How Tagging Pragmatics Influence Tag Sense Discovery in Social Annotation Systems

Thomas Niebler[1], Philipp Singer[2], Dominik Benz[3], Christian Körner[2], Markus Strohmaier[2], and Andreas Hotho[1]

[1] Data Mining and Information Retrieval Group, University of Würzburg, 97074 Würzburg, Germany
{niebler,hotho}@informatik.uni-wuerzburg.de

[2] Knowledge Management Institute, Graz University of Technology, 8010 Graz, Austria
{philipp.singer,christian.koerner,markus.strohmaier}@tugraz.at

[3] Knowledge & Data Engineering Group, University of Kassel, 34121 Kassel, Germany
benz@cs.uni-kassel.de

Abstract. The presence of emergent semantics in social annotation systems has been reported in numerous studies. Two important problems in this context are the induction of semantic relations among tags and the discovery of different senses of a given tag. While a number of approaches for discovering tag senses exist, little is known about which *factors* influence the *discovery process*. In this paper, we analyze the influence of *user pragmatic* factors. We divide taggers into different pragmatic distinctions. Based on these distinctions, we identify subsets of users whose annotations allow for a more precise and complete discovery of tag senses. Our results provide evidence for a link between tagging pragmatics and semantics and provide another argument for including pragmatic factors in semantic extraction methods. Our work is relevant for improving search, retrieval and browsing in social annotation systems, as well as for optimizing ontology learning algorithms based on tagging data.

1 Introduction

In social annotation systems, large groups of users freely annotate resources with tags. This social and dynamic process yields several interesting emergent phenomena, such as emergent classification of resources [1] or emergent tag semantics [2]. Early work in this area has identified three major challenges for social annotation systems [3], which still represent wide open research problems today: tag polysemy, tag synonymy and basic-level variation. In this paper, we want to focus on *tag polysemy*, which is the problem that words can have several different meanings or *senses* (e. g., "swing" might refer to the Java GUI Framework or to a dance style). In recent years, a number of methods have been published that focus on *tag sense discovery*, i. e., discovering different senses of a given tag automatically [4–6].

P. Serdyukov et al. (Eds.): ECIR 2013, LNCS 7814, pp. 86–97, 2013.
© Springer-Verlag Berlin Heidelberg 2013

Due to the open and social dynamics in social annotation systems, a wide variety of users and user behavior can be observed. At the same time, little is known about how different users and user behavior influence the emergent semantics that we can observe in such systems. In [2], we already worked on if the tags produced by certain user groups are more useful for yielding emergent semantics than others and if yes, what kind of users or user behavior is most useful. We now wanted to enhance that question to discover different senses in these emergent semantics. The overall objective of this paper now is to study a potential relationship between tagging pragmatics (how users use tags) and tag sense discovery in social annotation systems, and shed further light on the ways pragmatics influence semantics. Towards this end, we analyze if and how selected pragmatic factors influence the performance of a tag sense discovery task. While we do not aim to comprehensively analyze or explain the manifold ways in which pragmatics can influence this task, in this paper we are looking for a *signal*. We want to explore (i) **whether there is a link between pragmatics and tag sense discovery** at all and (ii) **if there is, how it might be explained**. For example, we want to find out what kind of user behavior (e.g. whether users use tags for categorization or description) yields more useful data for tag sense discovery. Such explanations would prove useful for future, more elaborate tag sense discovery methods that could leverage tagging pragmatics for better performance.

The results of this work are relevant for improving search, retrieval and browsing in social annotation systems, as well as for optimizing ontology learning algorithms based on tagging data. Polysemy, i. e., a word can have different meanings, clearly effects functions of social annotation systems such as information retrieval or browsing: Because the different senses of a tag may be semantically unrelated (e. g., *swing* is both a *programming library* and a *dance*), the user is presented with irrelevant content. If different reliable meanings of a tag can be discovered, this would greatly improve search and retrieval. Naturally, this problem is not restricted to social annotation systems, but it is present basically within all systems dealing with natural language; however, the open vocabulary as well as the lack of structure (compared to, e. g., the syntax of a written text) makes this issue in social annotation systems unique and interesting.

The paper is structured as follows. Section 2 introduces tag sense discovery and presents our applied disambiguation method. The pragmatic aspects are introduced in Section 3, including concrete measures to distinguish between different kinds of taggers. In Section 4, we analyze empirically how distinct usage patterns (as captured by the pragmatic measures) influence the process of tag sense discovery. We discuss related work in Section 5 before we conclude in Section 6.

2 Tag Sense Discovery

The goal pursued in this paper is best described as *tag sense discovery.* NLP approaches in this field like [7, 8] are typically applying clustering approaches to divide a suitable context of a given term into partitions which correspond to its senses. When transferring this idea to social annotation systems, the problem can be divided into the following subproblems, (i) *context identification*, i. e., how to construct a "suitable" context and (ii) *context disambiguation*, i. e., how to subdivide this context into senses. In the remainder of the paper, we will use the definition of folksonomies as social annotation systems as provided in [9].

2.1 Sense Context Identification

In prior work, [10] performed extensive studies on the characteristics of different context definitions for the task of tag sense discovery. The authors examined tag- and user-based document networks, as well as tag co-occurrence and similarity networks. It was found that tag similarity networks provided *"the most clear-cut results among all the network types"*. As similarity measure, we will use the tag context relatedness *cossim* as defined in [11] to depict the relations among the items present in the context of a given tag.

The next question is which tags to include in the context of a given tag t. The goal hereby is to choose a sample of context tags which are representative for t's main senses. Hereby we follow the procedure described by [12], who found that the *20 strongest first-order associations [...] are [...] a good mix of the two main senses for each word.* First-order associations correspond to tag-tag co-occurrence in our case. Although we do not necessarily target to discover *two* main senses, we follow these steps to construct a context for a given tag t:

1. Let $t \in T$ be a tag whose senses are to be discovered.
2. Let $SC_t = (V_t, E_t)$ be an initially empty undirected graph, whose edges are weighted by a weighting function $w : V_t \to \mathbb{R}$. We call this graph the *sense context graph* for t.
3. The vertices V_t are constructed by adding those 20 tags $t_i \in T, t_i \neq t, i = 1, \ldots, 20$ which co-occur most often together with t.
4. The edges are constructed by computing the pairwise tag context relatedness as described above among all $t \in V_t$; we add an edge between t_i and t_j if their similarity is greater than zero. The weights of the edges are given by the corresponding similarity value.

2.2 Sense Context Disambiguation

Given this graph representation of the context, the next problem is how to divide it into partitions which denote different meanings. We adopted hierarchical agglomerative clustering as used by [13] as a representative of a standard sense discovery algorithm. Based on the similarities among the context tags which form the edges of the sense context graph SC_t, the hierarchical clustering procedure can be directly applied to form "sense clusters".

It results in a so-called *dendrogram*, which graphically depicts the level of distance at which each merging step took place. We used *Ward's Method* [14] for computing the distance between two clusters. In order to derive clusters (which is desirable in our case), this dendrogram needs to be further parameterized.

One method is to "cut" the latter into a set of flat sense clusters by using a distance threshold k, which we determined empirically.

After clustering, we determined for each cluster the most similar tag in this cluster corresponding to t and used it as cluster label.

3 Tagging Pragmatics

The user population of social annotation systems and the behavior we can observe in such systems varies broadly. For example, in previous work we found that different types of tagging systems lend themselves naturally to different kinds of tagging motivation [15]. In the following we present an overview of various types of measures for detecting and characterizing different kinds of tagging pragmatics, i. e., different types of users and user behavior in social annotation systems. While there is a multitude of relevant distinctions, in this paper we will focus on the existing notions of categorizers / describers as well as generalists / specialists.

3.1 Categorizers and Describers

The notion of *categorizers* and *describers* was initially presented by Strohmaier et al. in [15] and further elaborated in [16] by introducing and evaluating different measures for tagging motivation. In this previous work, we found that a useful and valid measure for distinguishing between these two types of users is the *tag/resource ratio*. We will use this measure in our experiments to characterize user behavior. The tag resource ratio is defined as $trr(u) = \frac{|T_u|}{|R_u|}$ where $|T_u|$ denotes the number of tags a user has and $|R_u|$ the number of resources of the same user. The intuition behind this metric is that a categorizer would achieve a lower score when he uses a limited vocabulary of tags whereas a describer would receive a higher value due to the higher number of different tags used. This intuition has been validated in previous work [16].

3.2 Generalists and Specialists

In our work, we aim as well to distinguish between *specialists* who exhibit a narrow topical focus when annotating resources and *generalists* who exhibit an interest in a wide variety of topics. Although there is preliminary research on this distinction, no valid measures for making this distinction automatically are available today. For this reason, we adopt a set of four metrics – motivated by the work of Stirling [17] and others – that capture some high-level intuitions about generalists and specialists in social annotation systems in general. In this work we do not explicitly validate if these measures capture the ideas of general and special behavior perfectly, because for the anticipated experiments it is sufficient that they capture pragmatic factors. We leave the task of evaluating these measures in – for example – human subject studies to future work.

Mean Degree Centrality. This measure calculates the *mean degree centrality* (based on the tag-tag co-occurrence graph) of all tags in a personomy and is determined by $mdc(u) = \frac{\sum_{t \in T_u} deg(t)}{|T_u|}$. The calculation is based on the degree of a tag measured on the tag-cooccurrence vector space of the folksonomy. The sum of the degrees of all tags is divided by the total number of distinct tags T_u of this user. The intuition behind this measure is that generalists would use more tags that co-occur with many other tags throughout the folksonomy. Hence, generalists would get a high degree centrality whereas specialist would keep this measure low.

We also used a modification of the *mdc*, where we restricted T_u to the first quartile, i. e., the 25% most used tags per user. With this measure we want to remove the long tail of the tag usage vector of a personomy and just focus on the short head. We will call that measure in short *mqdc*.

Tag Entropy. The *tag entropy* characterizes the distribution of tags in a personomy and is defined by $ten(u) = -\sum_{i=1}^{|T_u|} p(t_i) log_2(p(t_i))$. It can help us to understand user behavior based on tag occurrence distribution. Each tag occurrence count in a personomy is normalized by the total number of occurrences and stored in the probability vector p. A user can either use the tags of her personomy equally often or can focus on some few tags very often. In the first case the tag entropy would be high which would indicate that the user is more of a generalist whereas in the second case the value would be lower and the person would provide more of a specialist behavior.

Similarity Score. The *similarity score* calculates the average similarity of all tag pairs of a personomy. The formula for this final measure is $ssc(u) = \sum_{t_1, t_2 \in T_u, t_1 \neq t_2} \frac{sim(t_1, t_2)}{|T_u| \cdot (|T_u| - 1)}$. The similarity of the tag pairs is measured by the cosine similarity of the tag co-occurrence vector space [11]. A high value would indicate that a person uses many closely related tags and this would display that she focuses just on a topical sub-field of the folksonomy leading to specialist behavior. In the other case the value would be low if a user uses very dissimilar tags and this would describe a typical generalist of such systems.

4 Do Tagging Pragmatics Influence Tag Sense Discovery?

In order to explore the effects of tagging pragmatics on the ability to discover senses in tags, we set up a series of experiments where we apply the previously introduced method for tag sense discovery. Then we segment the entire folksonomy in several sub-folksonomies based on the pragmatic measures for distinguishing between different types of users and user behavior. Subsequently, we evaluate the performance of different subpopulations on this task. We start by describing our experimental datasets and how we obtained a "ground truth" for evaluation from Wikipedia.

4.1 Datasets

Semantic Grounding Using Wikipedia. Clearly, identifying a representative and reliable ground truth dataset which captures (most of) the different senses of a particular tag is a difficult task. While expert-built dictionaries like WordNet [1] contain descriptions of different word senses, their coverage is limited (e. g., roughly 60% of top Delicious tags are present in WordNet). Furthermore, due to the dynamic nature of social tagging systems, "new" senses might emerge quickly which are not yet covered in the dictionary. For this reason, we have chosen the English version of Wikipedia[2] as ground truth, as its coverage is higher (89% for BibSonomy, and 85% for Delicious) and we expect the community-driven sense descriptions to be more complete compared to WordNet. The English Wikipedia provides about 4 million articles and covers a huge range of topics.

Our main source for sense descriptions are *disambiguation pages*. Disambiguation pages can either be identified by their URL (containing the suffix _ *(disambiguation)*), or via their membership to the Wikipedia category of disambiguation pages. For a polysemous term, they contain typically an enumeration of its senses in form of a bulleted list, with each list item containing a (typically one-sentence) description of the sense, and potentially a link to a sense-specific Wikipedia article. For a given term t, we first looked up its disambiguation page, and iterated over all contained bullet list items $b_1, \ldots, b_{n-1}$. Because the first paragraph preceding the bullet list often describes the "standard meaning", we added it as an additional item b_n. If no disambiguation page was available, we use the first paragraph of the corresponding article as a single sense description. The textual description for each item was then transformed into a bag-of-words representation by (i) splitting it using whitespace as delimiter, and (ii) removing stopwords and t itself. As a result, we obtain for each term t a set of Wikipedia sense descriptions $WP_t^1, \ldots, WP_t^n$, each being essentially a set of describing terms.

Tagging Datasets. We used two different datasets to evaluate our measures on real world data. The first dataset is a dump of the social annotation system BibSonomy[3], taken in November 2010. The second dataset we used was crawled from Delicious[4] in 2006[5]. Because the applied similarity metrics are less meaningful on sparser data, we restricted each dataset to the top 10,000 most often used tags to ensure more precise similarity judgments. Furthermore, we created two further variants by restricting only to users having tagged a sufficient amount of resources in order to be able to judge about their tagging behavior (via the pragmatic measures). For Delicious, we kept only users with at least 100 resources, for BibSonomy those having at least five ones (BibSonomy users own in general much less resources compared to Delicious).

[1] `http://wordnet.princeton.edu`
[2] `http://en.wikipedia.org`
[3] `http://www.bibsonomy.org`
[4] `http://www.delicious.com`
[5] `http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/DataSets/PINTSExperimentsDataSets/index_html`

4.2 Experimental Setup

For each dataset described we calculated all the pragmatic measurements introduced in section 3, i. e., the tag/resource ratio trr for discerning *categorizers* and *describers*, and for distinguishing *generalists* from *specialists* we used the two mean degree centrality variants mdc and $mqdc$, the tag entropy ten and the similarity score ssc. For each metric m, we finally obtained a list L_m of all users $u \in U$ sorted in ascending order according to $m(u)$.

All our measures yield low values for categorizers/specialists, while giving high scores to describers/generalists. This means that e.g. the first user in the mean degree centrality list (denoted as $L_{mdc}[1]$) is assumed to be the most extreme specialist, while the last one ($L_{mdc}[k], k = |U|$) is assumed to be the most extreme generalist.

Because we are interested in the minimum amount of users needed to provide a valid basis for disambiguation, we start at both ends of L and extract two folksonomy partitions CF^m_{10} and DF^m_{10} based on 10% of the "strongest" categorizers/specialists ($CatSpec^m_{10} = \{L_m[i] \mid i \leq 0.1 \cdot |U|\}$) and describers/generalists ($DescGen^m_{10} = \{L_m[i] \mid i \geq 0.9 \cdot |U|\}$). $CF^m_{10} = (CU^m_{10}, CT^m_{10}, CR^m_{10}, CY^m_{10})$ is then the sub-folksonomy of F induced by $CatSpec^m_{10}$, i. e., it is obtained by $CU^m_{10} := CatSpec^m_{10}$, $CY^m_{10} := \{(u,t,r) \in Y | u \in CatSpec^m_{10}\}$, $CT^m_{10} := \pi_2(CY^m_{10})$, and $CR^m_{10} := \pi_3(CY^m_{10})$. The sub-folksonomy DF^m_{10} is determined analogously. We extracted partitions CF^m_i and DF^m_i for $i = 10, 20, \ldots, 100$.

For each obtained folksonomy partition, we performed tag sense discovery as described in section 2, i. e., we created the sense context graph $SC_t = (V_t, E_t)$ for each contained tag t, and disambiguated the context via hierarchical agglomerative clustering. We determined the distance threshold k empirically as 0.55 for Delicious and 0.45 for BibSonomy. As an outcome, we got for each tag t a partition of its context tags E_t into "sense clusters" $SC^1_t, \ldots, SC^m_t$ with $\bigcup_{i=1,\ldots,m} SC^i_t = E_t$.

Based on the sense clustering $SC^1_t, \ldots, SC^m_t$ obtained for each tag t in each folksonomy partition, we evaluated the "quality" of each clustering by comparison with the corresponding Wikipedia senses $WP^1_t, \ldots, WP^n_t$ of t. A crucial question hereby is when a particular clustered sense SC^i_t "matches" a reference sense WP^j_t. We used a simple approach to this end and counted a "hit" when there existed an overlap between both sets, i. e., when $SC^i_t \cap WP^j_t \geq 1$. We refer with $matches(SC^1_t, \ldots, SC^m_t)$ to the set of clustered senses which match at least one Wikipedia Sense, and with $matches(WP^1_t, \ldots, WP^n_t)$ to those Wikipedia senses which match at least one clustered sense. While this represents only an approximate matching, inspection of a small sample of sense pairs revealed that the approach works reasonably well. Future research might focus on developing and applying more elaborate sense matching approaches. Based on these matches, we computed two measures inspired by precision and recall according to:

$$precision(\{SC^1_t, \ldots, SC^m_t\}, \{WP^1_t, \ldots, WP^n_t\}) = \frac{matches(SC^1_t, \ldots, SC^m_t)}{m} \quad (1)$$

$$recall(\{SC^1_t, \ldots, SC^m_t\}, \{WP^1_t, \ldots, WP^n_t\}) = \frac{matches(WP^1_t, \ldots, WP^n_t)}{n} \quad (2)$$

4.3 Results and Discussion

Figure 1 depicts the quality obtained for different disambiguation conditions for the Delicious dataset. Along the x-axis of each plot, users are being added, sorted by each pragmatic measure, respectively. This means that the folksonomy partitions are growing towards the size of the full dataset – which is the reason that all lines meet in their rightmost point. The y-axis measures precision and recall as defined above. The black solid line corresponds to the random baseline, in which users were added in random order.

When comparing with the baseline, a first observation is that most induced sub-folksonomies based on specialist and categorizer intuitions remain below the random baseline, with increasing quality towards the full dataset condition. This suggests that tagging data produced predominately by categorizers and specialists does not enhance performance of the tag sense discovery task.

For describers and generalists, the situation becomes more interesting: While many partitions based on generalists show a similar behavior and remain below the random baseline, those based on tag entropy (*ten*) and partially those based on mean degree centrality (1st quartile, *mqdc*) perform better, and score higher precision and recall values than the complete dataset. This effect is even more pronounced for partitions based on describers (using *trr*). It suggests that the pragmatics of tagging influence the performance of knowledge acquisition tasks such as tag sense discovery. *But how do the pragmatics influence tag sense discovery in detail?*

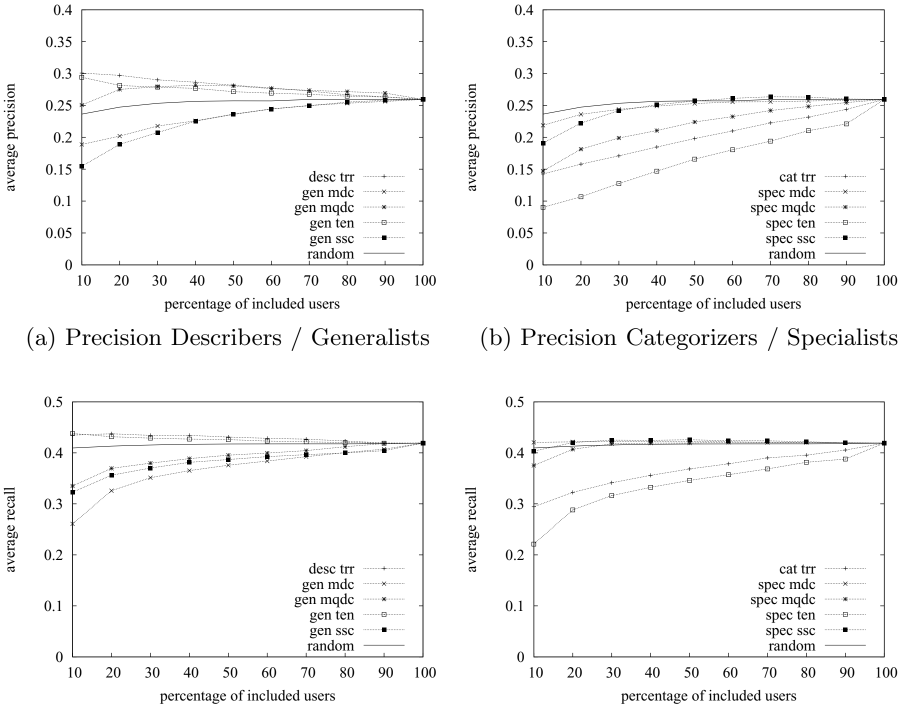
Our results offer preliminary *explanations*, identifying that particular types of behavior (such as extreme describers or extreme generalists) outperform other types of behavior (such as categorizers or specialists). On a general level, we can explain some ways in which tagging pragmatics influence tag sense discovery. For example, while categorizers and specialists in our experiments seem to negatively affect the ability to discover senses from tags, data produced by describers and generalists has demonstrated a potential to improve performance on this task. On a more specific level, we can observe that the best performance globally can be found for one of the smallest partitions, i. e., the one induced by 10% of describers. Their annotations (though technically consisting of much less data) seem to provide a better basis for discovering tag senses than the total amount of annotations in the system. One possible explanation lies in the intrinsic behavior of these users: Because their goal is to annotate resources with many descriptive keywords, it may not be surprising that they come closer to what Wikipedia editors do when "describing" word senses.

In order to verify the results of the Delicious dataset, we repeated our analyses on our second dataset (BibSonomy). The observations are consistent across our datasets, but we leave out the corresponding plots due to space limitations. Nevertheless, we provide further detailed results and plots online[6].

Understanding the ways in which tagging pragmatics influence tasks such as word sense discovery is appealing for several reasons. For example, using this

[6] `http://www.is.informatik.uni-wuerzburg.de/staff/niebler/ecir2013 supplementary_material`

kind of knowledge, very large datasets could be reduced to smaller datasets, which exhibit better performance on such tasks. Also, system engineers could provide incentives to stimulate a particular style of tagging (e.g., through tag recommender systems), which may help to foster the emergence of more precise semantic structures.



(a) Precision Describers / Generalists (b) Precision Categorizers / Specialists

(c) Recall Describers / Generalists (d) Recall Categorizers / Specialists

Fig. 1. Results for the Delicious dataset. The x-axis of each plot corresponds to the percentage of included users, ordered by the different metrics (different lines). The further to the right, the larger are the corresponding folksonomy partitions. The y-axis corresponds to precision / recall as defined in Section 4 by formulas 1 and 2. For the case of precision, higher values indicate a higher "correctness" of the discovered senses; for recall, higher values indicate a better "coverage" of Wikipedia senses. The solid line represents the random baseline. Most experimental cases stay close or below the baseline, i. e., they are not particularly well suited for disambiguation; An exception are small partitions consisting of describers (according to *trr*) and generalists (according to *ten* / *mqdc*).

5 Related Work

A first systematic analysis of emergent semantics in tagging systems was performed by [3]. One core finding was that the openness of these systems did not give rise to a "tag chaos", but led to the emergence of stable semantic patterns for a given resource. [18] presented an approach to capture emergent semantics from a folksonomy by deriving lightweight ontologies. In the sequel, several methods of capturing emergent semantics in the form of (i) measures of semantic tag relatedness [11], (ii) tag clusterings [19] and (iii) mapping tags to concepts in existing ontologies [20] were proposed. In our own previous work [2] we examined the effects of user behavior on emergent semantics in the Delicious system. We found that users called describer who try to describe things during the tagging are better candidates for the extraction of semantics from folksonomies. In [21] we evaluated a range of measures of term abstractness and concluded that centrality as well as entropy based measures are good indicators for measuring the generality level of tags. In [1] we explored the influence of tagging pragmatics on emergent social classification, finding that categorizers produce more useful tags than describers for this task.

Statistical natural language processing distinguishes between supervised, dictionary-based and unsupervised disambiguation [22]. Supervised approaches are based on labelled training data, and learn usually a classifier based on context features of a given word. Such approaches have rarely been applied to social annotation systems. Dictionary-based approaches rely on sense definitions defined in dictionaries or thesauri. [23] first identifies a set of candidate senses for a given tag within WordNet, interprets co-occurring tags as context and uses a measure of semantic relatedness to choose the most appropriate sense. In a similar manner, [4] uses cosine similarity between tag co-occurrence vectors and a bag-of-words representation of Wikipedia pages to identify the most suitable sense definition within DBPedia.[7] [24] also computes a relevance score between tags and Wikipedia articles for the same purpose.

While all of the related methods disambiguate senses in several ways, none of them focuses on the motivation of the users and its influence on the quality of the disambiguation process.

6 Conclusions

The overall objective of this paper was to look for a *signal* – we wanted to explore (i) *whether there is a link between pragmatics and tag sense discovery* and (ii) *if there is, how it might be explained.* Our results provide further evidence that in social annotation systems, knowledge acquisition tasks such as tag sense discovery can not be viewed in isolation from pragmatic factors, i. e., different kinds of users and user behavior. Our experiments demonstrate that tagging pragmatics can have an influence on the performance of tag sense discovery tasks. Our work also offers explanations, identifying the particular types of behavior (such as *extreme describers* or *extreme generalists*) that outperform other types of behaviors (such as categorizers or specialists). These findings represent an important stepping stone for future, more elaborate tag sense discovery methods that leverage pragmatic factors for improving performance. They also

[7] `http://www.dbpedia.org`

illuminate a way for engineers of social annotation systems to direct or influence user behavior in one or the other way to make their tagging data more amenable to a variety of knowledge acquisition tasks. In conclusion, our work further emphasizes the social-computational nature of social annotation systems, in which semantics emerge out of a combination of social user behavior with algorithmic computation.

References

1. Zubiaga, A., Körner, C., Strohmaier, M.: Tags vs shelves: from social tagging to social classification. In: Proceedings of the 22nd ACM Conference on Hypertext and Hypermedia, HT 2011, pp. 93–102. ACM, New York (2011)
2. Körner, C., Benz, D., Strohmaier, M., Hotho, A., Stumme, G.: Stop thinking, start tagging - tag semantics emerge from collaborative verbosity. In: Proceedings of the 19th International World Wide Web Conference, Raleigh, NC, USA. ACM (2010)
3. Golder, S.A., Huberman, B.A.: The structure of collaborative tagging systems. Journal of Information Science 32, 198–208 (2006)
4. Garcia-Silva, A., Szomszor, M., Alani, H., Corcho, O.: Preliminary results in tag disambiguation using dbpedia. In: Proceedings of the 1st International Workshop on Collective Knowledge Capturing and Representation, CKCaR 2009 (2009)
5. Au Yeung, C.M., Gibbins, N., Shadbolt, N.: Understanding the semantics of ambiguous tags in folksonomies. In: Proceedings of the International Workshop on Emergent Semantics and Ontology Evolution, ESOE 2007 (2007)
6. Si, X., Sun, M.: Disambiguating Tags in Blogs. In: Matoušek, V., Mautner, P. (eds.) TSD 2009. LNCS, vol. 5729, pp. 139–146. Springer, Heidelberg (2009)
7. Dorow, B., Widdows, D.: Discovering corpus-specific word senses. In: Proceedings of the 10th Conference on European Chapter of the Association for Computational Linguistics, EACL 2003, vol. 2, pp. 79–82. Association for Computational Linguistics, Morristown (2003)
8. Pantel, P., Lin, D.: Discovering word senses from text. In: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Canada, pp. 613–619 (2002)
9. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. In: Sure, Y., Domingue, J. (eds.) ESWC 2006. LNCS, vol. 4011, pp. 411–426. Springer, Heidelberg (2006)
10. Au Yeung, C.M., Gibbins, N., Shadbolt, N.: Contextualising tags in collaborative tagging systems. In: Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (HT2009), pp. 251–260. ACM, New York (2009)
11. Cattuto, C., Benz, D., Hotho, A., Stumme, G.: Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 615–631. Springer, Heidelberg (2008)
12. Rapp, R.: Word sense discovery based on sense descriptor dissimilarity. In: Proceedings of the 9th Machine Translation Summit, pp. 315–322 (2003)
13. Benz, D., Hotho, A.: Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. In: Proceedings of the 2nd Web Science Conference (WebSci 2010), Raleigh, NC, USA (2010)
14. Ward Jr., J.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58, 236–244 (1963)

15. Strohmaier, M., Körner, C., Kern, R.: Why do users tag? detecting users motivation for tagging in social tagging systems. In: International AAAI Conference on Weblogs and Social Media (2010)
16. Körner, C., Kern, R., Grahsl, H.P., Strohmaier, M.: Of categorizers and describers: An evaluation of quantitative measures for tagging motivation. In: 21st ACM SIGWEB Conference on Hypertext and Hypermedia, Toronto, Canada. ACM (2010)
17. Stirling, A.: A general framework for analysing diversity in science, technology and society. SPRU Electronic Working Paper Series 156, University of Sussex, SPRU - Science and Technology Policy Research (2007)
18. Mika, P.: Ontologies Are Us: A Unified Model of Social Networks and Semantics. In: Gil, Y., Motta, E., Benjamins, V.R., Musen, M.A. (eds.) ISWC 2005. LNCS, vol. 3729, pp. 522–536. Springer, Heidelberg (2005)
19. Zhou, M., Bao, S., Wu, X., Yu, Y.: An Unsupervised Model for Exploring Hierarchical Semantics from Social Annotations. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 680–693. Springer, Heidelberg (2007)
20. Angeletou, S.: Semantic Enrichment of Folksonomy Tagspaces. In: Sheth, A.P., Staab, S., Dean, M., Paolucci, M., Maynard, D., Finin, T., Thirunarayan, K. (eds.) ISWC 2008. LNCS, vol. 5318, pp. 889–894. Springer, Heidelberg (2008)
21. Benz, D., Körner, C., Hotho, A., Stumme, G., Strohmaier, M.: One Tag to Bind Them All: Measuring Term Abstractness in Social Metadata. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) ESWC 2011, Part II. LNCS, vol. 6644, pp. 360–374. Springer, Heidelberg (2011)
22. Manning, C.D., Schütze, H.: Foundations of statistical natural language processing. MIT Press, Cambridge (1999)
23. Angeletou, S., Sabou, M., Motta, E.: Semantically enriching folksonomies with flor. In: Proceedings of the CISWeb Workshop (2008)
24. Lee, K., Kim, H., Shin, H., Kim, H.J.: Tag sense disambiguation for clarifying the vocabulary of social tags. In: CSE, pp. 729–734. IEEE Computer Society (2009)