Automatic Identification of Tag Types in a Resource-Based Learning Scenario

Doreen Böhnstedt, Lasse Lehmann, Christoph Rensing, Ralf Steinmetz

Multimedia Communications Lab (KOM), Technische Universität Darmstadt,
Rundeturmstr. 10
64283 Darmstadt, Germany
{Doreen.Boehnstedt, Lasse.Lehmann, Christoph.Rensing, Ralf.Steinmetz}@kom.tu-darmstadt.de

Abstract. When users use tags they often have a rich semantic structure in mind, which can not be fully explicated using existing tagging systems. However, a tagging system needs to be simple in order to be successful, otherwise it will not be accepted by users. In our ELWMS.KOM system for the support of self-regulated Resource-Based Learning users can assign specific semantic types to the tags they use in order to manage their web-based learning resources. However studies have shown that most users would appreciate an automatic identification of tag types. In this paper we present a knowledge-based approach for the automatic identification of the tag types used in the ELWMS.KOM system. Evaluations conducted on different corpora show that the algorithm works with an overall accuracy of up to 84%.

1 Introduction

In an evolving world with fast changing circumstances and requirements learners are often required to learn self-directed in a demand-oriented manner. To close their knowledge gaps learners more and more use various resources from the WWW. This learning paradigm is called Resource-Based Learning (RBL) [8]. In order to be able to learn efficiently with resources gathered from the web, learners need to be supported accordingly: The overhead generated by tasks like search, organization and storage of the resources needs to be reduced to a minimum, and challenges like cognitive overload and disorientation caused by the vast amount of web resources and their hyperlinked structure need to be addressed. We have designed, implemented and evaluated ELWMS.KOM [2], a system for the support of self-directed Resource-Based Learning. The main goal of the ELWMS.KOM system is to give learners an accessible tool to organize their (web-)resources according to their cognitive model. To achieve this, we need to provide a means for the learners to make this model explicit, which usually is done by having the learners provide additional information about the web-resources. In recent years, researchers have perceived a mismatch regarding the creation of metadata. On the one hand, users generally avoid creating formalized metadata and do not want to fill out standardized forms to provide structured information about their documents [4]. On the other hand with the advent of the so-called Web 2.0, tagging systems have become popular, having users voluntarily generate vast amounts of tags (and thus metadata). One reason for the success of tagging in comparison to creation of formalized metadata is the fact that tagging generally is simple and does not follow formalized rules. Users basically can tag how and whatever they want.

Thus, we decided to choose tagging as a way for the users to make the model they have in mind explicit. In addition, we give them the possibility to optionally assign a semantic type (like event, person or location) to a tag – thus performing *semantic tagging*. For example, an attendee of the EC-TEL 2010 conference in Barcelona who heard a talk about the paper "*Extended Explicit Semantic Analysis for Calculating Semantic Relatedness of Web Resources*" by Philipp Scholl will very unlikely remember the full title of the paper when she wants to revisit the paper half a year later. However the probability that she remembers the event ("EC-TEL 2010"), the location ("Barcelona"), the person ("Philipp Scholl") or the topics ("TEL", "Semantic Relatedness") is very high. Thus if she or other users organized the named resource accordingly, she could access the paper later on very efficiently. The type of a tag, i.e. whether it is an event, a location, a person, a topic etc., constitutes important semantic information about a tag, which can be used to better structure and keep track of the information. An overview of related (semantic) tagging application is given in [3].

Our goal is to make the users explicate their semantic models in a more structured manner without loosing the simplicity and accessibility (and thus the success) of tagging systems. Although in our ELWMS.KOM system users have the possibility to specify the type of a given tag in a very simple and intuitive way, our final goal is to automatically identify the type of a given tag. The knowledge-based algorithm for the automatic identification we propose (Section 4) as well as its evaluation (Section 5) represent the main contributions of this paper. Further, we analyze related work in the area (Section 3) and – in the following section – shortly describe the ELWMS.KOM system, introduce the semantic tag types used and present studies that confirm our selection of types.

2 Semantic Tagging in a Resource-Based Learning Scenario

ELWMS.KOM is a system for the support of Resource-Based Learning consisting of an add-on for the browser Firefox to efficiently insert web resources into the semantic network by tagging and a web-based platform for search and retrieval [2]. User-based evaluations of the research and learning environment ELWMS.KOM have shown that it has the potential to support the resource management in RBL scenarios [10]. Learners are more active and plan their learning process better than without such a support. They are satisfied with their outcome while using ELWMS.KOM and its semantic tagging concept is broadly accepted.

The ELWMS.KOM prototype is the basis for the CROKODIL project[1] which focuses on the community aspects of Resource-Based Learning and provides a web-based community platform for the support of RBL. The tag types relevant for

[1] http://www.crokodil.de, Retrieved online: 2011-03-31

ELWMS.KOM (and CROKODIL) are detailed and enhanced with usage examples in the following, while Figure 1 shows an exemplary semantic network:

- Person / organization: e.g. author or referrer of a resource, a person the resource is connected with or is about
- Location: e.g. a location where a resource was found or is connected with
- Type: e.g. genre or mime type of a resource (e.g. blog, wiki)
- Event: e.g. a conference where a resource was presented
- Topic: what is a resource about
- Goal / Task: the goal (e.g. task or knowledge demand) a resource was searched for

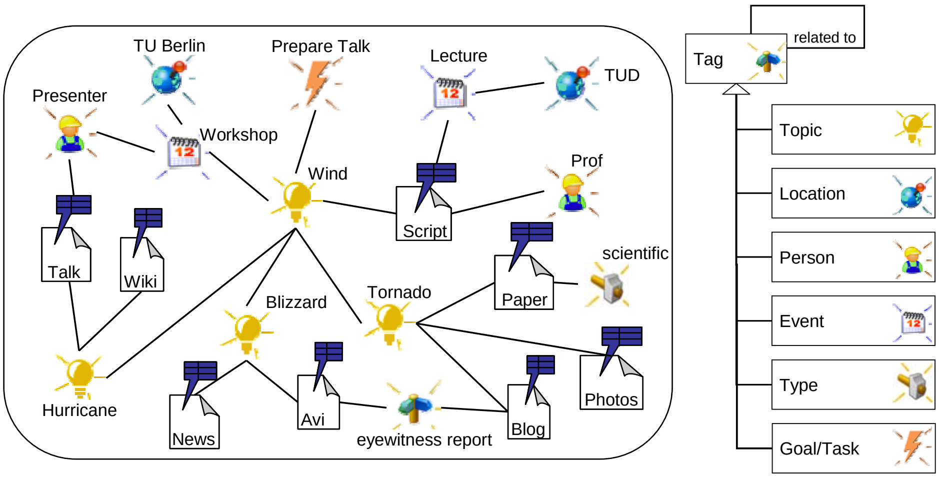

Figure 1. ELWMS.KOM – exemplary semantic network

This list of tag types was confirmed in two ways in user studies we have conducted. On the one hand, the users were asked whether the given types would satisfy their needs. This was approved by a vast majority of test persons. On the other hand, the tags that remained unassigned to one of the given tag types were inspected manually to judge whether an additional type could be identified, which was not the case. The goal type in the above list of tag types takes a special position because it is meant as an instrument for the users to organize their self-directed learning processes [10]. This goal type will always be set manually by the users and is thus excluded from the automatic identification. While the feature of tagging with tags of different types was generally well accepted in evaluations conducted so far, nearly all users stated that an *automatic identification* of the type of a tag would be great. Additionally, in the evaluations there was a significant amount of tags that remained without a type, even if it could have been assigned to one of the given tags. This issue could be addressed by an automatic identification approach, too.

Further, an analysis of tags and resources regarding the languages used has shown, that the ELWMS.KOM / CROKODIL system is used across languages. Even though most users in the evaluation have German as their native language, 75% of the resources and 30% of the tags were in English (named entities not counted).

Our goal is to provide a method to automatically detect the type of a tag in ELWMS.KOM and CROKODIL. Given this application scenario, the requirements for such an approach are as follows:

- The algorithm needs to classify a given tag into the types given above (plus a further category for tags that do not fall into one of the given categories).
- Since in our application scenario users use tags and resources across languages, the algorithm needs to work across at least German and English.

3 Related Work

In the following, we describe related approaches for the categorization of tag types and their automatic identification. Table 1 gives an overview of the most important approaches in the given context. The first three approaches also perform an automatic identification of the proposed tag types.

Table 1. Overview of tag categorization approaches

Wartena	**Bischoff**	**TagExplorer**	**Golder**	**ELMWS**
Topic	Topic	Thing	Topic	Topic
Attribute	Type		Type	Type
Author	Author/Viewer	Person/Group	Author/Owner	Person
Opinion	Rating		Rating	
	Usage Context		Task	Goal
Self reference	Self reference		Self reference	
	Location	Location		Location
			Refinement	
	Time	Time		
		Activity/Event		Event

All five approaches share *topics* and *persons* as tag types. Four of them share *events or dates* and the *type* of a given resource. Three of the systems identify *opinions* or *ratings* as well as *goals / usage context* and *location* as additional tag types.

Wartena [14] classifies tags used for books in the literature tagging system LibraryThing [13]. He defines two main categories *work* and *user*. The *topic* for the content of a book, *attribute* for genre and *usage context* of a book, and the *author* being assigned to the *work* main category, while *opinion* reflecting the personal opinion of a reader and *self reference*, which reflects the current status of a book (damaged, lent) assigned to the *user* main category. Wartena uses a machine learning approach for classification. Features used include eccentricity of tags (user-specific and for the whole corpus), frequency of tags, book ratings, author names and the containment of specific strings like "author", "reading", "great", "prize" etc. (see [14] for a complete list). For evaluation, 565 tags of the LibraryThing system were extracted and labelled manually with the categories on the two given levels (see Section 3.1). On the categorization task using the two top level categories the approach has a classification accuracy of 92% (with a baseline quality – which is according to the distribution of categories in the ground truth – of 71%). Using all categories, the accuracy of the method reaches 75% (baseline: 27%).

Bischoff et al. [1] analyzed tags in Flickr, LastFM and Delicious and identified different categories of tags in these systems (see Table 1). In the following, we describe the characteristics of the Delicious tags, since this categorization is the most closely related to our application scenario and has the most overlap with our approach. Tag types encompassing *opinions* or *ratings* are not needed in our system because the system provides other means for this. The automatic classification method proposed by Bischoff et al. relies on a twofold approach. The five tag types *time, self reference, location, author* and *type* are detected using a mix of known methods for named entity recognition, tag type specific heuristics as well as manually created or automatically extracted lists. Tags which can not be assigned to one of these five types by these methods are assigned to one of the three remaining types using a machine learning approach. Features used include the number of terms, number of characters, tag frequency, word type and the WordNet category a tag can be mapped to. The approach is evaluated on three corpora, extracted from the systems Delicious, LastFM and Flickr. All extracted tags have been labelled manually to create a ground truth. The corpus is dominated by topic tags, which make up ~67%. The approach achieves on the Delicious corpus, which is the most relevant in the given context, an average F-Measure of 69.2%.

In the **TagExplorer** system [12], Flickr tags are categorized into *location* (where the picture was taken), *subject* (corresponds to topic in other systems), *names* (corresponds to persons), *activity* and *time*. TagExplorer detects the tag type of Flickr tags using WordNet. Each tag type is mapped to specific WordNet categories. The tag label to classify is searched in WordNet and the first noun in the result list is used to determine the category. The authors assume that the WordNet categories are correct and the first noun search result represents the most probable type of that tag. Since for a direct mapping of tags to WordNet categories only a coverage of 52% of the tags can be achieved [11], so the ClassTag system [7] is used in addition. Here, information from Wikipedia is used as external source to provide a better coverage of WordNet categories. According to an extrapolation done by the authors, 69% of Flickr tags can be classified using this approach, while the mapping of tags to WordNet using the ClassTag system works with a precision of 72%.

Golder and Huberman [5], have analyzed Delicious tags and identified seven different tag types of which only the type "*Refining Category*" is not already present in other approaches. This category provides a means to further extend the functionality of Delicious by using tags to group tags and resources according to the alphabet or numbers.

The related work on tag categorization shows that while there is a common subset of categories used by most approaches, the exact combination of categories is always dependent on the specific use case and usage scenario. This also applies to the interpretation of a category itself: Where e.g. in TagExplorer a *location* states where a Flickr picture was taken, in other approaches the usage of this category is more free. On the other hand, the tag type *goal* is very relevant for a system supporting self-regulated RBL while it is not that important for tagging images in Flickr or music in LastFM. Since each approach uses different tag types and different corpora, a comparison or competitive evaluation of the different approaches is not feasible. However, the results give an idea on how well the approaches perform.

4 Automatic Identification of Tag Types

In our approach for automatic identification of tag types, we strongly rely on external corpora. In the following we shortly describe the external databases we use followed by a description of the algorithm in its entirety.

4.1 External Databases Used

Our algorithm consists of classifiers for each tag type we want to detect and most of the classifiers use one or more external databases and corpora for classification. In the following, we shortly describe the databases our algorithm makes use of.

Freebase[2] is a freely available database containing general knowledge. It contains 20 million entries for topics or objects like persons, events and others. The Freebase community maintains the data and aggregates them from other sources. Entries in Freebase possess structured information described by schemata. The main language is English although entries in different languages exist. Freebase ranks the results of a query according to its relevance using Lucene[3]. A freebase query returns ranked search results which are allocated to different Freebase categories. These categories can be mapped to the desired types.

GeoNames[4] is a location database that contains 10 million geographic names, representing 7.5 million topographic objects grouped in different main classes (e.g. state, waters, city…) and subclasses (airport, government building etc.). Each location is stored in the language of the country the object is located in as well as different other languages. English, however, is the language mostly used in GeoNames.

DBpedia[5] is a community project which aims to extract and provide structured information from Wikipedia. Partly the information is automatically extracted from Wikipedia info boxes contained in specific article types (like e.g. cities, states).

WordNet[6] is an English thesaurus containing verbs, nouns, adjectives and adverbs, grouped in so-called synsets (groups of synonyms). These synsets are allocated to categories (e.g. location, event, food …). Further, synsets possess different types of relations to other synsets, e.g. hyponymic, antonymic or holonymic relations.

Other databases used include a list of holidays[7], a database of first and last names [6], a list of web genres[8] as well as the dictionary BEOLINGUS[9]. The latter is used to

[2] http://www.freebase.com/, Retrieved online: 2011-06-20
[3] http://lucene.apache.org/, Retrieved online: 2011-06-20
[4] http://www.geonames.org/, Retrieved online: 2011-06-20
[5] http://dbpedia.org/, Retrieved online: 2011-06-20
[6] http://wordnet.princeton.edu/, Retrieved online: 2011-06-20
[7] http://de.wikipedia.org/wiki/Feiertage_in_Deutschland, Retrieved online: 2011-03-31
[8] http://www.webgenrewiki.org/index.php5/Genre_Classes_List, Retrieved online: 2011-03-31
[9] http://dict.tu-chemnitz.de/, Retrieved online: 2011-03-31

determine the lexical category of terms contained in a tag label. All databases and sources are freely available and none of them have been adapted or extended in the course of this work.

4.2 Tag Type Identification Algorithm

There are two different challenges in the automatic identification of a tag type: On the one hand we need to identify the set of types that are valid for a given tag, e.g. the tag "Merkel" can be a *person* meaning the German chancellor "Angela Merkel" or a *location* meaning Merkel, Texas. We determine the valid types by querying type specific databases like GeoNames for locations. On the other hand we need to decide which one of the valid tag types is plausible for a given tag. Thus in cases where different types are possible for a tag, we need a means to put the tag types in an order. This is what the generic Freebase database is used for. The main assumption we make is that the ranking of search results in Freebase reflects the probability for the conventionality of the semantic manifestation of the queried label represented by the search result. **Figure 2** shows the classification process. The only input to the system is the tag that needs to be classified, while the output is the type proposed by the algorithm, or optionally depending on the use case, a ranked list of types. The single steps of the process are described in detail in the following.

The **Preprocessing** step (1) involves filtering of special characters and detection of tag borders. The latter is necessary for the preprocessing of tags from the Delicious corpus only (cf. Sect. 5.1). Delicious tags are automatically divided into separate tags when a space is entered by a user, i.e. if a user gives two terms separated by a space these terms will always be treated as single tags. In order to have multi-term tags, users use various other delimiters instead. Often characters like '_' or '-' or a special notation like “camelCase” is used. Thus the preprocessing step tries to divide these tag compositions into single terms. Due to the fact that in the given scenario, many tags as well as resources are in German as well as in English but most of the entries in external databases are in English a **translation** is necessary for some of the tag types – specifically for topics. This is done using interlanguage links in Wikipedia. If an article matching the tag is present in the German Wikipedia and the corresponding article contains an interlanguage link to the English version of the article, the label of the English version is used as translation, otherwise no translation is used. We use this means for translation to avoid noise caused by translation ambiguity. The resulting tag and - if available - its translation are used as search terms in the Freebase processing step.

The original tag and - if available - its translation is used to conduct a **Freebase Query** (2). This query returns a list of Freebase entries as result. The tag types used in our scenario can be mapped to Freebase categories which are used to structure entries of the Freebase database. The Freebase entries that result from the query are matched with the tag used in the query. A Freebase entry matches, if the case-folded tag exactly matches the case-folded label of the Freebase entry, except for the tag type *person*, where a containment of the tag in the entry label is counted as matching as well. The top ten matching results are divided and each result is forwarded to the respective classifier together with its rank within the top ten. The ranks of translated tags are only forwarded to the topic, type and event classifiers. Since the person classifier doesn't use translations at all and the location classifier uses other sources for translation.

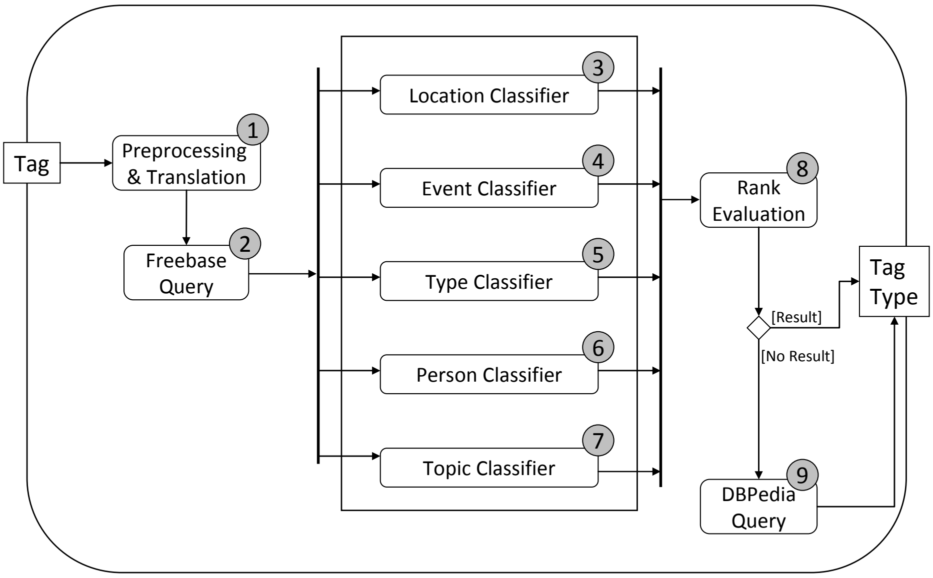

Figure 2. Overview of tag type identification approach

The **Location Classifier** (3) in the first instance queries the location database GeoNames with the original tag. When a tag label equals one of the German or English aliases of a result this result including all German and English aliases is stored in a list. Matching in that case means that the case-folded tag matches the likewise case-folded alias exactly, thus usually there is only one GeoNames result for which several different aliases are stored. The list resulting from GeoNames is then used to query Freebase and the minimal rank for all different aliases is returned by the classifier. Since a translation is already provided by GeoNames in the form of aliases, the Wikipedia translation of the original tag is not used by this classifier.

In addition to the ranked entries provided by Freebase, the **Event Classifier** (4) uses heuristics that give a bias towards the *event* type if they apply. If the tag label contains a four digit number (like 2011) or a combination of numbers between 1 and 31 and the name of a month, the classifier returns a rank of zero (i.e. the highest possible rank). If these heuristics do not apply, a database containing holidays is queried. If the tag label is contained in this database, again zero is returned. If none of these methods produce results the minimal Freebase rank of the original tag and its translation is returned.

The **Type Classifier** (5) first uses a list of web genres which was extracted, translated using Wikipedia and stored in a database. If the tag matches one of the entries in the database, rank zero is returned. Otherwise the classifier returns the minimum Freebase rank of the original tag and its translation.

The **Person Classifier** (6) uses a string-based heuristic to determine whether the tag is comprised of patterns like "B Gates" or "Bill G". If this heuristic applies rank zero is returned. As stated above, for classifying *persons*, the Freebase query in step (2) is done slightly different than for other tag types. On the one hand, no translation is used since persons are named entities and on the other hand, the matching between tag label and result does not have to be exact, but the containment of the tag in the result label is sufficient. Thus, the tag "Merkel" will match "Angela Merkel". Since Freebase only contains known personalities like politicians, musicians or writers, we need a means to identify names for unknown persons. Therefore an additional database containing German and English first and last names is used. The more frequently a name matching the tag label is, the higher the probability that the tag is a name and the lower the rank that is returned by the classifier. The minimal rank returned by this heuristic is one.

The **Topic Classifier** (7) classifies all Freebase entries as topic which can not be mapped to one of the other tag types and returns the respective rank. If the tag label is not found in Freebase but a matching Wikipedia entry is found, a rank of 11 (the maximal Freebase rank plus one) is returned. If this does not apply as well, WordNet and BEOLINGUS are used to determine the word types of the tag label. If most terms of the tag label are nouns, it is classified according to [1] as topic with a rank of 12.

During the **Rank Evaluation** (8) step, the ranks returned by the different classifiers are compared and the tag type with the lowest integer value rank is returned as the proposed tag type. It is possible that two types are assigned the same rank, due to the fact that tags are translated before querying the external data sources. In that case, the rank of the tag type determined by the original tag will always be prioritized over the type received with a translated tag.

The **DBpedia Query** (9) is an optional step and a fallback solution that is used in the unlikely case that the process so far did not lead to any result, i.e. none of the classifiers could identify the tag as the respective type. The tag is used to query DBpedia and the first result in the list that can be mapped to one of the tag types using the DBpedia category system is returned as the proposed type of the tag.

Due to the fact that the approach is solely based on heuristics and external sources (which can be cached locally), the algorithm can identify the type of a tag in near real-time, i.e. the average decision time is 416ms.

5 Evaluation

We have evaluated the proposed approach using two different corpora. The first one is comprised of tags taken from Delicious while the other corpus is constructed from usage data of the RBL system ELWMS.KOM described above.

5.1 Setting and Methodology

For both corpora used, the approach has to solve a categorization problem. Thus standard information retrieval measures like accuracy, precision, recall and F-measure can be used for evaluation.

The **Delicious corpus** consists of tags collected from the social bookmarking service Delicious[10]. To make sure that the corpus contains tags of all desired tag types like *locations*, *persons* and *events*, Delicious was searched for specific keywords (like "vacation" for getting *location* tags or "war" for getting *event* tags). 225 different resources from 143 users were retrieved. The corpus was then built out of all tags of resources found this way, excluding the tags initially searched for. All tags were then manually categorized by three human raters in order to construct a ground truth for the identification of tag types. The raters were able to inspect the tagged web resource before making their decision. In addition to the ELWMS.KOM tag types, the raters could assign a tag to an additional "not classified" category for tags that could not clearly be assigned to one of the named types. **Table 2** shows details for the Delicious corpus. The inter-rater agreement calculated using the free-marginal multi-rater variant of Fleiss Kappa [9] for this corpus was 0.66. Although there is considerable agreement among the raters, this also shows that even for human raters this classification task is not easy and the decision on the type of a tag can be subjective (see Sect. 5.2). Therefore mainly two subsets of the corpus are used for the evaluation: the subset of tags where the raters completely agreed on the type and the set where two of the three raters agreed.

Table 2. Details on evaluation corpora

	Delicious Corpus	**ELWMS Corpus**
Resources considered	225	422
Different users	143	21
Overall tags	1272	1161
Tags uniformly rated by all three raters	620 (100%)	493 (100%)
… classified as person	106 (17.1%)	168 (34.1%)
… classified as location	144 (23.2%)	16 (3.2%)
… classified as event	16 (2.6%)	56 (11.4%)
… classified as type	27 (4.4%)	19 (3.9%)
… classified as topic	288 (46.5%)	234 (47.5%)
… not classified	39 (6.3%)	0 (0%)
Tags with a 2/3 majority of raters agreeing	1018 (100%)	948 (100%)
… classified as person	141 (13.9%)	217 (22.9%)
… classified as location	159 (15.6%)	19 (2.0%)
… classified as event	25 (2.5%)	73 (7.7%)
… classified as type	58 (5.7%)	64 (6.8%)
… classified as topic	519 (51.0%)	567 (59.8%)
… not classified	116 (11.4%)	8 (0.8%)

The **ELWMS** corpus was generated during one of the user studies conducted for the ELWMS.KOM system. During the usage of the system, users searched for web-resources in order to complete tasks or to do research on specific topics. The users used the ELWMS.KOM Firefox add-on to tag the resources they found with tags of the given types (see Sect. 2). To construct the corpus, 1161 tags from 21 different

[10] http://www.delicious.com/, Retrieved online: 2011-06-20

users of the system were collected. The obtained tags were again manually categorized by two additional human raters to confirm the tag types and smooth out the subjectivity, resulting in three ratings per tag in total. On this corpus the agreement of the raters was 0.67. Again, two agreement subsets of this corpus were used for evaluation.

5.2 Results

Table 3 shows the confusion matrix when running our algorithm on the uniformly rated Delicious corpus. The fact that all three raters classified the tag types uniformly means that a classification of this subset is comparably straightforward. In this configuration, the approach reaches an overall accuracy – i.e. ratio of correctly classified tag types – and F-measure of 81.3%, showing that our approach works considerably well.

Table 3. Results for uniformly rated tags on Delicious corpus

a	b	c	d	e	f	← classified as
91	2	0	0	6	7	a = Person
2	136	0	0	4	2	b = Location
0	0	15	0	0	1	c = Event
1	0	1	8	9	8	d = Type
15	17	10	1	230	15	e = Topic
4	0	0	0	11	24	f = Not categorized
0.81	0.88	0.58	0.89	0.88	0.42	Precision
0.86	0.94	0.94	0.30	0.81	0.62	Recall
0.83	0.91	0.71	0.44	0.84	0.5	F-measure
					81.3%	F-measure (Ø)

Table 4. Results for tags with 2/3 majority on Delicious corpus

a	b	c	d	e	f	← classified as
108	4	0	0	17	12	a = Person
3	150	0	0	4	2	b = Location
0	0	20	0	3	2	c = Event
2	0	2	19	26	9	d = Type
43	26	14	11	389	36	e = Topic
22	4	5	1	40	44	f = Not categorized
0.61	0.82	0.49	0.61	0.81	0.42	Precision
0.77	0.94	0.80	0.33	0.75	0.38	Recall
0.68	0.87	0.61	0.43	0.78	0.40	F-measure
					71.2%	F-measure (Ø)

The confusion matrix shows that there is generally a high confusion between the *topic* type and the other tag types. This is due to the fact that the word type based heuristic in the topic classifier marks noun based tags automatically as topics even if they have not been identified as topics by the raters. Another reason is the freedom of raters to classify a *person* tag as topic, e.g. if a resource is *about* a person and thus the person can be seen as topic of the resource. This distinction is not made by our approach, which always classifies such a tag as *person*.

When taking into account tags where two of the three raters agreed on the type, the task becomes significantly harder, which is obvious since the human raters were not able to agree on a distinctive tag type as well. In such a scenario (shown in Table 4) the accuracy drops to 71.7% (the F-measure to 71.2%). However, when taking into account the rating of the third rater as a valid option, an F-measure of 78.2% is reached by our approach. Again the main source of errors is the classification of topics, specifically the distinction between topics and uncategorized tags.

When running the approach on the ELWMS corpus, it generally shows similar results. Again the overall accuracy and F-measure are significantly better on the "easy" classification tasks (where all raters agreed on the tag type). Here the approach reaches almost 84% (79.1% accuracy – see Table 5).

Table 5. Results for uniformly rated tags on ELWMS corpus

a	b	c	d	e	f	← classified as
140	2	0	0	2	24	a = Person
0	15	0	0	1	0	b = Location
0	0	53	0	3	0	c = Event
0	5	0	4	10	0	d = Type
15	2	1	0	178	38	e = Topic
0	0	0	0	0	0	f = Not categorized
0.90	0.63	0.98	1.0	0.92	0.0	Precision
0.83	0.94	0.95	0.21	0.76	0.0	Recall
0.87	0.75	0.96	0.35	0.83	0.0	F-measure
					83.7%	F-measure (Ø)

Table 6. Results for tags with 2/3 majority on ELWMS corpus

a	b	c	d	e	f	← classified as
169	5	1	0	13	29	a = Person
1	16	0	0	2	0	b = Location
6	0	62	0	4	1	c = Event
0	5	0	6	53	0	d = Type
33	7	11	4	415	97	e = Topic
0	0	0	0	3	5	f = Not categorized
0.81	0.48	0.84	0.60	0.85	0.04	Precision
0.78	0.84	0.85	0.09	0.73	0.63	Recall
0.79	0.62	0.84	0.16	0.79	0.07	F-measure
					74.0%	F-measure (Ø)

Again, the algorithm shows slightly weaker results on the harder version of the corpus. However, with 74%, the F-measure is still quite good. Again the major reason for errors is the confusion between *topic* tags and uncategorized tags.

In an additional evaluation we used a corpus solely containing English tags and named entities for both corpora to show the (possibly) negative influence of translations in the approach. In almost all cases the algorithm performed better with English tags (and named entities) only. This is an expected result due to the fact that nearly all external sources we used are mostly in English. On this version of the Delicious corpus the algorithm achieved an F-measure of 86.6% and 70.2% respectively and on the ELWMS corpus 85.4% and 73.6%. However, the difference is still small enough to conclude that the cross-language capabilities of our approach work well.

Overall, the results of our approach are comparable (if not better) to other approaches in the area. As said before, due to the differences in tag categories and usage scenarios, a competitive evaluation of approaches is not feasible. Although we have evaluated the approach only for corpora containing a mix of German and English tags and resources, theoretically the approach is applicable to other languages as well, as long as there is a Wikipedia language version that is suited to serve as translation corpus.

6 Conclusions

The semantic tagging approach based on tag types, shown in this paper, enables users of the Resource-Based Learning systems ELWMS.KOM and CROKODIL to easily build a semantic network for organizing web-resources according to their cognitive model. Furthermore, these tag types can be used to provide different visualizations, e.g. geographic maps with location tags and calendar views with event tags. In this paper, we have proposed a knowledge-based real-time-capable approach for the context-free identification of these tag types. On the tested corpora, our approach shows satisfying results that would allow for an automatic classification of tags without user interaction and with a high reliability. Thus, the algorithm reduces the manual effort for building the semantic network.

At the moment, the algorithm is context-free, i.e. it does not take into account the context of a given tag (e.g. the resource tagged, the user who tags or other tags for the same or similar resources) and makes a decision solely based on external knowledge. The evaluations have shown that this approach works very well. However, our approach will always return the most plausible type for a tag. That means that if for instance a user really means "Merkel, Texas" instead of "Angela Merkel" the algorithm will fail. The algorithm can theoretically work perfectly only if it takes into account the context of a tag, for example the content of the tagged resource. Thus our next step will be to match the content of the tagged resource with the content of the Freebase articles returned by the algorithm and adapt the ranking accordingly. One other possibility for further work is to utilize the ranked list of tag types to make other meanings of a tag visible to a user and thus broaden his horizon. Thus, the algorithm could also provide direct "learning support".

Acknowledgments. This work was supported by funds from the German Federal Ministry of Education and Research under the mark 01 PF 08015 A and from the European Social Fund of the European Union (ESF). The responsibility for the contents of this publication lies with the authors.

References

1. Bischoff, K., Firan, C.S., Kadar, C., Nejdl, W., Paiu, R.: Automatically identifying tag types. In: Huang, R., Yang, Q., Pei, J., Gama, J., Meng, X., Li, X. (eds.) Advanced Data Mining and Applications. LNCS, vol. 5678, pp. 31–42. Springer, Heidelberg (2009)
2. Böhnstedt, D., Scholl, P., Benz, B., Rensing, C., Steinmetz, R., Schmitz, B.: Einsatz persönlicher Wissensnetze im Ressourcen-basierten Lernen. In: Fischer, S., Seehusen, S., Lucke, U. (eds.) DeLFI 2008. Köllen, Bonn, GI. LNI, vol. P-132, pp. 113–124 (2008)
3. Böhnstedt, D., Scholl, P., Rensing, C., Steinmetz, R.: Collaborative semantic tagging of web resources on the basis of individual knowledge networks. In: Pianesi, G., Zancanaro, F., Houben, M., McCalla, G.-J. (eds.) UMAP 2009. LNCS, vol. 5535, pp. 379–384. Springer, Heidelberg (2009)
4. Duval, E., Hodgins, W.: A LOM research agenda. In: Proceedings of the twelfth international conference on World Wide Web, pp. 1–9 (2003)
5. Golder, S.A., Huberman, B.A.: Usage patterns of collaborative tagging systems. Journal of Information Science, 32(2), pp. 198–208 (2006)
6. Lokaiczyk, R.: Clustering und Klassifikation von Personennamen. diploma thesis, Universität Leipzig (2006)
7. Overell, S., Sigurbjörnsson, B., Zwol, R.: Classifying tags using open content resources. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining, pp. 64–73, New York, USA (2009)
8. Rakes, G.C.: Using the internet as a tool in a resource-based learning environment. Educational Technology, v36, n5, pp. 52–56 (1996)
9. Randolph, J.J.: Free-marginal multirater kappa: An alternative to fleiss' fixed-marginal multirater kappa. In: Joensuu Learning and Instruction Symposium (2005)
10. Scholl, P., Benz, B., Böhnstedt, D., Rensing, C., Steinmetz, R., Schmitz, B.: Einsatz und Evaluation eines Zielmanagement-Werkzeugs bei der selbstregulierten Internet-Recherche. In: Fischer, S., Seehusen, S., Lucke, U. (eds.) DeLFI 2008., Köllen, Bonn, GI, LNI, vol. P-132, pp. 125–136 (2008)
11. Sigurbjörnsson, B., Zwol, R.: Flickr tag recommendation based on collective knowledge. In: Proceedings of the 17th international conference on World Wide Web, pp. 327–336. New York, USA (2008)
12. Sigurbjörnsson, B., Zwol, R.: Tagexplorer: Faceted browsing of flickr photos. Technical Report, Yahoo! Research (2010)
13. Voss, J.: Librarything - Web 2.0 für Literaturfreunde und Bibliotheken. In: Mitteilungsblatt der Bibliotheken in Niedersachsen und Sachsen-Anhalt, pp.12-13 (2007)
14. Wartena, C.: Automatic classification of social tags. In: Lalmas, M., Jose, J., Rauber, A., Sebastiani, F., Frommholz, I. (eds.) Research and Advanced Technology for Digital Libraries. LNCS, vol. 6273, pp. 176–183. Springer, Heidelberg (2010)