Spatial Webs: Position Paper

Anthony G Cohn
School of Computing, University of Leeds, LS2 9JT, UK
agc@comp.leeds.ac.uk

]

I have a number of interests related to the topic of the workshop and I describe these briefly below. First I describe my general interest in spatial ontology, then I describe a new project on spatial data integration, and finally I discuss how work I have been conducting in cognitive vision may be relevant.

1. Spatial Ontology

Fundamental to any Spatial Information System, including a Spatial Web, is a characterisation of spatial entities and the relationships between them. This is an area I have worked in since the late 1980s, concentrating in particular on Qualitative Spatial and Spatio-temporal reasoning, particularly mereotopological relations over regions [5], but also on orientation [12]. The Leeds work on mereotopology is based on a fundamental notion of two entities being *connected*, and while the resulting calculus (called the Region Connection Calculus (RCC) has much in common with the 4 and 9-intersection calculi (e.g. [9]), the setting within a full first order language means that it is fundamentally more expressive, rather than simply being a *constraint language*, although for computational reasons, the reduced expressive power of a constraint language can be advantage. We have also developed a more expressive language, *Region Based Geometry* [3]. This gives a very expressive language, allowing complex shape information to be specified, but with a corresponding decrease in computational efficiency, as would be expected. An extension of RCC has been formulated to explicitly handle regions with indeterminate boundaries [6]. A review of the whole area of qualitative spatial reasoning can be found in [8].

Since many applications also have to consider how spatial entities change, and change their relationships to each other over time, we have also considered spatio-temporal calculi, in particular calculi based on spatio-temporal regions [7]. Fundamental to the notion of most notions of change is the idea that changes must, in some sense, be continuous. We have analysed a variety of notions of qualitative continuity, of varying strengths, including purely spatial, and purely temporal continuity [7].

2. Mapping the Underworld

Every year, in excess of four million holes are dug in UK roads to repair leaks, provide connecting services to new premises and to lay new cables and pipes[4]. Although recently installed assets may have been well mapped, location data on older services can be very poor, in some cases even non existent (except perhaps knowing the location of the terminating points). Some of the holes are unnecessary (dug in the wrong place owing to insufficient or wrong data), some cause third party damage to other underground services (or even first party damage!). More importantly, there are also considerable indirect costs owing to disruption on the roads caused by works, waste, and pollution. A recent Engineering and Physical Sciences Funding Council (EPSRC) initiative initiative is funding £1M of research to ameliorate the situation, ranging from better sensing devices to improved construction techniques to engendering better cultural relations and education of the relevant parties. The core of the problem is that there is at present insufficient and inadequate knowledge about *what is where*. The University of Leeds has received funding under this initiative to tackle this central issue. Even with improved technology in the future, it is unlikely that we will ever gain complete knowledge of the underworld (or indeed know that we have – proving a negative is always difficult), and thus the problem of ensuring that maximum benefit is gained from existing knowledge is crucial: the estimated annual total of indirect and direct costs of maintaining the nation's underground infrastructure is in excess of £3B [4] – thus even a small improvement could have great benefits.

Knowledge and data integration has long been recognised as an important practical problem, having interesting theoretical aspects too (e.g. [14, 13, 11]). Lenzerini [14] has identified some of the main problems in data integration: 1. Heterogeneity of sources (intensional and extensional level); 2. Limitations in the mechanisms for accessing the sources; 3. Materialized vs virtual integration; 4. Data extraction, cleaning and reconciliation; 5. How to process updates expressed on the global schema, and updates expressed on the sources; 6. The querying problem: How to answer queries expressed on the global schema; 7. The modelling problem: How to model the global schema, the sources, and the relationships between the two. Of particular interest to us here, is the integration of spatially referenced data. There are several reasons why the spatial integration problem is special, including that not all data is necessarily symbolic, and that spatial proximity influences the dependence of one object on another. We believe that constructing an *ontology* of the domain will be key to a successful integration. This term is used by philosophers to describe that branch of metaphysics that deals with the question of 'What things exist?', In recent years 'ontology' has become a buzz word in information science, where it refers to a rigorous formal specification of a vocabulary of concepts and relations. Such specifications are playing an increasingly important rôle in ensuring the integrity of data both within single, information-rich applications and in the transfer of data between applications. There have been a number of proposals for architectures for spatial data integration, e.g. [10, 1] and for specific techniques to integrate, or *conflate* spatial data, whether statistical, e.g. [18], fuzzy/rough-set, e.g. [2] or ontological, e.g. [10]; these techniques specifically attempt to address the issue that in practice data is imperfect, incomplete and noisy and this needs to be considered explicitly, e.g. [20]. Also relevant is research on integrating thematic information, such as [19].

3. Inferring Spatio-Temporal Behaviours

As part of a project in the area of *Cognitive Vision*, we have been engaged in building systems that can take low level, noisy, perceptual data, and autonomously build high level descriptions of the behaviours present in video sequences. The domain we have been considering is in "table top space" [16], learning simple table top card games, but in principle the techniques we have developed might be applied in other spaces, including geographic and environmental space. We cluster the low level perceptual data; each cluster then becomes a symbol which is used as part of a relational description of observed behaviour; and inductive logic programming system is then used to induce high level rules which generalise the observed behaviours. Further details can be found in [15, 17].

References

[1] D. J. Abel, Beng Chin OoI, Kian-Lee Tan, and Soon Huat Tan. Towards integrated geographical information processing. *Int. J. Geographical Information Science*, 12(4):353–371, 1998.

[2] O Ahlqvist, J Keukelaar, and K Oukbir. Rough and fuzzy geographical data integration. *Int J GIS*, 17(3):223–234, 2003.

[3] B. Bennett, Anthony G. Cohn, Paolo Torrini, and Shyamanta M. Hazarika. A foundation for region-based qualitative geometry. In W. Horn, editor, *Proceedings of ECAI-2000*, pages 204–208, 2000.

[4] M Burtwell, E Faragher, D Neville, C Overton, C Rogers, and T Woodward. Locating underground plant and equipment – proposals for a research programme, UKWIR, TR 03/wm/12/4.

[5] A G Cohn, B Bennett, J M Gooday, and N M Gotts. RCC: a calculus for region based qualitative spatial reasoning. *GeoInformatica*, 1:275–316, 1997.

[6] A G Cohn and N M Gotts. A mereological approach to representing spatial vagueness. In J Doyle L C Aiello and S Shapiro, editors, *Principles of Knowledge Representation and Reasoning, Proc. 5th Conference*, pages 230–241. Morgan Kaufmann, 1996.

[7] A. G. Cohn and S. M Hazarika. Continuous transitions in mereotopology. In *Commonsense-2001: 5th Symposium on Logical Formalizations of Commonsense Reasoning*, 2001.

[8] A G Cohn and S M Hazarika. Qualitative spatial representation and reasoning: An overview. *Fundamenta Informaticae*, 46(1-2):1–29, 2001.

[9] M Egenhofer and R Franzosa. Point-set topological spatial relations. *International Journal of Geographical Information Systems*, 5(2):161–174, 1991.

[10] Frederico T. Fonseca and Max J. Egenhofer. Ontology-driven geographic information systems. In *ACM-GIS*, pages 14–19, 1999.

[11] Richard Hull. Managing semantic heterogeneity in databases: a theoretical prospective. In *Proc. ACM Symposium on Principles of Databases (Invited Tutorial)*, pages 51–61, 1997.

[12] A. Isli and A.G. Cohn. A new approach to cyclic ordering of 2d orientations using ternary relation algebras. *Artificial Intelligence*, 122(1–2):137–187, 2000.

[13] M Lenzerini. Data Integration: A Theoretical Perspective. In *Proc. PODS 2002)*, 2002.

[14] Maurizio Lenzerini. Data integration is harder than you thought (extended abstract). In *Proc. CoopIS 2001*, volume 2172 of *LNCS*. Springer, 2001.

[15] D. Magee, C. J. Needham, P. Santos, A. G. Cohn, and D. C. Hogg. Autonomous learning for a cognitive agent using continuous models and inductive logic programming from audio-visual input. In *Proceedings of the AAAI Workshop on Anchoring Symbols to Sensor Data*, 2004.

[16] D Montello. Scale and multiple pyschologies of space. In I Campari A Frank, editor, *Spatial Information Theory: a theoretical basis for GIS*, number 716 in Lecture Notes in Computer Science, pages 312–321, Berlin, 1993. Springer Verlag.

[17] P. Santos, D. Magee, A. Cohn, and D. Hogg. Combining multiple answers for learning mathematical structures from visual observation. In *Proc. of the* 16^{th} *European Conference on Artificial Intelligence (ECAI-04)*, 2004.

[18] V Walter and D Fritsch. Matching spatial data sets: a statistical approach. *Int. J. Geo. Inf. Sci.*, 13(5):445–473, 1999.

[19] M. F. Worboys and M. Duckham. Integrating spatio-thematic information. In *GIScience*, 2002.

[20] M.F. Worboys and E. Clementini. Integration of imperfect spatial information. *J. Vis. Lang. & Comp.*, 12:61–80, 2001.