An evidence-based approach to collaborative ontology development

Introduction

Ontologies are increasingly visible in various disciplines, particularly in the area of knowledge management, where the encoding of static domain knowledge is a key process (Aldea et al, 2003). They are generally applied for a number of purposes, including (Noy and McGuinness, 2001):

- Sharing common understanding of the structure of information
- Enabling reuse of domain knowledge
- Making domain assumptions explicit
- Separate domain knowledge from the operational knowledge;
- Analysis of domain knowledge.

Methods for ontology generation identified can be described as

- Introspection, or self-reflection
- Collaborative development
- Data-driven or corpus-driven means

Linking two or more of these methods together is also possible, and collaborative systems are being employed.

Reports synthesised from expert knowledge have the advantage of very closely approaching the individual's own viewpoint; if then bolstered by discussion with others, the result may approach a consensus viewpoint. Such an approach does not take into account the visibility or availability of these features within the data that is available. Under some circumstances, this characteristic is not a defect for an ontology; however, if it is to be used for a data-driven or highly data-dependent application it is advantageous for the ontology to approach the dataset.

A Software-Engineering Approach to Ontology Development

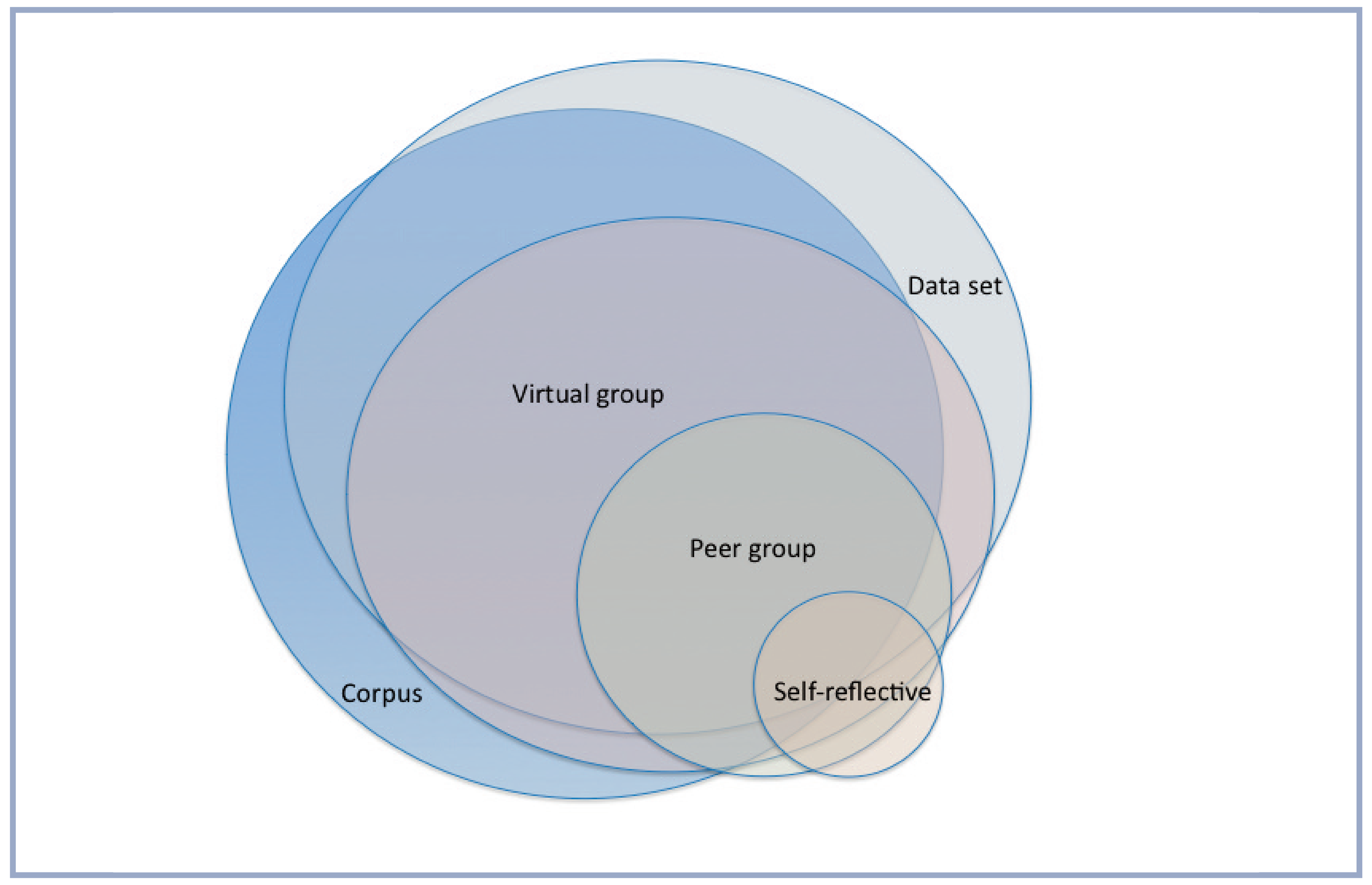
The practical importance in a collaborative context of an agile, open process was repeatedly emphasised during each case study, particularly in more interpretive developments. Agile development is supported by a number of new techniques, practices and tools; it tends to favour working solutions over future capabilities and encourages near-continuous engagement with users, non-specialist participants taking part in the development process, responding to changes in functional requirements as both the developer and the user increase their understanding of the problem space. This approach emphasises test driven development (TDD), frequent testing of software during development, rather than at the end of the process. Clear, measurable functional requirements are of importance, although the tests chosen may be revised flexibly at any point in the process. Outputs are tested frequently, during very short development iterations, which reduces the risk of 'drifting' away from core functional requirements.

Our work explores application of these principles to the development of structures designed for knowledge management (KM). Generally aimed at a clearly defined problem-space, functional requirements are derived from close association with users, domain information and evidence, in an ongoing agile process of test-driven ontology development.

Identifying Software Platforms for Ontology Development

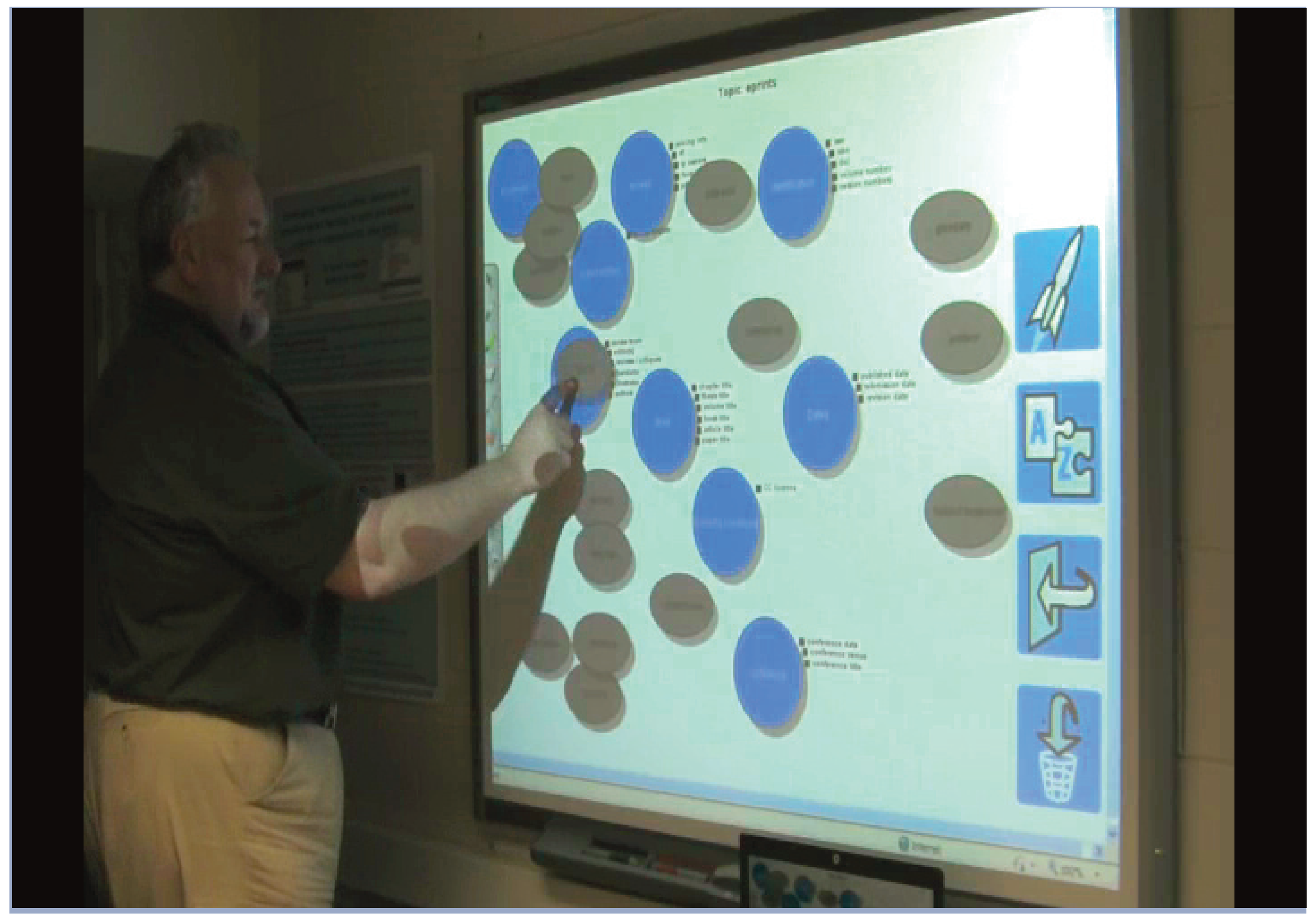
Various software have been evaluated and custom software developed for nonexpert use. A number of informal procedures for collaborative ontology development were sketched out, with particular focus on actively working with existing patterns of collaboration within the domains under investigation.

The use of a variety of tools implying very different levels of detail and technical accuracy or direct applicability implies a greater load on those responsible for completion of this progress. However, certain tools offered greater accessibility for the nonexpert. The corresponding benefit of this was that, given an accessible, simplified surrogate,



Ontology generation data sets and groups overlay with possible conflicts but bring resolution through collaboration

participants were able to quickly reach a level at which they were willing to discuss and contribute actively on issues such as the perceived quality, relevance or completeness of an ontology.



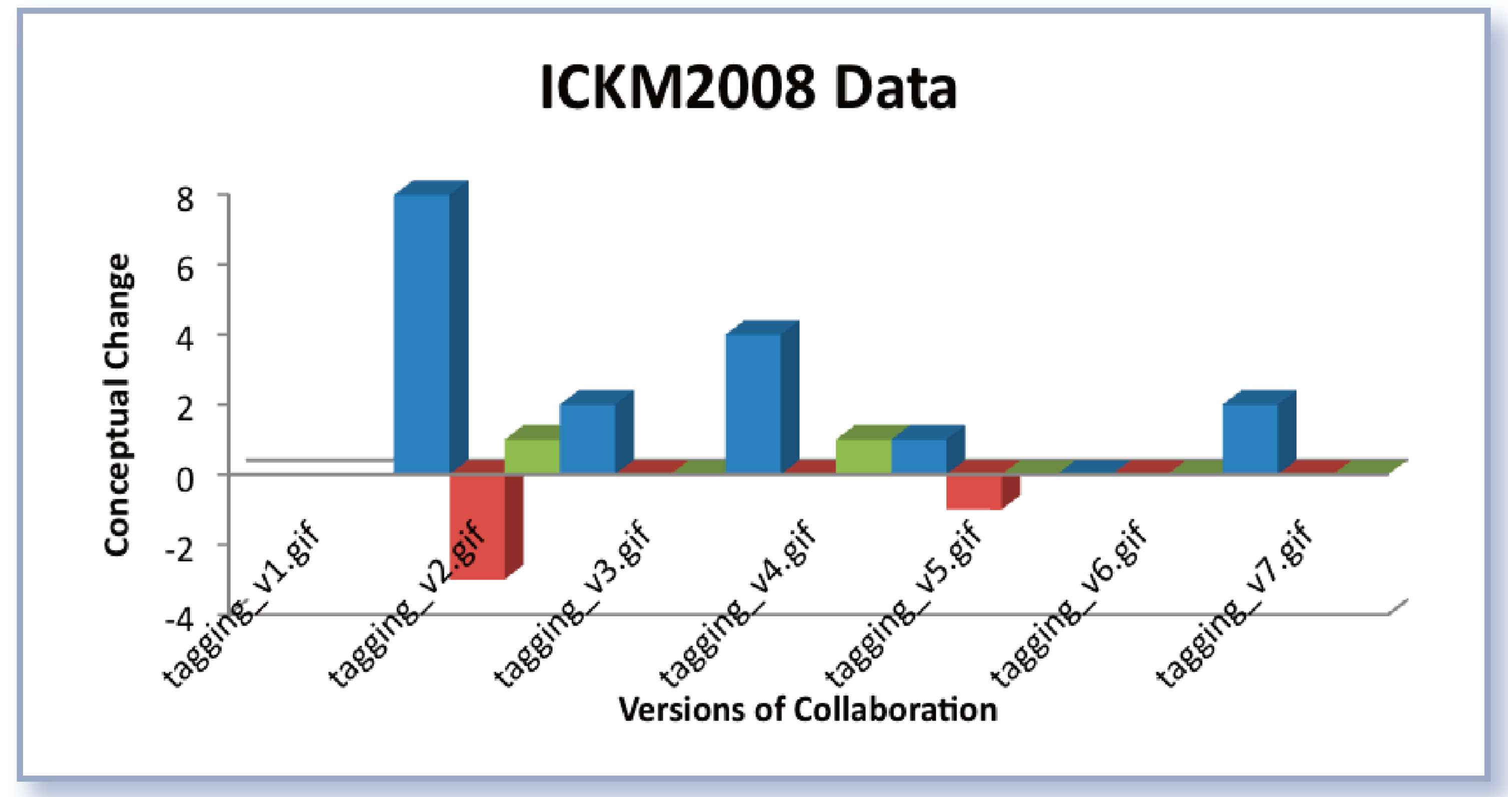
Term sorting on an electronic whiteboard

The strawman surrogates under discussion, and the results of the discussion, are usually ambiguous, incomplete or contain invalid statements or assertions, so the raw results are indicative rather than directly applicable.

Test, Metrics and Evaluation

Various methods were identified and used for evaluation of ontologies. These can be compared to the general typology offered by Brank et al (2005):

- Comparison to a gold standard
- The use of the ontology within an application
- Comparison with a source of data
- Evaluation by humans: how well an ontology meets predefined criteria



Data suggesting trend towards stability with additions, pruning & structural changes.

During the case studies, a trend towards increasing structural stability has additionally been observed, particularly in the latter case. That is, the quantity, severity and nature of requested changes has been observed in general to diminish over time, suggesting that, as Viegas et al (2004) observe in their exploration of the stability of collaboratively-created user content, the first users to work on a given structure generally sets the tone of the result and, therefore, their work usually has the highest survival rate (see graph above).

Conclusion

The work described here essentially attempts to explore how a specialised and complex process may be rendered more accessible to a wider audience of potential users and contributors. Exploring the boundaries between ontology development and related areas of research may, as suggested by Brank et al., permit participants in a development process may integrate insights from related domains. We found that prototyping approaches, using simple surrogates, are useful in encouraging discussion and input, and that the relevance of data sources is dependent on the area, aspirations and contexts of use. The usability engineer is seldom considered a key participant in ontology development, but in cases in I've which ontology development becomes a collaborative (usually computersupported) process, the participants be come a key element in the success or failure of that development process.

Emma Tonkin, **Heather D. Pfeiffer,** **and** **Andrew Hewson**

Research Officer
UKOLN
University of Bath
e.tonkin@ukoln.ac.uk

College Assistant Professor
Klipsch School of ECE
New Mexico State University
hdp@cs.nmsu.edu

Software Developer
UKOLN
University of Bath
a.hewson@ukoln.ac.uk

References

Aldea, A., Bañaresalcántara, R., Bocio, J., Gramajo, J. and Isern, D. (2003) An OntologyBased Knowledge Management Platform. In: Proceedings of the Workshop on Information Integration on the Web (IIWeb03) at the 18 th International Joint Conference on Artificial Intelligence.

Brank, J., Grobelnik, M., Mladenić, D. (2005). A Survey of Ontology Evaluation Techniques. Conference on Data Mining and Data Warehouses (SiKDD 2005), Ljubljana, Slovenia, 2005.

Noy, Natalya F. and McGuinness, Deborah L (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL0105 and Stanford Medical Informatics Technical Report SMI20010880, March 2001.

Viégas , Fernanda B., Wattenberg, M. and Dave, Kushal (2004). Studying cooperation and conflict between authors with history flow visualizations, Proceedings of the SIGCHI conference on Human factors in computing systems, p.575582, April 2429, 2004, Vienna, Austria.