THE DESIGN OF A SOCIAL SEMANTIC SEARCH ENGINE

Preserving Archived Collaborative Engineering Knowledge with Ontology Matching

Jörg Brunsmann
Faculty of Mathematics and Computer Science, Distance University of Hagen, D-58097 Hagen, Germany

Keywords: Semantic Web, Social Search, Semantic Search, Natural Language Processing, Ontology Matching, Onotoloy Alignment, Linked Data, Provenance, Digital Archives, Knowledge Management, Resource Description Framework, Ontologies.

Abstract: Private and business related knowledge acquisition is either performed via learning by doing or via human dialogue that includes transmission of social or collaborative questions and answers. Unfortunately it can be a time consuming task to find a trusted friend on the web for private recommendations or to find a qualified expert colleague in a (virtual) organisation for work-related questions or to find a suitable company contact person as a customer. Recently, such social question and answering is conducted with internet based technologies like social search engines which route a question to a appropriate human selected from a social or expert network. However, even if social search engines are involved, it is unlikely that existing social search approaches exploit machine-readable lightweight ontologies that enable classifying, publishing and sharing questions and answers to support subsequent semantic search without human involvement. This paper proposes the combination of semantic web and social search technologies in order to publish and archive social and collaborative generated knowledge for future reuse. Since knowledge classifying vocabularies evolve over time the paper also describes why archived knowledge may become obsolete and how ontology matching methods are used to migrate knowledge to conform to contemporary vocabularies.

1 INTRODUCTION

Probably without being aware of it, at some point everyone has been in touch with *social search* knowledge. Posting or answering a question in an internet forum, asking a colleague via phone in daily job activities, searching for responses to technical problems using a web search engine, asking a company agent as a customer for contract related help, writing a product review for a e-commerce web site or asking a friend about his private opinion via microblogging web sites, mobile phone or instant messenger are all valid examples of social search. In all these examples new explicit knowledge is created because one person asks or searches for knowledge from another person.

Because processing of natural language questions is not yet fully supported by traditional search engines, social search (Narasimhan, 2010) enables users to write down questions in natural language and let other users in their social or expert network answer the question. Selecting the user who is most competent to answer a question is based on *social rank* which reflects reputation and connectivity and other metrics (Hangal, 2010). While *page rank* selects a document based on authority, social rank selects a person based on intimacy (Horowitz, 2010) and trust (Morris, 2010).

During the search workflow (Evans, 2009), users try first to search on their own and use their social network only if the intial search was not successfull. To support this workflow, previosuly conducted questions and answers must be annotated with vocabularies based on the Ressource Description Framework (RDF). Such annotated documents are then published and archived for future reuse.

If metadata that represents contextualized social search knowledge is archived, it is necessary to maintain the annotated knowledge for future search and access. However, RDF based ontologies evolve and archived annotated knowledge must be migrated by processing ontology alignments.

The remainder of the paper is structured as follows. The next section provides a characterization of engineering knowledge. Section 3 describes

Brunsmann J..
THE DESIGN OF A SOCIAL SEMANTIC SEARCH ENGINE - Preserving Archived Collaborative Engineering Knowledge with Ontology Matching .
DOI: 10.5220/0003070402000205
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2010), pages 200-205
ISBN: 978-989-8425-29-4

semantic web knowledge representation technologies. Section 4 elaborates on classifications of social search and section 5 proposes the combination of social and semantic search technologies. The last section concludes with a description of future work.

2 KNOWLEDGE

Knowledge that is based on question and answer is articulated into a language and then transmitted and communicated to others. Knowledge as concept is formalized in the *Data-Information-Knowledge-Wisdom* (DIKW) model (Fricke, 2009). In the DIKW model, the data layer consists of raw elements whereas information provides declarative answers to who, what, where and when questions. Finally, knowledge provides answers to how (procedural) and why (causal) questions.

Knowledge acquisition processes by individuals and groups in enterprises are described by the famous *knowledge spiral* (Nonaka, 1995) as a conceptual foundation for enterprise knowledge management. *Explicit* knowledge is capable of being stored in machines whereas *tacit* knowledge is in person's heads and is very difficult to be represented in machines. Existing knowledge is *internalized* to create new tacit knowledge which is *socialised* afterwards, then *externalised* and so on. This process builds the knowledge spiral.

Mapping this spiral to social search activities, we find that during socialisation knowledge acquisition is done by verbal questions and answers. Annotating and publishing these questions and answers pairs is externalisation. Knowledge acquisition by searching published questions and answers is internalisation. Finally and especially, knowledge combination is performed during enterprise collaborations in the engineering industry.

2.1 Engineering Industry Knowledge

The SHAMAN digital preservation project (SHAMAN, 2009) investigates the knowledge preservation of different domains including the industrial design and engineering industry (Brunsmann, 2009). This industry use tools organized by product life cycle management (PLM) systems (SHAMAN, 2008) and strongly depend on heterogeneous knowledge resources like employees, processes, documents, databases (Kamara, 2002). Use cases for social engineering search knowledge include:

During *collaborative innovation processes* an idea is converted it into a sellable product by performing collaborative brainstorming sessions or by interviewing customers. The idea needs to be educated to colleagues, business partners and customers so that the partners can contribute their own ideas. Therefore it is necessary, that during brainstorming sessions and subsequent collaboration sessions questions and answers are recorded.

During *domain and enterprise collaboration*, co-operations between different enterprises (virtual organisations) and different engineering domains are formed. Such cross-enterprise and cross-domain collaborations exploit the specific knowledge area of each cooperation partner.

These two use-cases show that tacit "know-how" and "know-why" engineering knowledge is exploited during social search activities. For future enterprise benefit it can be made explicit, if it is expressed with machine-readable semantic web technologies which enable archiving and reuse of knowledge.

3 SEMANTIC SEARCH

The *semantic web* allows to reason over knowledge which is modelled as sets of assertions. In recent years the focus of the semantic web switched to publishing, integrating and retrieving *linked data* following the principle that web resources are identified with *interlinked resolvable HTTP URIs*. Linked data is modelled with lightweight vocabularies (ontologies) like SKOS, SIOC, FOAF, Dublin Core, vCard (YAHOO, 2010) expressed by the Resource Description Framework (RDF).

RDF triples can also be integrated into existing HTML pages by using RDFa, eRDF or Microformat. Search engine crawlers extract and store relevant RDF triples. However, semantic content that has been crawled is useless if it cannot be searched and accessed. Current search engines are keyword based and do not reduce the communication gap between a human and a computer.

Semantic web technologies promise that search engines will be able to answer natural language user queries. Current approaches either apply natural language processing to unstructured text or they assume the existence of structured statements over which they can reason. For example, (Lopez, 2007) describes an ontology-driven question answering system that takes an ontology and a natural language query as input and returns answers from a triple store.

Not only on the internet but also in enterprises, ontologies make tacit knowledge explicit. A shared and common meaning is modelled with ontology classes and properties which were formerly tacit in the head of employees so that they may be understood by other employees and partners.

Unfortunately, knowledge described with ontologies face the threat of syntactic and semantic heterogenity: the conceptualization of a domain varies from different author viewpoints, has different terminology (e.g. synonymy), overlap with other ontologies, cover different portions of a domain and can be represented by different ontology languages (e.g. RDF or Topic Maps).

In addition to such syntactic and semantic heterogeneitiy, it is very common that a real world domain is continually changing so that the ontology evolves as well. In order to keep the triples interpretable, ontology alignments can be used which are produced by ontology matching. Ontology matching is the process of relating two ontologies sharing one domain or two versions of one ontology. These ontology alignments can then be used to migrate ontology instances and to integrate different ontologies of the same domain.

4 SOCIAL SEARCH

While semantic search uses the contextual meaning of keywords to improve search results, *social search* describes the process of incorporating content generated by individuals and the individuals themselves into the generation of search results (Smyth, 2009). In contrast to social search, the term *people search* denotes the process of finding information about individuals across internet documents. The following sections give an overview of existing social search classification and lay out a use cased based social search classification.

4.1 Social Search Classification

Social search is the process of finding information online with the assistance of social resources. While searching for knowledge via keywords in a search engine, also results from the social network are presented, because it is likely that friends are more trusted than documents from the internet (Morris, 2010). *Collaborative search* is a social search where one or more individuals share an information need and work together to fulfill that need.

According to (Narasimhan, 2010) social search processes improve search results by

- Machine-based passive search using social media (user generated content). The social graph is used to filter trusted content and contextualize the response.
- Human-based active search using social graph (user interactions). The social graph is used to socialize the query and to classify user expertise which enables to route to incentivize participation.

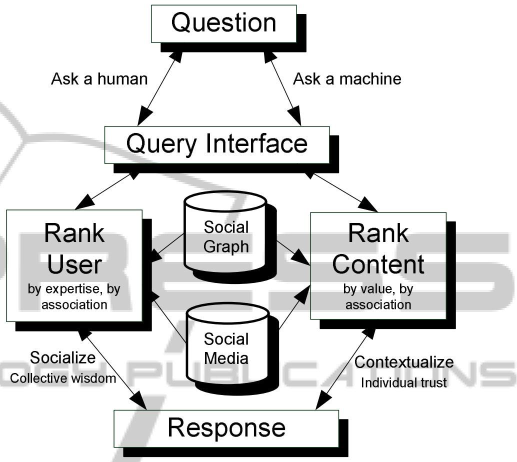

Figure 1: Social search according to (Narasimhan, 2010).

Figure 1 shows a social search system according to (Narasimhan, 2010) which can be regarded as:

- As sensor: the crowd as collective and real-time intelligence collector.
- As filter: trusted social relationships are used to extend search results.
- As router: social relationships are used to forward queries to a person with relevant expertise.

(Evans, 2009) defines three different types of social search.

- *Collective social search* is capturing real-time network trends or group wisdom.
- *Friend-filtered social search* is using the social network data exclusive or alongside traditional search results.
- *Collaborative search* (question answering) is when two or more users work together to find the answer to a problem.

4.2 Use Case based Classification

All existing classifications that were described above regard the term *social* as a network of friends. However, the full potential of social does not end at the list of friend for private reasons. It also includes colleagues, experts, agents and other actors. That means, social relationships not only exists in the private realm but also within enterprises or virtual organisations between colleagues. And social relationships exist in customer relationship environments (e.g. consumer to company agent).

This will get evident if one takes a closer look at the types and topics of questions and the motivations for questioning and answering that were described in (Morris, 2010). Table 1 maps question type and topic to use case.

Table 1: Type, topic and use case of questions.

Type	Topic	Use case
Recommendation	Technology, Restaurant	Private
Opinion	Ethics	Private
Invitation	Family	Private
Offer	Shopping	Business
Factual knowledge	Technology, Contract	Personal
Factual knowledge	Professional	Enterprise

Based on table 1, one can identify four different use cases:

Private social search often asks for recommendation and opinions. The motivation for the questioner is trust or failed search whereas the motivation for the answerer includes altruism, free time, to connect socially or ego motivations. The interaction can be regarded as “Friend2Friend”.

Personal social search include acquisition of factual knowledge for problems solving. A customer might have a question regarding the products that are distributed by a company. The motivation for the questioner is answer speed and quality and the motivation for the answerer might be customer satisfaction. The type of interaction is “Customer2CompanyAgent”.

Business social search enables a marketing unit of a company to contact a customer. E.g., the customer has extended his contract and the company wants the customer to recommend the product to his social network. The marketing unit can use the customers’ social network. The motivations for the questioner are business opportunities and the motivations for the answerer might incentives. The interaction type is “Company2Customer”.

Enterprise social search means to search for a work related answer which includes contacting a expert colleague. The enterprise collaboration and innovation process are examples of enterprise social search which might involve more than two persons and thus can be regarded as collaborative search. The motivation for the questioner is answer speed and quality and the motivation for the answerer might be expertise, ego and incentives. The interaction type is “Colleague2Colleague”.

Making the social and collaborative search knowledge explicit for others by attaching semantics is important for future knowledge reuse.

5 SOCIAL SEMANTIC SEARCH

(Evans et al, 2009) investigated people's search processes and preferences and found that they want to attempt to search on their own first or do not wish to interrupt their colleagues before they have tried to search on their own independently. Later, if the searcher did not find a satisfactory answer to a problem, they often turn to a colleague for help. Therefore, early social support should be passive.

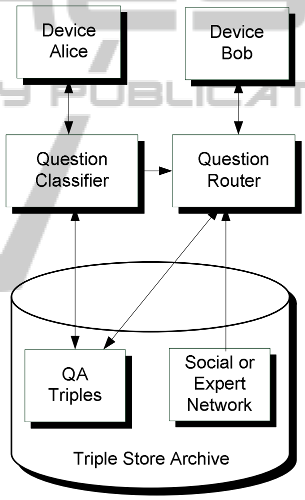

Figure 2: Social and semantic search.

Social search aims to find a human to answer a question whereas semantic search tries to find relevant documents that conforms to keywords or natural language queries. Combining these two complementary search strategies will provide a growing and real-time collective social semantic knowledge system that enables human/human interaction and human/machine interaction. Figure 2 shows a schematic view of such a system.

The system has access to semantically annotated documents and the user’s social network. The social semantic knowledge system does not force user to generate documents, it rather captures question-answering interaction, contextualize the social search results, publishes and archives the previously conducted questions and answers pairs.

Answering a question by verbal communication can be regarded as socialisation of knowledge. If the answering process is performed by social search processes it involves written words and thus can be seen as externalisation because it made available as document annotated with metadata. Externalized knowledge needs to be contextualized and attached with metadata and must be published. Such published documents can be found via semantic search and internalised by individuals.

5.1 Social Semantic Search Discussion

This section discusses requirements, advantages and disadvantages for social semantic search.

5.1.1 Requirements

The network of individuals has to be big enough in order to make social search effective. If only few individuals exists in the social network, it is likely that some questions remain unanswered which lowers the acceptance of the social search engine.

Experts must be proven and competent in their domain to avoid false answers. After receiving the answer the questioner must rate the answer in order to document the answer and answerer quality. In an enterprise, the answerer and questioner must be kept anonym to enable dispassionate ratings and prevent unmotivated answers.

Whereas the questioner has an immediate and present need or interest, the participation of the answerer has to be incentivized, so that also the answerer has a satisfactory experience.

Social search must be symmetric, e.g. a company must be able to submit an offer to a customer and a customer must be able to pose a question to a company.

The social semantic search system should identify if a user has difficulties in searching without human help. For example, if a user already searches for 10 minutes, an appropriate domain-specific expert could be suggested to chat with.

From a list of (anonym) individuals one should be able to select the answerer based on some criteria. In other situations the routing method should select the answerer on its own. Therefore, the social search system should support three different communication methods: route the question and answer synchron (e.g. select a user in instant messenger), semi-synchron (routing algorithm selects an appropriate answerer) or asynchron (like Yahoo answers).

5.1.2 Advantages

The social semantic search system increases the answer speed and reduces spam search results as it helps to generate answers from a trusted network of friends or experts so that answers have more relevance to the questioner.

Social search enables interaction via dialogue which helps the process of understanding and fosters the generation of implicit and explicit knowledge. Finding the right answer is faster compared to traditional search. In addition, the enterprise knowledge base gets better and is kept up to date as more people participate.

Semantic search capabilities increase the probability to find satisfactory results and thus reduce the probability that human involvement is needed. Finally, a company can improve the customer satisfaction by providing a real-time knowledge base to the customer.

5.1.3 Disadvantages

The questioner needs to trust the social ranking algorithm as probably non-experts will answer questions. In addition, blind trust can be misleading, since the answer of a close friend can still be wrong. Additionally, since the system is based on human contributions it is dependent on the input of the users and if the user base is small, it may not reach full acceptance.

Since the world is changing fast, experts need to keep up to data with knowledge explosion which is neither an easy nor cheap task. In addition, the answerer is interrupted in his normal work activities and receives incentives. All these aspects have to be evaluated by comparing costs and benefits of a social semantic search engine.

In enterprise social search the participation will decline if incentives are low. And finally, on the internet sooner or later spam will reach social search, which definitely will reduce the reputation of the social search engine.

6 CONCLUSIONS AND OUTLOOK

This paper described how knowledge is acquired by social and semantic questions and answers on the web, in the enterprise and in customer relationship management affairs. It also proposes to support the intuitive search workflow that first includes searching via machine and then involves a human from the social network. The initial non-human search is improved by semantic web technologies. The paper described social and collaborative use case scenarios in the engineering industry and elaborated on how to archive semantically annotated question and answers pairs for future reuse. Such archived engineering knowledge is exposed to threats like syntactic and semantic heterogeneity which could result in semantic obsolescence. Fortunately, ontology mappings help to overcome such issues. The contributions of this paper include:

- Combination of semantic web and social search technologies.
- Extension of social search for customer relationship management purposes and enterprise collaborations.
- Publishing and archiving of RDF annotated questions and answers pairs.
- Usage of ontology matching in archiving of RDF based question and answer knowledge.

Further investigations include a wide variety of research topics:

- Further evaluation of existing social search approaches and systems.
- Types of communication and dialogue workflow in private and business scenarios.
- User interface design for different usage scenarios (private, business).
- Evaluation of collaboration patterns (Pattberg, 2007) for usage in social search.
- Explore other social search use case scenarios (e.g. collaborative ontology engineering).
- Exploiting social network analysis metrics for social rank calculation.
- Detailed capturing of search workflow.
- Exploration of objective rating methods.
- Exploring incentive possibilities (real and virtual currencies).
- Description of multilingual problems of social semantic search.
- Evaluation of question analyzing methods.
- Appropriate ontology matchings technologies for evolving RDF vocabularies.
- Evaluation of enterprise social search costs and benefits.

ACKNOWLEDGEMENTS

This paper is supported by the European Union in the 7th Framework within the IP SHAMAN.

REFERENCES

Brunsmann, J., Wilkes W., 2009. Enabling product design reuse by long-term preservation of engineering knowledge. *International Journal of Digital Curation*, Vol 4, No 3, 17–28.

Evans B. M., 2009. 3 Flavors of Social Search: What to Expect.http://www.readwriteweb.com/archives/3_flavors_of_social_search_what_to_expect.php

Evans B. M., Kairam S., Pirolli P., 2009. Do Your Friends Make You Smarter? An Analysis of Social Strategies In Online Information Seeking. *Information Processing and Management*.

Fricke, M., 2009. The knowledge pyramid: A critique of the dikw hierarchy. Journal of Information Science, 35(2), April 2009, pp.131–142.

Hangal S., MacLean D., Lam, M. S., Heer J., 2010. All Friends are not Equal: Using Weights in Social Graphs to Improve Search. *Proceedings of the SIGKDD 4th Workshop on Social Network Mining and Analysis*.

Horowitz D., Kamvar, S. D., 2010. The Anatomy of a Large Scale Social Search Engine, WWW 2010 conference.

Kamara J. M., Augenbroe G., Anumba C. J. and Carrillo P. M., 2002. Knowledge management in the architecture, engineering and construction industry, Journal of Construction Innovation, Vol. 2, 53-67.

Lopez, V., Motta, E., Uren, V. and Pasin, M., 2007. AquaLog: An ontology-driven Question Answering System for organizational Semantic intranets, *Journal of Web Semantics*, 5, 2, pp. 72-105, Elsevier.

Morris, M. R., Teevan, J. Panovich, K., 2010. What do people ask their social networks, and why? A survey study of status message Q&A behavior. In *Proceedings of CHI*.

Narasimhan, N., 2010. The Evolution of Social Search, http://nityan.wordpress.com/2010/03/30/the-evolution-of-social-search/

Nonaka, I. and Takeuchi, H., 1995. The Knowledge-Creating Company, Oxford University Press.

Pattberg, J.; Fluegge, M., 2007. Towards an Ontology of Collaboration Patterns. In (Adam Pawlak et. al. Eds.): *Proceedings of 5th International Workshop on Challenges in Collaborative Engineering (CCE'07)*, pp.85-95, GI-Edition – Lecture Notes in Informatics.

SHAMAN Project, 2009. SHAMAN Homepage. http://shaman-ip.eu/

Smyth, B., Briggs, P., Coyle, M., O'Mahony, M., 2009. Google Shared. A Case-Study in Social Search. Proceedings of UMAP.

YAHOO, 2010. SearchMonkey documentation, http://developer.yahoo.com/searchmonkey/smguide/profile_vocab.html