7-2006

Adaptive Modeling of Workforce Domain Knowledge

Jian Qin
Syracuse University

Peter Creticos
Institute for Work and the Economy

Wen-Yuan Hsiao
ISONTO

Recommended Citation

Qin, J., Creticos, P., & Hsiao, W. (2006). Adaptive modeling of workforce domain knowledge. In: Proceedings of the Ninth International Society for Knowledge Organization Conference, July 4-8, 2006, Vienna, Austria.

Citation: Qin, J., Creticos, P., & Hsiao, W. (2006). Adaptive modeling of workforce domain knowledge. In: Proceedings of the Ninth International Society for Knowledge Organization Conference, July 4-8, 2006, Vienna, Austria.

Jian Qin [1] Peter Creticos [2] Wen-Yuan Hsiao[3]
[1] School of Information Studies, Syracuse University, Syracuse, NY 13244, USA
[2] Institute for Work and the Economy, Oak Brook, IL 60523, USA
[3]ISONTO, LLC, Jamesville, NY 13078

Adaptive Modeling of Workforce Domain Knowledge

Abstract

Workforce development is a multidisciplinary domain in which policy, laws and regulations, social services, training and education, and information technology and systems are heavily involved. It is essential to have a semantic base accepted by the workforce development community for knowledge sharing and exchange. This paper describes how such a semantic base—the Workforce Open Knowledge Exchange (WOKE) Ontology—was built by using the adaptive modeling approach. The focus of this paper is to address questions such as how ontology designers should extract and model concepts obtained from different sources and what methodologies are useful along the steps of ontology development. The paper proposes a methodology framework "adaptive modeling" and explains the methodology through examples and some lessons learned from the process of developing the WOKE ontology.

1 Introduction

Workforce development is a multidisciplinary domain in which policy, laws and regulations, social services, training and education, and information technology and systems are heavily involved. It is common practice in the United States that the federal government establishes the primary workforce development agenda: the Congress enacts legislation setting policies and ensuring the funding for workforce development programs. The Department of Labor, the Department of Education and other federal agencies write the administrative rules, establish programs and provide funds for national, state and local initiatives. Workforce organizations and state and local governments implement the programs by running various projects independently or in partnership. Stakeholders communicate to one another in this complicated process from their own standpoint in their own professional jargons.

Technology advances enable the Internet to serve as an open platform for workforce partners and government to collaborate on programs and projects and to deliver resources and services to the general public. The diversity of these partners and government agencies has mushroomed in the last decade. Some, such as general purpose one-stop centers that serve all job seekers and businesses looking to hire workers, employ generalists who must understand at least at the surface the full range of laws, programs and projects that define the workforce system. Others specialize in meeting the needs of particular groups, including: older workers, veterans, the disabled, workers who have been displaced because of global competition, migrants, minorities that have had uneven access to the workforce system, people receiving welfare assistance, and youth. The barriers for effective and efficient knowledge exchange over this open platform stem from the lack of a systematic modeling of the workforce knowledge domain.

We started investigating the problems and develop strategies to address the barriers three years ago. During this period, we developed a conceptual model that has been revised many times through consultations with and focus groups comprised of workforce professionals, researchers in the workforce field, educators, and officials at the Department of Labor (DOL). In addition, we held focus group meetings to solicit input on the ontology. An earlier version of the Workforce Open Knowledge Exchange (WOKE) ontology was described in Creticos & Qin (2004).

The Workforce Open Knowledge Exchange (WOKE) system is currently under development and a prototype has been shared with DOL and several workforce organizations. Their feedback on the system has been very positive. This paper summarizes the methodologies we used in developing the WOKE ontology and lessons learned from ontology modeling.

2 Development of Domain Specific Ontologies

Ontologies are considered to be the underpinning of Semantic Web. Research and development on ontologies started more than a decade ago. Broadly, there are two approaches for developing ontologies. One approach is to re-engineer part or all of the concepts and relationships in an existing thesaurus by following ontology construction principles. For example, the FAST project restructured the form subject headings in the Library of Congress of Subject Headings (LCSH) (O'Neill and Chan, 2004). Welinga et al. (2001) took the concepts in Western furniture and converted these terms into an ontology for managing the knowledge of antique furniture.

The other approach is to start from scratch. Some of such ontologies are large-scaled ontology projects, including Cyc (http://www.cyc.com/cyc/cycrandd/overview), WordNet (http://wordnet.princeton.edu/), and Unified Medical Language System (http://www.nlm.nih.gov/research/umls/). Their development methodologies are well documented in Fernández-López and Sunciógómez-Pérez (2002). As the Web is increasingly used as an information communication and exchange platform, domain specific ontologies are in great demand for organizations of all kinds. Since most thesauri and classification schemes are often too general to be deployed directly in Web-based systems, many such domain specific ontologies have to be built from scratch. The large number of publications in the past decade with "ontology-based" or similar terms in their titles demonstrates a strong research stream and active development in this area.

Strategies at various stages of building an ontology from scratch have been discussed in Ushhold and Gruninger (1996), Noy and McGuinness (2001), and subsequently in Fernández-

López and Sunciόngómez-Pérez (2002). Leo Obrst (2003) reworded the 7 steps proposed in Noy and McGuinness (2001) as:

1. Determine the domain and scope of the ontology
2. Consider reusing existing ontologies
3. Enumerate important terms in the ontology
4. Define the classes and the class hierarchy
5. Define the properties of classes
6. Define the additional properties related to or necessary for properties (i.e., cardinality, bidirectionality/inverse, etc.)
7. Create instances
8. Create axioms/rules

Most publications in ontology methodologies are written by computer scientists and software engineers, which show a clear orientation toward the engineering aspects of ontology development. Interactive conceptual modeling and the close engagement of subject experts and constituents for input were largely absent from these studies. While each of the steps relies heavily on various methods and tools, the validity and usability of ontologies is largely dependent on how well the ontologies fairly represent the users' conceptualization and contextual understanding of the knowledge domain. But achieving validity and usability requires a large amount of human effort which can be costly to anyone who wants to develop a domain specific ontology. Although Natural Language Processing (NLP) (Aussenac-Gilles et al, 2000), text mining (Maedche and Staab, 2000), query log analysis (Qin and Hernandez, 2006), and machine learning (Bournaud et al, 2000; Wiratunga and Craw, 2000) have been used to draw concepts and terms from texts, the initial modeling and scoping has to be done by humans.

Questions remain in developing domain specific ontologies from scratch: How should ontology designers extract and model concepts obtained from difference sources? What methodologies are useful along the steps of ontology development? The rest of this paper addresses these questions as we explain the "adaptive modeling" methodology used in the WOKE project and offer examples of some lessons learned from the process of developing the WOKE ontology.

3 Adaptive Modeling of Domain Concepts

"Adaptive modeling" is a term borrowed from computer science. In object-oriented programming, objects "have states and respond to events by changing state. The *Adaptive Object-Model* defines the objects, their states, the events, and the conditions under which an object changes state. If you change the object model, the system changes its behavior" (Yoder and Razavi, 2000).

The "objects" in the workforce development domain include concrete concepts such as *laws* that provide instructions and policies to state and local governments, program operators, and other relevant groups as well as appropriate funds, *programs* that implement workforce development policies, *projects* that execute the programs, *organizations* and *persons* of all types involved in programs and projects, and *resources* generated from or created for programs and projects. Abstract concepts are another type in the ontology. The abstract concepts represent the subject content of concrete concepts because they attach a meaning and context for the other objects. For example, projects targeted to youth obtain funds from associated programs that are established by associated laws. The subject of these projects, programs, and laws may be represented by terms such as “Competencies,” “Employable skills,” “Partnership in training,” and so forth. These abstract concepts can not be quantified but are important semantic labels for helping understand what the concrete concepts are about. In a search and browse scenario, this type of knowledge structures will serve as the semantic infrastructure for developing powerful search and browse functions.

We used a wide variety of methods and sources to gather information, to develop the ontology, and to refine our initial model for the workforce domain. One of the sources is the relevant terms in the Library of Congress Subject Headings (LCSH). An examination of LCSH quickly found that the vocabulary and concept relationships defined in LCSH were far from the needs of the workforce community and practices. The example in Figure 1 shows that the closest match for the core concept “workforce” is “labor supply”. However, the results from our focus group meetings with workforce professionals suggest that the concept “workforce” is a general term that produces an image in the mind’s eye of a user of one or many different groups of workers and jobseekers such as dislocated workers, veteran, youth, adult workers, farmers and migrated farm workers, etc. Various federal and state programs as well as projects run by organizations serve the workforce, and the meaning of what constitutes the workforce is established by the context of the program or project itself. E.g., any references made to “workforce” in program documents for an initiative targeting youth implies that “workforce” means job seekers and workers between the ages of 18 and 22 . A document examining the conditions of the labor market for a given area may use “workforce” to describe all who are able to work or who are working. Therefore, “workforce” as it is defined in the LCSH tends to be too general and macro-oriented and does not adequately reflect the more contextually driven meanings of the word.

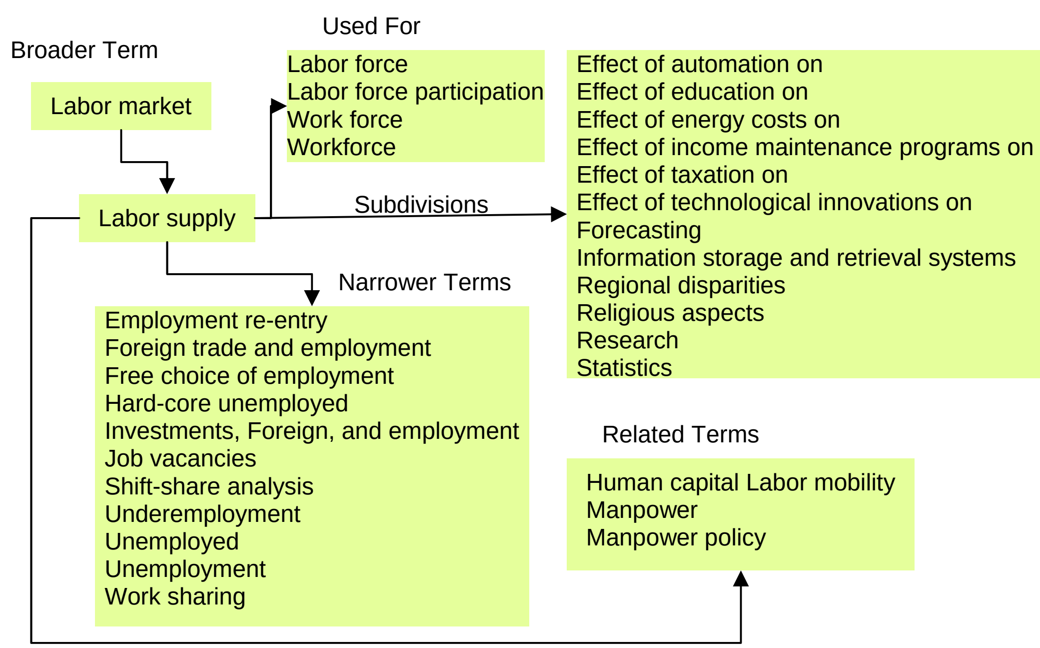

Figure 1. Concepts and relationships for Workforce in LCSH

The nature of the workforce development domain requires the WOKE ontology to be sensitive to the needs of a multidisciplinary, multi-sector user population. We determined that the WOKE ontology must be adaptive to 1) the real world knowledge structure, 2) users' working terminologies and habits, and 3) evolving workforce development policies and practices. Figure 2 describes the methodological framework we used in developing the WOKE ontology.

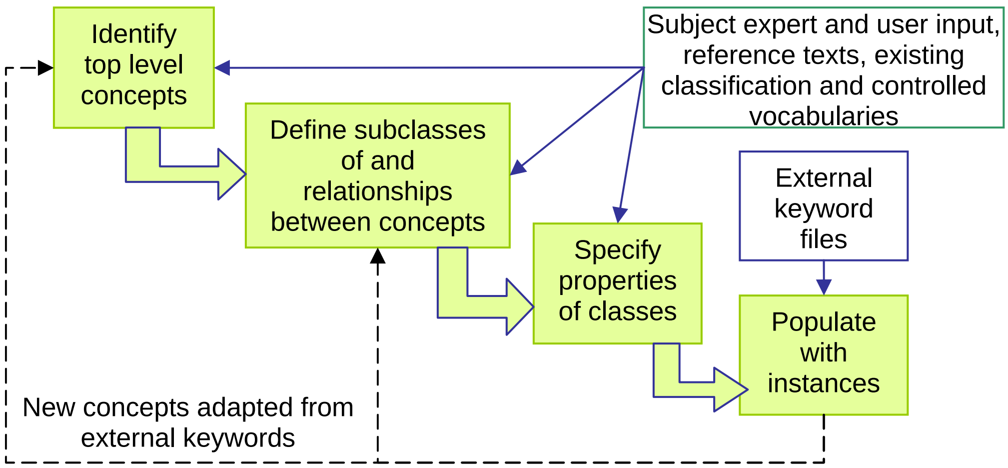

Figure 2. Methodology framework for developing the WOKE Ontology

3.1 Identifying top level concepts

The top level concepts play one of the three roles: 1) as an entity class that has instances conformed to the "is-a" relationship. Laws, resources, programs, projects, organizations, and persons belong to this group; 2) as a subject class that represents the knowledge body of workforce development domain; and 3) as an auxiliary or utility class that will be used as a value space for the entity class properties. The adaptive modeling produced three groups of top classes, each of which plays one of the three roles described here. It is also possible that a subject class plays the role of auxiliary and utility class. For example, "Industrial sector" is a top class representing the industry to which a workforce population belongs or a policy addresses, but at the same time it is also used as the value space for representing the subject content of workforce information resources, projects, and programs.

3.2 Defining subclasses of and relationships between concepts

Concepts are associated with one another through parent-child and sibling relationships in a hierarchical structure. Most top classes in WOKE ontology have two or three levels of subclasses. They came primarily from two sources: brainstorming with workforce professionals and information scientists and pools of keywords collected from constituents' databases. The brainstorming sessions through conference calls and face-to-face discussions resemble a top-down approach. By using this approach we clarified and defined boundaries between concept classes. It helped build the first and second levels of the concept hierarchy, which was then supplemented and enhanced by bottom-up approach – categorizing keywords contributed by workforce organizations against the hierarchy. If similar keywords for the same concept recurred, but there was no place in the hierarchy to fit it, a new class will be created to cover the emerging concept. All classes in a parent-child relationship followed the “a-kind-of” principle, i.e. a subclass is a kind of its parent class. The sibling classes followed the “mutually exclusive” principle, but there were exceptions. For instance, “Unemployed” and “Special classes of workers” are two sibling classes. While it is true that an unemployed worker may be a member of special classes of workers, the two overlapping groups are necessary because the federal programs and workforce projects are often targeted to workers in one group or the other. In this case, these two sibling classes are created according to workforce practices rather than the mutually exclusive rule.

3.3 Specifying properties of concept classes

Classes in an ontology fall into two categories: concrete and abstract. The concrete classes have properties and such properties can be used as a metadata model for an entity class. The properties for resource class, for example, may be modeled after the Dublin Core Metadata Element Set (DC). Based on the feedback from workforce staff, we adapted DC to fit the need in describing workforce resources by dropping the unnecessary elements and added more customized elements (properties). As a result, the resource class has properties using simple text string as the property type, including title, version, status, description, URL, format, keywords, and type. It also has properties that use class or instance of class as the property type, e.g. Activity area is a property of resource that references to the class “Activity areas” since the property type is class. Property definition is also a process of creating connections between related classes.

3.4 Populating the classes with instances

As Noy and McGuinness (2001) and Obrst (2003) point out, populating instances is an important step in developing ontologies. The instances for the WOKE ontology comprise those of entity classes and subject classes. However, these are two different types of instances. The entity class instances function as metadata records for resources, project, programs, and so forth. These instances are well defined in terms of their relationships with both concrete and abstract concepts, as well as their data types and value space. Although subject classes are abstract in nature and not quantifiable for the semantic meanings they represent, they may have synonyms, related terms, broader terms, or narrower terms, which can be treated as instances of subject classes. The ontology currently contains over 260 classes in three levels and more than 1600 terms that have been mapped to the 260+ classes.

The WOKE ontology was developed using an iterative process in which all classes were carefully weighted. We identified the top concepts in the first stage, and then specified the relationships between concepts. In each of the first three stages we consulted with subject matter experts and possible users on the conceptual model and subclasses. Existing taxonomies having broad acceptance by the workforce system were used to populate two classes. Concepts and terms from workforce documents, references, and existing subject categories and vocabularies were carefully examined and refined based on the discussion with subject experts.

4 Lessons Learned

The leadership with respect to the development of WOKE is comprised of subject matter experts and information scientists. This has resulted in a qualitatively different semantic framework than other organizational frameworks now employed in Web-based information systems serving the workforce community. WOKE has relied heavily on the initial involvement of users in identifying top-level concepts and subclasses and in populating the classes with instances. This approach has presented several special challenges, however.

First, the intended users of WOKE are rarely required to articulate a conceptualization of the workforce domain. Their efforts are focused on the immediate moment of delivering a service or in identifying a problem and addressing it through the design, development, implementation and evaluation of a policy, law, program or project. Consequently, it often was difficult to engage users in a broad open-ended discussion on the WOKE ontology. We found that it was necessary to establish a framework for that discussion by presenting an ontology for their reaction and assessment.

Second, we found that it was important to have a mix of users as part of any given discussion. A homogenous group often was too limited in its view of the workforce domain. For example, a group comprised of people delivering services exclusively to veterans would employ narrower definitions to terms shared by others in the workforce system and may identify only a small number of the relationships between instances. The discussion and interaction between members of a heterogeneous group not only revealed each member's understanding of a term and the relationships between instances, it also often produced broader conceptualizations of the workforce domain.

Third, the WOKE ontology offered users their first comprehensive view of the workforce domain. This often prompted new "discoveries." Relationships between classes or instances were not explicitly known until users were asked whether they existed. Once revealed, they prompted users to add new classes or instances. It often came down to a question of what is missing or lacking in the ontology. Overall, this process added both depth and complexity to the WOKE ontology.

Fourth, the complexity and depth of the WOKE ontology is constrained by the point in granularity of detail where the information ceases to be important to the user (i.e., when the detail becomes too fine) and when the perceived time it takes to apply the WOKE ontology in metatagging data becomes too costly in relation to the value of subsequent searches.

Finally, the development of the WOKE ontology is enhanced by the development of applications that demonstrate the utility and value of the ontology in retrieving and re-using knowledge within the workforce domain. The process becomes self-reinforcing as users comprehend the value of the ontology in organizing their understanding of the workforce system and in helping them gain new insights on policies and practices.

5 Conclusions

Creating an ontology for a multidisciplinary domain such as workforce development involves extensive discussions with constituents in various traditional fields. To extract and integrate concepts from difference sources, we developed an adaptive modeling methodology framework. Each iterative refinement of classes and relationships was based on the input from subject experts and frontline staff in order to adapt to the knowledge representation needs of a versatile user population. A prototype system has been developed that implemented conceptual model of WOKE ontology. The initial feedback from a number of major players shows positive comments for the system. As we continue to refine the ontology and populate it with more instances, the knowledge base cumulated will allow for developing more advanced applications.

References

Aussenac-Gilles, N., Biébow, B., and Szulman, S. (2000). Revisiting ontology design: a method based on corpus analysis. In: R. Dieng and O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2-6, 2000 Proceedings, pp. 172-188. Berlin: Spinger.

Fournaud, I., Courtine, M., and Zucher, J.-D. (2000). KIDS: an interative algorithm to organize relational knowledge. In: R. Dieng and O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2-6, 2000 Proceedings, pp. 217-232. Berlin: Spinger.

Creticos, P. and Qin, J. (2004). Open Knowledge Exchange (OKE) for workforce development. In: C. Bussler et al. (Eds.): WISE 2004 Workshops, LNCS 3307, pp. 73–81 (The Fifth International Conference on Web Information Systems Engineering November 22-24, 2004, Brisbane, Australia).

Fernández-López, M. and Sunciόngómez-Pérez, A. (2002). Overview and analysis of methodologies for building ontologies. The Knowledge Engineering Review, 17(2): 129-156.

Maedche, A. and Staab, S. (2000). Mining ontologies from text. In: R. Dieng and O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2-6, 2000 Proceedings, pp. 189-202. Berlin: Spinger.

Noy, N. F. and McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880, March 2001. Retrieved 2/22/2006 from: http://www.ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf

Obrst, L. (2003). 2 Issues: ontology methodology and upper ontology. Ontology-Forum. Retrieved 2/22/06 from: http://ontolog.cim3.net/forum/ontolog-forum/2003-04/msg00008.html

O'Neill, E. and Chan, L. M. (2004). FAST: a faceted LCSH-based subject vocabulary. Presentation at ALA Annual Conference. Retrieved 2/9/2006 from http://www.oclc.org/research/presentations/oneill/ALA2004FAST.ppt

Uschold, M. and Grüninger, M. (1996). Ontologies: principles methods and applications. Knowledge Engineering Review, 11(2): 93–137.
Wielinga, B. J., Schreiber, A., Wilemaker, J., and Sandberg, J. A. (2001). From thesaurus to ontology. In Y. Gil, M. Musen, & J. Shavlik (Eds.) Proceedings of the international conference on knowledge capture (K-Cap'01) pp. 194-201, New York: ACM.
Wiratunga, N. and Craw, S. (2000). Informed selection of training examples for knowledge refinement. In: R. Dieng and O. Corby (Eds.), Knowledge Engineering and Knowledge Management: Methods, Models, and Tools, 12th International Conference, EKAW 2000, Juan-les-Pins, France, October 2-6, 2000 Proceedings, pp. 233-248. Berlin: Spinger.
Yoder, J. W. and Razavi, R. (2000). Adaptive object-model. In: Proceedings of the ACM Conference on Object-Oriented Programming, Systems, Languages, and Applications October 15-19, 2000, Minneapolis, Minnesota. New York: ACM Press.