

Constructing Expert Profiles over Time for Skills Management and Expert Finding

Maryam Fazel-Zarandi
Department of Computer Science
University of Toronto
mfazel@cs.toronto.edu

Mark S. Fox
Department of Mechanical and Industrial Engineering
University of Toronto
msf@eil.utoronto.ca

ABSTRACT

Expert profiling and identification are important to both knowledge workers and organizations. To this end, skills management and expert recommender systems facilitate the management of skills and competencies and help find appropriate experts to meet a particular need. In this paper, we present a technique for generating evolving expert profiles of individuals composed of their skills and competencies using heterogeneous data from divergent sources of information. We use self-declarations, completed learning activities, and previous work experience to generate the initial profile. Recommendations, "wisdom of the crowd", direct observations of online and offline activities, and content generated by the individual both within the organization and on the Web are then used to enrich the profiles and to assess and validate skills and competencies over time.

Categories and Subject Descriptors

K.6.1 [Project and People Management]: Systems development;
H.3.4 [Systems and Software]: User profiles and alert services;
H.1.2 [User/Machine Systems]: Human factors;

General Terms

Design, Human Factors, Management.

Keywords

Expert profiling, expert finding, competency management, skills ontology

1. INTRODUCTION

Expert profiling and identification are important to both knowledge workers and organizations. In today's competitive business environment, companies need to understand the skills and competency of their human resources in order to best utilize them. This is particularly important for organizations that engage with multiple and changing clients such as consulting firms and software development companies since these organizations need to be able to flexibly respond to internal and external demands for

skills and competencies. From a knowledge worker's perspective, finding individuals with appropriate skills and knowing who to go to are important for accomplishing knowledge intensive tasks and solving complex problems since individuals usually rely heavily on others with knowledge and information.

Finding an individual with appropriate knowledge and skills, however, might not be an easy task for many reasons. Expertise is highly dynamic [19], difficult to qualify [7], and varying in level [7]. It is also difficult to validate other people's expertise [19] and to distinguish a good expert from a bad one. Furthermore, due to the complexity of some problems, the assistance of multiple experts may be required [7]. The difficulty of locating an expert increases in larger and more geographically distributed organizations and communities.

In order to facilitate the management of available human resources' competencies and perform qualitative and quantitative reasoning about available and required skills and competencies, the development of totally or partially automated techniques has received the attention of both researchers and organizations. The resulting systems (e.g., [1, 4, 5, 17, 18, 20]) either rely on individuals to provide accurate and comprehensive profiles of their competencies [7], or use mechanisms to automatically discover expertise information from secondary sources [23, 24]. Initially, the evidences considered were content created by the individual within the organization, enrollment in learning activities, and experiences related to the workplace. With the growth of the World Wide Web, however, digital media and communication networks have become an important medium for enabling new levels of interactions in organizations and communities. Many online communities and interactive collaboration spaces (such as forums and wikis) evolve into large-scale knowledge networks [14], providing additional evidence for expertise identification. In addition, the growing number of data published on the Web according to Linked Data principles [2] and using unambiguous vocabularies make automated data integration and reasoning possible for improving expertise identification [23].

In this paper we attempt to improve the selection and integration of different sources of information and types of data to better reflect individual's skills and competencies. The main contributions of our extension are: 1) relating skills to activity performance and specifying how to measure proficiency in different skills in terms of verifiable indicators; and 2) using different sources of information to infer and validate skills and competencies of individuals over time. The expert profiles can be used for a variety of tasks, ranging from locating and matching experts and requirements, to composing teams, gap analysis, and human resource development.

The remainder of this paper is organized as follows: Section 2 briefly presents related work on expert profiling. Section 3 is concerned with the definition and measurement of skills and competencies. Section 4 presents our approach for expert profiling followed by implementation details in Section 5. Section 6 concludes the paper.

2. RELATED WORK

Expert locator systems use implicitly or explicitly provided data about individual's expertise to identify appropriate experts. Creating comprehensive profiles of individuals based on their input, however, is a significant challenge. Self declarations of skills can be incorrect, inaccurate, or insufficient. In other words, people may not be aware of having a certain skill at a level of proficiency or they may lie on their descriptions of what they contributed or accomplished. In addition, individuals often lack motivation to add content to their profiles [11]. Furthermore, competencies are dynamic and an individual's knowledge and experience change over time. For these reasons, methods for extracting up-to-date expertise information from secondary sources have been considered.

To identify experts, automatic expertise finders use different evidences as indicators of expertise, some of which include artifacts created by experts, online and offline activities of experts, and reputation and authority of experts [23]. These methods typically build expertise profiles from implicit data using information retrieval techniques such as probabilistic language models [1, 4]. Although these systems solve the challenges of manually contributing and maintaining expertise profiles, these systems also have limitations. It is not always easy to relate the author of a particular content with the domains of expertise that it identifies [23]. In addition, some of the individuals' expertise may not be represented in their digital trace [11]. Furthermore, although generated content may indicate expertise in a field of interest, it is very difficult to determine the proficiency level of the individual.

In recent years, enterprise social networking has also been considered as a different approach to employee profiles. Some of these platforms, such as LinkedIn, allow individuals to create profiles of themselves and indicate their connections to other users. Others, such as IBM's Fringe Contacts, allow individuals to describe their colleagues by tagging them with keywords on their expertise and interests, thus, creating a publicly visible tag cloud characterizing the individual [3]. However, the resulting profiles lack commitment by the organization, especially with respect to the vocabulary used.

An important shortcoming of existing approaches is related to competency measurement and evaluation. Rating scales consisting of a numeric scale with a brief description of each number's corresponding meaning are the most widely used for specifying proficiency levels. Their disadvantage, however, lies in their inconsistent interpretations across users of a scale. In addition, it is important to know in each case whether completing a course, for example, is sufficient or actual experience is needed to achieve a certain level of proficiency. Thus, in order to have an objective evaluation mechanism, it is important to associate skills with specific activities and characterize them in terms of measurable and verifiable indicators.

3. SKILLS AND PROFICIENCY LEVELS

There are several definitions of competency present in the literature [6]. The definition we assume is the one given by the HR-XML Consortium work group¹: "a specific, identifiable, definable, and measurable knowledge, skill, ability and/or other deployment-related characteristic (e.g. attitude, behavior, physical ability) which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context." We adopt this definition for its emphasis on measurable skills and the connection between skills and activity performance, and use skill and competency interchangeably.

A skill suggests the possibility of performing an activity. However, not everyone possessing a skill can perform all the activities it enables. As such, it is necessary to specify the level of proficiency that is required. Proficiency in a skill may depend on different factors such as familiarity with the subject, the span of the activities one can perform, how much experience one has in performing the activities, etc. Different measurement scales exist for evaluating an individual against a skill. Rating scales, consisting of a numeric scale with a brief description of each number's corresponding meaning, are the most widely used. Their disadvantage, however, lies in their inconsistent interpretations across users of a scale. In addition, it is important to know in each case whether completing a course, for example, is sufficient or actual experience is needed to achieve a certain level of proficiency. Domain-specific metrics, such as number of errors found in a code for a programmer or number of failed projects for a contractor, are other metrics that can be used.

To specify proficiency levels, we use ideas from the Measurement Ontology of [16]. We consider proficiency in a skill as determined by attributes related to that skill that can be measured. For example, recency, years of experience, and average number of errors found are some of the attributes that can be used to measure proficiency in programming. This provides a very general representation which allows for measuring a variety of attributes. These attributes are input to the model (primitive types), can be objective or subjective and categorical or numeric. Each attribute takes up values from a specification set defined by its elements and an ordering between the elements. All measured attributes must have a specification set, and the value for an attribute must be an element of this set. We consider five different proficiency levels: *novice*, *advanced beginner*, *competent*, *proficient*, and *expert* and an ordering between them. For each skill, the required value of a particular measured attribute related to it is identified by domain experts for different levels of proficiency. Individuals are then assessed and/or observed and relevant attributes are measured. In other words, proficiency assessment is done through a series of activities that perform measurement.

In addition to relating skills to activities, we can also assume a taxonomy of skills in a specific domain of interest. It is also possible to define other relations between skills such as the *related-to* relation between two skills indicating that they are highly related to each other in the domain of interest (i.e., having one usually implies having the other). These relations can be used to infer further knowledge about individuals' skills and competencies in the absence of explicit data.

¹ <http://hr-xml.org>

4. EXPERT PROFILING

As stated earlier, there are many different sources of information that can be used for constructing expert profiles. Skill statements can be declared by individuals about themselves or by others, and/or can be derived from 1) activities performed by the individual either online or offline including enrollment in learning activities, experiences related to the workplace, and question-answering in online forums; 2) content created by the individual such as scientific publications, Wikipedia pages, and blog posts; and 3) recommendations and the “wisdom of the crowd”. We use these sources to construct employee profiles and to better understand and validate skills and competencies of individuals over time.

Using content generated by an individual as indicator of expertise has received much attention in previous research. For example, within academic communities, peer-reviewed scientific publications are high quality reliable evidences of expertise. Software codes and documents created in the workplace are other such evidences. With the growth of the Web, there is an additional abundance of publicly available (annotated) user generated content that can be considered. For example, contributing a reliable content to a Wikipedia page indicates that the contributor is knowledgeable on the topic of the page [23]. Corporate blogs are other examples of such user generated content. Generated content may also be useful in determining the proficiency level of an individual in the derived skill. For example, considering papers within the academic community, the impact factor of the journal or conference where the paper was published, and the number of citations it has received are factors that indicate a combination of novelty and significance which then indicate proficiency.

Other than generating or contributing content on the Web, online user activities such as answering questions on question-answer forums can also be used as indicator of expertise. If an individual provides information concerning a particular skill to an expertise seeker in internal forums or on the Web, then it can be stated that the person has that skill. In online forums, the level of proficiency of the information seeker can also be used to determine the proficiency of the information provider. If a skill statement about the same skill exists for the seeker, then it can be suggested that the provider has at least the proficiency specified in seeker’s skill statement. In addition, in question answering forums usually the receiver of the information can provide feedback on the usefulness of the information provided and there is usually a rating of the answer that can be used to imply proficiency.

An efficient and reliable way to gather and confirm skills is to use recommendations and peer-reviews. Recommendations are particularly useful when the recommender is trusted to make judgments about the skill. On the social Web, declarations are done through tagging others with keywords describing their expertise. We can reuse the already available tag data which are usually represented as triples where agent a tags resource or agent x with keyword k . However, since proficiency levels are missing in this data, it is reasonable to assume that the individual who was tagged is at least competent in a particular skill in order to be acknowledged by others. Furthermore, in many organizations, yearly reviews are performed for evaluating employees’ skills. These include performance appraisals where the employees are reviewed by their immediate managers, and 360-degree reviews where a combination of supervisor and peer ratings are considered. Feedback and reviews can only be considered as

useful and meaningful when reliable, and the use of unreliable ratings on performance and abilities can cause serious problems in the organization [12]. Different factors such as how long the rater has known the person, confidentiality of the reviews, and the relation between the rater and the person affect the accuracy of reviews. In a study of accuracy of 360-degree reviews, [8] reports that the “known for one to three years” group is the most accurate, and in general reviews by the supervisor are the most accurate followed by peers and direct reports, with self being the least accurate. [12] reports that although many researchers state that a minimum of three to five peer raters should be used in combination with one supervisor rating, they found that at least ten peers are required for a satisfactory level of reliability.

4.1 Model of the Expert

The unit of skills information about the individual that we consider is a *skill statement* which states that “individual p has skill s at level of proficiency at least l .” Each human resource will have a profile composed of a set of such skill statements.

The degree to which we can believe a skill statement depends on the evidence that supports it and how it was changed over time. For example, in many organizations, workflow management systems are used to track employees’ daily activities. If it is recorded that an individual has performed all the activities which a skill enables at a level of proficiency, then we can acknowledge that the individual has demonstrated that particular skill. On the other hand, if the only support for a skill statement is previous work experience, the credibility of the statement cannot be precisely determined. To take these factors into account, we define different states for skill statements. At any time a skill statement can be in one and only one of four states: *demonstrated*, *suggested*, *unknown*, or *refuted*. These states are core to how the understanding of skills evolves over time. Starting with less than accurate information about the individual (self-declarations with status *unknown*) skill statements are *suggested* or *refuted* as information becomes available until they are *demonstrated*.

To see how skill evaluation changes over time, consider the simple example of Lucy declaring that she is *proficient* in object-oriented programming. Since this is a self-declaration, the credibility of this statement cannot be determined and so it will have state *unknown*. However, if we know that she has previous experience as Java Developer which requires object-oriented programming at least at the *proficient* level and her former manager recommends her, the state of this statement can be changed to *suggested*. As more information becomes available on activities performed by Lucy, proficiency is measured and it is noticed that she does not meet one of the required criteria for being *proficient* in object-oriented programming, thus, the state is changed to *refuted*. Nevertheless, Lucy may have the skill at a lower level of proficiency. Thus, a statement referring to this skill at the *competent* level with state *unknown* is added to the knowledgebase. Since she has performed all the activities required for this level, the state is changed to *demonstrated*. As such, starting with less than accurate information about the individual, the state of a skill statement is changed, becoming closer to reality over time.

5. IMPLEMENTATION

Figure 1 illustrates the overall architecture of our system. We use self-declarations of expertise in addition to skills suggested from

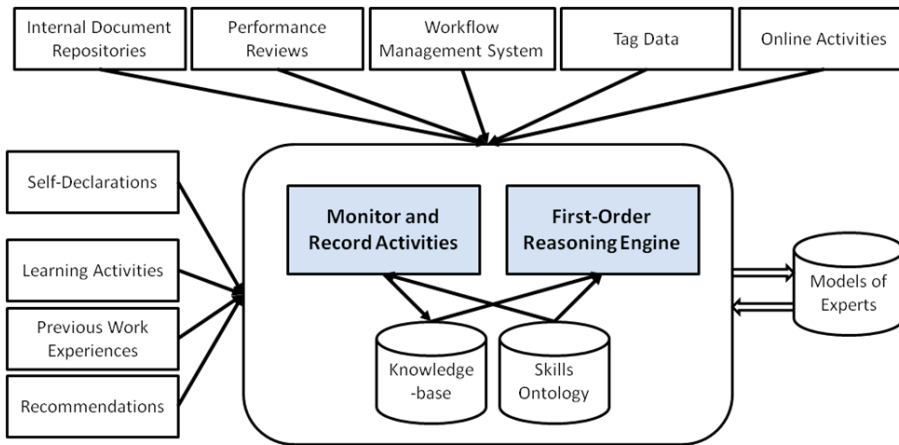


Figure 1. Overall system architecture for creating evolving expert profiles using different sources of information

credentials and work experiences to create an initial model of the individual. We then make use of recommendations, “wisdom of the crowd”, observations of online and offline activities, and content generated by the individual both within the organization and on the Web, in order to reason about individual’s skills and competencies in a dynamic environment. Table 1 presents the different categories used for inferring individuals’ expertise along with the conditions for changing the state of skill statements.

5.1 Representation

We use first-order logic (FOL) as the basis of our representation because of its expressive and declarative capability. FOL enables the characterization of reasoning tasks for inferring and validating skills and competencies in terms of classical notions of deduction and consistency. This enables exploitation of off-the-shelf systems such as existing FOL reasoning engines and database query

engines. Once inferred, the skill statements can be stored in a database for efficient querying.

In particular, we reuse: 1) the Process Specification Language (PSL) [10] which is a first-order language for modeling processes comprised of a layered collection of families of axioms; 2) the Organization Ontology [9] which formalizes the organizational structure; and 3) the Trust Ontology of [13].

Our ontology is an extension of PSL which provides predicates and axioms that enable representation of and reasoning about fluents, activities, activity-occurrences, and values of fluents before and after activity-occurrences. For example, the Activity-Occurrence Extension of PSL defines relations that allow the description of how activity-occurrences relate to one another with respect to the time at which they start and end; and the State Extension introduces the concept of state (before an activity-occurrence) and post-state (after an activity-occurrence) which we use to represent the state of skill statements.

Table 1. Different categories of information along with the conditions for changing the state of a skill statement

| Source | Condition | State |
|----------------------|---|--|
| Observation | | Demonstrated |
| Self Declaration | | Unknown |
| Learning Activity | From reputable institution | Suggested |
| | Otherwise | Unknown |
| Work Experience | Verified by a trusted source | Suggested |
| | Otherwise | Unknown |
| Assessment | Positive results | Demonstrated |
| | Negative results | Refuted at the particular level of proficiency. Unknown at a lower level |
| Content | | Suggested |
| Recommendation | Trusted in the related knowledge field | Suggested |
| | Otherwise | Unknown |
| Question Answering | Information about the skill statement of the requester is available in addition to positive feedback from the requester | Suggested |
| | Otherwise | Unknown |
| Positive Peer Review | Reliable sources trusted in the knowledge field. Number of votes greater than a predefined threshold | Suggested |
| Negative Peer Review | Reliable sources trusted in the knowledge field. Number of votes greater than a predefined threshold | Refuted at the particular level of proficiency. Unknown at a lower level |

Figure 2 illustrates the overall relation between the modules of our extension along with those from PSL². We refer to the entire set of modules as the Skills Ontology. For brevity, the axioms are

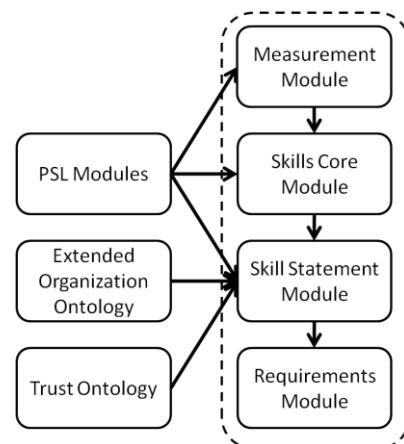


Figure 2. Modules of the Skills Ontology

² We use the following modules from PSL: psl-core, subactivity, complex activity, discrete states, occurrence tree, and fluent trees.

not included in this paper³.

5.2 Integrating Different Sources

The *Monitor and Record Activities* component in Figure 1 is responsible for monitoring individuals' activities and integrating heterogeneous data from divergent sources. To this end, the incoming data is mapped to a particular activity and added to the knowledge-base as an occurrence of that activity. Examples of such activities include *performs* denoting an individual x performing an activity in the workplace, *communicates* denoting x communicating information about a skill to another individual, and *creates-content-on* denoting x has created content related to a particular skill.

To map incoming data to a particular activity, we use existing vocabularies and ontologies. For example, we use the Dublin Core [21] which represents publication metadata for describing physical resources such as books, digital materials such as text files and source codes, and composite media such as web pages. In particular, the attributes *creator*, *contributor*, *subject* (represents the topic of the resource), and *type* (represents the nature or genre of the resource) are used to generate an occurrence of the *created-content-on* activity for the *creator*. Another example is the tagging ontologies of [15] and [22] which can be used for declarations of expertise by others. In this case, an occurrence of the *declares* activity is added to the knowledge-base.

In the absence of meta-data, the *Monitor and Record Activities* component relates documents and online activities to the domain of expertise and skills that they identify using typical information retrieval techniques (e.g., lexical pattern matching and keyword extraction, indexing). For example, if this component relates document d created by individual x to skill y , then an occurrence of the activity *created-content-on* is added to the knowledge-base. Based on this added activity and prior information, further knowledge about x 's skills and proficiency is then inferred and stored in the database for future querying.

6. CONCLUSION

To stay competitive within the market, organizations need to accurately grasp the competency of their human resources. This paper presented a technique for profiling experts and reasoning about skills and competencies in a dynamic environment. Starting with less than accurate models of individuals, additional skills are inferred based on the activities individuals participate in and the contents they generate. The competencies are then validated over time by using recommendations, "wisdom of the crowd", and peer-reviews. In other words, skill evaluation changes over time using the axioms as new information becomes available.

Future work will focus on the design and development of new systems that take advantage of the evolving expert profiles introduced in this paper to determine whether an individual satisfies a set of requirements, or to conduct gap analysis in order to determine who to train and what training programs to offer, and the evaluation of these systems in real world settings.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the constructive comments and helpful advice of Professor Michael Gruninger. This research is supported, in part, by the Natural Science and Engineering Research Council of Canada.

8. REFERENCES

- [1] Balog, K., Azzopardi, L. and de Rijke, M. 2006. Formal models for expert finding in enterprise corpora. In *Proceedings of the 29th annual international ACM SIGIR Conference on Research and Development in Information Retrieval* (Seattle, Washington, 2006). ACM, New York, NY, 43-50.
- [2] Bizer, C., Heath, T., Idehen, K. and Berners-Lee, T. 2008. Linked data on the web. In *Proceedings of the 17th International Conference on World Wide Web* (Beijing, China, 2008). ACM, New York, NY, 1265-1266.
- [3] Braun, S. and Schmidt, A. 2008. People Tagging & Ontology Maturing: Towards Collaborative Competence Management. In *Proceedings of the Proceedings of the 8th International Conference on the Design of Cooperative Systems (COOP)* (Carry-le-Rouet, France, 2008). Springer, 133-154.
- [4] Cao, Y., Liu, J., Bao, S. and Li, H. 2005. Research on Expert Search at Enterprise Track of TREC 2005. In *Proceedings of the 14th Text REtrieval Conference* (2005). National Institute of Standards and Technology, Boulder, CO.
- [5] Colucci, S., Di Noia, T., Di Sciascio, E., Donini, F. M., Mongiello, M. and Mottola, M. 2003. A formal approach to ontology-based semantic match of skills descriptions. *Journal of Universal Computer Science*, 9, 12 (2003), 1437-1454.
- [6] De Coi, J., Herder, E., Koesling, A., Lofi, C., Olmedilla, D., Papatreou, O. and Siberski, W. 2007. A Model for Competence Gap Analysis. In *Proceedings of the 3rd International Conference on Web Information Systems and Technologies (WEBIST)* (Barcelona, Spain, 2007).
- [7] Earl, M. 2001. Knowledge management strategies: Toward a taxonomy. *Journal of Management Information Systems*, 18, 1 (Summer, 2001), 215-233.
- [8] Eichinger, R. 2004. Patterns of Rater Accuracy in 360-degree Feedback. *Perspectives*, 27 (2004), 23-25.
- [9] Fox, M. S., Barbuceanu, M. and Gruninger, M. 1996. An organisation ontology for enterprise modeling: Preliminary concepts for linking structure and behaviour. *Computers in Industry*, 29, 1-2 (1996), 123-134.
- [10] Gruninger, M. and Menzel, C. 2003. Process Specification Language: Theory and Applications. *AI Magazine*, 24 (2003), 63-74.
- [11] Hansen, D. L., Khopkar, T. and Zhang, J. 2010. Recommender Systems and Expert Locators. *Encyclopedia of Library and Information Sciences, Third Edition* (2010), 4433 - 4441.
- [12] Hensel, R., Meijers, F., van der Leeden, R. and Kessels, J. 2010. 360 degree feedback: how many raters are needed for reliable ratings on the capacity to develop competences, with personal qualities as development goals? *The International Journal of Human Resource Management*, 21, 15 (2010), 2813-2830.
- [13] Huang, J. 2008. *Knowledge Provenance: An Approach to Modeling and Maintaining the Evolution and Validity of Knowledge*. Doctoral Thesis. University of Toronto, Toronto.
- [14] Huang, Y., Contractor, N. and Yao, Y. 2008. CI-KNOW: recommendation based on social networks. In *Proceedings of the 2008 International Conference on Digital Government Research*

³ The interested reader is referred to http://www.cs.utoronto.ca/~mfazel/skills_ontology/axioms.html

(Montreal, CA, 2008). Digital Government Society of North America, 1367840, 27-33.

[15] Kim, H., Passant, A., Breslin, J. G., Scerri, S. and Decker, S. 2008. Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces. . In *Proceedings of the 2nd International Conference on Semantic Computing* (San Francisco, USA., 2008).

[16] Kim, H. M., Fox, M. S. and Gruninger, M. 1999. An Ontology for Quality Management - Enabling Quality Problem Identification and Tracing. *BT Technology Journal*, 17, 4 (1999), 131-140.

[17] Liu, P. and Dew, P. 2004. Using Semantic Web Technologies to Improve expertise Matching within Academia. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies (I-KNOW)* (Graz, Austria, 2004), 370-378.

[18] MacDonald, C. and Ounis, I. 2006. Voting for candidates: adapting data fusion techniques for an expert search task. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (Arlington, Virginia, USA, 2006). ACM, New York, NY, 387-396.

[19] Maybury, M. T. 2006. *Expert finding systems*. MITRE Center for Integrated Intelligence Systems Bedford, Massachusetts, USA.

[20] Mochol, M., Wache, H. and Nixon, L. 2007. Improving the Accuracy of Job Search with Semantic Techniques. In *Proceedings of the 10th International Conference on Business Information Systems* (Poznan, Poland, 2007). Springer, New York, NY, 301-313.

[21] Nilsson, M., Powell, A., Johnston, P. and Naeve, A. *Expressing Dublin Core metadata using the Resource Description Framework (RDF)*. Dublin Core Metadata Initiative, Singapore, <http://www.dublincore.org/documents/dc-rdf/>.

[22] Passant, A. and P., L. 2008. Meaning of a tag: A collaborative approach to bridge the gap between tagging and Linked Data. . In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008)* (Beijing, China, 2008).

[23] Stankovic, M., Wagner, C., Jovanovic, J. and Laublet, P. 2010. Looking for Experts? What can Linked Data do for You? In *Proceedings of the WWW 2010 Workshop Linked Data on the Web (LDOW2010)* (Raleigh, NC, 2010).

[24] Yimam-Seid, D. and Kobsa, A. 2003. Expert-finding systems for organizations: Problem and domain analysis and the DEMOIR approach. *Journal of Organizational Computing and Electronic Commerce*, 13, 1 (2003), 1-24.