

Interrogazioni in linguaggio  
naturale a basi dati  
eterogenee: l'ontologia del  
sistema FuLL nei GIS

Ricerca & Sviluppo  
per l'Innovazione

AMFM 2006

*Prof. Maurizio Bombara*

[www.bcsoftware.it](http://www.bcsoftware.it)  
[info@bcsoftware.it](mailto:info@bcsoftware.it)

### Il progetto FuLL

“FuLL” (Fuzzy Logic and Language) è un progetto di ricerca e sviluppo ideato da BC e finanziato dal Ministero delle Attività Produttive nell’ambito del P.I.A. Innovazione (Misura 2.1. del P.O.N.), a cui collaborano istituti del CNR e dipartimenti universitari, aventi competenze altamente specialistiche nel settore.



L'innovazione di FuLL

Incrementare l'usabilità dell'interfaccia dei DBMS.

FuLL (Fuzzy Logic and Language) è un motore di ricerca intelligente, una tecnologia software che innova i sistemi di interrogazione dei database tramite l'utilizzo di una interfaccia in linguaggio naturale (NLIDB Natural Language Interfaces to Databases).

*La componente geografica è un "collante"*

La georeferenziazione dei dati permette di creare relazioni dinamiche fra i dati di un GIS per:

- incrociare informazioni di natura diversa
- ampliare la conoscenza su un determinato dominio
- generare nuova informazione

I dati spesso provengono da varie sorgenti.

### *Consultare basi dati eterogenee è difficoltoso*

Il reperimento delle informazioni diviene spesso problematico per disallineamenti nel:

- formato dei dati
- modello dei dati
- significato (semantica) dei termini

*La consultazione in NL acuisce il problema*

Ponendo le query in un linguaggio integrato e uniforme tutte le disomogeneità devono essere nascoste all'utente

Al cambiare della base dati l'utente si aspetta di poter interagire alla stessa maniera

### Caratteristiche del sistema FuLL

- interfaccia consistente: permette l'interrogazione su basi di dati eterogenee (ad esempio due db distinti) purché semanticamente equivalenti (cioè contenenti dati relativi allo stesso dominio applicativo)
- facile da utilizzare: disponibile anche all'utente non specializzato
- permette espressioni linguistiche non precise, dichiarative, l'uso di sinonimi e di espressioni di tipo qualitativo

*"Dammi tutti i negozi vicini alla stazione centrale"*

*"Indicami i comuni piccoli con distributori che erogano GPL"*

*L'ontologia ha un ruolo centrale*

Il modello ontologico

- racchiudere la conoscenza strutturata sul dominio
- fornisce una “vista” della porzione di dati che si vuole rendere consultabile



*Il modello ontologico guida le interrogazioni*

Si compone essenzialmente di due parti:

- OntoLex (parte linguistica)
- OntoOwl (parte di rappresentazione di dominio)

Ci concentreremo su OntoOwl...

Obiettivo:

L'ontologia di dominio deve essere facilmente integrabile su basi di dati geografiche eterogenee

*Ontowl deve soddisfare due opposte esigenze*

Per potersi integrare con DB eterogenei il modello deve:

- rappresentare concetti di dominio sufficientemente generali da poter essere condivisi da molteplici sorgenti dato
- rappresentare in maniera astratta entità effettivamente presenti nei database

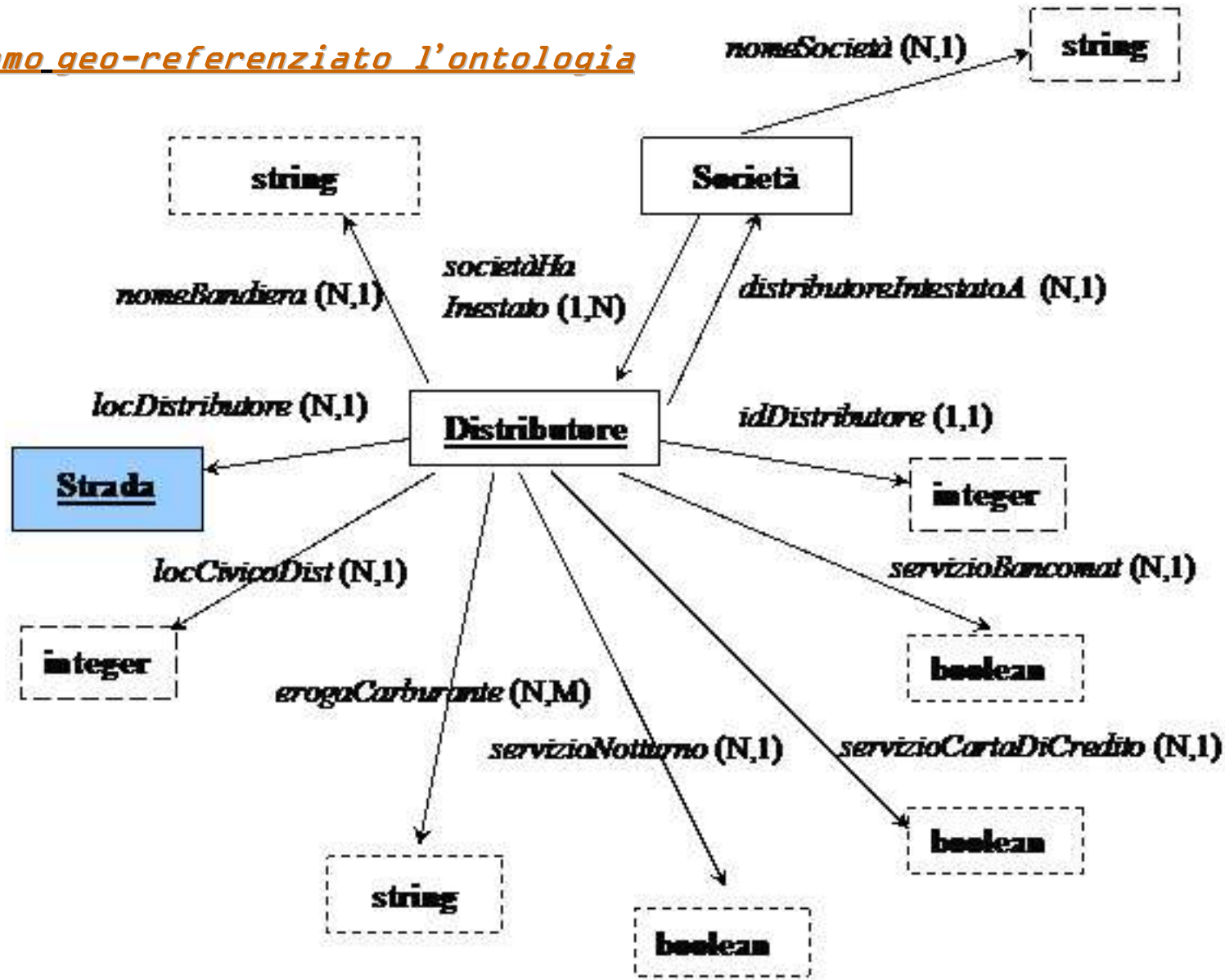
Quindi esso fonde aspetti di ontologia *applicativa* con aspetti di ontologia *fondazionale*

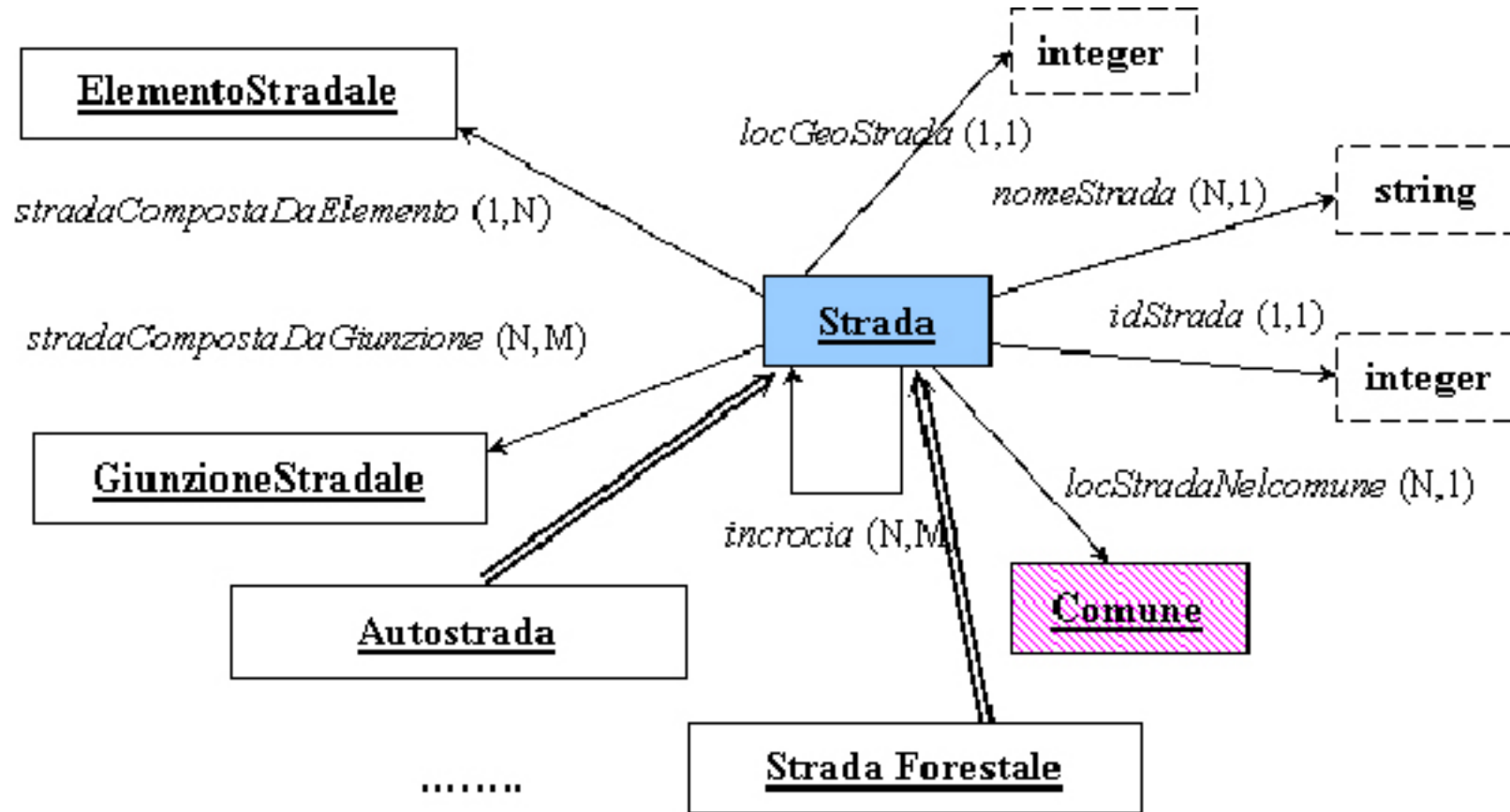
*Approccio ibrido top-down e bottom-up*

A questo scopo la definizione di concetti, relazioni e attributi è stata guidata da due approcci:

- dall'analisi dei dati sorgente ed astrazione di essi da tabelle fisiche in concetti e relazioni (bottom-up)
- dall'esigenza di rappresentare concetti e relazioni tipiche del dominio applicativo scelto (top-down)

Abbiamo geo-referenziato l'ontologia



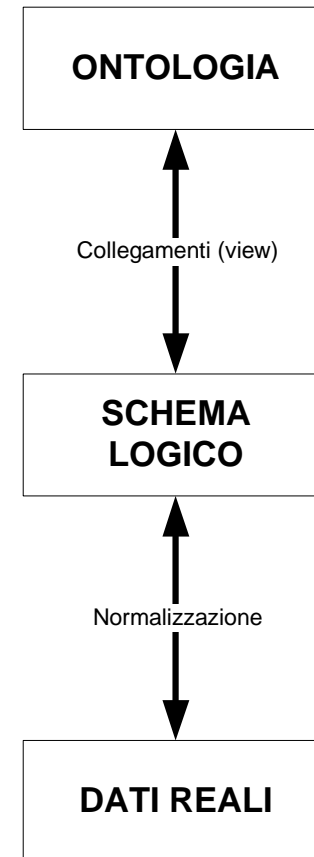


L'ontologia deve essere collegata ai dati

In ultima analisi i dati devono essere estratti dalla base di dati

Tale collegamento è stato strutturato in una architettura a più livelli

Lo scopo è quello di poter sostituire la base di dati sottostante senza bisogno di riscrivere l'ontologia



Confronto di FULL con interfaccia tradizionale

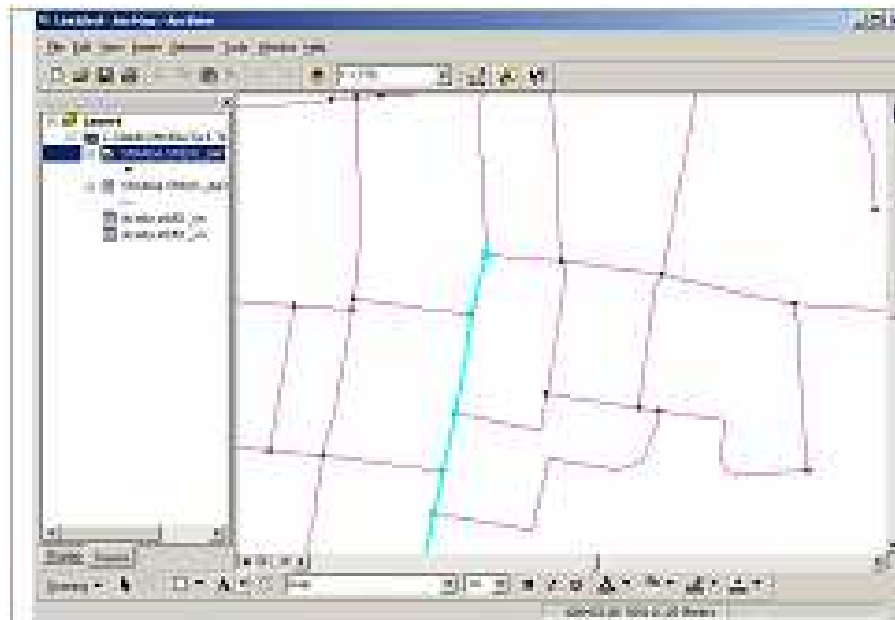


Figura 1: Individuare via Garibaldi con ArcGIS



Figura 2: Quali strade incrociano via Garibaldi?

La sequenza di operazioni da effettuare con il sistema di interfaccia tradizionale (rappresentata in Figura 1) è tipicamente lunga e articolata

# Risultati



Progetto di Ricerca e Sviluppo FuLL (Fuzzy Logic and Language) prototipo Beta

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

**FuLL**  
Fuzzy Logic and Language

Query in linguaggio naturale :  
 di sono comuni popolosi?

Reset    Invia richiesta a FuLL

Box di dialogo :

0 19000 38000 57000

Quali sono i comuni popolosi?

preview Tabella risultati

fuzzyrank	gid	id_comune	hapolazione2
100	51	32	64576
94	19	6	381160
56	25	11	33472

Partner scientifici:



Progetto di Ricerca e Sviluppo FuLL (Fuzzy Logic and Language) demo Catania

**IBC** SOFTWARE COMPANY In collaborazione con: **Provincia di Catania**  
 VII Dipartimento - TERRITORIO E AMBIENTE  
 1° Servizio - Pianificazione Territoriale e Trasporti

**FuLL**  
Fuzzy Logic and Language

Query in linguaggio naturale :  
 quali sono i comuni popolosi?

Reset    Invia richiesta a FuLL

Box di dialogo :

Risultati di FuLL :  
 Puntuale  
 Lineare  
 Poligonale

RecordCount: 6    **Mostra tabella risultati**

0 22000 44000 66000 88000 m

preview Tabella risultati

fuzzyrank	gid	id_comune	hapolazione2
100	48	313	305464
81	32	272	48601
74	38	295	44670

Partner scientifici:

info@bcsoftware.it  
www.bcsoftware.it



# Risultati



Progetto di Ricerca e Sviluppo FuLL (Fuzzy Logic and Language) prototipo Beta

In collaborazione con:  
**Provincia di Bologna**  
 Progetti ed applicazioni di Area Vasta e SIT

**FuLL**  
 Fuzzy Logic and Language

Query in linguaggio naturale :  
 quali sono le strade provinciali?

Reset    Invia richiesta a FuLL

Box di dialogo :

Dove sono le strade provinciali?

Preview Tabella risultati

gid	id_stradaprovinciale
104	37000100657
103	37000099901
102	37000063201

Partner scientifici:

Progetto di Ricerca e Sviluppo FuLL (Fuzzy Logic and Language) demo Catania

In collaborazione con:  
**Provincia di Catania**  
 VII Dipartimento - TERRITORIO E AMBIENTE  
 1° Servizio Pianificazione Territoriale e Trasporti

**FuLL**  
 Fuzzy Logic and Language

Query in linguaggio naturale :  
 dove sono le strade provinciali?

Reset    Invia richiesta a FuLL

Box di dialogo :

Risultati di FuLL :

- Puntuale
- Lineare
- Poligonale

RecordCount  
 448

Mostra tabella risultati

Preview Tabella risultati

gid	id_stradaprovinciale	id_comune	id_strada	nomecomune
448	1053	369	1053	MAZZARRONE
447	1330	369	1330	MAZZARRONE
446	17	366	17	LICODIA EUBEA

Partner scientifici:

info@bcsoftware.it  
 www.bcsoftware.it

*L'applicazione di FuLL ai GIS cambia l'utilizzo*

Mette a disposizione del cittadino, all'utente non specializzato l'ingente patrimonio di dati disponibile nei SIT

Migliora la qualità del risultato. Standardizza il risultato di query soggettive (qualitative): riduce il margine di soggettività nel manipolare parametri discrezionali.

Per gli utenti esperti: Abbatte drasticamente il tempo di elaborazione (interrogazioni) complesse.

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

- CNR, Istituto di Linguistica Computazionale (ILC)
- CNR, Istituto di Scienze e Tecnologie dell'Informazione (ISTI) - KDDLab
- Università di Pisa, Dipartimento di Linguistica

*e le pubbliche amministrazione che hanno reso disponibili i dati per la sperimentazione condotta*

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

*I nostri Partner per la Ricerca e lo Sviluppo*



Università di Catania



Dipartimento di Ingegneria  
Informatica e delle  
Telecomunicazioni  
Univ. Catania



Dipartimento di Ingegneria  
Civile e Ambientale  
Univ. Catania



Università di Pisa  
Dipartimento di Linguistica e  
Dipartimento di Informatica



Istituto di Linguistica  
Computazionale  
CNR



Istituto di Scienza e Tecnologie  
dell'Informazione  
CNR



Knowledge Discovery and  
Delivery Lab  
ISTI CNR & Dip. Inf. Univ. Pisa



Università di Milano Bicocca



Dipartimento di Informatica  
Sistemistica e Comunicazione  
Univ. Milano Bicocca



Università di Messina



Fraunhofer Institut  
Biomedizinische  
Technik  
Istituto Fraunhofer



BC S.r.l.

Catania

Via Caronda, 136

Cap 95128

Tel. +39 095 7286481

Fax +39 095 444199

Milano

Via Orobica, 3

Cap 20139

Tel. +39 - 02 52505300

[www.bcsoftware.it](http://www.bcsoftware.it)  
[info@bcsoftware.it](mailto:info@bcsoftware.it)