KNOWLEDGE ELICITATION TECHNIQUES FOR DERIVING COMPETENCY QUESTIONS FOR ONTOLOGIES

Lila Rao
Mathematics and Computer Science Department, The University of the West Indies, Mona, Kingston 7, Jamaica

Han Reichgelt
School of Computing and Software Engineering, Southern Polytechnic State University, Marietta, GA, U.S.A.

Kweku-Muata Osei-Bryson
Department of Information Systems & The Information Systems Research Institute
Virginia Commonwealth University Richmond, VA, U.S.A.

Keywords: Competency Questions, Knowledge Elicitation Techniques, Ontologies.

Abstract: This research explores the applicability of existing knowledge elicitation techniques for the development of competency questions for ontologies. This is an important area of research as competency questions are used to evaluate an ontology. The use of appropriate knowledge elicitation techniques increases the likelihood that these competency questions will be reflective of what is needed of the ontology. It thus helps ensure the quality of the ontology (i.e. the competency questions will adequately reflect the end users requirements).

1 INTRODUCTION

Knowledge elicitation (KE) involves the gathering of knowledge from experts (Shadbolt and Burton, 1989). There is a number of existing knowledge elicitation techniques (e.g. 20 questions, card sort, repertory grid and laddering) (Hickey and Davis, 2003). These techniques have been used extensively for various types of applications (e.g. expert systems) (Nakhimovsky et al., 2006, Reichgelt and Shadbolt, 1992, Bryrd et al., 1992).

Ontologies have been identified as important components of a number of information systems (Guarino, 1998, Pinto and Martins, 2004) such as knowledge management systems (Sicilia et al., 2005, Rao and Osei-Bryson, 2007), e-business applications (Lee et al., 2006, Fensel et al., 2001, Papazoglou, 2001) and data warehouses (Critchlow et al., 1998, Shah et al., 2005). Therefore, the quality of the overall system is likely to be highly dependent on the quality of the ontology.

There are many different definitions of the term "ontology" and different proposals for what should be represented in the ontology. However, most agree that it is some formal description of a domain, which can be shared among different applications and expressed in a language that can be used for reasoning (Noy, 2004).

As ontologies grow in size and complexity because of the increasing number of demands are being placed on them, ensuring their quality is an important consideration in the development of these systems. Quality is a multi-dimensional concept (Wang et al., 1995, Wand and Wang, 1996), and, in order to assess the quality of the ontology a set of dimensions should be defined. These dimensions can be used to derive metrics that can be used not only to assess the quality of the ontology but also to determine whether proposed quality improvement techniques are actually effective. One of the proposed quality dimensions is *coverage/ completeness* (Jarke et al., 1999) which has been defined as the extent to which the ontology covers the domain of interest (Rao and Osei-Bryson, 2007). This can be measured as the difference between what is required of the ontology and what is available in the ontology.

Rao L., Reichgelt H. and Osei-Bryson K. (2008).
KNOWLEDGE ELICITATION TECHNIQUES FOR DERIVING COMPETENCY QUESTIONS FOR ONTOLOGIES.
In *Proceedings of the Tenth International Conference on Enterprise Information Systems - ISAS*, pages 105-110
DOI: 10.5220/0001678701050110

One of the most commonly used techniques to evaluate ontologies is competency questions (Staab et al., 2001, Sure et al., 2002). Competency questions define the ontology's requirements in the form of questions that the ontology must be able to answer (Gruninger and Fox, 1994, Gangemi, 2005). These competency questions are actually providing an approach for measuring the coverage of the ontology as the percentage of the total set of competency questions posed that can be answered by the ontology is indicative of coverage. However, in order for the measure to be accurate we must ensure that the set of competency questions are complete. It would be misleading to measure the coverage of the ontology using this set of competency questions if it is not likely that this set of questions is complete. Thus, appropriate techniques are needed for identifying competency questions. If we ensure that the techniques used are likely to lead to a complete set of competency questions then the coverage measure will be more dependable, which will help ensure the quality of the ontology and hence the quality of the overall system.

We will demonstrate the applicability of these elicitation techniques by building an ontology and a set of corresponding competency questions for a university's information technology (IT) infrastructure domain. Knowledge about this domain is routinely used to solve a number of different problems, ranging from troubleshooting to network redesign and decisions about software to acquire to server administration. Any ontology for this domain can therefore be shared by a range of different users solving of different problems.

The rest of the proposal is organised as follows. Section 2 provides a review of the literature that is relevant to this research, including the ontology literature (Section 2.1) and various knowledge elicitation techniques (Section 2.2). Section 3 describes the applicability of various knowledge elicitation techniques to the development of the competency questions and provides an illustrative example using a specific domain. Finally, section 4 provides some concluding remarks and some directions for future research.

2 LITERATURE REVIEW

2.1 Ontology and Competency Questions

An ontology has been defined as "a formal description of entities and their properties, relationships, constraints, behaviors" (Gruninger and Fox, 1995). A number of approaches have been proposed for developing ontologies (Gruninger and Fox, 1995, Staab et al., 2001). Gruninger and Fox (1995) propose an approach to engineering ontologies that consists of three steps:

1) Defining an ontology's requirements in the form of questions that an ontology must be able to answer (i.e. competency questions). This is known as the competency of the ontology (Fox et al., 1998).
2) Defining the terminology of the ontology - its objects, attributes and behaviours. In this way the ontology provides the language that will be used to express the definitions in the terminology and the constraints required by the application.
3) Specifying the definitions and constraints on the terminology.

Staab et al. (2001) describe an ontology development process consisting of 5 phases (i.e. the feasibility study, the kickoff phase for ontology development, refinement, evaluation and maintenance).

Competency questions thus provide an important tool to validate an ontology as they can be used to evaluate the ontological commitments that have been made, and are indeed generally accepted as a verification technique for ontologies (Kim et al., 2007). Staab et al. (2001) recommend using these competency questions for the evaluation phase of their proposed ontology development process. Thus the evaluation process is highly dependent on the competency questions that are formulated, and it is therefore imperative that the process of deriving the competency questions is thorough, and it is therefore crucial that one identify a set of techniques for reliably eliciting all the competency questions.

2.2 Knowledge Elicitation Techniques

There is a number of existing knowledge elicitation techniques such as interviews (e.g. structured, unstructured and semi-structured), case studies, prototyping, sorting (e.g. card sorting), triad analysis, 20 questions, laddering and document analysis (Shadbolt and Burton, 1989, Nakhimovsky et al., 2006).

Laddering is used to construct a graphical representation of the concepts and relations in a domain. The elicitor makes use of prompts to explore the expert's understanding of the domain. A graph, consisting of a number of nodes and labelled arcs, is constructed in the presence of the expert.

This technique involves three main steps. The first step involves asking the expert to identify a starting point (seed item) (i.e. a concept that is important in the domain). The next step involves moving around the domain using various prompts (i.e. asking questions to move down, across and up the expert's domain knowledge). The final step involves the elicitation of attributes for the various concepts (Reichgelt and Shadbolt, 1992).

The card sort, triad analysis and twenty questions techniques assume that the knowledge engineer has some prior knowledge of the domain under consideration. This initial knowledge can be obtained through available documentation as well as by conducting unstructured interviews. The available documentation can be used to get a sense of the domain under consideration (i.e. some of the basic concepts and relationships within the domain). Once the knowledge engineer has some understanding of the domain, unstructured interviews can then be used for providing high level knowledge of the domain. Unstructured interviews suit the early stages of elicitation when the knowledge engineer is trying to learn about the domain but does not know enough to set up indirect or highly structured tasks (Cooke, 1999).

Card sort entails the use of a given set of cards with the names of relevant domain elements or problems written on them. Experts are used to sort the cards into several piles according to whatever criteria they choose. This process is repeated until the expert the expert has exhausted the ways to partition the elements (Shadbolt and Burton, 1989). Card sort is useful when the aim is uncover the different ways that an expert sees the relationships between a set of concepts (Reichgelt and Shadbolt, 1992).

Triad analysis requires that the expert is given or asked to generate a set of important elements. The interviewer randomly selects three of these examples and asks the expert to distinguish between them such that two of the examples in the triad have a common property not possessed by the third (Ryan and Bernard, 2000). This distinguishing property is known as the construct. This process continues with different triads of elements until no further discriminating constructs can be identified by the expert (Reichgelt and Shadbolt, 1992).

20 questions require that the knowledge engineer chooses an element from the domain or a problem. The domain expert is then required to determine what the element or problem is but is only allowed to ask questions that the knowledge engineer can answer either yes or no (Kemp, 1996). This allows the knowledge engineer to determine the heuristics that an expert uses in his or her problem solving process.

3 KNOWLEDGE ELICITATION TECHNIQUES FOR DERIVING

Although competency questions are seen as a viable way to evaluate an ontology (Gruninger and Fox, 1994), (Staab et al., 2001), (Sure et al., 2002), (Gangemi, 2005), (Kim et al., 2007) there is limited work describing appropriate techniques for developing them. Gruninger and Fox (1995) state that motivating scenarios should be used for generating informal competency questions (see Figure 1). However they do not elaborate on how these motivating scenarios will be identified.

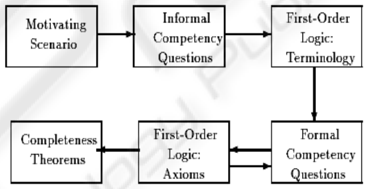

Figure 1: Procedure for Ontology Design and Evaluation (Gruninger and Fox, 1995).

Sure et al. (2002) stress the importance of the domain expert as a valuable source of knowledge for structuring the domain. Personal interviews are a commonly used method for knowledge acquisition from domain experts, thus, they propose that the competency questions should be derived from interviews with the domain expert. However, they do not elaborate on how to conduct these interviews. They use these competency questions to create the initial version of the semi-formal description of an ontology as well as for the evaluation of the ontology. Noy and Hafner (2007) also point to the need for interaction between the knowledge engineer and domain expert for the development of competency questions but do not mention any techniques that can be used for facilitating this interaction.

It seems fair to say that the applicability of existing knowledge elicitation techniques to the development of competency questions has not been fully explored. However, given the fact that researchers have reported great success with the use of more structured knowledge elicitation techniques, such as laddering, card sort etc, in knowledge elicitation for expert systems (Shadbolt and Burton, 1989, Wang et al., 2006), it seems reasonable to expect that such knowledge elicitation techniques will also prove useful in the elicitation of competency questions. We will therefore explore the applicability of three knowledge elicitation techniques that seem applicable to the development of competency questions, namely card sort, triad analysis and 20 questions. In particular, we will explore the use of laddering for the elicitation of an initial ontology, and 20-questions, triad analysis and card sort for the development of the competency questions.

Given the nature of the knowledge obtained through laddering, it will be clear why we use laddering to elicit the initial ontology. The questions that the domain expert generates in the twenty questions technique are questions that he/she considers important in that domain and therefore should be able to be answered by the ontology. Using card sort, the ontology engineer will be able to determine criteria that are important to the domain expert for grouping similar cases, and thus form the basis of the competency questions as the expert will expect that the ontology can answer queries about these concepts. Similar considerations apply for triad analysis.

It is likely the case that a combination of the existing techniques (Shadbolt and Burton, 1989, Harper et al., 2003) may actually be most effective for eliciting competency questions. As mentioned previously, each of the three techniques requires some knowledge of the domain which can be captured by reviewing available documentation and using unstructured interviews. To get a more detailed description of the domain the use of card sort, triad analysis and 20-questions will be explored.

A number of domain experts will be used in this exercise. Multiple experts will help to ensure that as many competency questions as possible can be identified. Various groups of experts are likely to be concerned with specific tasks within the domain and therefore the knowledge elicited will be specific to those tasks. Multiple experts will provide a consensus of the important concepts and relationships within the domain.

There have been a number of problems with using these existing elicitation techniques. These include, for example, the experts being adverse to some of these techniques, the techniques being time consuming and costly, combining the knowledge of multiple experts, choosing the appropriate technique (Cooke, 1999). However, tools and techniques have been and are being developed to help address these problems (Hickey and Davis, 2003, Harper et al., 2003, Major and Reichgelt, 1990, Nakhimovsky et al., 2006).

4 EXAMPLE DOMAIN

We will explore the usefulness of different knowledge elicitation techniques to the development and evaluation of ontologies by applying them to a particular domain, namely the IT infrastructure domain at a university campus in Jamaica. Knowledge about the university's IT infrastructure can be used to solve various types of problems (e.g. disaster recovery planning/business continuity planning, security and risk management, training and network design). Having a formal description of the entities, relationships, constraints and behaviours (Gruninger and Fox, 1995) in the domain ensures that all the decisions are being made with the same information. Additionally, various entities within the university may need to communicate in order to solve particular problems related to the IT infrastructure. Having an ontology as a reference will facilitate this communication as one of the main purposes of an ontology is to formally describe the domain of discourse so as to provide a common language for all entities to communicate, thus reducing the potential for ambiguity. Once developed this ontology could then be used by other universities that require the same types of problem solving.

One of the problems requiring access to information about the IT infrastructure domain is disaster recovery planning (DRP). The aim of DRP is to ensure that entities (i.e. the university) function effectively during and following a disaster (Bryson et al., 2002). A well-organized disaster recovery plan will directly affect the recovery capabilities of an entity. The contents of the plan should follow a logical sequence and be written in a standard and understandable format (Wold, 2002). For example, in the case of the university campus in Jamaica there is an annual threat of hurricanes. Therefore, the disaster recovery plan should include procedures that need to be followed in the event that a hurricane becomes a threat to Jamaica. These procedures would include, for example, the systems that would need to be shut down, where they are physically located, who is responsible for them being shut down, who uses them so will be affected by their shut down. Having this information readily available in the ontology will make it possible to establish the plan more effectively.

Additionally, the IT infrastructure domain knowledge can be used to develop security plans for the university's systems. Both the information stored in the system and the system as a whole needs to be secured. In order to establish these security plans decision makers would need to know, for example, what information is stored in each system, the tasks that the information is being used for, decisions that these tasks are being used to make, the threat of possible risks to the various systems. Again, if this information is stored in an ontology then it will be readily available to decisions makers, in a consistent, standardised format.

The ontology for the IT infrastructure domain will be developed using an extended form of laddering as the main knowledge elicitation technique. A number of the employees of the university, playing various roles, will be used in the knowledge elicitation process to develop this domain ontology.

In order to evaluate the quality of the ontology (i.e. the completeness/coverage of the ontology) a set of competency questions for the IT infrastructure domain will be developed. The ontology will be considered to be of a high quality if it is able to answer the competency questions. It is therefore crucial that good methodologies for creating these competency questions are found. Forced answer techniques, such as twenty questions and card sort may be good ways of doing this and their use will be explored. The employees of the university will be used as the domain experts for this process.

5 CONCLUSIONS

The evaluation of an ontology relies heavily on the competency questions formulated, and the issue of using appropriate knowledge elicitation techniques to competency questions is therefore of central importance in ontology development. This research addresses this issue by exploring the applicability of three specific knowledge elicitation techniques (i.e. 20 questions, triad analysis and card sort) to the development of competency questions for an ontology.

If these techniques prove useful then this work will help improve the quality of ontologies and in so doing improve the quality of the systems (e.g. knowledge management systems, e-commerce systems and data warehouses) that they are a part of. They will help derive a measure for the coverage of the ontology which can help assess its quality.

In the future we will explore how the techniques proposed in this paper can be used to develop an approach for the development, representation and evaluation of high quality ontologies.

We will also explore the additional benefits that the knowledge elicitation process may provide. For example, the process may help to identify the various user groups within the domain. Identifying these groups and their needs will help identify the various user groups of the ontology. Those users that formulate similar competency questions can be classified as belonging to a particular group of users. Based on these groups of users the ontology can then be designed in a way (e.g. using subontologies) that can maximise the efficiency of access to the ontology. Further, as the system is used metadata will be generated that will reflect the usage of the system. This can be analysed to track the usage (i.e. types of queries on the ontology) to determine if they ontology needs restructuring (e.g. adding an additional subontology for a frequently requested type of query that was not identified in the initial design). Therefore, the ontology will be maintained as it is used. Thus, when a query is being processed by the system, depending on the type of the query, the appropriate subontology will be identified and used for processing. Thus, the entire ontology will not have to be searched. This can have significant benefits as ontologies are becoming larger and more complex.

REFERENCES

Bryrd, T., Cossick, K. & Zmud, R. (1992) A Synthesis of Research on Requirements Analysis and Knowledge Aquisition Technqiues. *MIS Quarterly*.

Bryson, K.-M., Millar, H., Joseph, A. & Mobolurin, A. (2002) Using Formal MS/OR Modeling to Support Disaster Recovery Planning. *European Journal of Operational Research,* 141**,** 679-688.

Cooke, N. J. (1999) Knowledge elicitation. IN DURSO, F. T. (Ed.) *Handbook of Applied Cognition.* UK: Wiley.

Critchlow, T., Ganesh, M. & Musick, R. (1998) Automatic Generation of Warehouse Mediators Using an Ontology Engine. *5th International Conference on Knowledge Representation Meets Databases* Seattle, Washington.

Fensel, D., Van Harmelen, F., Universiteit, V., Horrocks, I., Mcguinness, D. & Patel-Schneider, P. (2001) OIL: An Ontology Infrastructure for the Semantic Web. *IEEE Intelligent Systems,* 16**,** 38-45.

Fox, M. S., Barbuceanu, M., Gruninger, M. & Lin, J. (1998) An Organization Ontology for Enterprise Modeling. IN PRIETULA, M., CARLEY, K. & GASSER, L. (Eds.) *Simulating Organizations:*

Computational Models of Institutions and Groups. Menlo Park, CA., AAAI/MIT Press.

Gangemi, A. (2005) Ontology Design Patterns for Semantic Web Content. *4th International Semantic Web Conference (ISWC 2005).* Galway, Ireland.

Gruninger, M. & Fox, M. S. (1994) The Design and Evaluation of Ontologies for Enterprise Engineering. *Workshop on Implemented Ontologies, European Conference on Artificial Intelligence (ECAI).* Amsterdam, NL.

Gruninger, M. & Fox, M. S. (1995) Methodology for the Design and Evaluation of Ontologies. *IJCAI'95, Workshop on Basic Ontological Issues in Knowledge Sharing.* Montreal.

Guarino, N. (1998) Formal Ontology and Information Systems. *First International Conference on Formal Ontologies in Information Systems.* Trento, Italy, IOS Press.

Harper, M. E., Jentsch, F. G., Berry, D., Lau, H. C., Bowers, C. & Salas, E. (2003) TPL–KATS-card sort: A tool for assessing structural knowledge. *Behavior Research Methods, Instruments, & Computers,* 35**,** 577-584.

Hickey, A. M. & Davis, A. M. (2003) Elicitation Technique Selection: How Do Experts Do It? *Requirements Engineering Conference.* California.

Jarke, M., Jeusfeld, M., Quix, C. & Vassiliadis, P. (1999) Architecture and Quality in Data Warehouses: An Extended Repository Approach. *Information Systems,* 24**,** 229-253.

Kemp, E. A. (1996) The Role of the Individual Project in Teaching Knowledge Acquisition. *International Conference on Software Engineering: Education and Practice (SE:EP '96).* Dunedin, New Zealand.

Kim, H. M., Fox, M. S. & Sengupta, A. (2007) How to Build Enterprise Data Models to Achieve Compliance to Standards or Regulatory Requirements (and share data). *Journal of the Association for Information Systems,* 8, 105-128.

Lee, T., Lee, I., Lee, S., Lee, S., Kim, D., Chun, J., Lee, H. & Shim, J. (2006) Building an Operational Product Ontology System. *Electronic Commerce Research and Applications,* 5**,** 16–28.

Major, N. & Reichgelt, H. (1990) ALTO: An Automated Laddering Tool. *Current Trends in Knowledge Acquisition.* O C S L Press

Nakhimovsky, Y., Schusteritsch, R. & Rodden, K. (2006) Scaling the Card Sort Method to Over 500 Items: Restructuring the Google AdWords Help Center. *Conference on Human Factors in Computing Systems CHI '06* Montréal, Québec, Canada

Noy, N. (2004) Semantic integration: A survey of ontology based approaches. *SIGMOD Record,* 33**,** 65-69.

Papazoglou, M. P. (2001) Agent-Oriented Technology in Support of E-Business. *Communications of the ACM,* 44**,** 71-77.

Pinto, H. S. & Martins, J. P. (2004) Ontologies: How Can They Be Built? *Knowledge and Information Systems* 6**,** 441-464.

Rao, L. & Osei-Bryson, K.-M. (2007) Towards Defining Dimensions of Knowledge Systems Quality. *Expert Systems with Applications,* 33**,** 368-378.

Reichgelt, H. & Shadbolt, N. (1992) ProtoKEW: A Knowledge-Based System for Knowledge Acquisition. IN SLEEMAN, D. & BERNSEN, O. (Eds.) *Research Advances in Cognitive Science volume 5: Artificial Intelligence.* Hillsdale, NJ., Lawrence Erlbaum.

Ryan, G. W. & Bernard, H. R. (2000) Data Management and Analysis Methods. IN DENZIN, N. & LINCOLN, Y. (Eds.) *Handbook of Qualitative Research.* Thousand Oaks, CA, Sage Publications, Inc.

Shadbolt, N. & Burton, A. M. (1989) The Empirical Study of Knowledge Elicitation Techniques. *ACM SIGART Bulletin***,** 15-18.

Shah, S., Huang, Y., Xu, T., Yuen, M., Ling, J. & Ouellette, F. (2005) Atlas – A Data Warehouse for Integrative Bioinformatics. *BMC Bioinformatics* 6**,** 34.

Sicilia, M.-A., Lytras, M., Rodriguez, E. & Garcia-Barriocanal, E. (2005) Integrating descriptions of knowledge management learning activities into large ontological structures: A case study. *Data & Knowledge Engineering,* article in press.

Staab, S., Schnurr, H.-P., Studer, R. & Sure, Y. (2001) Knowledge Processes and Ontologies. *IEEE Intelligent Systems,* 16**,** 26-34.

Sure, Y., Erdmann, M., Angele, J., Staab, S., Studer, R. & Wenke, D. (2002) OntoEdit: Collaborative Ontology Development for the Semantic Web. *First International Semantic Web Conference (ISWC 2002).* Sardinia, Italy.

Wand, Y. & Wang, R. Y. (1996) Anchoring Data Quality Dimensions in Ontological Foundations. *Communications of the ACM,* 39**,** 86-95.

Wang, R. Y., Storey, V. C. & Firth, C. P. (1995) A Framework for Analysis of Data Quality Research. *IEEE Transactions on Knowledge and Data Engineering,* 7**,** 623-640.

Wang, Y., Sure, Y., Stevens, R. & Rector, A. (2006) Knowledge Elicitation Plug-in for Prot´eg´e: Card Sorting and Laddering. *Asian Semantic Web Conference (ASWC'06).* Beijing, China.

Wold, G. H. (2002) Disaster Recover Planning Process *Disaster Recovery Journal* 5**,** 29-34.