BUILDING AN EMPLOYEE-DRIVEN CRM ONTOLOGY

Céline Van Damme*, Stijn Christiaens°, Eddy Vandijck*
*Vakgroep MOSI, Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
{cvdamme, eddy.vandijck}@vub.ac.be

°STARLab, Vrije Universiteit Brussel
Pleinlaan 2, 1050 Brussel, Belgium
stijn.christiaens@vub.ac.be

ABSTRACT

The information overload in today's web and information systems has a negative impact on the information retrieval process. Most information is stored in an unstructured format, as is the case of valuable customer information in CRM systems. A CRM domain ontology can help in overcoming the problem of unsatisfying search results and ameliorate the customer knowledge creation process. We present an integrated visual approach, where the users themselves can take part in the ontology creation process. In this three-phase approach, we support the users in the process of adding extra information to their content. Our approach is based on an ”employonomy”, a folksonomy created by the employees, text mining and user feedback.

KEYWORDS

Ontology, Folksonomy, CRM, Visual approach

1. INTRODUCTION

Managing qualitative knowledge is considered as a critical success factor for Customer Relationship Management or CRM. The more knowledge the company has over its customers, the better it can understand and serve them. This will enhance their loyalty to the company (Gebert 2002).

The first stage in the CRM process is mainly based on processing the information extracted from the front office business processes (marketing, sales, service). This kind of information is stored in an operational CRM system in a structured as well as in an unstructured format (Dyché 2001). This latter format can only be retrieved through the use of a full text search engine[1]. In general, simple full text search engines employed in large document collections do not generate satisfying search results: only 1 out of 5 relevant documents are retrieved (Blair and Maron 1985; Blair and Kimbrough 2001).

We believe this customer knowledge creation process can be improved by building a CRM domain ontology. Citing Gruber (1993), ”an ontology is an explicit specification of a conceptualization”. All the concepts, instances and relations (between concepts and instances) of a domain are formally described in a machine interpretable language. An ontology allows more accurate search results in unstructured information retrieval. However, building ontologies is expensive, time

[1] This is the case in SugarCrm, Salesforce.com and IExtensionsCRM.

consuming (Navigli 2003; Hepp 2007) and does not reflect the actual vocabulary of the whole user community (Hepp 2007).

A folksonomy, a portmanteau word combining folk and taxonomy (Smith 2004), is a new categorization technique on the WWW. Users may categorize all their information -identified with a unique URL - with freely chosen keywords or tags. This additional categorization increases relevance in the information retrieval process. The aggregation of all user-generated tags or metadata leads to a bottom-up taxonomy with its own strengths and weaknesses. On the one hand folksonomies are unable to cope with issues such as synonymy, polysemy and basic level variation (Golder and Huberman 2006). On the other hand they are a direct reflection of the user’s vocabulary and are created at a high pace and low cost (Quintarelli 2005). Therefore, semantic web researchers are trying to offset the weaknesses of formal semantics (ontology) with the strengths of informal semantics (folksonomy) and vice versa in order to create better ontologies for the semantic web (Spyns et al. 2006; Christiaens 2006; Van Damme et al. 2007).

We believe that a folksonomy can also be employed for creating a corporate domain ontology, in this case a CRM domain ontology. However, introducing a folksonomy in an enterprise, an ”employonomy” as we coined the word, engenders a motivation problem: employees have to participate in the tagging process. We deem that this problem can be diminished (1) by using an employee intervention in a time efficient way and (2) by providing a user friendly interface.

In this paper, we present an integrated visual approach for building an CRM ontology that is based on (1) an ”employonomy”, a folksonomy created by the employees, (2) text mining and (3) user feedback . In section 2 we present work that is related to our research. We then explain our integrated visual approach in section 3, followed by conclusions and directions for future work in section 4.

2. RELATED WORK

Our work is related to the research of folksonomy refinement and community-driven ontology engineering.

Recently, there is a vast interest on folksonomies on the WWW. As mentioned in the previous paragraph, this latter technique is quite valuable, although it has its weaknesses. Therefore more and more research is performed on enriching folksonomies. Bar-Ilan, Shoham et al. (2006) proved through a tagging experiment on images that structured tagging generates more value than unstructured tagging. Other experiments show how a taxonomy is induced out of tags through statistical natural language processing techniques and an algorithm based on cosinus similarities (Schmitz 2006; Heymann and Garcia-Molina 2006). Van Damme et al. (2007) are suggesting an overview of various approaches (e.g. existing web resources such as Google and Wikipedia, online lexicons and human feedback) for turning folksonomies into ontologies. The latter approach contains similarities with our research since we are also creating an ontology out of informal semantics. It differentiates from ours since (1) we are building an ontology for a CRM system, (2) we are not relying on existing formal semantics or web resources, (3) we are including text mining techniques, and (4) we are explaining how the resources (employonomy, text mining and user feedback) can be employed.

The last few years more and more research has been carried out on community-driven ontology engineering. The DOGMA-MESS system presented in de Moor et al. (2006) shows how ontology engineering can be scaled down in complexity and up in the number of participating domain experts. In Hepp et al. (2006) the authors have proven that the web pages of the wikipedia community can be employed as ontological concepts. The authors have even shown that the general Wiki technology has a low technology barrier for ontology engineering. This low barrier is also present when using Concept Maps: their graphical user interface facilitates the ontology construction process (Hayes et al. 2005). The authors have demonstrated that using Concept Maps allows a collaboratively capture of knowledge in ontologies. The approach we are presenting in this paper is related with this visual approach: we are offering an interface that minimizes the intellectual efforts. In that way we hope to dispel any possible employee resistance.

3. AN INTEGRATED VISUAL APPROACH

We propose an integrated visual approach that creates a CRM domain ontology out of an "employonomy". However, letting employees tag their content leads to a flat space taxonomy. This kind of taxonomy is insufficient for creating the concepts and its relationships in the ontology. Using text mining techniques such as calculating the tf-idf weight might be a solution, but this kind of approach lacks human interference. Therefore, we propose an approach based on an intertwined usage of both techniques.

3.1 Elaboration of the ontology creation

When an employee generates unstructured content (e.g., addition of notes) or uploads a document in the CRM system, she will be asked to provide some tags that categorize the content. We will process the new content or text and mine it for keywords based on the tf-idf (term frequency inverse document frequency) weight. The weight is calculated for every word in the text that remains after removing stop words, articles and performing word stemming. The tf-idf formula: tf-idf= n/N * log(corpus/D_n) (with n = frequency of the word in the document, N = total number of words in the document, corpus = collection of documents and D_n = document frequency of the word) multiplies the word's document frequency by the logarithm of its inverse document frequency in the corpus (Salton and Mcgill 1986). The second part of this expression computes how common the term is in the corpus. In our situation the corpus consists of corporate documents related to customer business processes. A high tf-idf weight represents a rarely used word in the corpus but frequently employed in the specific document and consequently becomes a good descriptive document keyword. After calculating the weights, we will rank all the words by descending weight and then pick out the first two terms, since we assume that two keywords are enough for describing this kind of content. Testing in a corporate environment will render feedback.

Our approach will use the sum of the keywords determined by the tf-idf weights and tags as input for the following phases. We claim that there is a high probability that these terms will be easily related to each other, as they should be representative for the actual meaning in the text. The number of terms will be a first measure to determine the continuing actions. The total number of terms (number of tags + number of keywords) can be restricted to a maximum (e.g., 10 in order to reduce user complexity). The maximum number of relations[2] between these terms can be found by calculating a combination without repetition: number of combinations = n!/ (n-r)! (with r = 2, the number to be chosen and n = the number of terms).

[2] We assume a maximum of one relation between two terms.

As a result, the possible number of relations increases quickly (almost exponentially). With 3 terms, we find a maximum of 3 relations, 4 terms result in 6 relations and 5 leaves the user with 10 relations to find. It is clear that the required user effort increases equally with the number of possible relations. As we stated in the beginning of this paper, asking too much feedback at once will deter the cooperation of the employee. We will tackle this complexity by dividing the task over two phases: bagging and relating. The number of relations determines the next phase. We will use 7 as the threshold number, as this is also the number of short-term memory locations (Matlin 2003). In case the maximum number of relations is above 7, we direct the user to the bagging phase. If the maximum number of relations exceeds 7, we skip the bagging phase and point the employee immediately to the relating phase.

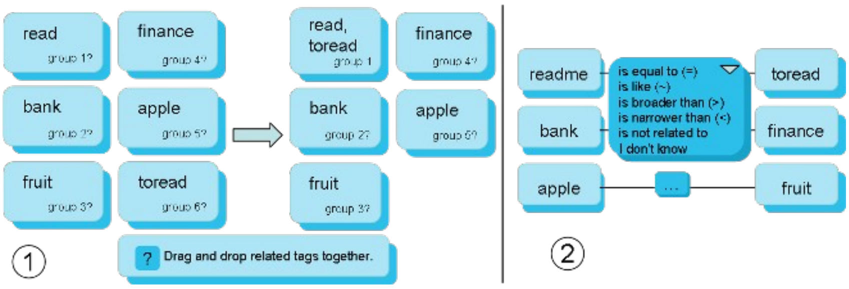

Figure 1. Phase 1 (bagging) and phase 2 (relating)

3.2. Phase 1: bagging

The first phase is the bagging phase, which is the easiest of the two phases. It starts whenever the maximum number of relations is above 7. As the complexity of trying to capture a too high number of relations will certainly scare the employees away, we limit this phase to making sets of terms. The employee will be able to drag and drop related terms into bags. The drag and drop mechanism ensures that this operation runs smoothly and without much input effort. A bag of terms is a visual cue that represents a certain (currently further unspecified) relation. The operation requires little conceptual overhead, as the user does not have to worry about the details of the relation. If she believes there is any reason to group these terms together, she can do it. In this way the SKOS[3] "is related to" relation is captured without bothering the user with this explicit representation. The frequency of co-occurrence of two terms in a bag will be stored in the database and used as one of the measures in our approach. Once a co-occurrence threshold is reached for two terms, they can be transported into the next phase (relating) for further information gathering. Figure 1.1 displays how this should work.

3.2. Phase 2: relating

In the second phase, we try to collect more details on the relation between terms. As this is a step that requires more conceptual overhead, we need to facilitate the user as much as possible. Our first reduction of complexity results from the fact that this phase is only entered when the number of possible relations is below 7. This number of relations is manageable, partially caused by the fact that everything can be

[3] SCOS core Guide.2005. [online]. [Accessed 20th February 2007]. Available from World Wide Web: <http://www.w3.org/TR/2005/WD-swbp-skos-core-guide-20051102/ >

handled by short-term memory. A second complexity reduction can be achieved by visually aligning the appropriate terms. Terms in the same bag (grouped by either the user himself, or by other users) are placed on the same line. A third reduction of complexity is the use of a limited option set. We currently limit our approach to 4 relations, namely equality, broader than, narrower than and kinship (is like). We combine these with two other options: (1) ”is not related to” and (2) ”I don’t know” to allow the user freedom to manoeuvre. We believe that more complex (but richer) relations also require more conceptual overhead. For instance, the difference between the subsume (is a) and mereology (part of) relation will not always be sufficiently clear. E.g., what is the exact relation between head and eye? Detailed knowledge on the definition of these relations clearly dictates that this would be mereology (eye is part of head), but our users cannot be expected to know these detailed definitions. The conceptual overhead of knowing, understanding and applying these definitions will surely scare many users away. Figure 1.2 shows how low-complexity relating can be achieved.

4. DISCUSSION, CONCLUSIONS AND FUTURE WORK

In this paper we have proposed an early-stage integrated visual approach that creates a CRM domain ontology out of an ”employonomy”, a folksonomy generated by employees, text mining and user feedback. At this point, we are convinced that our approach is capable of enriching the informal semantics of a folksonomy (by adding clear relations), thus assisting in bridging the gap between folksonomy and ontology. Currently, this is a theoretical approach that needs to be tested in a corporate environment. We are planning to implement this in a company in the near future in order to validate our model and underlying assumptions. Furthermore, the concept definitions and its relationships have to be created and extended. We deem this can be done by including additional phases in the model (e.g., phase 3 concept definition and phase 4 relation specialization). It is important, however, to always limit interaction to one phase at a time in order to avoid overwhelming the employee.

Creating an ontology for an operational CRM system will not only ameliorate the information retrieval processes in this system. It can also be a solution for patching this system with the back office systems. In that way an overall view of the customers can be constituted.

We believe that this kind of employee-driven ontology engineering process will be applicable in other company systems and departments. Nowadays corporate semantics are very important, not only for retrieving the right information and connecting systems, it is also a way for improving the communication process between different departments.

ACKNOWLEDGEMENT

The research described in this paper was partially sponsored by the EU IP 02905 Prolix project, the IWT-Tetra 50115 PoCeHRMOM project and by the Brussels-Capital Region (IWOIB PRB 2006). We would like to thank our colleagues for their interesting feedback and discussions.

REFERENCES

Bar-Ilan J. et al., 2006. Structured vs. unstructured tagging - a case study. *Proceedings of the Collaborative Web Tagging Workshop (WWW'06)*. Edinburgh, UK.

Blair, D.C. and Maron, M.E., 1985. An evaluation of retrieval effectiveness for a full-text document retrieval system. In *Communications of the ACM,* Vol. 28, No. 3, pp. 289–299.

Blair D.C. and Kimbrough S.O., 2001. Exemplary documents: a foundation for information retrieval design. In *Information Processing and Management*, Vol. 38, No. 3, pages 363–379.

Christiaens S., 2006. Metadata mechanisms: from ontology to folksonomy ... and back. *Proceedings of the OTM Workshops 2006*, Montpellier, France, pp. 199–207.

de Moor A., et al., 2006. DOGMA-MESS: A Meaning Evolution Support System for Interorganizational Ontology Engineering. *Proceedings of the 14th International Conference on Conceptual Structures (ICCS 2006)*. Aalborg, Denmark, pp. 189-203.

Dyché, J., 2001. *The CRM Handbook: A Business Guide to Customer Relationship Management.* Addison-Wesley, USA.

Golder S. and Huberman B. A., 2006. Usage patterns of collaborative tagging systems. In *Journal of Information Science,* Vol. 32, No. 2, pages 198–208.

Gebert, H. et al, 2002. Towards customer knowledge management- integrating customer relationship management and knowledge management concepts. *Proceedings of ICEB 2002 Conference*. Taipei, Taiwan.

Gruber Thomas R., 1993. *A Translation Approach to Portable Ontology Specifications*. In *Knowedge Acquisition,* Vol. 5, No. 2, pp.199–220.

Hayes, P. at al., 2005. Collaborative knowledge capture in ontologies. *Proceedings of the 3rd international conference on Knowledge capture*. New York, USA, pp. 99–106.

Heymann P. and Garcia-Molina, H., 2006. Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical Report 2006-10, Stanford University.

Hepp, M. et al., 2006. Harvesting wiki consensus - using wikipedia entries as ontology elements. *Proceedings of the First Workshop on Semantic Wikis – From Wiki To Semantics, Workshop on Semantic Wikis*. Budva, Montenegro, pp. 124-138.

Hepp. M., 2007. Possible ontologies: How reality constrains the development of relevant ontologies. In *IEEE Internet Computing,* Vol. 11, No. 7, pp. 96–102.

Margaret W. Matlin., 2003. *Cognition*. John Wiley and Sons, Inc. SUNY Geneseo.

Navigli, R. et al, 2003. Ontology learning and its application to automated terminology translation. In *IEEE Intelligent Systems,* Vol 18, No. 1, pp. 22–31.

Quintarelli. E., 2005. Folksonomies: power to the people, June 2005. Paper presented at the ISKO Italy-UniMIB meeting. Available from: http://www-dimat.unipv.it/biblio/isko/doc/folksonomies.htm [cited 20 February 2007].

Salton G. and McGill, Michael J., 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, USA.

Schmitz, P., 2006. Inducing ontology from flickr tags. *Proceedings of the Collaborative Web Tagging Workshop (WWW'06)*. Edinburgh, UK.

Smith, G. 2004. Folksonomy: Social Classification. 3 August 2007. Gene Smith: Blog [online]. [Accessed 20 February 2007]. Available from World Wide Web: <http://atomiq.org/archives/2004/08/folksonomy_social_classification.html>

Spyns, P. et al, 2006. From folksologies to ontologies: how the twain meet. *Proceedings On the Move to Meaningful Internet Systems 2006: CooPIS, DOA and ODBASE*, Montpellier, France, pp. 738–755.

Van Damme, C., et al., 2007. FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies. *Proceedings of the ESWC 2007 Workshop "Bridging the Gap between Semantic Web and Web 2.0"*, Innsbruck, Austria, *forthcoming*.