

INTERROGAZIONE IN LINGUAGGIO NATURALE DI BANCHE DATI GIS DELLA PROTEZIONE CIVILE

Antonio CONDORELLI ^(*), Giuseppe MUSSUMECI ^(*), Orazio MARLETTA ^(*)
Maurizio BOMBARA ^(**), Davide CALÌ ^(**), Ivana CALÌ ^(**), Vincenzo CARTELLI ^(**),
Giuseppe TROPEA ^(**)

^(*) Dipartimento di Ingegneria Civile e Ambientale, Università di Catania V.le A.Doria 6, 95125 Catania,
tel. 095/7382212 - fax 095/7382247 - e-mail gmussume@dica.unict.it - acondor@dica.unict.it

^(**) BC S.r.l. Software Company – Via Caronda n.136 – 95128 Catania
Tel. +39 0957286481 - fax +39 095444199 - info@bcsoftware.it - www.bcsoftware.it

Sommario

Il problema della facilità di interazione fra utenti e applicazioni GIS è ampiamente conosciuto ed assume una valenza di particolare importanza quando, come nel caso delle banche dati GIS della Protezione Civile, i potenziali fruitori sono caratterizzati da livelli di qualificazione professionale notevolmente differenziati (ad es. dal tecnico specializzato al volontario). La possibilità di utilizzare le applicazioni GIS attraverso un'interfaccia in linguaggio naturale consente di estendere al massimo la fruibilità dei Sistemi Informativi Territoriali grazie ad un accesso semplice ed immediato alle banche dati.

Il lavoro che qui si presenta riguarda dunque un'applicazione della tecnologia “*FuLL (Fuzzy Logic and Language)*”, che consente proprio l'interrogazione di banche dati geografiche in linguaggio naturale, su una base dati geografica sviluppata nel territorio della Provincia di Catania per uno studio di prevenzione e previsione ai fini della Protezione Civile.

Abstract

The problem of an easy interaction between users and GIS applications is well known. It becomes a critical issue in cases, as the one of Civil Protection GIS data banks, where users are characterized by different levels of professional qualification (i.e. specialized technicians vs. volunteers). Thus, the possibility to use a natural language interface on GIS applications becomes, in cases similar to the one we introduce, a very important feature for the informative system, extending the limits of its usability and ensuring a simple and user friendly interface. This paper shows “*FuLL (Fuzzy Logic and Language)*” technology, which allows querying geographical databases in natural language, applied onto a geographical database developed on the territory of the Province of Catania for a study about the prevention and forecasting in the field of Civil Protection.

1. Introduzione

L'approccio moderno alla pianificazione e gestione delle emergenze di Protezione Civile deve necessariamente basarsi sui concetti di semplicità, flessibilità ed immediatezza degli interventi. In quest'ottica, la recente introduzione dei GIS, già da tempo utilizzati nelle attività di previsione e prevenzione dei rischi, anche come sistemi di supporto delle decisioni nella gestione dell'emergenza, va certamente nella giusta direzione. Sussistono tuttavia alcune problematiche che ostacolano un'ampia diffusione della tecnologia GIS nelle diverse strutture operative della Protezione Civile, riconducibili a due cause fondamentali:

- mancanza di standard per la strutturazione e l'armonizzazione dei dati di interesse;

- difficoltà (in particolar modo da parte di alcune specifiche tipologie di operatori) nella fruizione delle informazioni e nell'interrogazione dei dati, generalmente non gestite tramite interfacce amichevoli e di immediata comprensione.

Il primo aspetto richiede la definizione preliminare del “dominio” di interesse, ovvero del patrimonio informativo necessario, e la definizione di standard e modelli di riferimento per organizzare le banche dati, alla luce anche delle particolarità e complessità dei modelli di valutazione e simulazione generalmente utilizzati.

Il secondo aspetto è invece strettamente legato alla specifica qualificazione del personale addetto alle emergenze, fortemente eterogeneo e non sempre dotato delle necessarie competenze per l'utilizzo dei DB e degli strumenti GIS a disposizione, né di una adeguata conoscenza del territorio d'intervento.

Relativamente a quest'ultimo aspetto, l'adozione di un'interfaccia in linguaggio naturale, che consenta, ad esempio, di interagire con il GIS mediante interrogazioni del tipo “quali sono le strade più sicure da percorrere per i mezzi di emergenza in caso di sisma?” ottenendo rapidamente visualizzazioni cartografiche e risposte alfanumeriche, rappresenta un passo molto importante ai fini della realizzazione degli obiettivi proposti.

A tal fine in questo lavoro si presentano i primi risultati di un approccio metodologico integrato che, con riferimento allo scenario campione della Provincia di Catania ed al rischio sismico della rete stradale, prevede:

- la modellizzazione concettuale del dominio della Protezione Civile, in termini di entità coinvolte e di relazioni fra di esse, ai fini della costruzione dell'“ontologia” dello stesso;
- l'estrazione e l'utilizzazione di dati contenuti nel SIT realizzato dalla Provincia Regionale di Catania per la previsione e prevenzione dei rischi territoriali (2001) e dei relativi modelli per l'analisi funzionale del sistema viario in caso di evento sismico;
- l'impiego della tecnologia innovativa “FuLL (Fuzzy Logic and Language)”, ideata e sviluppata dalla Software Company BC S.r.l., che, integrata come componente di interfaccia utente nei motori Web-GIS, permette di consultare ed estrarre le informazioni dalla banca dati, traducendo ed elaborando automaticamente le domande poste dall'utente in linguaggio naturale, tramite tecniche basate su logica fuzzy e con un approccio orientato al dialogo.

2. La tecnologia “FuLL”

Il sistema “FuLL (*Fuzzy Logic and Language*)”, ideato e sviluppato dalla Software Company BC s.r.l., consiste in una specifica tipologia d'interfaccia che permette l'interrogazione in linguaggio naturale di basi dati geografiche eterogenee, purché semanticamente equivalenti.

Tra gli elementi innovativi e le principali caratteristiche tecniche di FuLL vi sono:

- Ontologia : un passo di fondamentale importanza è la definizione di una adeguata ontologia, che assume un ruolo centrale nell'architettura dell'intero sistema garantendo la possibilità di porre semplicemente *query* in un linguaggio naturale rendendo, al tempo stesso, trasparenti all'utente tutte le disomogeneità di sintassi, terminologia, modelli e formati delle basi dati sorgenti. L'ontologia di FuLL è dunque il componente progettato per racchiudere la conoscenza strutturata sul dominio.
- Tecniche Fuzzy per l'interpretazione linguistica: sono utilizzate a supporto e come integrazione dell'analisi semantica delle domande poste dagli utenti. Queste tecniche infatti permettono di garantire maggiore flessibilità al linguaggio e danno al sistema, come ulteriore strumento, la possibilità di rappresentare e comprendere anche aspetti soggettivi e qualitativi. Una caratteristica comune delle interrogazioni formulate in linguaggio naturale è il modo impreciso e vago in cui la maggior parte di queste vengono poste dal richiedente. L'impiego di tecniche basate sulla logica fuzzy rende possibile e facile ottenere una corrispondenza fra queste richieste ed i valori codificati in maniera numericamente precisa nel database di interesse.

- Collegamento ai dati: altro elemento progettuale importante del sistema *FuLL* è quello concernente il collegamento dell'ontologia alla base di dati dalla quale i dati devono essere estratti. Tale collegamento è stato strutturato in una architettura a più livelli, partendo da un livello più astratto sino ad uno schema logico a più basso livello, a ridosso dei dati stessi del database. In questo modo si ottiene la possibilità di sostituire la base di dati sottostante, a parità di dominio, senza bisogno di riscrivere l'ontologia.

La tecnologia *FuLL* è in corso di industrializzazione ed è stata sviluppata da BC nell'ambito di un progetto a cui hanno collaborato partner scientifici del C.N.R. (Istituti I.L.C. e I.S.T.I.-KDDLAb) e universitari (Dipartimento di Linguistica dell'Università di Pisa), oltre che pubbliche amministrazioni in qualità di produttori e utilizzatori di dati in ambito GIS (Provincia di Bologna, Provincia di Catania).

3. L'area di studio e le banche dati utilizzate

Il Sistema Informativo Territoriale utilizzato nella presente applicazione è stato sviluppato tra il 1999 ed il 2001 quale base dati per la "*Valutazione delle funzioni strategiche e della vulnerabilità della rete infrastrutturale dei trasporti*" del "*Programma Provinciale di Previsione e Prevenzione ai fini di Protezione Civile*" della Provincia Regionale di Catania, territorio notoriamente classificato come una delle aree a più elevato rischio sismico in Europa.

I tematismi implementati riguardano principalmente i centri abitati e la rete viaria, suddivisa in *feature class* distinte per appartenenza amministrativa (autostrade, strade statali, provinciali, altre strade). Le singole infrastrutture sono suddivise in tronchi in funzione delle intersezioni stradali, in modo tale che ciascun arco risulti privo di nodi all'interno. Ciascun tronco di ogni *feature class* del tema "strade" è stato dunque caratterizzato, nel relativo database, da una serie di campi per lo più legati alle caratteristiche geometriche dell'opera (larghezza della carreggiata, pendenza media del profilo longitudinale, raggio minimo tra le curve presenti) e finalizzati alla valutazione d'idoneità della stessa infrastruttura a servire correnti di traffico dei mezzi speciali della Protezione Civile.

Adattando al caso specifico alcuni modelli di analisi funzionale presenti in letteratura e applicandoli al database delle caratteristiche dei vari tronchi stradali, sono stati individuati ed "estratti" dalla rete viaria complessiva una serie di percorsi strategici (la sotto-rete delle *lifelines*) che rappresentano i collegamenti migliori per ciascun centro abitato della Provincia verso i possibili centri erogatori di servizi di Protezione Civile presenti nel territorio. Su tali percorsi, tenendo conto della popolazione potenzialmente servita e di altri parametri che caratterizzano i diversi centri abitati, è stato valutato il livello di esposizione sismica.

Per quanto riguarda la vulnerabilità, sono state censite e caratterizzate mediante apposite campagne di rilevamento in situ circa 200 opere d'arte, principalmente ponti e viadotti, che interessano la rete viaria oggetto di studio. Sulla base delle informazioni acquisite (ad es. altezza e luce della campata, altezza delle spalle e delle pile, materiali utilizzati, schema strutturale adottato, anno di progettazione, etc.) è stata implementata una specifica *feature class* nell'ambito della quale, utilizzando specifici modelli presenti in letteratura, è stato possibile effettuare una valutazione del livello di vulnerabilità di ciascun opera. Correlando questi indicatori con il livello di pericolosità sismica previsto in ciascun sito, così come valutato nell'ambito del lavoro sopra citato da altri esperti del settore, è stato possibile effettuare una stima del danno atteso sulle opere d'arte censite.

4. Il dominio della Protezione Civile: potenzialità di interrogazione in linguaggio naturale

L'obiettivo primario della ricerca è stato quello di valutare le potenzialità applicative della tecnologia *FuLL* al dominio della Protezione Civile e, in particolare, all'ambito del rischio sismico. La possibilità di accedere ai dati d'interesse mediante un'interfaccia semplice ed interamente basata sul linguaggio naturale è senza dubbio elemento di grande innovazione e conferisce valore aggiunto a un SIT incrementandone notevolmente il livello di fruibilità. Tale possibilità assume una particolare importanza per gli Organismi di Protezione Civile, costituiti da personale parecchio eterogeneo dal punto di vista delle competenze tecniche e che, tuttavia, necessita (in particolar modo durante le

emergenze) di interagire proficuamente e rapidamente con le banche dati per poter orientare correttamente il proprio operato.

Il primo passo per raggiungere il suddetto obiettivo è consistito nella costruzione di un modello ontologico appropriato: è stato necessario non solo inserire nell'ontologia le nuove entità tipiche del dominio "rischio sismico" (ad esempio gli indici di vulnerabilità e rischio, le tipologie dei materiali, le lifeline, i ponti, gli scenari di scuotimento sismico, ecc.), ma anche integrare i possibili attributi con le entità preesistenti, ovvero già modellate nell'ontologia di *FuLL* per altri ambiti operativi. In altre parole, si è effettuata una integrazione con i domini già esistenti in ambito GIS all'interno di *FuLL*, ovvero "Mobilità e Trasporti" e "Sistema Insediativo Territoriale". Nella figura seguente è rappresentato un frammento del nuovo modello ontologico realizzato.

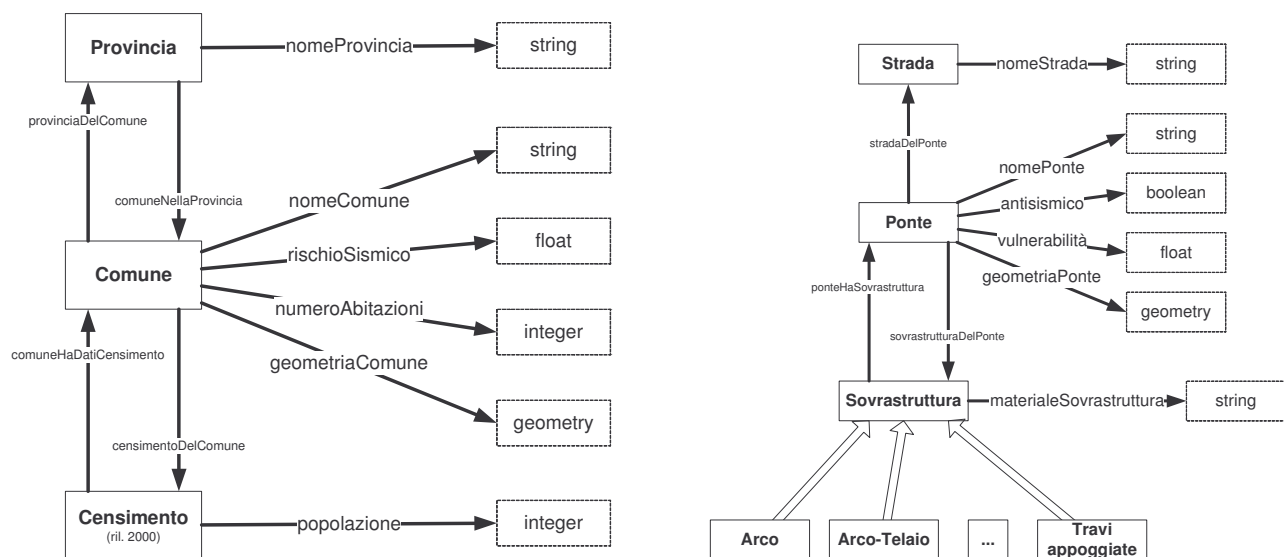


Figura 1 – Frammenti di ontologia

Il secondo passo è consistito nel collegare il nuovo modello ontologico così costruito sia col database sopra descritto relativo all'analisi del rischio sismico, sia col database della Provincia di Catania (rivedendo i collegamenti pre-esistenti). In questo particolare scenario, la difficoltà principale è legata alla presenza di due basi di dati distinte, eterogenee per provenienza dei dati ed impostazione.

L'approccio seguito è basato sull'utilizzo di un livello logico di astrazione intermedio fra l'ontologia ed i dati "grezzi", all'interno del quale effettuare le necessarie normalizzazioni e collocare le nuove tabelle di relazione fra entità eterogenee. Un importante risultato conseguito è stato quello di impiegare il modello ontologico per guidare l'integrazione del database sul rischio sismico con il database della Provincia di Catania contenente i dati relativi a "Mobilità e Trasporti" e "Sistema Insediativo Territoriale". Mediante l'ontologia è stato possibile individuare con estrema facilità gli attributi da normalizzare e le relazioni da costruire ex-novo.

Sul modello ontologico *FuLL* è in grado di restituire tutte le caratteristiche di interesse nell'ipotesi di interrogazioni del sistema sia da parte di tecnici esperti che di utenti meno informati sulle connotazioni tecniche del settore. Ad esempio il sistema potrebbe rispondere sia ad una *query* tecnica del tipo "Quali ponti hanno sovrastruttura in cemento armato precompresso?", sia a *query* più generali come "Quali sono i tratti della strada provinciale SS194 che costituiscono lifeline in caso di sisma?" I concetti che sono stati sottoposti ad interpretazione con tecniche *fuzzy* sono quelli di particolare interesse ai fini di una stima intuitiva sia in situazioni che necessitino prontezza ed immediatezza di intervento sul territorio sia in situazioni di studio ai fini di un intervento preventivo. Ci si è concentrati sulla definizione di appropriati *fuzzy-set* per i concetti di "indice di rischio sismico" ed "indice

di vulnerabilità”. Il prototipo realizzato è in grado di comprendere interrogazioni in linguaggio naturale, e quindi di supportare l’interazione con il GIS, del tipo seguente:

- “Quali sono le strade più sicure da percorrere per i mezzi di emergenza in caso di sisma?”
- “Quali sono i ponti non antisismici?”
- “Quale comune ha il più elevato indice di rischio?”
- “Quali sono i comuni ad alto rischio sismico?”
- “Quali sono le lifeline con pendenza media bassa?”

I risultati, alcuni dei quali illustrati nelle figure seguenti, consistono sia nella visualizzazione cartografica che in informazioni alfanumeriche.

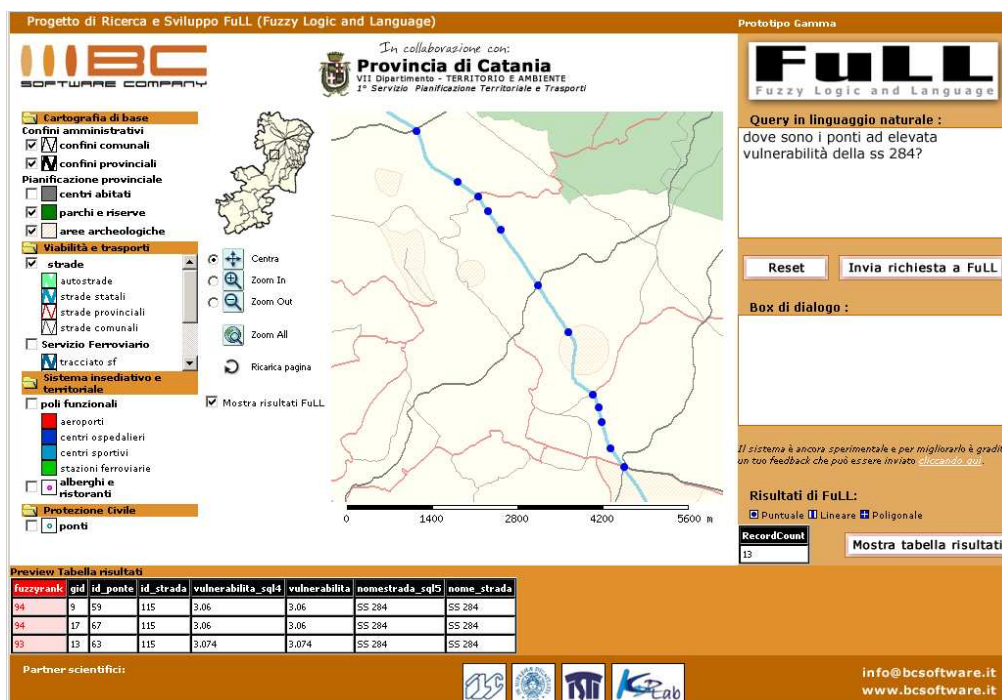


Figura 2. Risultati della query: “Dove sono i ponti ad elevata vulnerabilità della SS 284”?

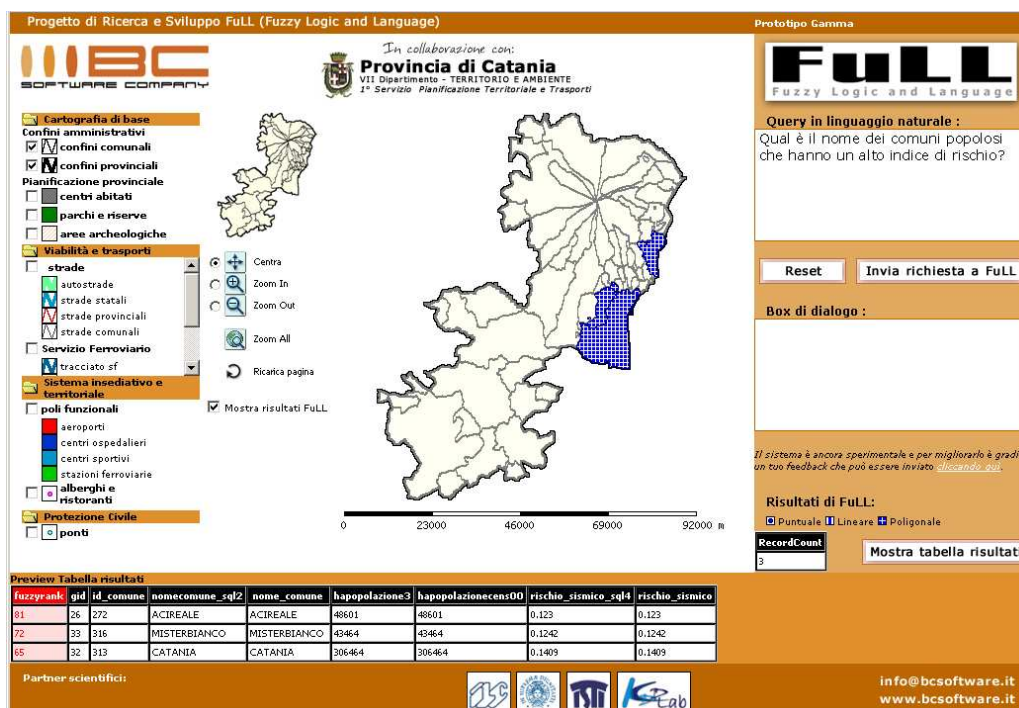


Figura 3. Risultati della query: “Quali comuni popolosi hanno un alto indice di rischio?”

5. Conclusioni

Nel lavoro è stata presentata un'applicazione su una banca dati geografica sviluppata per finalità di Protezione Civile che consente, mediante la tecnologia *FuLL*, l'interrogazione dei DB in linguaggio naturale. Un aspetto particolarmente originale della ricerca consiste nell'applicazione della tecnologia *FuLL* su un dominio non di carattere generale ma assolutamente specialistico e ad alto contenuto tecnico, come quello della Protezione Civile e della pianificazione delle emergenze rispetto al rischio sismico.

I risultati ottenuti, particolarmente incoraggianti, hanno fornito un'ulteriore conferma della versatilità dello strumento *FuLL* e hanno delineato grandi potenzialità di sviluppo per il futuro, nell'ottica di un incremento della fruibilità di banche dati geografiche specialistiche anche da parte di un'utenza non tecnica (caso tipico del personale della Protezione Civile, particolarmente eterogeneo quanto a formazione e competenze). L'interfaccia si presta inoltre ad ulteriori possibili integrazioni con altri ambienti (internet-web) e strumenti di controllo, quali dispositivi portatili o palmari collegati in rete via GPRS o UMTS o anche semplici telefoni cellulari, mediante i quali è verosimile prevedere la possibilità di inviare interrogazioni in linguaggio naturale anche tramite SMS.

Bibliografia

- S.Cafiso, R.Colombrita, A.D'Andrea, G.Mussumeci, A.Condorelli, F.Colombrita (2001) "Un modello di GIS per la valutazione del rischio sulle infrastrutture stradali nelle emergenze della protezione civile" - *XI Convegno S.I.I.V.* – Verona, 28 - 30 Novembre 2001.
- S.Cafiso, A.Condorelli, G.Mussumeci (2005) "Functional analysis of the urban road network in seismic emergency. GIS application on Catania city." – volume "Seismic prevention of damage. A case study in a Mediterranean city" (editor M.Maugeri) series: *Advances in earthquake engineering*, vol.14 (ISBN:1-84564-004-7).
- A. D'Andrea, S. Cafiso, A.Condorelli (2005) "Methodological considerations for the evaluation of seismic risk on road network" – *Pageoph (Pure and Applied Geophysics)*, volume 162 No.4, pp. 767-782, *Pageoph Topical Volumes "Messina Seismological Observatory Memoriale Volume"* editing di A.Bottari, D.Mayer-Rosa, J.Ibanez, M.Maugeri; edizioni Birkhauser, Berlino.
- Medyckyj-Scott, D. & Hearnshaw, H. (1993). *Human Factors in Geographical Information Systems*, Belhaven Press, London .
- Max J. Egenhofer, A. & Rashid, B. & Shariff, M., 1998. "Metric details for natural-language spatial relations", *ACM Transactions on Information Systems (TOIS)* Volume 16, Issue 4, pp. 295 – 321.
- Bartolini R., Caracciolo C., Giovannetti E., Lenci A., Marchi S., Pirrelli V., Renso C., Spinsanti L., 2006. "Creation and Use of Lexicons and Ontologies for Natural Language Interface to Databases", in *proceedings of 5° conference on Language Resources and Evaluation (LREC 2006)*.
- Guarino N., 1997. "Understanding, building, and using ontologies: A commentary to "Using Explicit Ontologies in KBS Development", by van Heijst, Schreiber, and Wielinga." *International Journal of Human and Computer Studies* 46: 293-310.
- P.P Chen., March 1976. "The entity-relationship model - toward a unified view of data", *ACM Transactions on Database Systems*, vol. 1, no. 1, pp. 9-36.
- Hugh W.Calkins, January 1996. "Entity Relationship Modeling of Spatial Data for Geographic Information Systems", *International Journal of Geographical Information Systems*.
- Zeiler, Michael, 1999. "Modeling Our World – The ESRI Guide to Geodatabase Design", *ESRI Press*.
- ILC CNR: Bartolini R., Caracciolo C., Giovannetti E., Marchi S., Pirrelli V. – ISTI CNR: Renso C., Spinanti L. – Dip. Linguistica Univ. Pisa: Lenci A., 2006. "Creation and Use of Lexicons and Ontologies for NL Interface to databases", Fifth international conference on Language Resources and Evaluation, LREC 2006.