# Ontologies for Scrutable Student Modelling in Adaptive E-Learning

Judy Kay, Andrew Lum

School of Information Technologies,
University of Sydney, NSW 2006
Australia
{judy, alum}@it.usyd.edu.au

**Abstract.** This paper discusses the problem of inference about core terms in a student model representing a learner's knowledge, using ontological inference. We try to address the challenge of modelling the student's high level knowledge, based upon user modelling evidence that is associated with fine-grained elements of the model, drawn from the learner's activity in an online learning system. We also describe the critical role that ontologies have the potential to play in supporting scrutable user modelling, where the learner can explore the student model and the processes underlying it so that we can support learner reflection on their knowledge as a foundation for them controlling their learning. In particular, we describe how an ontology over the concepts within the student model has a critical role in supporting visualisation of large student models. We describe how the ontological structure can be used to define graphs with useful properties for visualisation. We report on our experiments with constructing and visualising ontologies in the e-learning domain as well as an analysis of their structure and effectiveness.

## 1 Introduction

For effective adaptation to occur in hypermedia e-learning systems there is a need for user modelling. The student model stores the system beliefs about the learner, and by making inferences, the system can customise and adapt its services to cater for a learner's needs.

It is important to be able to scrutinise student models [1]. Adaptive e-learning systems need to be able to provide explanations as to why elements and relationships in the student model have a particular value or why adaptation happened and also lead to self-reflection [2]. It is also useful for the student to be able to determine what inferences can be made from certain elements in the student model. This is very important in cases where, for example, parts of the student model can be made public [3]. In this case, it should be possible for a learner to define the levels of inference that they are willing for a system to make about them.

There is a problem in how to effectively structure a student model. Student models can potentially contain hundreds or even thousands of elements. One solution is to use an ontology to structure the student models. The ontology fulfills several roles in our student models. Firstly it defines the terms and relationships. This means that the student model has a common vocabulary with other parts of the system, in particular metadata. Secondly, it provides a structure to the student model data, giving us an immediate mechanism for doing inference. Thirdly, the ontology structure means that we can take advantage of existing graph visualisations which require a graph structure over the data to be visualised. They are critical if we are to build substantial student models about the learner from the small amounts of information that are readily available at the interface, especially early in the learning process.

We have been exploring ways to automatically construct an ontology from existing documents and, in particular, from existing glossary sources. Section 2 describes the construction of these ontologies and the enhancements required for turning them into effective student models. Section 3 reports on our own experiments in creating ontologically structured student models and how we have enhanced them. Section 4 discusses the issues of using ontologies for visualisation and our approaches to addressing problems with these ontologies. The last section provides a discussion of the issues addressed and our conclusions.

## 2 Scrutable Student Models and Ontologies

In this section, we describe the way that we have tackled the creation of ontologies for use in scrutable student modelling. Essential to our approach is that we want to be able to explain all aspects of the student models and the underlying processes to the learner. We consider this important for ensuring that learners maintain a sense of control of their own learning. In the case of ontological reasoning, this means that we want to be able to explain the ontology and its construction to the learner.

As a foundation to this approach, we have been exploring the automated construction of ontologies from existing glossaries and dictionaries [4]. As we discuss below, there is a clear benefit in using ontologies that are automatically constructed. In the case where scrutability is a priority, there is a significant additional benefit in that dictionaries are written expressly for the purpose of explaining the meanings of concepts to people. Therefore, they serve as a natural means of explaining the underlying ontology of a student model that has been derived from the dictionary.

### 2.1 Automatic Construction of Ontologies

In general, ontologies are time consuming to construct [5]. On this aspect alone, there is clear appeal in exploiting existing documents which capture the structure of a relevant ontology and automatically build the ontological structure. OntoExtract [6] and Text-To-Onto [7] are examples of such systems. We have been using Mecureo [4] to build ontologies in the broad area of Computer Science based on

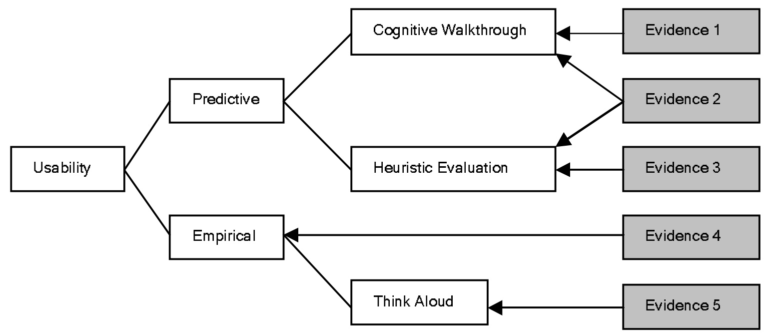


**Fig. 1.** A student model with fine-grain evidence for learner knowledge of concepts in the HCI domain. Evidence may feed into a single concept (such as sources 1, 3, 4 and 5) or multiple (source 2). Evidence may feed into any level of the ontology, in this case, source 4 feeds into a mid-level concept rather than a leaf concept.

FOLDOC [8], an online dictionary, as well as in the area of HCI [9]. These are lightweight ontologies based on lightweight parsing of dictionaries and glossaries to define the set of concepts and relationships in a domain.

### 2.2 Enhancement of Ontologies for Student Modelling

Automatically generated ontologies are, more often than not, incomplete. This incompleteness may not only be missing terms, but also may have missing or inappropriate relationships. Enhancement of the ontology is required for it to be useful in modelling the critical learning elements in a course. We call this the *restricted-ontology problem*.

Consider the issue of missing terms. For the case of the HCI Glossary, we discovered higher level terms, such as *novice*, were mentioned in many definitions but never defined itself as a term in the glossary.

There are also the cases where the ontology uses slightly different terminology to what is used in the application domain. This is also a form of the *restricted-ontology problem*.

### 2.3 Reasoning about Core Terms in the Student Model

One reason for using an ontology is the ability to make inferences about the student's knowledge of the terms in it. The system stores evidence which contribute to the student's knowledge level for each term, though this evidence is often of a fine granularity and may not directly contribute to the higher level terms. For example, a student correctly solving a set problem gives evidence of precise skills for that problem and we now want to reason about higher level skills.

This means there is a need to be able to infer core terms from the finer grained terms. For example, consider the terms in Fig. 1. From the fine-grained evidence supporting student knowledge of the low level concepts, we want to be able to infer the level of student knowledge about higher level concepts. For example, we want to infer the level of student knowledge about *predictive* usability, based upon the fine-grained evidence for *heuristic evaluation* and *cognitive walkthrough*. We also want to be able to infer the student's knowledge about *empirical* evaluation from evidence about methods such as *think aloud*. In addition, we want to be able to infer about the learner's knowledge in the broad area of *usability*, based on their knowledge of both *predictive* and *empirical* usability evaluation. There are many existing numerical methods for probabilistic reasoning that can be applied to the student model to make these inferences such as Dempster-Schafer theory or Bayesian methods.

If an ontology is well built, this should be a straight forward task as the relationships between the terms will have sensible and hopefully scrutable meaning as in the example above. In the case of automatically constructed ontologies, this may not be the case.

Some of the difficulties in making these inferences follow from the elements of the inference process:

- the amount of evidence available about each concept. If we have just one piece of evidence about the learner's knowledge of an aspect like cognitive walkthrough, this would seem to be a weaker indication of learner knowledge than would be the case if there were ten pieces of evidence.
- much of our evidence is positive in that we track when the student appears to have listened to a lecture. This is typical of much work involving the web, where positive data dominates.
- the strength of the ontological inference needs to be taken into account.

## 3 Experiments in Ontological Inference

We have been experimenting with the challenges of exploiting ontologies for scrutable student modelling. The context of our work has been a course in user interface design and programming. We create detailed student models by collecting evidence of student learning at the course web site. In particular, there are 20 online lectures, each with a series of visual elements, much like overhead slides, and audio lecture content with each such slide. These enable students to listen to lectures flexibly. At the same time, we are able to monitor the duration of time between each slide access. By comparing this with the duration of the audio, we are able to determine whether the full audio could have been played. We also require students to type lecture notes in conjunction with slides. Other student modelling information includes lab marks and other assessed elements.

In order to build student models, we needed to define the elements in it. We did this in terms of an automatically constructed ontology which was based on the Usability Glossary by Usability First [5].

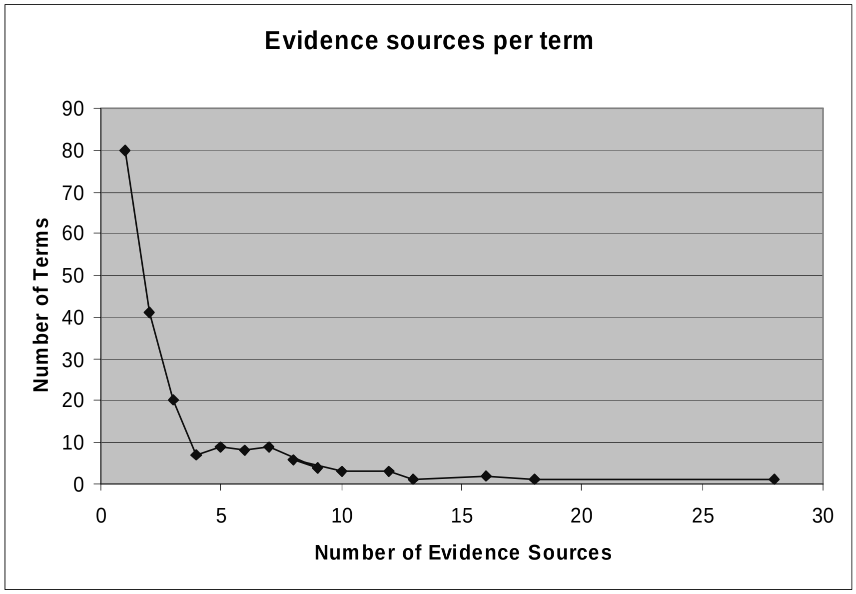


**Fig. 2.** The amount of evidence sources per term – we can see that the majority of terms have less than 10 evidence sources, though there are a high number with only one evidence source.

The glossary contains 1129 terms and categories. In addition, we defined an additional 105 local definitions which were used to enhance the ontology, giving us a total of 1234 elements. We use a subset of these terms (195 out of the 1234) as metadata terms on each of the learning objects and assessment elements. We then have corresponding elements in the student model.

As described above, the automatically constructed ontology falls short of our student modelling needs. We have developed an extremely simple approach to addresses this problem, while maintaining the benefits of a human-readable dictionary as the basis for the ontology: we define additional terms in a local extension of the dictionary [10].

From the ontology construction and enhancement phases, we have an ontology which contains all the terms of the student model and learning object metadata that the course teacher considered necessary. As the metadata uses the same vocabulary as the ontology, we treat the accessed pages where the terms are used as metadata as evidence sources for the student knowing that term in the student model. For our course, we have added metadata to the learning objects and used a simple heuristic comparing the amount of time a user spent on a slide in a learning object to the audio time for that slide to generate a value of their knowledge for that term.

Fig. 2. shows the distribution of the number of evidence sources per term. There is a high proportion of terms with less than 3 evidence sources (62%), and a small number with many evidence sources. The number of terms with a low amount of evidence poses difficulties in being able to make inferences about the student's

knowledge as mentioned in Section 2.3. The values in the user model are not useful in modelling the user's actual knowledge for these terms since the low number of evidence sources means that these terms are not taught in as much depth as a term with a high number of evidence sources.

By having the evidence sources as web page accesses, we only have positive evidence for the terms in the student model. We have weekly homework, laboratory and quiz marks for the students in the course. These will allow us to include additional evidence sources that can have negative values.

In our current work, we do not make any ontological inference on the student models to infer new higher level evidence from the fine grain evidence. We use the Ontology Web Language (OWL) [11] to represent our student models and is used as an input to our visualisation. This could support inference of the student model. We plan to explore this in the future. The ontology used in the student models also has relationship strengths generated by Mecureo during the ontology creation process that can be utilised to contribute to the strength of inference when reasoning about higher level skills. These relationships are typed based on the way the terms are used in the definitions. Types include parent/child, synonym/antonym, and sibling relationships. For example, *auditory feedback* is a synonym of *sound*, and *predictive metric* has parent *usability*.

## 4 Ontologies for Visualisation of Student Models

As already mentioned, we want to be able to provide learners with a visualisation of their student models. One reason, as noted, is that this can support reflection, an important foundation for improved learning. If the visualisation makes use of an ontology, it can serve a second important role in helping learners to see how their knowledge of one concept in the course can affect or be affected by other concepts.

However, there are additional properties that ontologies should exhibit if they are to be suitable for visualisation. The ontologies should have a modest fan-in and fan-out, that is, the number of relationships leading to and from any term in the ontology. We call this the peerage. Peerage is important because concepts with a number of peers will be isolated from the rest of the ontology in the visualisation. In the experiments we have run with our own visualisation, we found that it was hard to navigate to these concepts and see how they relate to the rest of the ontology. Equally, a large fan-out poses serious problems since it indicates that one concept is related to a large number of others. In a visualisation, this leads to serious clutter and interface problems.

In the case of the ontology for the UIDP course, the basic process described in the last section gives the peerage properties summarized in table 1. This shows that, in the current ontology, user defined terms are for more weakly connected to the rest of the ontology; they have an average of 3 peers, compared with 9 for terms in the foundation glossary (see last column of Table 1).

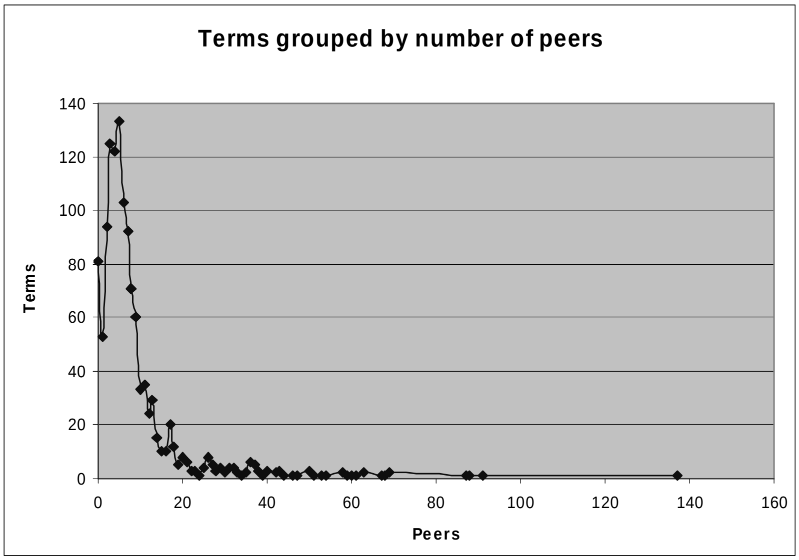


**Fig. 3.** The distribution of terms in our ontology grouped by the number of peers (fan-out) they have. There are a significant number of terms having a low peerage or an extremely high peerage.

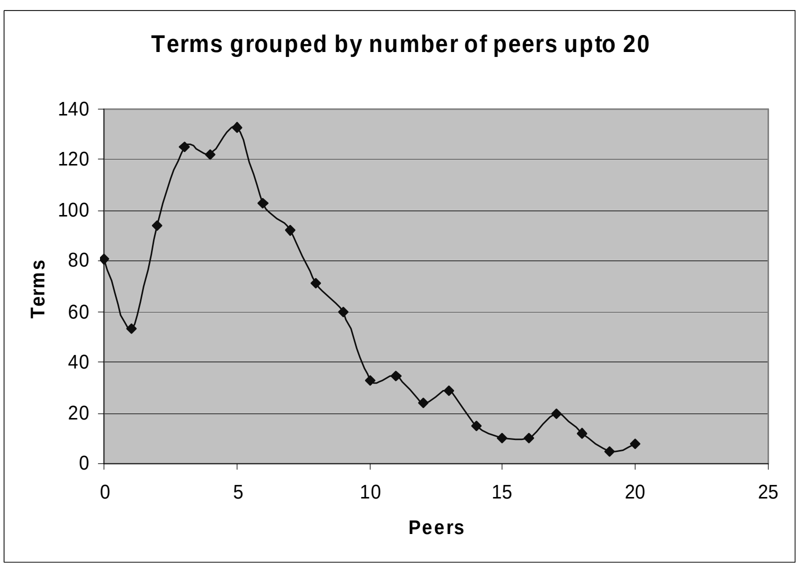


**Fig. 4.** A close-up of Fig. 3, with cut-off at 20 peers.

**Table 1.**

| | Concepts | Added as course metadata | Peers | Average peers per concept |
|---|---|---|---|---|
| **Defined locally** | 105 | 105 | 345 | 3.29 |
| **Defined in glossary** | 1129 | 90 | 10345 | 9.16 |
| **Total** | 1234 | 195 | 10690 | 8.66 |

In Fig. 3, we show some analysis of this distribution. It indicates that although the majority of terms have less than 20 peers, we still have a significant number with no peerage at all, and some outliers having over 100 peers. This extremity in peer distribution results in our visualisation suffering from the problems mentioned above.

We believe visualisation will be most effective if each concept has around 5 to 20 peers. Accordingly, we focus on this region of Fig. 3, as shown in Fig. 4. Clearly, we have significant numbers of terms with very low peerage. In fact, 475 terms (almost 40%) have less than 5 peers. Notably, on average, these are dominated by the user defined terms, as can be seen from Table 1. This is particularly serious since these are the concept terms that are sufficiently important for the course that we added them to the dictionary as. An important way to address this problem is to improve the added terms. We can do this quite easily, because, at present, we have only added new terms without providing a definition for them. In [10] we showed how adding a full definition increased the peerage of the added terms.

Another serious problem relates to concepts that have too many peers. There are 99 terms (8%) that have more than 20 peers in our ontology. The best way to address this problem is to make use of the weights that Mecureo places on relationship links. We have yet to explore this.

## 6 Discussions and Conclusions

We have identified several issues in reasoning about a learner's knowledge where ontological inference can provide a solution. In particular we want to infer about core terms for a domain when we only have fine grain evidence for lower level concepts. Although there are existing systems such as VisMod [12] that allows exploration of a student model and reasoning about a node and it's neighbours, or Summary Street [13] that provides feedback to users summarising text by comparing their summaries to a semantic space generated by latent semantic analysis, our approach relies on an underlying ontology that not only provides ontological relationships between terms to aid in reasoning, but a readily exploitable structure for visualisation.

In our own experiments, we have created an ontology for our course automatically from an online glossary source. However we have identified several deficiencies in our student model:

- there is a large proportion of terms in the student model with only one or two evidence sources

- the homogeneity of our evidence sources means that we only have positive evidence contributing to the values of the terms
- we need a way to infer evidence about the higher level terms that take into account the relationship weightings.

We have also analysed and discussed obstacles in visualising ontologies and the desirable characteristics. From our own experiments, we have decided that a peerage of 5-10 is desirable to avoid the effects resulting from terms with extremely low or high peerage.

The approach we have described shows promise if we can overcome the issues discussed in this paper. We have discussed two critical roles for ontologies in student modelling. The first is to support inference about core terms. The second is to provide a graph structure for use in visualisation.

## 7 Acknowledgements

We thank Hewlett-Packard for supporting this research.

## References

1. Kay, J., *A scrutable user modelling shell for user-adapted interaction*, University of Sydney (1999)
2. Bull, S., Brna, P., and Pain, H., *Extending the scope of the student model*. In User Modeling and User-Adapted Interaction. Vol. 5(1) (1995) 44-65
3. Lum, A. *Scrutable User Models in Decentralised Adaptive Systems*. In: P. Brusilovsky, A. Corbett, and F.d. Rosis, Editors. *9th International Conference on User Modelling.* Springer, (2003) 426-428
4. Apted, T. and Kay, J., *MECUREO Ontology and Modelling Tools Special Issue on Concepts and Ontologies.* In WBES of the International Journal of Continuing Engineering Education and Lifelong Learning. (Accepted 2003, to appear)
5. Fensel, D., *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*: Springer (2001)
6. Reimer, U., et al., *Ontology-based Knowledge Management at Work: The Swiss Life Case Studies*, in J. Davis, D. Fensel, and F.v. Harmelen, Editors, *Towards the Semantic Web: Ontology-driven Knowledge Management*. 2003, John Wiley & Sons: West Sussex, England. p. 197-218.
7. Maedche, A. and Staab, S. *Mining Ontologies from Text*. In: *Knowledge Engineering and Knowledge Management. Methods, Models, and Tools: 12th International Conference, EKAW.* Springer-Verlag Heidelberg, (2000) 189-202
8. *FOLDOC: The Free Online Dictionary of Computing*, Available at http://www.foldoc.org/. (2001)
9. *Usability Glossary*, Available at http://www.usabilityfirst.com/glossary/main.cgi. (2002)

10. Apted, T., Lum, A., and Kay, J. *Supporting metadata creation with an ontology built from an extensible dictionary*. In: P.D. Bra and W. Nejdl, Editors. *Adaptive Hypermedia and Adaptive Web-based Systems*. Springer, (2004)
11. *OWL Web Ontology Language Overview*, Available at http://www.w3.org/TR/owl-features/. (2003)
12. Zapata-Rivera, J.D. and Greer, J., *Inspecting and Visualizing Distributed Bayesian Student Models*, in G. Gauthier, C. Frasson, and K. VanLehn, Editors, *Intelligent Tutoring Systems*. 2000. p. 544-553.
13. Kintsch, E., et al., *Developing Summarization Skills through the Use of LSA-Based Feedback.*, In Interactive Learning Environments. Vol. 8(2) (2001) 87-109