standard technologies and approaches (as the OpenGIS-compliant extensions of PostgreSQL), the important goal of a full portability towards different spatial DBMS has been reached.

## INTRODUZIONE

FuLL (Fuzzy Logic and Language) è un progetto di ricerca finanziato dal Ministero delle Attività Produttive (misura PIA Innovazione), avente l'obiettivo lo sviluppare una tecnologia software che innova i sistemi di interrogazione dei database tramite l'utilizzo di un'interfaccia in linguaggio naturale (NLIDB Natural Language Interfaces to Databases).

Il progetto è stato ideato e realizzato dalla Software Company BC, in collaborazione con vari partner scientifici e sperimentali che si ringraziano, e si è concluso con successo nel giugno 2006.

I primi risultati del progetto FuLL sono relativi allo sviluppo di un prototipo software dimostrativo, specializzato per la consultazione in lingua italiana di banche dati geografiche di applicazioni SIT-GIS, che è attualmente in fase avanzata di industrializzazione.

L'innovazione tecnologica di FuLL consiste nella concezione di un motore di ricerca intelligente che facilita l'utente nella consultazione di informazioni contenute in banche dati strutturate, traducendo ed elaborando in modo automatico le richieste poste dall'utente in linguaggio naturale, in query complesse SQL, tramite tecniche fuzzy e con un approccio

orientato al dialogo. FuLL è scalabile e personalizzabile in varie lingue e contesti/domini applicativi, in base al DB Cliente ed al Target di utenza.

Dal punto di vista del progetto FuLL il SIT-GIS non è altro che un particolare schema di dati da interrogare, che rispetto alle basi dati tradizionali presenta l'aspetto della geo-referenziazione dei dati e la possibilità di richiamare funzioni di calcolo spaziali.

Fra gli obiettivi di progetto risultava fondamentale l'arrivare in tempi brevi ad un prototipo funzionante della tecnologia da poter sperimentare su banche dati reali, anche di dimensioni abbastanza estese.

A questo fine, sin da subito, si è cercato di individuare all'interno dell'architettura funzionale i moduli che avrebbero potuto essere mappati su tecnologie consolidate e già esistenti. In particolare tool e funzionalità per le quali certamente era proficuo sfruttare soluzioni esistenti riguardavano:

- gli standard sintattici da seguire per codificare l'ontologia di dominio (ovvero la conoscenza che FuLL ha del dominio di interesse) e gli strumenti per costruire e manipolare la stessa;
- il modulo di interfaccia grafica web per far comunicare l'utente con FuLL e mostrare i risultati;
- il DBMS deputato a gestire i dati di FuLL e dell'applicazione target, purché dotato di estensioni SQL spaziali.

In tutti questi ambiti si è deciso di procedere con tecnologie open-source disponibili immediatamente. Questa scelta ha consentito un risparmio in termini di risorse, di tempo e di sforzo realizzativo complessivo, consentendo di arrivare ad un prototipo altamente rifinito e garantendo inoltre la progettazione e lo sviluppo di moduli espandibili e strutturalmente compatibili anche con tecnologie proprietarie.

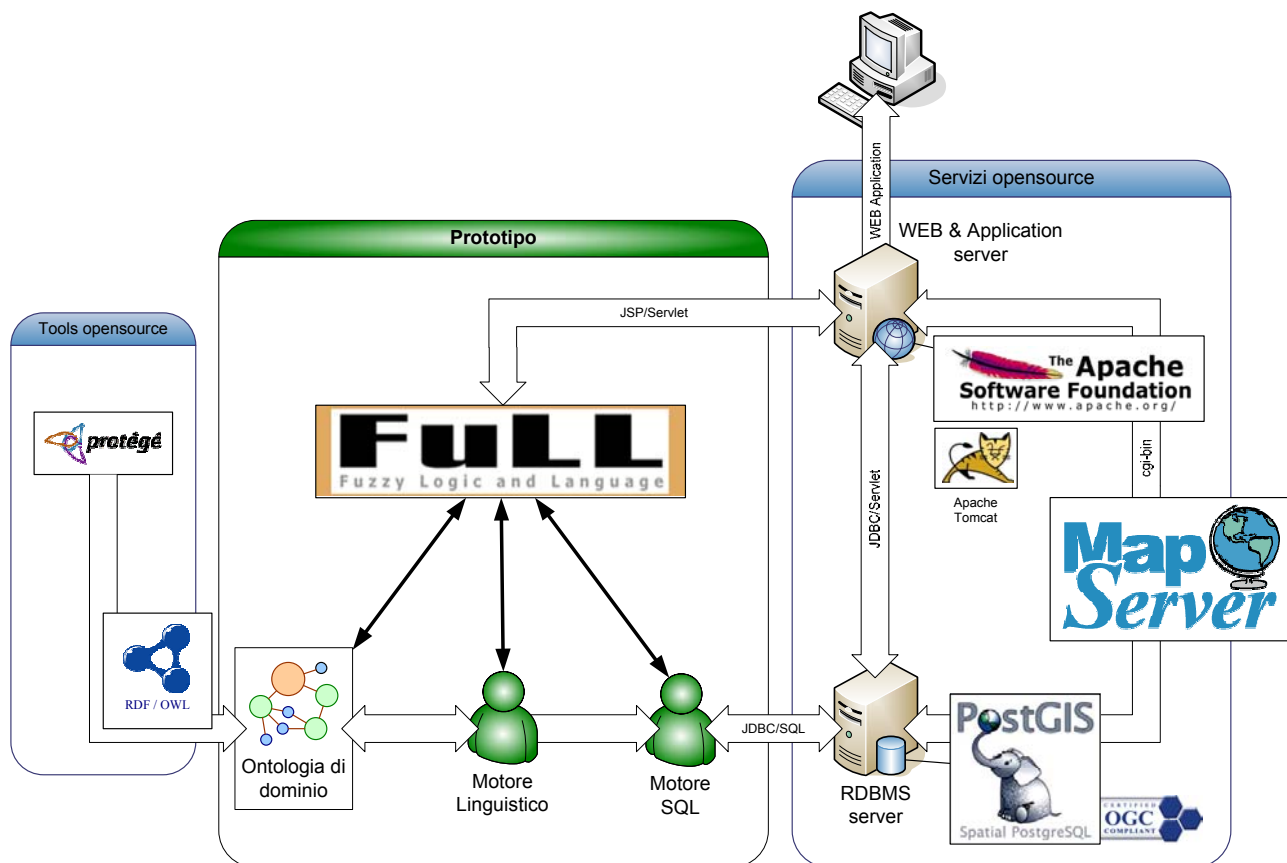


Figura 1 – Schema delle tecnologie opensource nel progetto FuLL

## SVILUPPO E GESTIONE DELL'ONTOLOGIA

L'ontologia di dominio nel progetto FuLL rappresenta il punto di collegamento tra l'interfaccia in linguaggio naturale e il database. Da un punto di vista progettuale si è tenuto conto di sviluppare un'ontologia strutturata in maniera tale da permettere la maggiore integrazione ed interoperabilità possibile di dati tra applicazioni. In questo senso, la ricerca di un linguaggio *sintatticamente* basato su XML e *semanticamente* sulla logica descrittiva ha condotto naturalmente alla scelta di OWL (Web Ontology Language) assecondando le raccomandazioni del consorzio W3C del 2004, anche in virtù della sua espressività e della disponibilità di risorse ad esso associate.

OWL permette una gamma di applicazioni descrittive come la gestione di portali Web, gestione di collezioni, ricerche basate sul contenuto, abilitando agenti intelligenti, servizi web e *ubiquitous computing*. Si basa su "RDF Model and Schema" e aggiunge un vocabolario più ampio per descrivere proprietà e classi: tra le altre, relazioni tra classi (ad esempio disgiunzione), cardinalità (ad esempio "esattamente uno"), uguaglianza, tipizzazione più ricca di proprietà, caratteristiche di proprietà (ad esempio simmetria) e classi enumerate. Tuttavia non supporta come primitivi dati ed operazioni spaziali: questo ha aperto interessanti spunti di ricerca nell'ambito del progetto, in quanto si è trattato di investigare i modi migliori per rendere possibile la rappresentazione dei concetti spaziali.

Da un punto di vista realizzativo, nel progetto sono stati distinti due domini o "classi" di applicazioni SIT, relativi ai temi "mobilità e trasporti" e "sistema insediativo e territoriale".

Protegè è una piattaforma free ed open source per la costruzione e la gestione di ontologie e basi di conoscenza, che fornisce varie funzionalità grafiche per l'editing e la visualizzazione. In particolare esiste un'estensione per il formato OWL (<http://protege.stanford.edu/plugins/owl/>) che è stata utilizzata per lo sviluppo dell'ontologia di dominio.

## GESTIONE DATI ED SQL SPAZIALE

Nel progetto FuLL la scelta realizzativa per la sperimentazione del prototipo riguardo allo schema logico del DB target è caduta su un DBMS spaziale, ovvero in grado di supportare un linguaggio di interrogazione SQL spaziale. Questo per rimanere quanto più coerenti con la filosofia del progetto di interrogazione in linguaggio naturale a basi di dati il più possibile generali. Per quanto motivo è stata esclusa nella prima fase una possibile connessione diretta ad un software GIS proprietario tramite interfacce proprietarie. Questa scelta permette di rendere tutte le metodologie sviluppate per FuLL indipendenti dal dominio applicativo specifico spaziale. La scelta importante è quella di basarsi su soluzioni standard, così da permettere di non legare il sistema ad una specifica tecnologia proprietaria.

Il software DBMS spaziale scelto per la prima sperimentazione è stato PostGIS (<http://www.postgis.org>), estensione GIS di Postgres (<http://www.postgresql.org>), un DBMS open source tra i più usati. Inoltre il linguaggio SQL di PostGIS rispetta gli standard OpenGIS (<http://www.opengis.org>) di SQL spatial, consentendo di rimanere nell'ambito di una sintassi SQL generale. Un altro motivo per la scelta di PostGIS è la compatibilità con Mapserver (utilizzato per l'interfaccia web-GIS) un software *open source* per la visualizzazione di dati geografici su web.

Risultato importante di queste scelte rivolte a tecnologie standard è la scalabilità e la portabilità verso altri sistemi DBMS spaziali. La portabilità può essere considerata piena verso altri Spatial DBMS basati su OpenGIS SQL. Mentre nel caso di SQL spaziali proprietari (come Spatial Oracle), il porting potrà avere un impatto sulla sintassi SQL e quindi sull'algoritmo di mapping che potrà essere leggermente modificato nella costruzione della sintassi SQL. Ritentiamo che a livello di funzionalità di base adottate in FuLL (distanza, overlay) la metodologia possa ritenersi assolutamente generale e che sia necessario un porting solo sintattico.

## INTERFACCIA WEB-GIS

All'interno del progetto è stata sviluppata anche una web application che si pone come interfaccia per facilitare la sperimentazione del prototipo e consente il suo facile utilizzo tramite un box per l'inserimento delle richieste ed una apposita sezione dedicata alla visualizzazione ed esplorazione dei risultati ottenuti. L'applicativo web è stato realizzato integrando diversi moduli software open source disponibili per sistema operativo Windows e Linux, come dettagliato nella Tabella 1.

Server GIS e layer cartografia	Mapserver (opensource)
Web server	Apache (opensource)
Integrazione Mapserver-Apache	tecnologia CGI-BIN
Pagine dinamiche	tecnologia JSP
Servlet container	Tomcat (opensource)
DBMS	Postgres + PostGIS (opensource)

Tabella 1. Tecnologie open source per l'interfaccia web del prototipo FuLL

Al fine di consentire la consultazione di dati geografici dall'applicazione si è scelto di utilizzare il modulo Mapserver (<http://mapserver.umn.edu>), integrato con Apache. Mapserver genera la cartografia in funzione dei layers attivati tramite CheckBox in legenda (sulla colonna sinistra della pagina) e consente la navigazione delle mappe tramite gli strumenti di zoom e pan; si tratta di un modulo open source che sfrutta la tecnologia CGI-BIN, quindi eseguibile tramite il WEB server Apache (open source), per generare una pagina HTML contenente tutte le immagini generate, nonché tutti i controlli di navigazione della mappa (show/hide di layers, zoom, pan, ...; vedi Figura 2).

Progetto di Ricerca e Sviluppo FuLL (Fuzzy Logic and Language) Prototipo Gamma

**IBC** SOFTWARE COMPANY In collaborazione con:  
**Provincia di Bologna**  
Progetti ed applicazioni di Area Vasta e Sit

**FULL**  
Fuzzy Logic and Language

**Query in linguaggio naturale :**  
dove sono i distributori di gpl vicinissimi all'aeroporto?

Reset Invia richiesta a FuLL

**Box di dialogo :**

**Feedback :**  
La risposta è soddisfacente?  
 Sì  No  
Se non è soddisfacente lascia un commento, grazie.

Invia Feedback


**Risultati di FuLL:**  
 Puntuale  Lineare  Poligonale

#Record Mostra tabella risultati

2

**Preview Tabella risultati**

fuzzyrank	gid	distance3	id_distributore	erogacarburante	nome_strada	civico_distributore
100	18	751.784799260394	0370060039	GPL	VIA PERSICETANA VECCHIA	25
62	1	2651.75023324546	0370060056	GPL	STRADA STATALE VIA EMILIA	205

Partner scientifici: 

info@bcsoftware.it  
www.bcsoftware.it

Figura 2 – Interfaccia utente di FuLL prototipo gamma (es. “dove sono i distributori di gpl vicinissimi all’aeroporto?”)

Tutte le funzionalità GIS dell’applicativo WEB sono quindi demandate a questo modulo che opera generando dinamicamente un Mapfile. Esso definisce i rapporti fra gli oggetti, indirizza MapServer all’individuazione dei dati e definisce come devono essere disegnate le entità.

Al primo accesso all’applicazione web, il file map usato contiene esclusivamente i collegamenti ai dati d’origine relativi al GIS da interrogare, tramite la definizione di oggetti di tipo “layer”, con le opportune sotto-classificazioni visibili nella legenda.

All’atto della visualizzazione di risultati generati da un’interrogazione a FuLL, viene creato un Mapfile esteso contenente la definizione di tre nuovi layers che effettuano il collegamento alle entità di tipo puntuale, lineare e poligonale della tabella risultati, diversamente colorati in modo da distinguerli dalla grafica dei layers sottostanti.

Mapserver gestisce in modo nativo la connessione a database Postgres con estensioni PostGIS, per cui la coordinazione di tali moduli operativi è risultata immediata ed avviene tramite la definizione di un oggetto layer secondo lo schema di seguito riportato:

LAYER

```

CONNECTIONTYPE postgis
CONNECTION "user=<user> password=<password> dbname=<dbname> host=<host>"
DATA "<geom_field> from <table_name>[ using SRID=-1]"
STATUS <ON|OFF|DEFAULT>
TYPE <POINT|LINE|POLYGON>

```

...

END

La tabella risultati contiene un campo di tipo geometrico popolato con le geometrie (punti, linee e poligoni) dei risultati della query e visualizzati, tramite il server GIS MapServer, come layer “on top” rispetto tutti gli altri all’interno del frame GIS dell’applicativo web; tramite un opportuno CheckBox è possibile attivare o disattivare il layer dei risultati. Gli altri campi sono popolati con il contenuto informativo non grafico dei risultati; i primi tre record sono visualizzati a piè di pagina principale mentre la tabella è visualizzabile nella sua interezza tramite l’apposito button “Mostra tabella risultati” (vedi figura 2).

## **CONCLUSIONI**

E' fondamentale sottolineare come tutta l'architettura funzionale del progetto, pur contenendo moduli ad alta innovazione tecnologica che, naturalmente, non sono con codice aperto, è stata progettata tenendo in conto l'obiettivo della massima integrazione con le tecnologie esistenti. Questo obiettivo è stato centrato grazie alla possibilità di accedere a strumenti di gestione e sviluppo presenti nel mondo opensource. In particolare la sperimentazione condotta su Postgres consente di portare con sforzo minimo il modulo di traduzione SQL anche su DBMS commerciale come Oracle, mentre l'adozione di schemi di rappresentazione dei dati basati su OWL/XML consente di mappare le strutture dati di FuLL anche in ambienti operativi differenti rispetto a quelli nei quali è stata condotta la sperimentazione.

## **RINGRAZIAMENTI**

Si ringraziano tra i partner scientifici che hanno collaborato al progetto FuLL

- C.N.R., Istituto di Linguistica Computazionale (I.L.C.)
- Università di Pisa, Dipartimento di Linguistica

e le pubbliche amministrazioni che hanno partecipato alla sperimentazione del prototipo

- Provincia di Bologna, Progetti ed Applicazioni di Area Vasta e SIT
- Provincia di Catania, 1° Servizio Pianificazione Territoriale e Trasporti del VII Dipartimento