Recommending Concepts to Experts: An Exploration of Recommender Techniques for Collaborative Ontology Engineering Platforms in the Biomedical Domain

Simon Walk [1]*, Markus Strohmaier [1,2], Tania Tudorache [2], Natalya F. Noy [2], Csongor Nyulas [2] and Mark A. Musen [2]

[1]Knowledge Management Institute, Graz University of Technology, Inffeldgasse 21a/II, 8010 Graz, Austria
[2]Stanford Center for Biomedical Informatics Research, 1265 Welch Road, Stanford, CA 94305-5479, USA

ABSTRACT

Biomedical ontologies such as the 11th revision of the International Classification of Diseases and others are increasingly produced with the help of collaborative ontology engineering platforms that facilitate cooperation and coordination among a large number of users and contributors. While collaborative approaches to engineering biomedical ontologies can be expected to yield a number of advantages, such as increased participation and coverage, they come with a number of novel challenges and risks. For example, they might suffer from low participation, lack of coordination, lack of control or other related problems that are neither well understood nor addressed by the current state of research. In this paper, we aim to tackle some of these problems by exploring techniques for recommending concepts to experts on collaborative ontology engineering platforms. In detail, this paper will (i) discuss different recommendation techniques from the literature (ii) map and apply these categories to the domain of collaboratively engineered biomedical ontologies and (iii) present prototypical implementations of selected recommendation techniques as a proof-of-concept.

1 INTRODUCTION

In the field of biomedical research, an increasing number of ontologies are created collaboratively by a large group of people. Examples include biomedical ontologies such as the Gene Ontology (GO), the Ontology of Biomedical Investigations (OBI), the National Cancer Institute Thesaurus (NCI) or the 11th revision of the International Classification of Diseases (ICD-11). While collaborative approaches to engineering biomedical ontologies can be expected to yield a number of advantages, such as increased participation and coverage, higher acceptance or improved quality, they come with a number of novel challenges and risks. For example, recent research on collaborative authoring environments indicates that the quality of collaboratively constructed products depends on the number of active participants, the ability to direct qualified participants to relevant content, amongst other factors (Kittur and Kraut, 2008). In addition, collaborative ontology engineering projects might suffer from a lack of coordination, lack of control, low quality and other related problems that are neither well understood nor addressed by the current state of research. These problems hinder progress and have the potential to jeopardize success of future ontology engineering projects in the biomedical domain. To tackle these challenges, new approaches for coordinating work and for supporting contributors are needed.

*To whom correspondence should be addressed: s.walk@student.tugraz.at

One way of augmenting users in collaborative systems is to provide them with adequate support and guidance for contributing their expertise (Ling *et al.*, 2005). In systems such as Wikipedia (Cosley *et al.*, 2007), recommender techniques are already used to help coordinate collaborative tasks and support users in identifying articles to work on. In the context of collaboratively engineering biomedical ontologies, no such tools exist yet.

The main focus of this paper is to explore recommender techniques for collaborative ontology engineering platforms in the biomedical domain. The paper is structured as follows: In Section 2 we will discuss related work on collaborative engineering of biomedical ontologies as well as existing work on recommender techniques. In Section 3 we will map different recommender techniques to the biomedical ontology engineering domain. In Section 4 we will provide a short introduction to an exemplary collaborative ontology engineering project from the biomedical domain: the ICD-11 project. In Section 5 we will present results from three proof-of-concept recommender implementations. In section 6, we will conclude by discussing our approaches and point to future work.

The overall contributions of this paper are a high level mapping of recommender techniques to collaborative ontology engineering platforms in the biomedical domain, and a proof-of-concept in the form of implementations in the context of the ICD-11 project.

2 RELATED WORK

2.1 Collaborative Authoring & Ontology Engineering

In the context of collaboration platforms, many factors are known to influence the motivation and activity of individuals. For example, we know that transparent and well defined goals can affect groups and their performance. Making contributors aware of the utility of their contributions represents another important factor (Ling *et al.*, 2005). Restructuring the payoff function, i.e. reducing the costs and increasing the benefits of contributions, has also been identified as a potential intervention to increase participation (Cabrera and Cabrera, 2002).

An increasing number of biomedical ontologies, such as the Gene Ontology, the National Cancer Institute Thesaurus, or the ICD-11, are created using collaborative ontology engineering platforms. Requirements for collaborative ontology engineering platforms have been discussed by, for example, (Noy and Tudorache, 2008). Examples of existing platforms include OntoEdit, different forms of Wikis such as Wiki@nt and OntoWiki or WebProtégé (Tudorache *et al.*, 2011). While many tools put an emphasis on collaboration, we know little about how to effectively coordinate and shape collaborative ontology engineering projects. iCAT Analytics (Pöschko *et al.*, 2012) represents a first attempt to provide a detailed analysis of collaborative ontology engineering processes.

2.2 Recommender systems

The main objective of recommender technology is to provide personalized suggestions that help an individual, or a group of individuals, to find objects or items of interest (Burke *et al.*, 2011). Historically, recommendations were used on e-commerce websites to enhance impulsive buying behavior of customers.

In the literature, a distinction between three basic recommendation strategies can be identified: item/content based, collaborative filtering, and knowledge-based recommender techniques (Burke *et al.*, 2011). While content based recommender strategies focus on recommending items that are similar with regard to their content, collaborative filtering strategies focuses on recommending items that are similar with regard to behavioral patterns of similar users. Knowledge-based recommender strategies focus on identifying similar items using background/domain knowledge.

In the context of collaborative authoring systems, such as Wikipedia, recommender systems can be useful not only to help finding items of interest, but also to increase participation (Cosley *et al.*, 2007). While ontologies have been used as source of domain knowledge for generating recommendations (Sieg *et al.*, 2010), applying recommender techniques to recommend concepts to experts is a novel problem.

3 RECOMMENDING CONCEPTS TO EXPERTS

In the following, we aim to explore how the three identified recommender strategies map onto collaborative ontology engineering platforms in the biomedical domain.

3.1 Recommending concepts based on content

The intuition behind content based recommender techniques is to find and identify similar items or concepts by calculating and comparing similarity between content-related features of each concept.

Content in an ontology can be defined as features of concepts, textual properties such as titles and descriptions, notes and discussions or ratings. In the context of collaboratively engineered biomedical ontologies, these properties can be term names, textual definitions of the concepts, clinical descriptions such as related/affected body parts, synonyms, signs and symptoms, investigation findings such as lab activities or measures needed to diagnose a disease or even treatment plans.

Similarity for content based recommender systems is usually calculated on a common set of features or properties that all items or concepts share, using similarity or correlation measures such as Pearson correlation, cosine similarity or the Jaccard coefficient. Other textual similarity measures that could be used include the Levenshtein distance, or simple overlap of textual properties of concepts. Depending on the environment, different similarity measures can yield different results when presented with the same input.

Potentials & Limitations for content based recommender systems are closely tied to the properties and content of the concepts in the ontology. An advantage of content-based recommendations is that they can be generated even in the absence of social usage data (i.e. they do not suffer from the ramp-up problem). A lack of rich textual content however might impair the overall usefulness of content-based recommendations.

3.2 Recommending concepts via collaborative filtering

The intuition behind collaborative filtering is to find concepts or items based on similar user behavior. This is accomplished by identifying behavioral patterns or usage patterns of users, and by grouping them according to their similarity (Sarwar *et al.*, 2001; Goldberg *et al.*, 1992).

Usage patterns are patterns that define the interest of a user for items or concepts. They either can be explicitly entered information such as ratings or implicit measures deducted from the amount of previously viewed, bought, or changed items by a single user. In the context of collaboratively engineered biomedical ontologies, usage patterns could be defined by grouping different behavior of users on a collaborative platform such as adding, editing or even moving or deleting a concept, property or individual. Notes can be used as an indication for interest as well as viewing patterns or viewing times.

Similarity for collaborative filtering is usually calculated by identifying users with common interests which can be done by calculating the similarity between their usage patterns. In collaborative systems, interest is often modeled by explicit item rankings entered by the users. Since this kind of information is typically not available in biomedical contexts (users do not rate their favorite concepts), other features, such as the number of times a concept has been viewed or changed or even other properties assigned to users and concepts, can be used. To calculate similarity, a series of different similarity measures, including Pearson correlation or cosine similarity (Sarwar *et al.*, 2001), is available.

Potentials & Limitations for collaborative filters are closely tied to the extent that usage data is available. Collaborative filtering approaches are particularly prone to the early phases of collaborative ontology engineering projects, where little data about user interactions is available. However, once sufficient data is collected, collaborative filtering approaches can recommend concepts that are not necessarily related content-wise, but through other usage pattern based characteristic.

3.3 Recommending concepts using domain knowledge

The intuition behind knowledge based recommender systems is to find similar concepts based on specific domain knowledge. They represent a sub-class of content based recommender systems and differ from them by using domain knowledge to create rules to determine the best item or concept to recommend, instead of simple properties.

Domain knowledge is specific knowledge extracted from the environment of the system or the system itself. The biggest challenge in creating knowledge based recommendations is identifying viable domain knowledge, that will produce good results when used to calculate similarity. Recommendations can be produced by traversing along the edges in an ontology to identify related sub, super or sibling concepts. In addition, linkages between ontologies can be exploited to generate knowledge based recommendations as well.

Similarity for knowledge based recommender techniques can be calculated using properties that could either be actively collected, by querying users for input, or implicitly by analyzing previous behavior of a user.

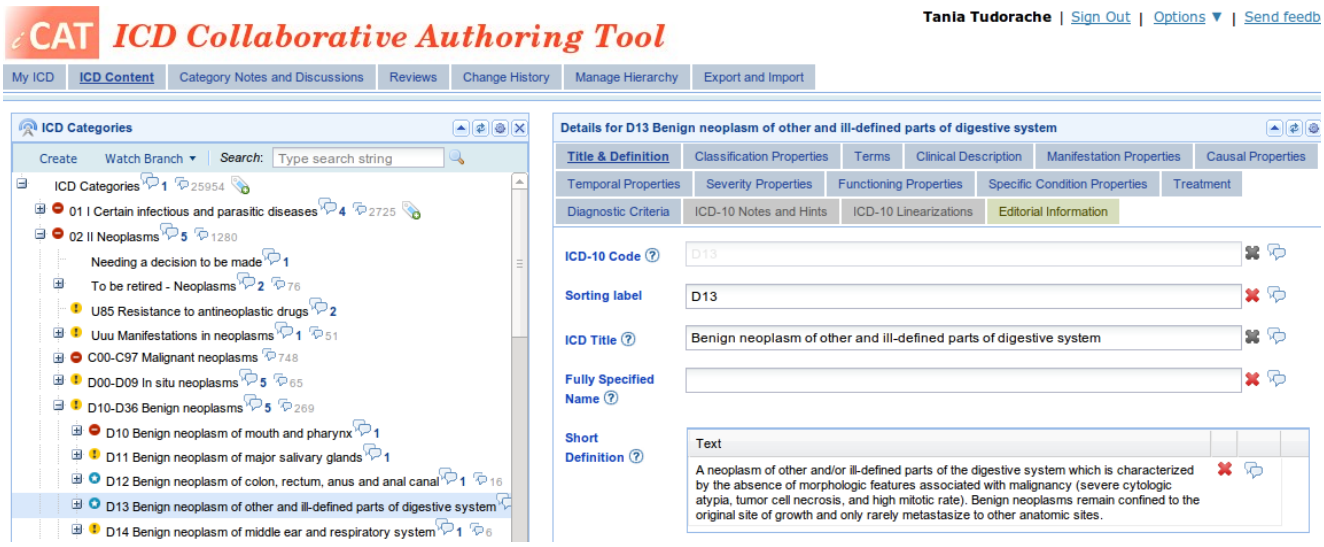

Fig. 1. Example of a collaborative ontology engineering platform: The iCAT user interface

Potentials & Limitations for knowledge based recommender systems are mostly related to the problem of distinguishing between basic content and domain knowledge. In addition, not every domain knowledge property will provide an equal basis for good recommendations. An advantage of knowledge based recommender techniques is that, at least in some way, they are less dependent on the quantity of content and contributions.

4 THE ICD PROJECT AND iCAT

In the following, we will briefly introduce the ICD-11 project. We will use the project later as an example to illustrate the adoption of recommender techniques for collaborative ontology engineering platforms in the biomedical domain. The International Classification of Diseases is a taxonomy maintained by the World Health Organization and is updated to a new revision around every decade. It is used worldwide for monitoring health related expenses, to inform policy makings, and to collect disease statistics. ICD-10 and all other predecessors of the ICD-11 were created by selected international experts; the production process was closed to the public. For ICD-11, the WHO decided on a more open, collaborative approach. This new approach allows experts all over the world to contribute to ICD-11, using a web based collaboration platform called the ICD-11 Collaborative Authoring Tool (iCAT, as depicted in Figure 1) (Tudorache *et al.*, 2010). There are currently around 100 international experts working on ICD-11.

Next, we will present a number of proof-of-concept implementations of recommender techniques aiming to demonstrate how recommenders could be applied to collaborative ontology engineering platforms in the biomedical domain.

5 VALIDATION: PROOF OF CONCEPT

To study the general feasibility of recommending concepts to experts, we implemented three selected recommendation techniques for the ICD-11 project as a proof-of-concept. In the following we will illustrate the implications of different recommender techniques through examples using actual interaction data obtained from selected ICD-11 users. We will refer to these users as LB, AR and RC from here on.

5.1 Content-based concept recommendations

We used real data excerpts, extracted from the ICD-11 and its log of changes, to demonstrate how content based recommender systems can be applied to collaborative ontology engineering platforms in the biomedical domain.

For all users U and the set of all concepts C, we extracted their previously changed concepts $C_u \subseteq C$, together with all words from the title and the words included in the definition of a concept $c \in C_u$.

Before doing so, we performed three additional tasks: (i) stop word (e.g.: is, as, and, so etc.) removal, (ii) stemming, a mechanism used in natural language processing to reduce words to their stem and (iii) data cleaning, i.e. we have removed special characters from the textual properties of the concepts. For this example, we used the stop word list available from the Natural Language Toolkit.

For similarity calculations, we used cosine similarity, as there is evidence that it provides good results for existing collaborative environments (Adomavicius and Tuzhilin, 2005). Results range from 0 (= completely unsimilar) to 1 (= identical).

$\vec{W}_{LB} = \{disease: 906, skin: 125, contact: 33, acute: 97\}$
$\vec{W}_{AR} = \{disease: 386, skin: 34, contact: 0, acute: 39\}$
$\vec{W}_{RC} = \{disease: 272, skin: 841, contact: 399, acute: 65\}$

$\vec{V}_{LZ1} = \{disease: 1, skin: 1, contact: 0, acute: 0\}$
$\vec{V}_{L56} = \{disease: 0, skin: 1, contact: 0, acute: 1\}$
$\vec{V}_{Z20} = \{disease: 1, skin: 0, contact: 1, acute: 0\}$

Table 1. $\vec{W}_u$ and $\vec{V}_c$ (left) displaying excerpts of processed word-count lists (right) from users and concepts used for cosine similarity calculations.

$\vec{W}_u$ represents all words and their respective number of appearances in the title and definition of all $d \in C_u$ and $\vec{V}_c$ collects all words and word counts per concept $c \in C$. Table 1 shows excerpts of $\vec{W}$ and $\vec{V}$ depicting word lists for the users LB, AR and RC as well as excerpts of the concepts $LZ1$ ("LZ1 Impairment of normal functioning resulting from skin disease"), $L56$ ("Other acute skin changes due to ultraviolet radiation") and $Z20$ ("Contact with and exposure to communicable diseases"), after stemming and stop word removal.

Cosine similarity was calculated for every pair $cos(\vec{W}_u, \vec{V}_c)$ where $u \in U$ and $c \in C$ and stored in the user-concept similarity matrix $M_{U,C}$ (see Table 2).

	LZ1	L56	Z20
LB	$0,792159$	$0,170571$	$0,721471$
AR	$0,762570$	$0,132542$	$0,700839$
RC	$0,809721$	$0,659126$	$0,488160$

Table 2. The user-concept similarity matrix $M_{U,C}$ depicts the similarity between users and concepts. The higher a value, the more similar a concept is to previously changed concepts of the user.

In Table 2, recommendations for LB can be generated from $M_{U,C}$ by suggesting those concepts that have the highest similarity values. In our example, this approach would recommend $LZ1$ and $Z20$.

5.2 Collaborative filtering-based concept recommendations

To demonstrate collaborative filtering-based recommendations, we used log data from iCAT to calculate similarity between all user $u \in U$. In this approach, two users are similar if they have modified similar concepts. The concepts $c \in C$ that have been changed by u from November 2009 to 30$^{\text{th}}$ August 2011, are denoted as sets C_u (as seen in Table 3). For the users LB, AR and RC these concepts are $H40.1$ ("Primary open-angle glaucoma"), $BPNCS$ ("Benign proliferations, neoplasms and cysts of the skin"), XII ("Diseases of the skin") and DBS ("De Barsy syndrome").

C_{LB} = {H40.1, BPNCS, XII}
C_{AR} = {DBS}
C_{RC} = {BPNCS, XII, DBS}

Table 3. The set C_u (left) represents an excerpt of the set of all concepts (right) modified by u from November 2009 to 30$^{\text{th}}$ August 2011 that was used for calculating similarity between users.

Based on this matrix, we used the Jaccard coefficient to calculate similarity between users based on the set of all concepts c_u for all $u \in U$ resulting in a user-user similarity matrix $M_{U,U}$ (see Table 4) as $M_{i,j} = J(u_i, u_j)$.

	LB	AR	RC
LB	1	0	0,5
AR	0	1	0,33
RC	0,5	0,33	1

Table 4. The user-user similarity matrix $M_{U,U}$ lists the Jaccard coefficient similarity values between all pairs of users.

We introduced an arbitrary threshold $minSimilarity$ set at 0.0001 which excludes user pairs with very little similarity. Based on that modified table, we start to count the number of changes done by u_j on every single concept $c \in C_{u_j}$ and store the results in the matrix $N_{U,C}$ (see Table 5).

	$H40.1$	$BPNCS$	XII	DBS
LB	2	1	12	0
AR	0	0	0	20
RC	0	12	45	14

Table 5. $N_{U,C}$, the user-concept change count matrix lists the number of changes done by every user $u \in U$ to every concept $c \in C$.

The values for the user concept similarity matrix $O_{U,C}$ are calculated as depicted in Equation 1.

$$O(i,j) = \sum_{k=0,k\neq i}^{n} N(k,j) + N(k,j) * M(i,k) \qquad (1)$$

The final results are illustrated in Table 6. Collaborative filtering recommends concepts for a user based on the concepts that similar users found interesting. In our example, the user LB and RC have a high similarity rating of $0,5$, and user RC has contributed to DBS many times. Hence, the concept DBS is recommended (similarity of 21) to LB.

	$H40.1$	$BPNCS$	XII	DBS
LB	-	-	-	21
AR	0	15,96	59,85	-
RC	3	-	-	-

Table 6. User-concept similarity matrix $O_{U,C}$ holds the similarity results according to Equation 1 for the collaborative filtering approach. The higher the similarity the likelier it is, that the user is interested in that concept.

5.3 Knowledge-based concept recommendations

Our final proof-of-concept implementation will assume that users are most likely interested in concepts that are ontologically related to the ones they have already shown interest in. The more a concept is interlinked or referred to by previously changed concepts in the ontology, the more related it is to the interests of that specific individual. An excerpt of the ICD-11 ontology, represented as directed graph using the is-a relationships, is depicted in Figure 2.

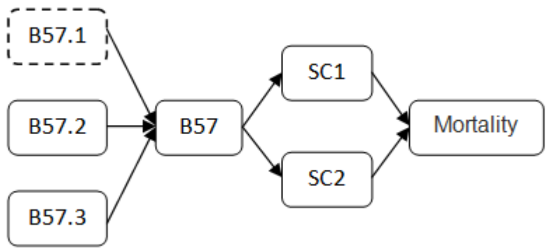

Fig. 2. Representation of an ICD-11 excerpt as a directed graph. Nodes refer to concepts while edges represent isA relationships. Dotted lines indicate modified concepts.

We assumed that user LB has only changed the concept $B57.1$ ("Acute Chagas disease without heart involvement"). To explore related concepts, we followed links in the ontology until either a predefined depth level was reached or enough highly interlinked concepts were discovered. Table 7 depicts all values for expansions of concepts $B57$ ("Chagas' disease"), $B57.2$ ("Chagas disease (chronic) with heart involvement"), $B57.3$ ("Chagas disease (chronic) with digestive system involvement"), $SC1$ ("Selected Cause is Remainder of certain infectious and parasitic diseases in the Condensed and Selected Infant and child mortality lists"), $SC2$ ("Selected Cause is Trypanosomiasis") and $Mortality$ ("Tabulation list for mortality").

Next we traversed along all paths, as shown in Figure 2 and Table 7, from all previously edited concepts, counting the number of encounters of each concept. The higher the number of encounters, the more weight it will receive in a ranking of concepts for user u.

Rk.	Content Based	Score	Collaborative Filtering	Score	Knowledge Based	Score
1	L02.9 'Cutaneous abscess, furuncle and carbuncle, unspecified'	0.381792	II Neoplasms	9.120824	'Selected Cause is Remainder of certain infectious and parasitic diseases in the Condensed and Selected General mortality lists'	42
2	L02.8 'Cutaneous abscess, furuncle and carbuncle of other sites'	0.372091	VI 'Diseases of the nervous system'	9.119071	'Ectodermal dysplasia syndromes'	34
3	'Chronic ulcer skin'	0.359489	XI 'Diseases of the digestive system'	8.155308	'Chromosomal disorders affecting the skin '	28
4	'Congenital skin anomaly other'	0.359119	E09-E1B 'Diabetes mellitus'	8.136958	'Genetic, chromosomal and developmental disorders affecting the skin '	24
5	'Pediculosis/skin infestation other'	0.356631	V 'Mental and behavioural disorders'	8.117054	XII 'Diseases of the skin'	23
6	'Fear of skin disease other'	0.350870	IX 'Diseases of the circulatory system'	8.106256	'Genetic syndromes affecting nails'	19
7	'Malignant neoplasm of skin'	0.345841	A15 'Respiratory tuberculosis, bacteriologically and histologically confirmed'	8.100946	'Tabulated - Other diseases of the skin and subcutaneous'	18
8	'Dysplasia syndromes with skin/mucosae involvement'	0.338079	I21 'Acute myocardial infarction'	8.058171	L20-L30 'Dermatitis and eczema'	17
9	'Tabulated - Other diseases of the skin and subcutaneous'	0.333487	H25 'Senile cataract'	7.100532	'Dysplasia syndromes with premature ageing appearance'	17
10	'Tabulated - Other malignant neoplasms of skin'	0.327885	VII 'Diseases of the eye and adnexa'	6.137702	'Parasitic infestations affecting the skin'	16

Table 8. Ranked concept recommendations for the user RC according to our three different recommender techniques. The higher the score, the better the concept is ranked for recommendation. Every approach provides different results regarding level of detail, scope and sub-domain.

Depth	B57.1	B57.2	B57.3	B57	SC1	SC2	Mortality
1	1	0	0	1	0	0	0
2	1	0	0	2	1	1	0
3	1	0	0	2	2	2	2
4	1	0	0	2	2	2	3

Table 7. A listing of the number of encounters on different depth levels for concepts C_{LB} of user LB when path traversing Figure 2 to find most interlinked concepts for calculating knowledge based recommendations.

6 DISCUSSION & FUTURE WORK

The final recommendations for RC (see Table 8) produced by our different proof-of-concept recommendation algorithms differ significantly with regard to level of detail, scope and sub-domain. Recently or currently browsed concepts could be included in the similarity calculations to increase the scope of the generated recommendations. While we believe that concept recommender systems could represent a useful instrument to augment user experience on collaborative ontology engineering platforms in the biomedical domain, we have not performed evaluations of our implementations yet. Understanding what kinds of recommender techniques are useful in what kind of contexts, including a more sophisticated analysis of user-behavior similar to (Kern *et al.*, 2010), represent important next steps for future research. We believe that understanding the utility of recommender systems to steer and augment user activity in collaborative ontology engineering projects represents an exciting avenue for future biomedical research. This work is relevant for the future design of collaborative ontology engineering platforms, and for operators of such systems as well as for users and contributors.

REFERENCES

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, **17**, 734–749.

Burke, R. D., Felfernig, A., and Gker, M. H. (2011). Recommender systems: An overview. *AI Magazine*, **32**(3), 13–18.

Cabrera, A. and Cabrera, E. F. (2002). Knowledge-Sharing dilemmas. *Organization Studies*, **23**(5), 687–710.

Cosley, D., Frankowski, D., Terveen, L., and Riedl, J. (2007). Suggestbot: using intelligent task routing to help people find work in wikipedia. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 32–41, New York, NY, USA. ACM.

Goldberg, D., Nichols, D. A., Oki, B. M., and Terry, D. B. (1992). Using collaborative filtering to weave an information tapestry. *Commun. ACM*, **35**(12), 61–70.

Kern, R., Körner, C., and Strohmaier, M. (2010). Exploring the influence of tagging motivation on tagging behavior. In *Proceedings of the 14th European conference on Research and advanced technology for digital libraries*, ECDL'10, pages 461–465, Berlin, Heidelberg. Springer-Verlag.

Kittur, A. and Kraut, R. E. (2008). Harnessing the wisdom of crowds in wikipedia: Quality through coordination. In *CSCW '08: Proceedings of the ACM 2008 conference on Computer supported cooperative work*, pages 37–46. ACM.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., Rashid, A. M., Resnick, P., and Kraut, R. (2005). Using social psychology to motivate contributions to online communities. *Journal of Computer-mediated Communication*, **10**(4).

Noy, N. F. and Tudorache, T. (2008). Collaborative ontology development on the (semantic) web. *Nature Biotechnology*.

Pöschko, J., Strohmaier, M., Tudorache, T., and Musen, M. A. (2012). Pragmatic analysis of crowd-based knowledge production systems with iCAT Analytics: Visualizing changes to the ICD-11 ontology. In *Proceedings of the AAAI Spring Symposium 2012: Wisdom of the Crowd*. Accepted for publication.

Sarwar, B., Karypis, G., Konstan, J., and Reidl, J. (2001). Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th International Conference on World Wide Web*, pages 285–295, New York, NY, USA. ACM.

Sieg, A., Mobasher, B., and Burke, R. (2010). Ontology-based collaborative recommendation. *Computing*.

Tudorache, T., Falconer, S. M., Nyulas, C. I., Noy, N. F., and Musen, M. A. (2010). Will Semantic Web Technologies Work for the Development of ICD-11? In *Proceedings of the Ninth International Semantic Web Conference*.

Tudorache, T., Nyulas, C., Noy, N. F., and Musen, M. A. (2011). WebProtégé: A Distributed Ontology Editor and Knowledge Acquisition Tool for the Web. *Semantic Web Journal*, **11-165**.