

Primi elementi di un quadro conoscitivo per la modellazione delle conoscenze geografiche

Robert Laurini
LIRIS – INSA de Lyon – F69621 Villeurbanne
Robert.Laurini@insa-lyon.fr

Nell'ambito dei SSIITT, è comune parlare di dati geografici ed anche delle conoscenze spaziali allorchè abbiamo bisogno per la gestione e la pianificazione del territorio di conoscenze geografiche dette anche territoriali. I dati geografici stoccano non solo i dati spaziali (coordinate, ecc.) ma anche i dati detti non-spaziali o attributi alfanumerici, cioè le altre caratteristiche generalmente nei tipi alfanumerici come la popolazione, l'uso del suolo, ecc. Una conoscenza può essere definita come un'informazione che serve a risolvere un problema. Di più, definisco le conoscenze spaziali come costruite solo sui dati spaziali allorchè le conoscenze geografiche sui dati non-spaziali. "*La città di Belluno è localizzata al nord di Venezia*" è una conoscenza spaziale; ma "*a Venezia, ci sono più abitanti nella terraferma della parte storica*" è una conoscenza geografica.

Ma come definire e organizzare le conoscenze geografiche? Lo scopo di questo paper sarà quello di definire i primi elementi di un quadro conoscitivo, cioè di elencare i principi elementari necessari alla costruzione di questo quadro. Di fatto, con questo quadro sarà possibile veritabilmente di organizzare un ragionamento geografico. In questo paper, non tratterò degli aspetti temporali.

Questo paper sarà organizzato come segue. Dopo aver presentato i principi della modellazione delle conoscenze geografiche, parlerò di alcuni modi per estrarre queste conoscenze. In fine, i problemi relativi alla rappresentazione matematica saranno sviluppati. E concluderò sulla necessità di raffinare questo quadro conoscitivo.

1. Principi della modellazione delle conoscenze geografiche

In questa sezione, saranno presentati i concetti di base necessari per la costruzione dei principi.

1.1. Concetti di base

Avanti di elencare questi principi, devo stabilire il vocabolario.

La rappresentazione numerica degli oggetti geografici è generalmente basata sulla geometria euclidiana sul piano o sulla sfera. Quando conosciamo i confini, gli oggetti sono modellati come poligoni, altrimenti la geometria sfumata (fuzzy set) può anche essere utilizzata. Nell'insieme degli oggetti geografici, possiamo considerare:

a/ **gli oggetti geodetici**: sono alla base delle coordinate sul geoide, e comprendono l'equatore, i poli Nord e Sud, i meridiani ed i paralleli. Sono modellati con punti, e cerchi.

b/ **gli oggetti amministrativi** (regioni, provincie, comuni, particelle, parchi naturali, ecc.): sono generalmente descritti come poligoni 2D non-connessi e generalmente formano tesselazioni.

c/ **gli oggetti antropici**: cioè gli edifici, i ponti, le gallerie, ecc. Sono oggetti euclidiani tridimensionali. Ma, nei SSIIT classici sono rappresentati solo a 2D.

d/ **gli oggetti naturali**: non conosciamo generalmente i loro confini; per un fiume, dobbiamo considerare l'alveo di magra e l'alveo di piena; non si sa dove incomincia o finisce una montagna; il mare si muove, ecc. Per questi oggetti, il modello fuzzy può essere considerato. Per i fenomeni naturali, come la temperatura, la pressione, il vento, i modelli basati sui campi continui sono apprezzati (Couclelis, 1992).

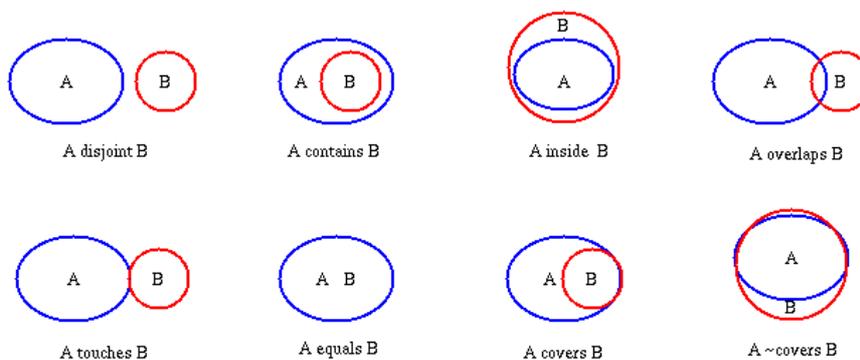
In questo quadro conoscitivo, propongo di avere due rappresentazioni, un'unica rappresentazione per lo stoccaggio nel database con la massima precisione. Per la visualizzazione, propongo di generare le altre rappresentazioni con algoritmi di generalizzazione (del tipo Douglas e Peucker, 1973).

Le relazioni spaziali tra gli oggetti geografici sono di parecchie categorie.

a/ **Relazioni topologiche**: quando i confini sono conosciuti, le relazioni (Allen 1983, Egenhofer 1994, ecc.) vigiono. Per gli oggetti sfumati, è un po' più complicato. Un modello interessante è quello dell'uovo-tuorlo (egg-yolk) (Cohn e Gotts, 1996) in cui ogni oggetti a due confini, il "cuore"

(tuorlo), e l'estensione massima (uovo).

Fig. 1 – Relazioni topologiche di Egenhofer (1994).



b/ **Relazioni proiettive** legate ai punti cardinali (Nord, Sud, Est, Ovest) che hanno proprietà speciali; niente al nord del Polo Nord, e transitività est-ovest legata alla sfera.

c/ **Relazioni di distanza:** due oggetti possono essere vicini o remoti. Ad esempio: un aeroporto è a poca distanza di una città.

d/ **Relazioni mereologiche:** tra le parti non connesse di un poligono. Ad esempio, la Sicilia ha una relazione mereologica con l'Italia continentale.

1.2. Principi

Ho identificato sette principi. Ma all'inizio, devo ricordare la legge di Tobler (Tobler, 1970) come principio 0.

Principio 0: Legge di Tobler.

"Everything is related to everything else, but near things are more related than distant things." Cioè, *"tutte le cose hanno relazioni con altre cose, ma le cose vicine sono più relate delle cose remote"*.

Una conseguenza è che si deve anche stoccare alcune informazioni all'esterno del sistema. Ma la maggioranza dei Sistemi Informativi Territoriali che esistono si limitano ai dati interni alla loro giurisdizione, e niente fuori. Ad esempio, al confine della Regione Veneto, sarebbe bene di dire che una parte confina al mare, una parte all'Austria, una parte alla Lombardia, perchè questi parti del confine non hanno le stesse caratteristiche.

Principio 1: Origine delle conoscenze

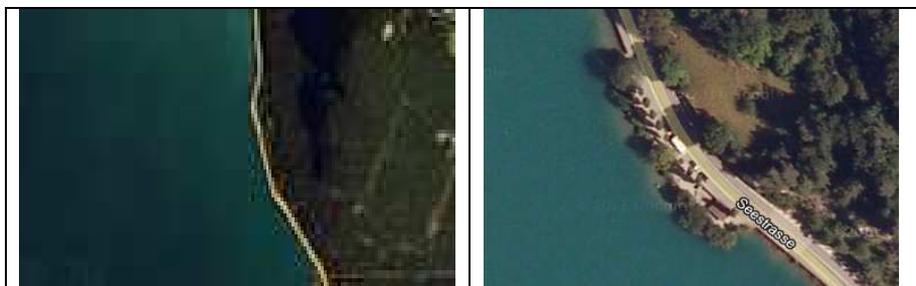
"Le conoscenze spaziali sono nascoste nelle coordinate, e in più, le conoscenze geografiche negli attributi non-spaziali". E' utile di aggiungere che queste conoscenze sono spesso implicite. Ad esempio, se prendiamo le coordinate della città di Belluno e quelle della città di Venezia, possiamo derivare che "Belluno sta al nord di Venezia". Quindi, dobbiamo intavolare la domanda di sapere se dobbiamo stoccare questa conoscenza o guardarla come implicita, cioè di derivarle "on-demand".

Gli indicatori sono visti come conoscenze derivate dagli attributi generalmente numerici.

Principio 2: Variazione delle relazioni spaziali secondo le scale

"Le relazioni spaziali tra gli oggetti geografici variano secondo le scale". Nel linguaggio corrente, si dice facilmente che una strada costeggia il mare". Ma in realtà, la strada non tocca sempre l'acqua del mare; in certi luoghi, ci sono spiagge, può darsi edifici, ecc. Dal punto di vista matematico, a una scala, il mare tocca la strada, ma a un'altra, sono disgiunti. Matematicamente, con le relazioni di Engenhofer, abbiamo sia touches, sia disjoint.

Fig. 2 – Esempio della strada che costeggia il lago. Secondo le scale, la strada tocca o non tocca il lago.



Principio 3:

"Non è indispensabile enumerare tutte le conoscenze spaziali". Se abbiamo n oggetti, allora n^2 relazioni Nord-Sud. Di fatto, possiamo derivarle tutte automaticamente. Allora, propongo di limitarsi alla regola di Tobler.

Principio 4: Passaggio della sfera al piano

"Su piccoli spazi, basta una rappresentazione planaria (coordinate cartesiane); per gli spazi più grandi, si deve tenere conto della rotondità

della terra". Ma come definire un piccolo o un grande spazio ? Propongo di prendere una soglia, ad esempio in Italia, una provincia. Quindi, per i comuni basta una proiezione planaria; ma definitivamente no per una regione.

Principio 5: Visualizzazione e acutezza visiva

"La rappresentazione di visualizzazione è legata all'acutezza visiva". In cartografia classica, la limite sta tra mm 1 e 0.1. Ad esempio, se prendiamo una strada di m 10 di larghezza, a certe scale deve essere visualizzata come una superficie, ad altre come una linea, e dopo sparisce. Lo stesso ragionamento è anche valido per una città o un piccolo paese, una superficie; dopo un punto; dopo niente.

Ma se un piccolo oggetto sta nella query, allora lui deve essere visualizzato sebbene non sarebbe visto alla scala. Alla scala 1:50 000, una strada di m 10 di larghezza dà mm 0.4, cioè diventa una linea allorchè un vicolo di m 4 che dà mm 0.05 quindi sparisce.

Principio 6: Vincoli di visualizzazione.

"Le relazioni topologiche devono rimanere sempre salvo quando sparisce un oggetto". Nella figura 3 è rappresentata la costa francese del Mediterraneo tra Italia e Spagna. Se generalizziamo la costa in una linea tra questi due paesi, Nizza si ritrova in medio al mare, ma Marsiglia e Montpellier stanno nella terra ferma. In più, abbiamo perso il delta del Rodano. Per stabilire le relazioni topologiche, dobbiamo muovere questi porti in modo che le relazioni touches tra essi e il mare stanno corretti. Lo stesso ragionamento vige per collegare il Rodano al mare.

Un altro esempio sarebbe tra la Francia e la città di Ginevra: in questo caso, Ginevra deve sempre rimanere al fuori dalla Francia.

Principio 7: Influenza degli oggetti vicini

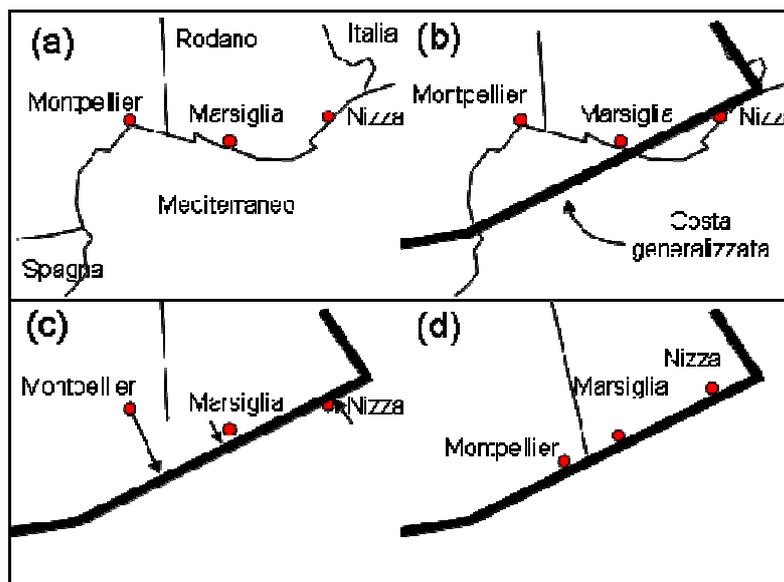
"In una base di conoscenze spaziali, non dimenticare gli oggetti geografici vicini che possono avere una grande importanza". Ad esempio, la città di Ginevra sebbene non appartiene alla regione francese Rhône-Alpes, ha molto importanza sullo sviluppo economico e sociale nella regione Rhône-Alpes.

2. Fonti delle conoscenze geografiche

Da dove vengono queste conoscenze? Negli anni '80, era comune parlare di sistemi esperti in cui gli esperti stessi erano la fonte principale delle conoscenze. E' sempre una possibilità, pero' adesso esistono altre

fonti molto più efficaci, cioè il data mining spaziale e l'analisi dei documenti.

Fig. 3 – Conseguenze della generalizzazione della costa: i porti si devono muovere e il confluente del Rodano deve essere corretto.



2.1. Data mining spaziale

L'analisi sistematica dei dati stoccati nei database è il modo privilegiato per estrarre le conoscenze. Molti ricercatori hanno lavorato sull'estrazione delle conoscenze spaziali (Shekhar et al., 2003) cavando soprattutto le entità co-localizzate. Nel suo lavoro, Cimbra (2008) ha identificato le conoscenze seguenti:

- a/ i fatti su luoghi
- b/ gli aggruppamenti tra luoghi (cluster)
- c/ i flussi tra luoghi
- d/ la colocalizzazione di fatti
- e/ i vincoli tra i luoghi (rivolgersi al principio 6)
- f/ fatti importanti nei luoghi confinanti (principio 7).

2.2. Analisi di documenti

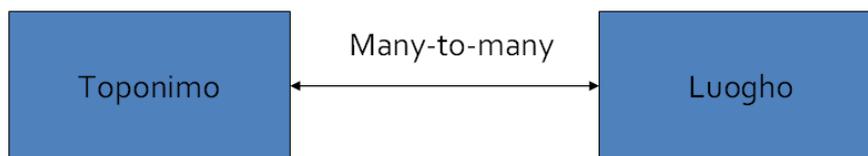
Un'altra fonte di conoscenze geografiche è basata sui documenti testuali e multimediali. In questo paper, tratterò solo i documenti testuali. Il

GeoParsing è conosciuto come l'analisi per estrarre l'informazione geografica utilizzando due risorse informazionali, le ontologie ed i gazetteers.

Le ontologie sono reti semantiche di vocabolario generalmente con relazioni semantiche *is-a* e *part-whole*. Esistono ontologie geografiche con queste relazioni semantiche, ma adesso alcune relazioni spaziali sono anche integrate (Laurini 2012).

Un gazetteer è un elenco dei nomi di luogo (toponimi). Nel passato era solo una lista, ma adesso abbiamo veritabilmente un database. Tra i luoghi e i nomi, ci sono relazioni di tipo "many-to-many". Di fatto, Mississippi è il nome di uno stato ed anche di un fiume. Ci sono parecchie città nel mondo che si chiamano Córdoba. Secondo le lingue, la città di Venezia é anche conosciuta come Venice, Venise, Venecia, Venedig, Benetke, Benátky, ecc. C'è anche il cambio di nome nel tempo: Bisanzio, Costantinopoli, Istanbul. E il confine varia col tempo: ad esempio la Roma di Romulus e quella di oggi.

Fig. 4 – Relazione "many-to-many" tra i luoghi ed i toponimi.



3. Rappresentazione digitale

Per rappresentare le conoscenze, esistono quattro modelli, le asserzioni testuali, la logica descrittiva, i dialetti di XML ed i coremi.

3.1. Asserzioni testuali

Presentare le conoscenze testualmente con frasi è un modo molto comune e molto interessante, ma questa rappresentazione non è facilmente "machine-treatable", cioè ci sono difficoltà per avviare ragionamenti automatici. I soli esistenti strumenti sono per la visualizzazione di queste conoscenze. Esempio: "*dove ci sono un lago e una strada che porta al lago, allora c'è un ristorante alla vicinanza*".

3.2. Logica descrittiva

Con la logica descrittiva, si può facilmente scrivere queste conoscenze.

Sono machine-treatable ma l'utente finale ha alcune difficoltà non solo per capirle, ma anche per scriverle. Esempio:

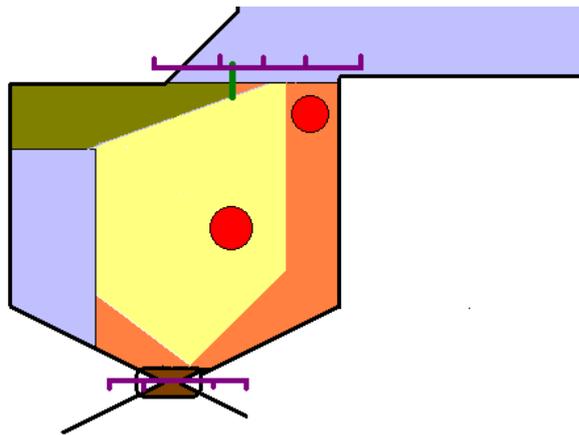
$$\begin{aligned} &\forall l \in \text{Laghi} \wedge \forall s \in \text{Strade} \wedge (\text{touches}(l, s)) \\ &\Rightarrow \\ &\exists r \in \text{Ristoranti} \wedge (\text{distanza}(r, l) < 100 \wedge (\text{distanza}(r, s) < 100) \end{aligned}$$

Devo aggiungere che questa rappresentazione è totalmente machine-treatable.

3.3. Dialetti di XML

Un modo particolare è l'uso di un dialetto di XML, cioè è machine-treatable.

Fig. 5 – Corema della Spagna.



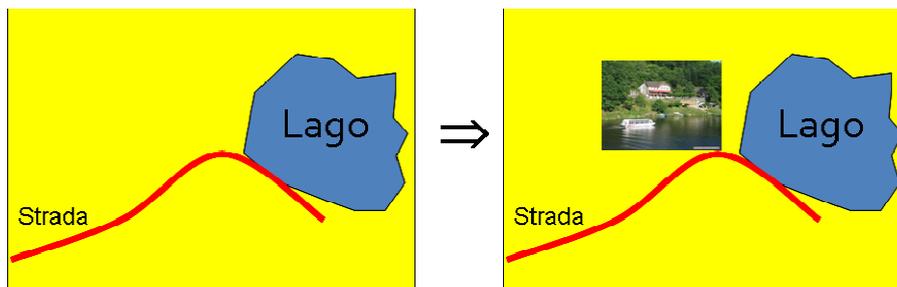
3.4. Coremi

Un nuovo modello è una rappresentazione visuale: i coremi creati da Brunet (1980) è una rappresentazione schematizzata di un territorio cioè che illustra solo le cose importanti. Ma questi coremi possono anche DelFatto et al. 2007).

La Figura 5 dà un corema della Spagna dove si può vedere che il paese è semplificato, ci sono due città principali, c'è un deserto nel centro, il nord ovest con le colture, e le coste orientale e meridionale sono saturate.

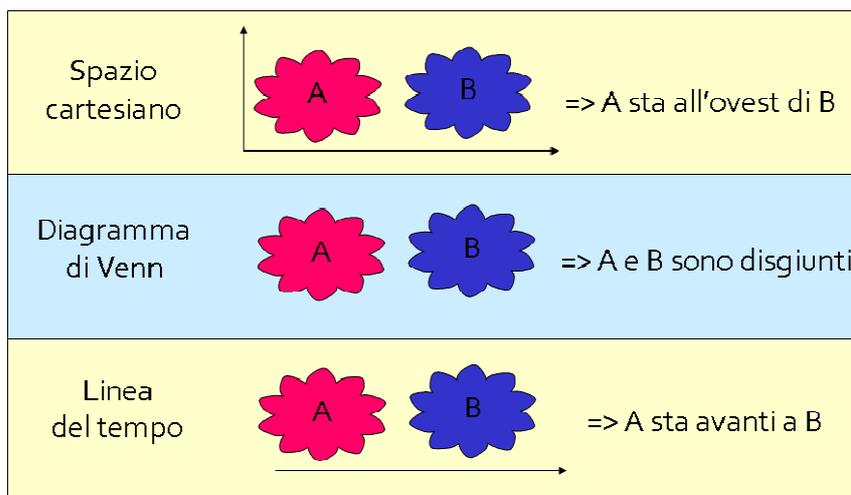
La figura 6 ci dà la rappresentazione coremica del esempio.

Fig. 6 – Corema per rappresentare la conoscenza tra il lago, la strada e il ristorante.



Ma per il loro uso, è necessario definire correttamente lo spazio d'interpretazione. Di fatto, esistono al minimo tre spazi e la Figura 7 illustra questi spazi.

Fig. 7 – Contesti di interpretazioni dei coremi.



- a/ lo spazio cartesiano in cui l'oggetto A sta all'ovest di B.
- b/ un diagramma di Venn in cui i due oggetti sono disgiunti.
- c/ una linea temporale in cui A sta avanti a B.

Nel caso della Figura 6, siamo nel caso del diagramma di Venn.

4. Conclusioni

Più di vinti anni fa, avevo scritto un libro (Laurini e Milleret-Raffort 1989) sulle conoscenze spaziali, ma adesso, dobbiamo ripensare questo dominio. In questo paper, i primi elementi di un quadro conoscitivo per le conoscenze geografiche sono accennati. Questo quadro deve essere validato e ammegliorato. Può darsi esistono altri principi che devono essere aggiunti per coprire tutti gli aspetti.

Di più, sebbene la logica descrittiva sia sufficientemente potente per modellare le conoscenze geografiche, l'interesse per le altre rappresentazioni grandisce.

Bibliografia

- Allen J.F. (1983) *Maintaining knowledge about temporal intervals*. In: Communications of the ACM. 26 November 1983. ACM Press. pp. 832-843, ISSN 0001-0782
- Brunet R. (1980) *La composition des modèles dans l'analyse spatiale*, in L'Espace géographique, n° 4, 1980
- Cohn A. and N. Gotts. (1996) *The "Egg-Yolk" Representation of Regions with Indeterminate Boundaries*. in: P. Burrough and A. Frank (Eds.) *Geographic Objects with Indeterminate Boundaries*: 171-187. London, Taylor & Francis.
- Coimbra A.R. (2008). *ChorML: XML Extension For Modeling Visual Summaries of Geographic Databases Based on Chorems*, Master project, INSA Lyon.
- Couclelis H. 1992) *People Manipulate Objects (but Cultivate Fields): Beyond the Raster-Vector Debate in GIS*. *Spatio-Temporal Reasoning 1992*: 65-77.
- Del Fatto V., Laurini R., Lopez K., Loreto R., Milleret-Raffort F., Sebillo M., Sol-Martinez D., Vitiello G. (2007) *Potentialities of Chorems as Visual Summaries of Spatial Databases Contents*, VISUAL 2007, 9th International Conference on Visual Information Systems, Shanghai, China, 28-29 June 2007. Edited by Qiu, G., Leung, C., Xue, X.-Y., Laurini, R., Springer Verlag LNCS, Volume 4781 "Advances in Visual Information Systems", pp. 537-548.
- Douglas D., Peucker T. (1973) *Algorithms for the reduction of the number of points required to represent a digitized line or its caricature*, *The Canadian Cartographer* 10(2), 112-122 (1973)
- Egenhofer M. (1994) *Deriving the Composition of Binary Topological Relations*. *Journal of Visual Languages and Computing* 5(2): 133-149.
- Laurini R. (2012) *Importance of spatial relationships for geographic ontologies*. In "Planning Support Tools: Policy Analysis, Implementation and Evaluation" Proceedings of the Seventh International Conference on Informatics and Urban and Regional Planning INPUT 2012 (Editors: Michele Campagna, Andrea De Montis, Federica Isola, Sabrina Lai, Cheti Pira, Corrado Zoppi) (ISBN code: 9788856875973), pp. 122-134.
- Laurini R. Milleret-Raffort F., (1989) *L'ingénierie des connaissances spatiales*. Paris, HERMES, Collection Technologies de pointe, n° 30.
- Shekhar S., Zhang P., Huang Y., Vatsavai R. (2003) *Trend in Spatail Data Mining*, as a chapter to appear in *Data Mining: Next Generation Challenges and Future Directions*, H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha(eds.), AAAI/MIT Press, 2003,
- Tobler W., (1970) *A computer movie simulating urban growth in the Detroit region*". *Economic Geography*, 46(2): 234-240.