Turning a Corporate Folksonomy into a Lightweight Corporate Ontology

Céline Van Damme[1], Tanguy Coenen[2] and Eddy Vandijck[1]

[1] MOSI
[2] STARlab
Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussel, Belgium
{celine.van.damme,tanguy.coenen,eddy.vandijck}@vub.ac.be

Abstract. Companies use company-specific terminology that may differ from the terminology used in existing corporate ontologies (e.g. Tove) and therefore need their own ontology. However, the current ontology engineering techniques are time-consuming and there exists a conceptual mismatch among developers and users. In contrast, folksonomies or the flat bottom-up taxonomies constituted by web users' tags are rapidly created. In this paper, we present an approach that cost-efficiently derives a lightweight corporate ontology from a corporate folksonomy. We tested it on the folksonomy of a European company and first results are promising: it shows that it creates additional value to the company.

Key words: ontology, folksonomy, company, added value

1 Introduction

It has been stated, e.g. in [1, 2] that ontologies improve the communication among humans or machines since they provide a shared understanding of a domain. This makes that ontologies are very useful for companies. For instance they can help to improve the communication between employees and to integrate different information systems.

At this moment, there exist several corporate ontologies, for instance Tove [3] and Enterprise ontology [4]. These ontologies describe general concepts and relations related to enterprise and process modeling. We believe these kinds of ontologies may not be useful for every enterprise since companies have a corporate-specific terminology and consequently have their own concepts. In our opinion, an enterprise may need its own corporate ontology.

However, building ontologies with the current ontology engineering techniques have disadvantages. First of all, it is a very time-consuming process[5] and secondly the actual users are not involved in the developing process. As a consequence there exists a conceptual mismatch between the developers and the actual users' vocabulary [6].

These disadvantages are not present in the relatively new categorization method called tagging and its resulting folksonomy. Following the Web2.0

paradigm, a growing number of websites incorporate a tagging/folksonomy mechanism. They allow users to refer to resources (bookmarks, pictures or scholarly publications) on the web with freely selected keywords or tags. The users are not restricted to a controlled vocabulary produced by a group of experts. Users can enter any words that enter their mind. This makes them active participators in creating new tags. Aggregating this user created meta data leads to a flat, bottom-up taxonomy, also known as a folksonomy.

Despite the strengths, tagging has its weaknesses: no conceptual meaning or hierarchical relations are added to the tags. As a consequence, tags have no synonyms or homonyms. Furthermore, specialized as well as general tags can be used to annotate the same resource [7, 8]. These weaknesses can be solved by (1)giving the users tools that enable them to add more information to their tags (e.g. cluster tags as on Delicious) [8] and/or (2) trying to generate more information on the tags by employing text mining, statistical techniques and asking additional feedback from the community.

The last few years, we observe a growing attention of the semantic web community for tagging and its resulting folksonomies. At the one hand, we observe researchers that try to enrich the flat ambiguous tags with existing online resources (e.g. Google, Wordnet, existing ontologies) [9] and on the other hand, there are researchers that consider this user created meta data as a valuable source to develop ontologies [10].

In this paper, we argue that cost-efficiently deriving a lightweight ontology from a folksonomy is also applicable to a corporate folksonomy. We regard a lightweight ontology as the simplest form of an ontology: an ontology where only one relation is included or a taxonomy as described by [11]. We propose a 6-step approach which includes several techniques such as the Levenshtein metric, co-occurrence, conditional probability, transitive reduction and visualization. Although, some suggestions have already been made on how a corporate ontology can be built from a corporate folksonomy, no research results have been published so far. We implemented our approach on a corporate folksonomy of a large European distribution company in which Dutch and French are the two official company languages. We obtained the simplest form of an ontology, a lightweight ontology, visualized with the open source tool Graphviz[1]. First results show that our approach is promising and we also detected a number of benefits for the company.

The paper is structured as follows: we provide an overview of related work in section 2. In section 3, we discuss all the techniques and resources suggested in literature to enrich folksonomies or derive ontologies. We explain whether they are suitable to turn an enterprise folksonomy into a lightweight corporate ontology. The most appropriate ones are consequently included in the 6-step approach which is presented in section 4. In section 5, we discuss the general results of our experiment and describe some benefits for the company. Section 6 discusses our findings and presents our future research. A conclusion is provided in section 7.

[1] `http://www.graphviz.org/`

2 Related Work

At the time of writing, few papers have been written on discussing the use of folksonomies in a company. The authors in [12] present a social bookmarking tool, called Dogear, that lets employees tag their bookmarks from the corporate intranet and the world wide web. The advantages of collaborative tagging in the enterprise is discussed in [13]. The authors suggest that tagging can be used as an expert location tool that facilitates the process of organizing meetings with experts in the company. Tags are a reflection of people's interest and/or knowledge and can as a consequence be seen as a tool to detect experts and their domain of expertise.

However, the authors in [12, 13] do not explain how to make the tags less ambigious nor turning them into an ontology. This is discussed in [14]. The authors propose to derive a CRM[2] ontology from a corporate folksonomy. They suggest an integrated visual approach that integrates text mining techniques, tags and user feedback. Each time the employee adds a message or note to the CRM system, tags are required. At the same time, automatic keywords are detected based on the tf-idf score. The tf-idf score is calculated by multiplying the word's document frequency by the logarithm of its inverse document frequency in the set of relevant company documents. The higher the score, the more descriptive the keywords are [15]. In a first phase the user has to indicate whether there exists a relationship between the tags and the keywords with the highest tf-idf score. The relationship has to be specified in a second phase. In this approach, the human effort as well as the implementation time is very high. We also have to point out that the proposed approach still has not been tested.

Literature on folksonomies enrichment or turning folksonomies into ontologies is currently more common in the domain of the World Wide Web. In [16] tags of the photo-sharing site Flickr were used in an experiment to induce a taxonomy, the simplest form of an ontology [11]. The approach of [16] is based on statistical natural language processing techniques where a subsumption or hierarchical relation was deducted. The authors of [9, 10] both suggest to include different techniques as well as the wealth of existing online web resources such as Wordnet, Wikipedia, Google, online dictionaries and existing ontologies. The authors in [9] present an approach to enrich tags with semantics to make it possible to integrate folksonomies and the semantic web. The authors use online lexical resources (e.g. Wordnet, Wikipedia, Google) and ontologies to map tags into concepts, properties or instances and determine the relations between mapped tags. However, the resources are tapped in one way (e.g. Wikipedia is used as spelling checker for tags) and the community is not involved to confirm the semantics obtained from existing ontologies and resources. Consequently, tags that reflect new concepts, relations or instances or new relations between tags are neglected. On the contrary, the opposite is suggested in [10]: ontologies are derived from folksonomies. Online lexical resources are suggested to be exploited in several ways. For instance Wikipedia is suggested as a spelling checker

[2] CRM or Customer Relationship Management

as well as a tool for finding concepts and homonyms. Furthermore, the authors suggest to involve the community.

However, a corporate folksonomy differs from a folksonomy created on the World Wide Web. The users, their underlying motivations and the environment can be different. In case of a corporate folksonomy the user or employee is known and will not always tag voluntarily. An employee may be enforced to tag or may be given an incentive by the company. As a consequence, the amount of additional feedback asked from the users to create a lightweight ontology should be reduced. Labor costs are very high and therefore the number of employees involved with the feedback process should be minimized. In contrast to web communities it is far easier to ask the cooperation of the community: community members have a different mindset than employees and are more willing to participate in additional processes. However, in most cases they are anonymous. Company-specific terminology is mostly used in a closed company environment which makes it hard to include web resources in the ontology construction process. The terminology may contain terms which have a specific meaning for only a small group of employees. In the next section, we will explain how the techniques and resources proposed in literature are applicable to the creation of a lightweight ontology by means of corporate folksonomy.

3 Overview of Techniques and Resources

We can distinguish 2 important steps in the literature to enrich folksonomies or to turn folksonomies into ontologies: (1) finding similar tags and (2) finding concepts and relations between tags. In this section, we briefly describe the different techniques and resources and how they can be applied in each step. We also explain how they can be employed to the creation of a lightweight ontology.

3.1 Step 1: Finding Similar Tags

Stemming algorithms A stemming algorithm reduce a word to its stem or root. The algorithm removes suffixes and hereby e.g. reduces the words *linked* and *links* to *link* [17]. The algorithm includes rules that are language dependent. Company-specific language can be lost because of the stemming algorithm. These words can differ from the general spelling rules or they can be abbreviations. Some languages, such as Dutch, incorporate English words in the vocabulary without adjustments to the Dutch language. When stemming algorithms are used, there should be a way to determine the language of the tags and whether it involves corporate-specific language.

Levenshtein metric The Levenshtein metric is a text similarity metric which calculates the distance between two words. More specifically, it counts how many letters have to be replaced, deleted or inserted to transform one word into the other [18]. Dividing this sum by the total number of letters in the word, gives us the Levenshtein metric. It is a valuable technique to verify the similarities of

two tags. In order to calculate the distance, first all possible tag pairs have to be made. In [9] a threshold value of 0.83 is used to indicate that two tags are similar. Yet tests showed us that a threshold value of 0.83 excluded a number of similar tags. For instance, the Dutch adjective *groen* or *green* in English, depending on its function in a sentence, can be *groen* or *groene*. Both tags are the same and their Levenshtein similarity is lower than 0.83. We believe this technique should be employed at a lower threshold value and include human feedback, a representative employee that is very well aware of all the terminology used in the company, to confirm or reject the similarity.

Resources In [9, 10] the use of online resources such as Google, Wikipedia, online dictionaries is suggested. These resources can be regarded as spelling checkers. However, the company-specific terminology makes it though hard to use them. For instance, a company had a *gara* tag, used as the abbreviation for the Dutch word *garage*. When using *gara* as a search term for Google, we did not find any link referring to the correct meaning of the term. On Wikipedia, we found a page describing the term, but the concept or description attributed to it was incorrect. On Wikipedia, *gara* is a Basque word and the name of a Spanish newspaper. This causes problems. We have to know whether the tag belongs to the specific terminology of the company or not. In order to find this out, human feedback is necessary.

3.2 Step 2: Finding Concepts and Relations between Tags

Co-occurrence For each tagged resource all the tag pairs are determined. The tie strength between a tag pair is increased each time two tags are used together. It is interesting to know which tags are often used together in a corporate environment.

Clustering Techniques These techniques can be used to cluster related tags. In [9] clustering techniques are applied to the co-occurrence matrix of the tags. They calculated the cosine angle between the tag vectors that represent the tags. The smaller the angle, the more similar the vectors are. This can be an interesting technique to find a group of related terms in the company.

Conditional Probability A rule based on the conditional probability definition was proposed in [16]. More specifically, the rule tries to find out whether one of the tags in the pair can be defined as broader and the other one as narrower term. By applying the definition of the conditional frequency, the conditional probability is calculated by dividing the co-occurrence of the tag pair by the frequency of the individual tag's. Results vary between 0 and 1. The higher the result, the more the term is used in combination with the other term and consequently the more dependend it is of the other term. When the difference between the two results exceeds a certain threshold value, in [16] the threshold value is set to 0.8, a subsumption relationship is found. Finding broader and narrower terms is important to derive hierarchical relations.

Social Network Analysis These techniques make clusters of people based on their shared objects and or tags. In [10], these techniques are suggested to analyze the ontology of a community. This could be useful to a company, to analyze the ontology of a team.

Transitive Reduction In [16] the authors remove the roots that are logically above the parent nodes. However, we believe transitive reduction, a technique from graph theory, is far more interesting. Transitive reduction reduces the edges of a graph G to a graph G' by keeping all the paths that exist between the nodes in Graph G [19]. The edges are consequently removed because of the implied transitivity.

Visualization Techniques The use of visualization is proposed in [14] to lower the barriers to participate in naming the relations between concepts. In literature, several approaches for visualizing tags and lightweight ontologies are described. In [20] CropCircles are suggested to help people understand the complexity of a class hierarchy. We hypothesize that visualizing the lightweight corporate ontology may facilitate the validation process of the approach and might generate added value to the company.

Resources The resources mentioned in 3.1 as well as the existing ontologies are suggested as a mean to detect concepts and find relations between tags [10]. As mentioned in 3.1, it is hard to use them for a corporate ontology.

4 Deriving a Lightweight Corporate Ontology

In this section, we first describe the corporate folksonomy under study and then proceed to elaborating our approach.

4.1 Folksonomy in the Company

We have implemented our approach in a large European distribution company with headquarters in Belgium in which Dutch and French are the two official company languages. The company employs more than 15.000 people across Europe.

Tagging has been used on all their communication messages for more than 20 years. Messages such as letters and faxes that are not sent electronically are manually scanned, tagged and archived into an information system. Tags replace the subject line of the message. Tagging is completely integrated in the corporate culture. The messages can be created manually, automatically and semi-automatically. The automatic and semi-automatic messages have default tags. In case of semi-automatic messages, the author has to add complementary tags. Manually created messages require user created tags.

Initially, tags were introduced to solve the information retrieval problem since full text search engines were not available at the time. Tagging has remained part

of the communication messaging system. However, the ambiguity of the flat tags and the information overload obstructs the search process. The company introduced some tag rules such as a minimum number of tags, no stop words, no plurals and no conjugated verbs, but only a minority of the employees in the company obeys all these rules.

Even though the tagging system at this company is somewhat different from current web-based tagging practices, the 20-years worth of tagged messages represented a real opportunity to test out the approach in a real-life case. Such cases are rare, as not many organizations have adopted tagging in a way which allows the analysis of a large body of tags. Tagging is so widely adopted and part of the corporate culture we believe the tags can be made to represent a non-toy lightweight ontology.

4.2 Statistics of the Corporate Folksonomy

In 2006, more than 7.000.000 messages were created and roughly 60.000.000 tags in total were used. 91% of the messages are created by Dutch speaking employees. We focused our analysis on the tags added to Dutch messages. More specifically, we analyzed 2 different message types individually: quick internal messages and notes since these are the most important communication channels in the company.

4.3 Our Approach

In this paragraph, we present our approach to derive a corporate ontology. The approach includes 6 steps. We included the following techniques: Levenhstein metric, co-occurrence, conditional probability function,transitive reduction and visualization techniques.

Step 1: Selection of the Tags First, we made a list with all the Dutch stop words[3] and removed the stop words from the database. However, there were not so many stop words in the tags, on average 2% of the tags had to be removed. After filtering the messages with fewer than 2 tags, we had to withdraw the less frequently used tags. We ranked the tags in an absolute frequency table and noticed that the quick internal messages and notes have a group of tags (approximately 150) with a very high frequency (between 5000 and 147.000). We decided it would be interesting to use a top down approach and start to grasp the meaning and interrelations of these frequently used tags.

Step 2: Clean the Tags Since most of the tag rules, imposed by the organization, are not obeyed, tags are polluted (e.g. plural and singular tags) and need to be cleaned up. We used the Levenhstein similarity metric combined with human feedback. In the Dutch language, there exists some confusion about the spelling

[3] Based on the list available at `http://snowball.tartarus.org/algorithms/dutch/stop.txt`

of certain words. The letter *c* and *k* are sometimes pronounced in the same way. In the last decade the spelling rules changed several times and as a consequence people do not know whether words should be written with a *c* or *k*. For instance the English word *contract* is sometimes written in Dutch as *contract* or *kontrakt*. Humans, more specifically employees, are able to detect these types of keyword similarity without erasing the company-specific terms.

Based on a trial and error method, we decided to take 0.65 as a threshold value. All the tag pairs that reach a Levenhstein similarity of 0.65 are presented to the user and when two keywords are similar, the user has to check the corresponding check button. Then, the tag with the lowest frequency is replaced with the one with the highest frequency. We decided to implement this rule since we believe that the tag with the highest frequency determines how the word should be written by the wisdom of the crowds in the company [21]. After the adjustment, we reselected the tags in the same way as described in the first step.

Step 3: Co-occurrence For each message we made all the tag pairs. Then, we counted the frequency of each unique tag pair. The more two tags are used together, the higher this frequency or co-occurrence value. We noticed some tag pairs with a high frequency. Again, we decided to include only the ones with the highest frequency (a frequency of more than 5000) to find the most frequent relations.

Step 4: Finding Broader/Narrower Relations We wanted to derive the simplest form of an ontology and therefore needed to find the broader/narrower relations between the terms, for instance the relation between *animal* and *dog*. We applied the conditional probability function as described in previous section. Therefore, we divided the co-occurrence of the tag pair by the frequency of the tag itself. We did some manual tests deciding on 0.70 as the most appropriate threshold value. The higher the threshold value, the broader and the less deep the resulting ontology will be. Some parts of the ontology are logically interpretable and we were therefore able to do some logical trial and error trials.

Step 5 & 6: Transitive Reduction and Visualization Some of the relations are removed by the transitive reduction and they are consequently visualised with the Graphviz tool.

5 First Results

We applied our approach to 2 different message types: quick internal messages and notes. When visually comparing the output of the two message types, we noticed that the 2 generated lightweight ontologies contain different terms. This means that the tag usage between the two message types differs. Consequently, we will need to find a way to map the different partial results into a complete ontology.

We noticed that we have captured other relations than merely broader/narrower or *a kind of* relations. For instance the relation between the tags *name of shop*

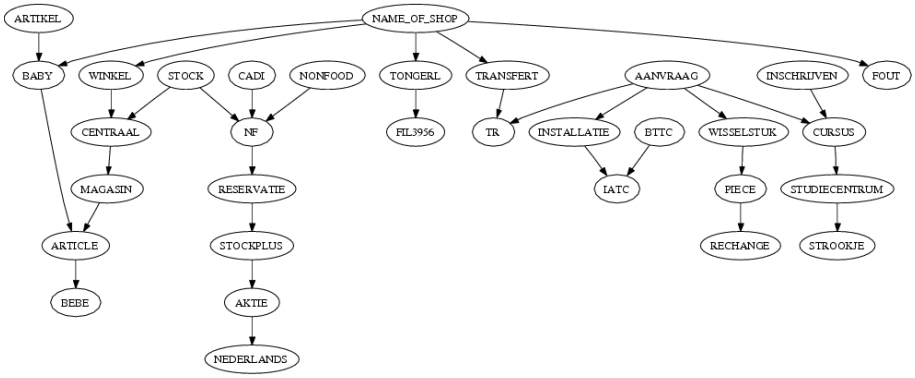

Fig. 1. Part of the lightweight ontology based on tags of quick internal messages

and *baby*, can not really be considered as a *a kind of* relation but more like a *is related to* relation. It provides more information regarding a stock item of the shop. Therefore, it would be interesting to find a way to capture these different kinds of relations and also check whether we may still apply transitive reduction.

We also observed that the graphs, as in figure 1, include some tags corresponding to the French language such as *article*, *bebe*, *magasin*, *piece*, *rechange*. When having a closer look at the data set, we noticed that there are some bilingual messages with bilingual tags. The tags can not be directly filtered from the database since there is no unique identifier. Looking at the results, we observed a pattern: the same tag relation exists between the Dutch and French tag pair e.g. in figure 1 (artikel, baby) and (article, bebe). We also observed this in the other results which are not visually included in this paper.

Tests with the Levenhstein metric, revealed that we can eliminate some French tags due to the close similarity among both languages e.g. *factuur* in Dutch and *facture* in French. In this way, the Levenshtein metric can reduce the pollution by French tags.

5.1 Added Value for a Company

The tags of the corporate folksonomy under study are assigned to all the communication messages sent in the company. We believe that tags assigned to communication messages reflect the actual business processes in the company. This contrasts with tags used in a social bookmarking tool such as presented in [13]: tags represent the knowledge or interests of the employee. By applying our approach to these tags, we have reduced their tag's weaknesses as described in 1, summarized and visualized all the actions that have taken place in the company over a time period.

Based on the analysis of the visual output generated through our approach, we noticed a number of benefits for the company. As we will explain in the next paragraphs, we believe the visualization obtained from the approach could be

used as a tool for management, follow-up tool for new terminology and as a tool for the creation of new teams.

Management Tool By observing figure 1, we noticed two remarkable relations. On the one hand, we saw that there exists a link between the *name of shop*[4] and the tag *fout* or *mistake* in English. On the other hand, we found a relationship between the *name of shop* and the tags *Tongerl* and *Fil3965*. The tag *Tongerl* is used as the abbreviation for a Belgian city and *Fil3965* is the ID of one of the shops. The first mentioned relationship could be a signal that something is wrong and that the relationship between these tags should be further investigated. The latter one could indicate that the shop *Fil3965* has a high sales revenue or high customers complaints. By taking the time factor into account, these results could be compared over different time periods. Therefore, the approach presented in this paper might be an interesting tool for high-level managers in the company. High-level managers are more focused on higher level company's issues such as corporate strategy and are not always aware of all the things that are going on in the company. The visualization of the lightweight ontology obtained through our approach could support them in their daily work and help them in decision making. Therefore, we regard it as a kind of tool for decision making or a sort of add-on for an existing business intelligence tool. This technique could be a valuable extension to existing datamining techniques. At this moment the company is not doing any analysis of its unstructured information.

Follow-up Tool for new Terminology The proposed approach could be valuable as a follow-up tool for new corporate terminology. It reveals how new terms are utilized and interpreted. In case of company acquisition, such a tool could be very interesting. When a company gets acquired by another company, the acquired company will have to apply new terminology to improve the communication process between both of them. Again, the time factor can be included in the process to evaluate and compare the results.

Creating Teams When new teams have to be set up, the approach might helpful to choose the most appropriate employees. This visualization shows how tags are combined with other ones. By selecting all the terms that are related to a concept, the corresponding employees could be selected for the creation of a new team.

6 Discussion and Future Research

The approach is briefly validated by presenting the results to the IT-director and the communication system's analyst of the company. They verified the results by looking at the visualizations and checking the tags in the communication messaging system. They both confirmed their validity.

[4] We renamed this tag to guarantee the anonymity of the company

We also discussed the added value of the visualization. In their opinion, the first and third benefit would be most interesting to their company. They even suggested a visual search tool as an additional added value. Such as tool could be an extension of the suggested management tool. When the manager finds an interesting hierarchical relation or cluster, he should be able to click on it to retrieve the corresponding messages.

We plan to expand our tests to other message types and include tags with a lower frequency to verify the added values we deduced from our current results. In addition, we should set up focus groups with employees of the company where the added value can be extensively discussed.

Further, we will try to find a method to map the ontologies obtained by applying the approach to different message types.

The approach should be further extended and include more techniques and algorithms such as clustering techniques. In this way, more relations might be included in the ontology. However, we believe a cost-benefit analysis should also be built-in in the approach to evaluate whether a more extended version of the ontology will generate the necessarily return on investment. Currently, the approach minimizes the human input and in this way a lightweight-ontology is cost-efficiently derived from the corporate folksonomy.

6.1 Conclusion

In this paper we proposed a 6-step approach to turn a corporate folksonomy into a corporate taxonomy. We implemented the approach on a folksonomy of a European distribution company where tags are assigned to communication messages. We applied the approach to two different kinds of communication messages: quick internal messages and notes. Based on the results, we concluded that the approach summarizes all the actions that have taken place in the company and reflect the actual business processes. The visualization of the results facilitated the validation process and allowed us to detect some benefits for the company: management tool for managers, a follow-up tool for new terminology and a tool which supports the creation of teams. The first and third mentioned benefits were validated by the IT-director and the communication system's analyst of the company. They even suggested to extent the management tool with information retrieval functionality.

Though the approach should be extended and further tested in the near future, we may conclude that the first results look promising.

References

1. Uschold, M., Gruninger, M.: Ontologies: Principles, Methods, and Applications. Knowledge Engineering Review. 11(2), 93–155 (1996)
2. Fensel, D. Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce. Springer, Heidelberg (2003)

3. Fox, M.: The Tove Project towards a common-sense Model of the Enterprise. In: Belli, F., Radermacher, F.J. (eds.) IEA/AIE. LNCS, vol. 604, pp. 25–34. Springer, Heidelberg (1992)
4. Uschold, M., King, M., Moralee, S., Zorgios, Y.: The Enterprise Ontology. Knowledge Engineering Review. 13, 31–89 (1998)
5. Paslaru Bontas, E., Tempich, C.: Ontology Engineering: a Reality Check. In: Meersman, R., Tari, Z., et al (eds.) The 5th International Conference on Ontologies, Databases, and Applications of Semantics. LNCS, vol. 4275, pp. 836–854. Springer, Heidelberg (2006)
6. Hepp, M.: How Reality Constraints the Reality of the Development of Relevant Ontologies. In: IEEE Internet Computing, vol. 11, nr. 1, pp.96–102. IEEE Press, New York (2007)
7. Golder, S., Huberman, B.A.: Usage of Patterns of Collaborative Tagging Systems. Journal of Information Science. 32(2), 198–2008 (2006)
8. Guy, M., Tonkin, E.: Tidying up Tags? D-Lib Magazine, 12(1), 1082–9873 (2006)
9. Specia, L., Motta, E.: Integrating Folksonomies with the Semantic Web. In: Franconi,E., Kifer, W., May, W. (eds.) Proceedings of the Semantic Web Conference. LNCS, vol. 4519, pp. 624–639. Springer, Heidelberg (2007)
10. Van Damme, C., Hepp, M., Siorpaes, K.: Folksontology: An Integrated Approach for Turning Folksonomies into Ontologies. In: Proceedings of Bridging the Gap between Semantic Web and Web 2.0, pp.71-84. Innsbruck (2007)
11. Uschold, M., Japsper, R.: A Framework for Understanding and Classifying Ontology Applications. In: Proceedings of the IJCAI99 Workshop on Ontologies and Problem-Solving Methods. Stockholm (1999)
12. Millen, D.R., Feiberg, J., Kerr, B.: Dogear: Social Bookmarking in the Enterprise. In: Proceedings of Collaborative Web Tagging Workshop at WWW2006. Edinburgh (2006)
13. John, A., Seligmann, D.: Collaborative Tagging and Expertise in the Enterprise. In: Proceedings of Collaborative Web Tagging Workshop at WWW2006. Edinburgh (2006)
14. Van Damme, C., Christiaens, S., Vandijck, E.: Building an Employee-driven CRM Ontology. In: Proceedings of the IADIS Multi Conference on Computer Science and Information Systems (MCCSIS): E-society2007. pp. 330-334. Lisbon (2007)
15. Salton, G., McGill, J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1986)
16. Schmitz, P.: Inducing Ontology from Flickr Tags. In: Proceedings of Collaborative Web Tagging Workshop at WWW2006. Edinburgh (2006)
17. Porter, M.F.: An Algorithm for Suffix Stripping. Program. 14(13), 130–137 (1980)
18. Levenshtein, V.I.: Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady. 10(8), 707–710 (1966)
19. Aho, A.V., Garey, M.R., Ullman, J.D.: The Transitive Reduction of a Directed Graph. SIAM J. Comput. 1(2), 131–137 (1972)
20. Parsia, B., Wang, T., Goldbeck, J.: Visualizing Web Ontologies with CropCircles. In: Proceedings of the End User Semnatic Web Interaction Workshop at ISWC. Galway (2005)
21. Surowiecki, J.: The Wisdom of the Crowds. Anchor, New York. (2005)