

COSTRUZIONE DI UN'ONTOLOGIA PER LO SVILUPPO DI APPLICAZIONI SEMANTICHE IN CAMPO GEOGRAFICO

Gianni DORIA (*), Mauro VELLUTO (*)

(*) CSI-Piemonte-Direzione SIT, c.so Tazzoli 215/12B – 10137 Torino,
tel. +39-011-3169711, fax +30-011-316821

Sommario

L'interoperabilità tra SI richiede uno sforzo finalizzato alla condivisione del significato dei dati scambiati. Il presente articolo propone un percorso che, partendo dall'utilizzo dell'analisi ontologica per la costruzione di una base di conoscenza territoriale condivisibile, arriva a descrivere alcuni elementi utili per un suo utilizzo concreto attraverso una sperimentazione nell'ambito della protezione civile.

Abstract

Interoperability of Information Systems requires a commitment aimed at sharing the meaning of exchanged data. The proposed paper suggests a way that, starting from the use of ontological analysis aimed at sharing territorial knowledge bases, describes some elements suitable for concrete testing in a specific field: the Civil Protection.

Introduzione

Una delle maggiori criticità legate alla progettazione e alla gestione di Sistemi Informativi (SI) deriva dal fatto che spesso i processi di produzione di informazioni utili, richiedono in partenza l'utilizzo di dati eterogenei da diversi punti di vista. Vogliamo considerare un livello di eterogeneità insito nel fatto che ogni entità del mondo reale assume significati diversi a seconda dell'interesse o del punto di vista di chi la considera. Questo ovviamente vale anche per tutti gli attori di un SI dal gruppo di progettazione agli utenti, siano essi privilegiati, sporadici o potenziali. Tra questi significati, solo alcuni sono resi formalmente espliciti nelle informazioni del SI. Alcuni concetti di livello 'generale' appartengono al cosiddetto senso comune, e risultano noti a chiunque, anche se non formalizzati in alcun modo. Immaginiamo un SIT che rappresenti la base conoscitiva dello stato delle risorse idriche di un territorio: esso permette ad un utente, che faccia riferimento ad un proprio bagaglio di nozioni di senso comune di leggere come espliciti alcuni significati in realtà impliciti: l'incrocio di un tratto di fiume con una strada rivela la presenza di un attraversamento non esplicitamente rappresentato, così come la non inclusione di un terreno edificabile all'interno delle aree di rispetto può smascherare una *forzatura* amministrativa.

Ovviamente quello stesso tratto di fiume nella realtà è qualcos'altro e su di esso possiamo dire che ha alcune caratteristiche fisiche, una dimensione lineare, una larghezza media, una portata media, la sua superficie ha un certo colore, le sue acque hanno una determinata composizione chimica, ecc. Ognuna di queste caratteristiche non è statica, ma cambia nel tempo, ed esercita un'influenza biunivoca sugli altri elementi del territorio (pensiamo alla composizione chimica). Possiamo aggiungere che esso possiede caratteristiche non fisiche, di tipo emotivo: la sua vista può suscitare sentimenti in chi lo guarda, lo attraversa, o ne ricorda l'aspetto passato. Tutti questi sono significati di senso comune, né formalizzati, né resi in alcun modo espliciti dal nostro SI. Altri concetti

potrebbero non trovar formalizzazione perché troppo complessi. Cos'è la pressione antropica e come agisce sul fiume? Quale rapporto esiste tra lo sviluppo del territorio e l'azione politica? Quali sono i fattori *esterni* che influenzano quest'ultima?

Certo non è possibile formalizzare ed esplicitare tutti i significati collegabili ad un'entità del mondo reale in ogni possibile situazione. Ciò che invece può essere utile fare è cercare di definirne e rappresentarne formalmente quelli più generali, indipendentemente da ogni contesto specifico, in termini di concetti e relazioni primitive di livello talmente alto da essere universalmente espliciti. Da questa formalizzazione generale, se necessario, si possono far discendere descrizioni di significati di quella stessa entità in contesti particolari, attraverso specializzazioni di dominio, fino a costruire basi di conoscenza coerenti e condivisibili.

L'analisi ontologica

Una via possibile per ottenere questo risultato è un processo che partendo da un'analisi ontologica, porti ad una formalizzazione dei suoi risultati in una ontologia vera e propria.

In questo lavoro non entreremo nel dettaglio delle basi teoriche su cui poggiano le ontologie utilizzate nel ambito dei sistemi informativi. Rimandiamo gli approfondimenti di questi aspetti alla vasta letteratura scientifica di riferimento da cui prenderemo in prestito alcune definizioni utili per fissare alcuni concetti e chiarire meglio i passaggi successivi.

Analisi ontologica: un processo di individuazione e definizione delle distinzioni rilevanti e delle relazioni legate alla natura delle entità appartenenti ad un certo dominio, con l'obiettivo di eliminare l'ambiguità di termini che hanno diverse interpretazioni in diversi contesti. Un'analisi ontologica è propedeutica alla definizione di un'ontologia

Ontologia: è uno specifico artefatto che esprime il significato inteso di un vocabolario in termini di categorie primitive e relazioni, che descrivono la natura e la struttura del dominio di un discorso. Essa è esprimibile in un linguaggio formale (dal punto di vista informatico).

Ontologia Top Level: descrive concetti molto generali, come lo spazio, il tempo, la materia, gli oggetti, ecc. che sono indipendenti da un particolare problema o contesto.

Ontologie di dominio: descrivono un vocabolario legato ad un certo dominio (la protezione civile, i trasporti, la medicina, ecc.) o ad un obiettivo (la ricerca di documenti sul web, la gestione di un magazzino, ecc.), specializzando gli elementi introdotti in un'ontologia *top level*.

Ontologie applicative: descrivono concetti che dipendono sia da un certo dominio, sia da un obiettivo particolare. (Guarino 1998).

Tutto questo comporta uno sforzo di analisi supplementare che può apparire non giustificabile rispetto ai benefici ottenibili nell'economia complessiva di un sistema informativo: perché allora sviluppare una ontologia (Noy,McGuinness 2001) e perché farlo all'interno del dominio geospaziale?

Per condividere i significati e la struttura delle informazioni fra fruitori:

Il panorama dei servizi informativi oggi disponibili varia da quelli basati su Web, come ad esempio i WFS definiti dalla OGC a quelli in cooperazione applicativa definiti da CNIPA. Se diverse fonti informative che trattano argomenti correlati condividono la stessa ontologia alla base dei termini che usano, sarà allora possibile realizzare nuovi servizi in grado di estrarre e aggregare informazioni da queste fonti eterogenee. Nuove applicazioni potranno utilizzare queste informazioni aggregate. La costruzione di basi di conoscenza condivisibili è il primo passo per costruire sistemi realmente interoperabili. In linea generale l'interazione tra sistemi implica problemi di tipo sintattico e semantico. Le questioni sintattiche sono risolvibili tecnologicamente anche grazie a molti standard a supporto; ma per quanto riguarda la condivisione del significato delle informazioni soffriamo la mancanza di standardizzazione. Su questo filone di attività CNIPA intende promuovere la definizione di ontologie alla base dei servizi in cooperazione applicativa erogati dalle PA centrali e locali.

Per consentire il riutilizzo delle conoscenze di dominio

Il riutilizzo della conoscenza di dominio è una delle motivazioni principali della attuale ricerca sulle ontologie. Rispetto a questa esigenze molte sono le iniziative di definire in modo formale e coerente gli aspetti più generali in modo da poter essere quindi riutilizzati in ambiti più specifici: diversi domini necessitano di rappresentare il concetto di tempo comprendendo le nozioni di intervalli di tempo, i punti in tempo, misure di tempo, e così via, analogamente per il concetto di spazio con le relative nozioni geometriche e topologiche. Nella progettazione della ontologia intendiamo integrare diverse ontologie pre-esistenti che descrivono porzioni del dominio: se queste ontologie sono fondate su una visione coerente dei concetti comuni, il pericolo di introdurre incongruenze nella base di conoscenza è mitigato. Purtroppo il filone di progetti orientato a definire ontologie di alto livello è così ricco da rendere dispersivo anche il mero elenco dei progetti più significativi.

Per rendere esplicite le assunzioni di dominio

Rendere esplicite le assunzioni di dominio che sottostanno una particolare applicazione rende possibile cambiarle facilmente qualora la nostra conoscenza cambiasse. Viceversa codificare tali assunzioni in procedure rende tali assunzioni non solo difficili da scoprire e comprendere, ma anche onerose da modificare. Una specificazione esplicita della conoscenza di dominio è utile agli utenti che devono imparare cosa il vocabolario di dominio significhi.

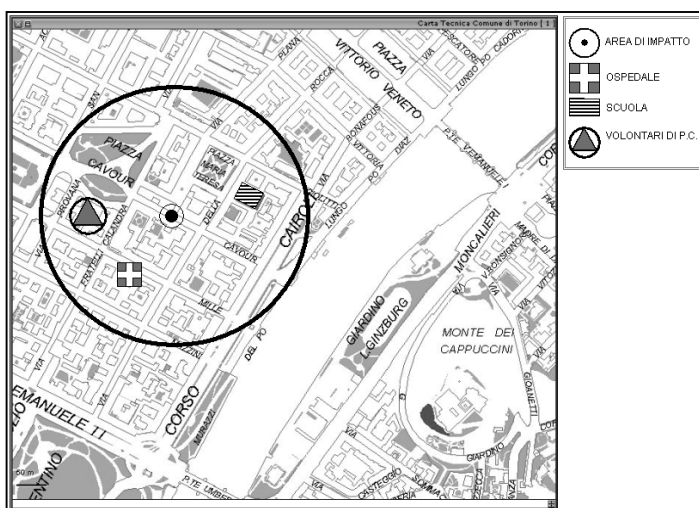
Per separare la conoscenza del dominio dalla conoscenza operativa

La conoscenza operativa è quella che permette di svolgere una particolare compito, la conoscenza di dominio è invece l'ambito in cui il compito è svolto: se è possibile descrivere il compito in forma dichiarativa indipendentemente dall'ambito in cui questo è condotto allora potremmo costruire agenti generalizzati che sappiano svolgere lo stesso compito in ambiti diversi. Compiti di ricerca di cammini in un grafo o il calcolo delle relazioni spaziali potrebbero essere descritti come forme di conoscenza operativa ricorrenti nei contesti geografici.

Per analizzare la conoscenza di dominio

L'analisi della conoscenza di dominio è possibile una volta che sia resa disponibile una specificazione dichiarativa dei termini. L'analisi formale dei termini è estremamente preziosa sia per il riutilizzo delle ontologie sia per estenderle.

Una sperimentazione nel campo della Protezione Civile



Prendiamo ora in considerazione uno specifico dominio: la Protezione Civile. Un caso particolarmente interessante rispetto a quanto detto finora perché tipicamente i Sistemi Informativi Territoriali ad essa dedicati attingono dati da altri SI (quindi acquisiti e utilizzati in contesti diversi) ponendo in l'overlay, per la costruzione dello scenario di un evento calamitoso, le informazioni territoriali riferite a: ambito interessato dall'evento, caratteristiche fisiche del territorio colpito, fonti di rischio, bersagli e risorse. Consideriamo una mappa prodotta dalla componente geografica di un

siffatto SIT, come quella riportata in figura, che rappresenta lo scenario di rilascio di una sostanza tossica in un'area urbana. L'ambito è rappresentato da una porzione di carta tecnica comunale, da cui è possibile ricavare il grafo della viabilità della zona coinvolta, mentre all'interno dell'area di rilascio è evidenziata la presenza di alcuni elementi di rilievo: un ospedale (di cui conosciamo il numero complessivo di posti letto e la presenza di un pronto soccorso), una scuola (di cui conosciamo ordine e numero di studenti iscritti), la sede di un'associazione di volontari di protezione civile (di cui conosciamo il numero di aderenti, e l'elenco dei relativi materiali e mezzi disponibili). I significati espliciti ricavabili sembrano delineare una situazione piuttosto chiara per un generico lettore della carta: siamo in presenza di un evento che ha colpito bersagli umani importanti da mettere in sicurezza o evacuare al più presto, utilizzando la viabilità disponibile: gli alunni della scuola e i pazienti dell'ospedale innanzitutto, oltre a tutti i residenti nella zona dell'impatto. Per fare questo disponiamo di una risorsa, costituita dai volontari, impiegabili nelle prime fasi successive all'evento, vista la posizione della loro sede. Ma, mettendoci nei panni di un ipotetico utente privilegiato del SI, ad esempio un esperto di materia, responsabile del coordinamento dei soccorsi, scopriamo alcuni limiti importanti dell'informazione prodotta, che rendono la mappa in questione complessivamente poco utile. Ad esempio, nel caso in cui l'incidente sia avvenuto durante un giorno festivo, oppure di notte, non dovremmo più preoccuparci di evacuare gli alunni, perché la scuola sarebbe vuota. L'ospedale poi, una volta trasferiti i pazienti (cosa più semplice se la struttura ha un proprio piano di emergenza specifico), potrebbe fornire attrezzature e personale per un eventuale posto medico avanzato, posizionato al di fuori dell'area di impatto per fare un primo *triage* dei feriti e stabilizzare i più gravi prima del trasporto in strutture più sicure. I volontari poi non è detto che si rivelino una risorsa: nel caso in cui essi non siano attrezzati per affrontare un evento di questo tipo, ognuno di essi è di fatto un bersaglio, mentre nel caso di rilascio di sostanze non particolarmente pericolose, i residenti nella zona, se tempestivamente informati dai mezzi di comunicazione, restando in casa e adottando alcune semplici precauzioni possono anche non correre pericoli. La viabilità infine è un vero rebus: le strade coinvolte sono risorse, ovvero vie utilizzabili per evacuare i bersagli o far affluire i soccorsi, oppure fonti di pericolo, nel caso in cui, bloccate dal traffico si rivelino un impedimento alle operazioni di soccorso? In sintesi, possiamo dire che ognuno degli elementi territoriali rappresentati dalla mappa ha caratteristiche proprie indipendenti dall'incidente, ma in presenza di un evento può rivestire uno o più ruoli differenti (bersaglio, risorsa, fonte di pericolo), in funzione delle proprie caratteristiche peculiari, delle relazioni topologiche con altri elementi, della natura dell'incidente, del tempo, fino ad arrivare a fattori apparentemente lontani, come la sensibilità delle amministrazioni pubbliche, che portano ad investire in previsione e prevenzione (il piano di emergenza dell'ospedale, o le procedure di informazione alla popolazione). Ciò che manca al SI per rendere realmente utili dati nati in contesti diversi in un'applicazione di questo tipo è uno schema complessivo, che espliciti e formalizzi ciò, legando coerentemente le caratteristiche generali degli elementi territoriali con le regole e i significati più specifici del dominio in questione. L'idea progettuale alla base della sperimentazione consiste quindi nel costruire una base di conoscenza attraverso una gerarchia di ontologie partendo da un'ontologia *top level*, integrata con una ontologia geografica, descrivendo poi il dominio attraverso l'utilizzo di una specifica ontologia di settore, riuscendo ai nostri scopi quanto più possibile dell'analisi ontologica compiuta in seno alle comunità.

Le scelte ontologiche

Di seguito descriviamo le ontologie scelte per la sperimentazione:

Ontologie *top level*: abbiamo scelto di riferirci a DOLCE (Descriptive Ontology for Linguistic and Cognitive Engineering), realizzata dall'ISTC-CNR -Istituto per le Scienze e le Tecnologie Cognitive (<http://www.loa-cnr.it/DOLCE.html>) che offre garanzie di rigore e completezza tali da renderla un interessante punto di riferimento di livello 'alto' per esplicitare assunzioni presenti in differenti ontologie di dominio. Nel nostro caso risultano particolarmente interessanti alcune classi e relazioni descritte. La classe *perdurant*, ad esempio contiene entità costituite da parti spaziali e

temporali che possono evolvere, esattamente come nel caso di un evento calamitoso in protezione civile. La classe dei *physical objects* contiene invece elementi reali, dotati di identità univoca, che possono avere qualità fisiche, non necessariamente costanti nel tempo, esattamente come nel caso degli elementi territoriali. Tra queste qualità non costanti possiamo pensare al 'ruolo' assunto durante un evento a cui gli elementi partecipano attraverso una delle relazioni definite in Dolce (*participant in*). Altra classe importante è quella di *abstract*, che ha tra le sue sottoclassi quella definita come 'regione spaziale', che nel nostro caso può contenere la descrizione, in termini di classi e relazioni, dello spazio topologico 'cartografico', i cui elementi sono le rappresentazioni cartografiche degli elementi territoriali.

Le Geo ontologie Una delle caratteristiche salienti dei SIT è la loro capacità di sfruttare le relazioni topologiche tra entità, al fine di far emergere nuove relazioni. Come le relazioni topologiche influenzano il significato dell'informazione? Qual'è lo status ontologico delle relazioni topologiche? Ha senso considerare come delle proprietà primitive o è più conveniente o più coerente considerare la Topologia come una componente operativa della conoscenza? (Dolbear e Hart 2008). Una topologia persistente è ancora la rappresentazione di una conoscenza operativa? La posizione fisica o la forma sono proprietà essenziali di ogni entità reale: in quali condizioni possono essere usate come costituenti l'identità? In sintesi, ci sono molti interrogativi a cui speriamo di dare una risposta senza pretese di universalità ma con il requisito della riutilizzabilità. Nella indagine circa il reperimento di componenti riutilizzabili per la definizioni di ontologie abbiamo cercato le controparti semantiche agli standard geografici di riferimento quali, ad esempio l'Iso tc211, e a quelli promossi dall'OGC poiché questi definiscono, anche se non formalmente, un vocabolario condiviso dei concetti comuni del dominio geospaziali. È notevole l'iniziativa del Geospatial Semantic Web promossa da OGC (<http://www.opengeospatial.org/projects/initiatives/gswie>) i cui obiettivi sono quelli di avviare la formulazione di una ontologia geospaziale e di estendere le specifiche dei servizi WFS affinché possano erogare informazioni codificate in OWL e accettare richieste sotto forma di interrogazioni semantiche.

All'interno di W3C il Geospatial Incubator Group (<http://www.w3.org/2005/Incubator/geo/>) ha affrontato i problemi relativi all'ubicazione geografica. I campi di attività del gruppo consistono nell'aggiornamento del vocabolario W3C GEO, e la formulazione di proposte per raccomandazioni per una più completa ontologia geospaziale. L'ente cartografico britannico Ordnance Survey (<http://www.ordnancesurvey.co.uk/oswebsite/ontology/>) ha attivato una iniziativa volta a definire gli aspetti ontologici delle informazioni spaziali e pubblica alcune ontologie con lo scopo di migliorare fruizione delle informazioni territoriali distribuite.

In ambito strettamente italiano all'interno dell'iniziativa Intesa Gis la questione di una possibile ontologia geospaziali è stato affrontato in (IntesaGIS 10013) che sostiene che sia "necessaria una attività di approfondimento dell'ontologia dei contenuti e soprattutto di quella delle diverse realtà del mondo dei Sistemi Informativi Territoriali e dell'elaborazione geografica per la realizzazione di un approccio ontologico alla gestione del Data Base, sia a fronte del suo stendersi alle varie applicazioni, sia ad una sua apertura alla comunità europea e mondiale." E' chiaro che le specifiche di contenuto Intesa potrebbero costituire una base comune ai fruitori e fornitori nazionali su cui costruire servizi fortemente interoperabili.

Ontologia di dominio: per specializzare l'ontologia *top level* descrivendo concetti e relazioni tipici del dominio della protezione civile, abbiamo utilizzato un'ontologia definita dall'Istituto per la Protezione e Sicurezza dei cittadini del Centro Comune di Ricerca dell'Unione Europea (JRC) nell'ambito di un accordo quadro stipulato con la Regione Piemonte, finalizzato alla redazione del "Programma Regionale di Previsione e Prevenzione dei Rischi". L'obiettivo generale del progetto era la definizione di uno schema generale di acquisizione delle informazioni sul quale basare lo sviluppo di un Sistema Informativo a supporto delle attività di protezione civile. Il percorso del progetto, dettato da un interessante schema metodologico, è partito con un'analisi particolareggiata del contesto in cui agisce la Protezione Civile e delle sue funzioni, basata sullo studio del corpo normativo riguardante le disposizioni in materia. Questa prima fase, assimilabile in qualche modo

ad un'analisi ontologica, seppur molto orientata all'obiettivo del progetto, ha portato ad una definizione del dominio e all'identificazione di alcune condizioni al contorno che hanno permesso di individuare le tipologie di informazioni rilevanti e formalizzarle in un'ontologia che risulta particolarmente interessante per i nostri scopi in virtù di un'organizzazione delle informazioni ispirata allo schema logico di definizione di rischio, in particolare l'aggregazione degli elementi nelle classi *componenti del rischio* e *valutazione del rischio* (Di Mauro et al. 2005). Della prima abbiamo utilizzato:

- la struttura delle sottoclassi della classe 'fenomenologia incidenti per la stima degli effetti primari', in cui gli eventi calamitosi sono organizzati tassonomicamente secondo le fenomenologie con cui si presentano e non, come di solito accade, secondo una classificazione delle relative fonti di rischio. Nella nostra ontologia generale, quindi, le classi fenomenologiche dell'ontologia di JRC sono diventate sottoclassi della classe '*perdurant*' di DOLCE, ereditandone tutte le proprietà.

- la struttura delle sottoclassi della classe 'elementi esposti', nella nostra ontologia generale sono diventate le sottoclassi della classe '*physical object*' all'interno delle quali sono organizzati gli elementi territoriali.

Della seconda invece la classe 'analisi del rischio' è diventata, nell'ontologia generale una sottoclasse di non *physical enduring*, che permette la descrizione degli elementi concettuali che contribuiscono alla definizione del concetto di rischio, tra cui aspetti psicologici e politici.

Il collegamento con i dati Una volta descritto il nostro dominio di conoscenza resta la necessità di convertire l'informazione contenuta nei dati in una forma di conoscenza coerente in modo che un motore di interrogazione semantico o un ragionatore possa accedervi. Il modello logico per la rappresentazione dell'informazione usato nel web semantico consiste essenzialmente di un'insieme di triple <oggetto,predicato,oggetto>. Tale modello implementato (RDF) è alternativo al modello relazionale usualmente impiegato per la gestione delle basi dati. Per far confluire nella base di conoscenza i dati *legacy* è necessario quindi una loro ricodifica. In seguito al maturare delle tecnologie del Web Semantico vi è una crescente necessità di accedere ai contenuti di basi dati relazionali senza dover replicarne l'intero contenuto. Le soluzioni oggi disponibili sono generalmente basate su un formalismo che permette di descrivere come le informazioni contenute in una base dati relazionale siano interpretabili come triple RDF. La specifica di conversione da tabelle a RDF è assimilabile ad una descrizione delle informazioni contenute nella base dati ed è facilmente esprimibile attraverso una specializzazione RDF: si veda ad esempio (Bizer e alt.2007), (Barrasa e alt.2004). Le triple estratte dalla base dati possono essere virtualmente importate nella ontologia e costituire una base di conoscenza completa su cui è possibile condurre interrogazioni o inferenze.

Riferimenti e risorse

J.Barrasa, Ó.Corcho, A.Gómez-Pérez 2004 "*R2O an Extensible and Semantically Based Databases-to-ontology Mapping Language*"

C. Bizer ,R. Cyganiak ,J. Garbers , O. Maresch 2007 "Treating Non-RDF Relational Databases as Virtual RDF Graphs"

Di Mauro C. et al. 2005, "*Definizione degli standard di acquisizione delle informazioni a supporto delle attività di Protezione Civile – Rapporto Finale*", pp 17-21 (http://www.regione.piemonte.it/protezionecivile/images/PDF/rapporto_apr05.pdf)

C.Dolbear, G.Hart 2008 "*Ontological Bridge Building - using ontologies to merge spatial datasets*" in proceedings of the AAAI Spring Symposium

N.Guarino 1998, "*Formal Ontology in Information Systems*". Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998", Amsterdam, IOS Press, pp. 3-15.

[IntesaGIS 10013] "Stato dell'arte e linee di sviluppo" <http://www.cnipa.gov.it/>

N.F.Noy, D.L.McGuinness 2001 "*Ontology Development 101: A Guide to Creating Your First Ontology*"