The use of conceptual maps for two ontology developments: nutrigenomics, and a management system for genealogies.

Alexander Garcia[1,2,3,4], Susanna-Assunta Sansone[1], Philippe Rocca-Serra[1], Chris Taylor[1], & Mark A. Ragan[2,4]

[1] The European Bioinformatics Institute
EMBL Outstation – Hinxton
Wellcome Trust Genome Campus, Cambridge CB10 1SD, UK
{agarcia, sansone, rocca}@ebi.ac.uk
[2] Institute for Molecular Bioscience
The University of Queensland
Brisbane, Qld 4072 Australia
(a.garcia, m.ragan)@imb.uq.edu.au
[3] Australian Center for Functional Plant Genomics
[4] ARC Centre in Bioinformatics

Abstract. We briefly describe the methodology we have adopted in order to develop ontologies. Because our scenarios involved domain experts distributed geographically, the domain analysis and knowledge acquisition phases used different independent technologies that were not always integrated into the Protégé suite. Groupware capabilities were thus achieved. From these experiences we identify conceptual maps (CMs) as an important collaborative and knowledge acquisition tool for the development of ontologies. Direct manipulation and collaborative facilities that currently exist in Protégé can be improved with those lessons learnt from this and similar experiences. Here we describe our scenario, competency questions, results, and milestones for each methodological stage, use of CMs, vision for a collaborative environment for ontology development. This presentation is based on two different sets of experiences, one within nutri-genomics and the other one in plant genealogy management systems.

1. Introduction

When developing an ontology involving geographically distributed domain experts, the domain analysis and knowledge acquisition phases may become a bottleneck due to difficulties in establishing a formal means of communication (*i.e.* in sharing knowledge). Conceptual maps (CMs) have been demonstrated to be an effective means of representing and communicating knowledge [1].

Traditionally, ontologies have been built by highly trained knowledge engineers with the assistance of domain specialists. It is a time-consuming and laborious task. Ontology tools are available to support this work, but their use requires training in knowledge representation and predicate logic [2]. Bio-ontologies are developed primarily by biologists. Domain experts are rarely available in one place, so the development of bio-ontologies is usually a distributed effort in which teleconferences, email, commentary-tracking systems, and videoconferences are used at all stages. During our ontology efforts, we identified the lack of an integrated environment in which at least some of these technologies come together to facilitate both knowledge representation and sharing as a major bottleneck. CMs may help to overcome these issues.

Conceptual maps are graphs that consist of nodes, with connecting arcs that represent relationships between nodes [3]. The nodes are labeled with descriptive text representing the "concept", and the arcs are labeled (sometimes only implicitly) with a relationship type. We used CMs in two stages of our process, the capture of knowledge and testing the structure of the ontology. Initially we started to work with informal CMs; although they are not computationally enabled, for humans they appear to have greater "computational efficiency" than other forms of knowledge representation, *e.g.* EXCEL™ spreadsheets or Microsoft Word™ tables. As our models gained semantic richness, the CMs evolved and became more complex by formalizing the knowledge in our ontologies.

We found that the CMs made it possible for domain experts to identify and represent concepts, and to declare relations among them. More importantly, they helped clarify the difference between the ontological model, ER (Entity relationship) models and the possible object model (OM). For biologists, ontologies have a concrete representation in dictionaries, whereas they view object models as being more related to implementation. Implementation details were thus separated from ontologically related issues. We used CMAP (http://cmap.ihmc.us/) [1] as a CM editor.

The ontologies we are developing asymmetric and complementary. In one we want to ease the process of accurately capturing nutrigenomics data *via* web-forms, whereas on the other hand we want to facilitate the building of queries over large genealogy databases (http://cropwiki.irri.org/icis/index.php/Germplasm_Ontology). Two different experiences with similar problems, and a common bottleneck, knowledge acquisition. From both ontologies we identified the importance of cognitive support over the groupware facility.

This paper is organized as follows. In Section 2 we present our methodology, and describe how we used CMs not only to capture knowledge, but also to share it in a distributed environment. Section 3 presents the development of a CM plug-in for Protégé. Brief discussions, conclusions, and an outline of our future work, are presented in section 4.

2. Methodology

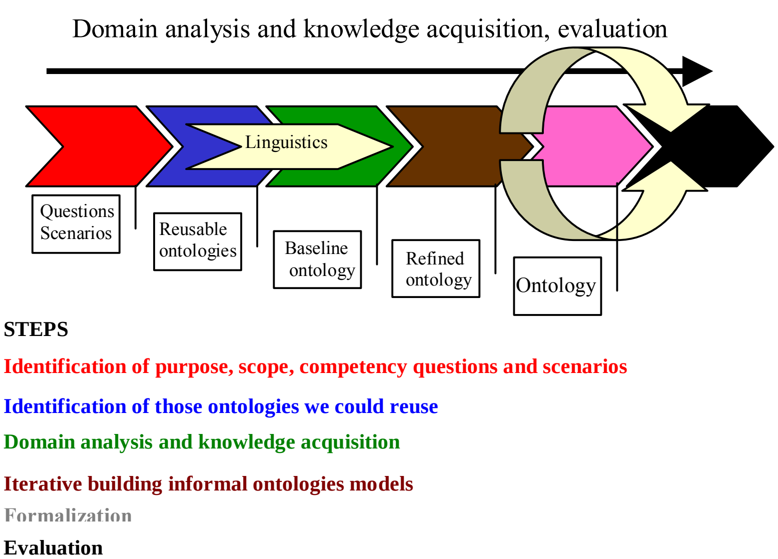

Fig. 1. Methodology, milestones, and phases.

For our particular purposes we decided to adapt some previously reported methodologies in order to enable communication among domain experts and with the ontologist, effectively reuse other ontologies, and provide to the extent possible a high-level conceptual scaffold so other ontologies could be integrated later. We extended the methodology proposed by Mirzaee *et al.* [4]. Figure 1 schematizes the methodology we followed.

Domain analysis is a process in which information used in a particular domain is identified, captured, and organised for the purpose of making it reusable. We hosted a series of meetings during which domain experts agreed on terminology, and on how to structure the reporting of an investigation. We view domain analysis as an iterative process, taking place at every stage. We focused our discussions on specific descriptions of what the ontology should support, and sketched the intended area of application that the ontology was to capture. Our goal was also to guide an ontology engineer, and involve him or her in a more direct manner; so we also made decisions about inclusion, exclusion and the first draft of the hierarchical structure of concepts in the ontology.

An important outcome from this phase was the consensus that we reached on terms that could potentially have a meaning for our intended users. The main aim of these informal linguistic models was to build a explanatory dictionary; some basic relations were, as well, established between concepts.

We built different models throughout our analyses of available knowledge sources and information gathered in previous steps. First a "baseline ontology" was gathered, *i.e.* a draft version containing few but seminal elements of an ontology. Typically, the most important concepts and relations were identified somewhat informally. We could assimilate this "baseline ontology" into a taxonomy, in the sense of a structure of categories and classifications. We consider a taxonomy as "a controlled vocabulary which is arranged in a concept hierarchy", and ontology as "a taxonomy where the meaning of each concept is defined by specifying properties, relations to other concepts, and axioms narrowing down the interpretation." As the process of domain analysis and knowledge acquisition evolves, the taxonomy takes the shape of an ontology. During this step, the ontologist worked primarily with only very few of the domains experts; the others were involved in weekly meetings. In this phase the ontologist sought to provide the means by which the domain experts he or she was working with could express their knowledge. Some deficiencies in the available technology were identified, and for the most part were overcome by our use of CMs.

For subsequent steps (*i.e.* formalisation and evaluation), different needs may be identified.

3. CM plug-in for Protégé

Our knowledge acquisition phase took place in different stages, for some of which the domain experts were not together. CMs proved very useful in facilitating the visualisation and discussion, and in providing domain experts with a tool that could be used to declare the primary elements of their knowledge. OWLviz [5] was initially tested to support domain experts in this task, but this plug-in did not provide direct manipulation (DM) capabilities over the concepts and the relations among them. We also tested Jambalaya [6] before deciding to use two separate tools (*i.e.* Protégé [7] and the CMAP tools). Since CMs support the declaration of nodes and relationships, it was easy to assimilate these to classes and properties. The conversion was a straightforward, albeit manual, process.

The main feature we identified from our work with CMs was the DM capability provided to us by the software. This functionality had several advantages, which we list below. Interesting, all of these advantages had previously been identified by Shneiderman:

- Novices can learn basic functionality quickly, usually through a demonstration by a more experienced user.
- Experts can work extremely rapid to carry out a wide range of tasks, even defining new functions and features.
- Knowledgeable intermittent users can retain operational concepts.
- Error messages are rarely needed.
- Users can see immediately if their actions are furthering their goals; if not, they can simply change the direction of their activity.
- Users have reduced anxiety because the system is comprehensible and because actions are so easily reversible.

We are currently starting to develop the CM plug-in. Basically, it facilitates the declaration properties and classes, writing to the OWL file. Some of the formal requirements we have identified for our plug-in are:

- Graphic manipulation of classes and properties *via* contextual menus.

- Direct publication over the web of the CMs we generate.
- Drag-and-drop capabilities.
- Relationship between concepts and their concrete representations, and annotation features (*e.g.* text, colors, graphics, and even files).
- Manipulation of the same file by different users, with a mechanism to track changes.
- Availability of a chat window.
- Possibility for moderated or un-moderated sessions. This is particularly important for situations in which more than four people are working online on the same file.
- The user interface should be non-intrusive.
- The user should be presented with an empty canvas on which concepts, linking phrases and properties can be declared by a direct click.

4. Conclusions, and future work.

Since our methodology involves participatory design activities, it is important for the tool to support this range of activities. We consider that CMs may play a crucial role in assisting users in these activities. Our development is inheriting many of those features already available in CMAPTOOLS; we are extending it so we may additionally also allow users to "discuss" on-line, while at the same time manipulating the OWL file. We are thus extending the capabilities currently available in Protégé, not just to enhance browsing but more deeply to promote a collaborative environment for the development of ontologies. Since protégé was mainly developed as a desktop tool its web implementation lacks some group-ware features. In order to implement an integrated web-ontology development environment human computer interaction studies need to be conducted.

5. References

1. Canas., A.J., et al. *CMAPTOOLS: A Knowledge Modeling and Sharing Environment*. in *Concept Maps: Theory, Methodology, Technology*. 2004. Universidad Pública de Navarra: Pamplona, Spain: Universidad Pública de Navarra.
2. Seongwook., Y., et al., *Survey about Ontology Development Tools for Ontology-based Knowledge Management.* 2003.
3. Lambiotte., J.G. and D.F. Dansereau., *Multirelational semantic maps.* Educational Psychology review, 1984. **1**(4): p. 331-367.
4. Mirzaee, V., L. Iverson, and B. Hamidzadeh. *Towards ontological modelling of nistorical documents.* in *The 16th International Conference on Software Engineering and Knowledge Engineering (SEKE).* 2004.
5. Knublauch., H., *OWLviz a visualisation plugin for the Protege OWL plugin.*, http://www.co-ode.org/downloads/owlviz/OWLVizGuide.pdf.
6. Storey., M.-A., et al. *JAMBALAYA: Interactive visualization to enhance ontology authoring and knowledge acquisition in Protege*. in *Workshop on Interactive Tools for Knowledge Capture, K-CAP-2001*. 2001. Victoria, B.C. Canada.
7. Geniari., H., John.,, et al., *The evolution of Protege: An environment for knowledge-based systems develpment.* International Journal of Human Computer Studies, 2003. **58**(1): p. 89-123.

6. Acknowledgements

The authors would like to thank Robert Stevens and Karim Nashar for the useful discussions and collaboration. A. Garcia is supported by Institute for Molecular Bioscience, Australian Center for Functional Plant Genomics, the ARC Centre in Bioinformatics and the EMBL-EBI. SA Sansone is supported by the ILSI-HESI Genomics Committee and Philippe Rocca-Serra by the European Commission NuGO project.