Categorising Social Tags to Improve Folksonomy-based Recommendations

Iván Cantador[1,3], Ioannis Konstas[2,3], Joemon M. Jose[3]

[1] *Departamento de Ingeniería Informática, Universidad Autónoma de Madrid, 28049 Madrid, Spain*

[2] *School of Informatics, University of Edinburgh, EH8 9AB Edinburgh, United Kingdom*

[3] *Department of Computing Science, University of Glasgow, G12 8QQ Glasgow, United Kingdom*

Abstract

In social tagging systems, users have different purposes when they annotate items. Tags not only depict the content of the annotated items, for example by listing the objects that appear in a photo, or express contextual information about the items, for example by providing the location or the time in which a photo was taken, but also describe subjective qualities and opinions about the items, or can be related to organisational aspects, such as self-references and personal tasks.

Current folksonomy-based search and recommendation models exploit the social tag space as a whole to retrieve those items relevant to a tag-based query or user profile, and do not take into consideration the purposes of tags. We hypothesise that a significant percentage of tags are noisy for content retrieval, and believe that the distinction of the personal intentions underlying the tags may be beneficial to improve the accuracy of search and recommendation processes.

We present a mechanism to automatically filter and classify raw tags in a set of purpose-oriented categories. Our approach finds the underlying meanings (concepts) of the tags, mapping them to semantic entities belonging to external knowledge bases, namely WordNet and Wikipedia, through the exploitation of ontologies created within the W3C Linking Open Data initiative. The obtained concepts are then transformed into semantic classes that can be uniquely assigned to content- and context-based categories. The identification of subjective and organisational tags is based on natural language processing heuristics.

We collected a representative dataset from Flickr social tagging system, and conducted an empirical study to categorise real tagging data, and evaluate whether the resultant tags categories really benefit a recommendation model using the Random Walk with Restarts method. The results show that content- and

context-based tags are considered superior to subjective and organisational tags, achieving equivalent performance to using the whole tag space.

Keywords: Social tagging, recommender systems, ontologies, Semantic Web, W3C Linking Open Data

1 Introduction

1.1 Motivation

During the last few years, we have been witnessing an unexpected success and increasing popularisation of social tagging systems. In these systems, users create or upload content (items), annotate it with freely chosen words (tags), and share it with other users. The whole set of tags constitutes an unstructured knowledge classification scheme that is commonly known as *folksonomy* [32]. This implicit classification is then used to search and recommend items. The nature of tagged items is manifold: photos in Flickr[1], audio tracks in Last.fm[2], video clips in YouTube[3], and Web documents in Delicious[4], among others. A user can usually create (upload) items, and annotate them with tags he considers appropriate. In some folksonomies, the user can also tag items he did not create.

The main advantage of folksonomies is that users are not requested to rely on a priori agreed knowledge structure or shared vocabulary, and thus are not imposed any constraint in the tagging process and information management. Nevertheless, this issue implies a number of limitations on the content retrieval mechanisms. Social tags may explicitly describe the content of an item, e.g. by listing physical objects that are shown in a photo or a video, or by giving keywords that appear in a Web document or a song lyric. They may also provide contextual information about the annotated item, e.g. by identifying the place a photo was taken, or the date a video was recorded. Furthermore, they may express subjective opinions and qualities (`nice picture`, `rock music`, `dark movie scene`, `incomprehensible text`), or self-references and personal tasks (`my wife`, `to read`, `work`). This suggests that users have different intentions when tagging, and not all the tags available in a folksonomy are related with the content of the annotated items [3].

Current folksonomy-based search and recommendation engines do not take into account the above distinction of tags, and run their content retrieval algorithms in the entire tag space. The problem is that

---

1 Flickr – Photo sharing, http://www.flickr.com

2 Last.fm – Personal online radio, http://www.last.fm

3 YouTube – Video sharing, http://www.youtube.com

4 Delicious – Social bookmarking, http://delicious.com

although useful subjective and organisational tags are for the purposes (intentions) of an individual, still they may fail to be of benefit when recommending items to other users. As a result, mixing these with the rest content- and context-based tags may not add or even deteriorate the overall quality of the recommendations. We hypothesise that distinguishing and considering purpose-oriented categories of tags could be extremely valuable to improve the accuracy of recommendation approaches. Hence, the corroboration of this assumption represents the main challenge to address in the work presented herein.

In order to achieve such tag categorisation, we first have to understand the meaning of each social tag. For example, to determine that `kilt` can be categorised as a "content-based" tag, it has to be identified that a `kilt` is a Scottish piece of cloth, i.e. a "physical entity". Similarly, to categorise `glasgow` as a "context-based" tag, it has to be identified that Glasgow is a city in Scotland, i.e. a "location".

It is our objective to study the role of various tag categories for item recommendation. However, categorising a set of general purpose tags is not trivial. In this context, we have identified the following research questions:

- **RQ1:** ***Is it possible to find out the underlying meanings of social tags in a general way?***

  We should 1) identify the meanings of social tags independently of the domains covered by the folksonomies they belong to; and 2) be aware of contemporary terminology that continuously appears in our daily lives (`web 2.0`, `podcast`, `diy`).

- **RQ2:** ***Is it possible to automatically categorise social tags based on their intention?***

  The transformation of semantic concepts into purpose-oriented categories could be done by exploiting external knowledge bases such as thesauri, taxonomies and ontologies.

- **RQ3:** ***Is a purpose-oriented categorisation of social tags useful for folksonomy-based recommendation strategies?***

  To validate the utility of the purpose-oriented categories, these should be evaluated in a real folksonomy-based recommender system.

1.2 Contributions

In this work, we address the research questions listed above, and make the following contributions:

- We have developed a mechanism that automatically processes and maps social tags to semantic concepts depicted in external structured knowledge bases.

- Exploiting the semantic relations provided by the above knowledge structures, we have designed a novel strategy to automatically infer the semantic classes of a given concept that allow determining the intension of the corresponding social tag.
- We have conducted an empirical study to evaluate the effect of various tag categories in photo recommendation. The experiments have been performed with a dataset obtained from Flickr, a multi-domain tagging system where photos are freely annotated by their owners. The results show there are certain tag categories that are superior to others in terms of recommendation performance, and even equivalent to that obtained when using the whole tag space.
- We have collected a dataset from Flickr system, which we have made available to the research community.

1.3 Structure of the paper

The rest of the paper is organised as follows. Section 2 describes related works that have motivated our research. Section 3 presents an overview of our approach to automatically categorise social tags based on their purpose. Section 4 explains in more detail the approach, describing how the semantic concepts underlying social tags are identified, and how they are mapped to a set of predefined purpose-oriented categories. Section 5 describes the folksonomy-based recommendation model with which we have evaluated our tag categorisation proposal. Section 6 presents the conducted experiments, and Section 7 provides a discussion of the obtained results. Finally, Section 8 contains conclusions and future research lines.

2 Related Work

2.1 Categorisation of social tags

A prior goal in many social tagging systems is to meet the needs of individual users, e.g. by allowing personal organisation of items and their subsequent retrieval. Nonetheless, social tags should help other people to browse and find items. Furthermore, being a mechanism of community-based item description, they should also facilitate information sharing and discovery (recommendation). Marlow et al. [22], and Ames and Naaman [3] discuss an exhaustive list of incentives expressing the range of potential motivations that influence tagging. Among them, content management and retrieval are shown as two of the most important incentives to tag resources. Our work is based on this observation, and aims to identify which social tags are more useful for content retrieval and recommendation.

To provide such functionalities, it is not obvious how social tags can be best exploited. Suchanek et al. [35] show that user-generated tags present significant semantic noise more than terms extracted from Web page contents or search queries. When tagging, people not only introduce misspellings (`barcelona`, `barclona`), and use different synonyms (`car`, `automobile`), acronyms (`nyc`, `new york city`) and morphologic derivations (`blog`, `blogs`, `blogging`) for a given concept [36], but also include tags that express personal assessments (`funny`, `to print`), or even are unintelligible to another person (`#####`) [35]. We deal with these issues making use of a tag processing and filtering approaches presented in a previous works [36][37], and mapping the resultant tags to semantic concepts described in external knowledge bases (KB), similarly to [9], where social tags are linked to ontology classes and instances.

The purposes of tagging and consequently the types of social tags are manifold. Recent works have analysed this fact, aiming to identify which are the social tags relevant for knowledge management and information retrieval. Apart from describing the content of the items, social tags may represent contextual information [3], subjective opinions and qualities [15][30], or self-presentation and organisation aspects [41]. We consider these purpose-oriented tag categories, and propose a more fine-grained categorisation within them, in order to study which types of tags are useful for content retrieval tasks.

Motivated by the previous works, Bischoff et al. [7] manually classify a number of tag collections obtained from different social tagging systems (Flickr, Delicious, Last.fm) in several tag types, and study the distributions of tags assigned to each type, analysing their usage implications on search tasks. The obtained results provide insight into the use of different kinds of tags for improving search. Here we go a step beyond attempting to categorise the tags automatically. In this case, the evaluation of the tag categorisation is assessed with a recommendation model [17], which does not depend on a specific domain. In this paper, we have conducted experiments with a dataset obtained from Flickr, where photos are freely tagged by the owners in a multi-domain scenario.

To achieve such tag categorisation, the meanings of social tags have to be found beforehand. We propose to map them to semantic concepts described in external KBs, such as thesauri and ontologies. Halpin et al. [16] show that tagging distributions tend to stabilise into power law distributions, which is an essential aspect of what might be user consensus around the categorisation of information driven by tagging. The authors state that it is quite plausible that folksonomies and ontologies are fundamentally compatible. We follow this principle, and attempt to integrate social tags into YAGO [34], a large ontology that covers WordNet [24] and a significant part of Wikipedia[5].

Constructing and linking folksonomies with structured semantic KBs is indeed a problem that has attracted much attention recently [28]. Mika [23] is recognised as one of the first authors to extend the traditional bipartite resource-concept model of ontologies with the social dimension. He presents a graph

---

[5] Wikipedia encyclopaedia, http://wikipedia.org

based approach to construct a network of related tags, projected from either a user-tag or resource-tag association graph. Applying clustering techniques to tags, and using their co-occurrence statistics, he produces conceptual hierarchies. Specia & Motta [33] present a combination of pre-processing strategies and statistical techniques together with knowledge provided by ontologies available on the Semantic Web to generate clusters of highly related tags that correspond to ontology concepts. As explained in subsequent sections, we shall also make use of tag processing and filtering techniques, similar to those presented in [33]. Angeletou [4] proposes a semantic enrichment of folksonomies by exploiting online ontologies, thesauri and other knowledge sources to make explicit the semantic relations between social tags. Instead of inferring such semantic relations between tags, we use those explicitly defined in YAGO. This is enough for our approach since our goal is to categorise the tags, and we only have to exploit hierarchical relations between them.

2.2 Folksonomy-based recommender systems

Collaborative tagging systems allow a user to search for the content that he has tagged using a personal vocabulary. As users with similar interests tend to have a shared vocabulary, tags created by one user may be useful to others, particularly those with similar interests. This is in fact the essence of recommender systems. In these systems, a user does not usually declare explicitly his information needs (e.g., by means of a keyword-based query). In contrast, he is presented with items that may be interested for him according to his profile (content-based approaches), or to the profiles of "similar" people (collaborative filtering approaches). The reader is referenced to [1] for an overview of the state of the art in recommender systems. In the following, we focus our attention on recommendation approaches that exploit folksonomy information.

Au Yeung et al. [6] describe a strategy that clusters the items tagged by the users. In the item-tag space, given a network of items, a graph-based clustering algorithm to obtain sets of related items is applied. As the different clusters should contain items that are related to similar topics, a cluster can be considered as corresponding to one of the interests of the user. Moreover, the experiments presented in the paper show that the obtained groups of tags and items seem to correspond to the different meanings of ambiguous tags. In this work, we also use a graph-based algorithm on the item-tag space. In our case, using a Random Walk strategy we aim to identify related tags and items relevant for the user.

Similarly, Gemmell, Shepitsen, Mobasher and Burke explore in several works [14][31] strategies that cluster the entire space of tags to obtain sets of (semantically) related tags. These clusters may represent coherent topic areas. By associating a user's interest to a particular cluster, the user's interests in the topic are surmised. As discussed in the last section of the paper, this type of clustering techniques could be incorporated into our approach in order to enhance the automatic categorisation of ambiguous social tags, according to the context of the user profile in which a given tag appears.

Instead of implicit clusters, other personalization and recommendation approaches aim to exploit explicit, and more structured representations of folksonomies. Quintarelli et al. [29] propose a personal multi-facet categorisation of tags, which allows the exploitation of taxonomic relations to enhance content retrieval. In a series of previous works [8][9][36], we have investigated recommendation approaches that make use of ontology-based user profiles. Social tags are automatically transformed into ontology concepts (classes and instances) using semantic knowledge bases like WordNet and Wikipedia. Arbitrary ontology relations between these concepts are exploited to expand the user profiles, and personalise search and recommendation results. In this work, we also attempt to map social tags to semantic concepts. As explained in subsequent sections, in this case, we propose to use YAGO ontology, aiming to join and contribute to the W3C Linking Open Data initiative.

Recent works have focused on exploiting folksonomies as sources of semantic information, integrating them with content-based and collaborative filtering (CF) recommendation approaches.

De Gemmis et al. [11] present a hybrid strategy that learns the profile of the user from both static content and tags associated with items rated by him, instead of relying on tags only. The authors propose to include in the user profile not only his personal tags, but also the tags adopted by other users who rated the same items as him. Since the main problem lies in the fact that tags are freely chosen by users, and their actual meaning is usually not very clear, they suggest to semantically interpreting tags by means of WordNet. Our tag categorisation also follows this idea, but extends the use of WordNet to Wikipedia, allowing the consideration of social tags related to proper nouns and contemporary terms not available in a dictionary such as WordNet.

Tso-Sutter et al. [38] describe a generic method that allows tags to be incorporated into standard heuristic-based CF algorithms, such as user- and item-based CF, by reducing the three-dimensional (user, item, tag) correlations to three two-dimensional correlations, and then applying a fusion method to re-associate these correlations. The integration of folksonomy information into CF is also studied by Zhen et al. [44]. In this case, the authors propose to use the model-based CF algorithm based on probabilistic matrix factorization. Differently to these approaches, as explained in the paper, our tag-based recommendation model follows the CF paradigm by means of applying Random Walk algorithm on the global graph formed by users, items, tags and their explicit relations.

3 Overview of the approach

Our goal is to automatically categorise social tags based on their intention, considering the following four main categories:

- **Content-based**. Social tags that describe the content of the items, such as the objects and living things (animals, plants) that appear in a photo or video, or are mentioned in a text document or a song lyric. Some examples of tags belonging to this category are `vehicle`, `dog` and `tree`.
- **Context-based**. Social tags that provide contextual information about the items, such as the place where a photo was taken, the date or period of time when a video was recorded, etc. Examples of this kind of tags are `madrid`, `mountain`, `summer` and `holidays`.
- **Subjective**. Social tags that express opinions and qualities of the items. Some examples of these tags are `happy`, `sunny` and `contemporary art`.
- **Organisational**. Social tags that define personal usages and tasks, or indicate self-references. Examples within this category are `to look at`, `scan from print`, `myself` and `our best friend`.

These tag categories are similar to those identified in the literature. Bischoff et al. [7] compare several categorisation schemas. Table 1 summarises this comparison, and includes our categorisation, which fits with previous schemas.

[Table 1 about here]

In contrast to previous studies, we attempt to automatically determine the most suitable category for a given social tag. Figure 1 depicts the whole categorisation process. Depending on the nature of the input tag, we distinguish two different cases. If the tag can be mapped[6] to a semantic concept of an external KB, then it will be assigned to either content- or context-based categories (stages 1a, 2a and 3a in the figure). In this case, we assume a semantic concept corresponds to a physical or non-physical entity related to content or contextual information of an item: objects, living entities, locations, time references, etc. On the other hand, if the tag is not found in the available KBs, we employ Natural Language Processing (NLP) techniques and categorisation heuristics to determine whether the tag can be assigned to subjective or organisational categories (stages 1b, 2b and 3b in the figure). In the following, we briefly describe the above cases. More details are given in Section 4.

[6] In Section 4.2, we explain in detail how a tag is mapped to a semantic concept of an external KB. At this point, the reader is asked to assume the existence of an automatic mechanism that links a tag with "names" of taxonomy categories, ontology classes or instances, etc. available in the KB.

[Figure 1 about here]

Content-based and context-based categories

In principle, social tags belonging to content- and context-based categories are nouns denoting physical and non-physical entities whose definition can be found in dictionaries, encyclopaedias or thesauri. Thus, the first step is to process and map an input tag to a concept existing in a KB (stage 1a in Figure 1). Let us suppose that the input tag is `nyc`, which is the acronym for the city of New York, USA. Looking for this term in KBs, we could obtain references to semantic entities related to that concept. For example, in Wikipedia, New York city is identified by the URLs `http://en.wikipedia.org/wiki/NYC` and `http://en.wikipedia.org/wiki/New_York_City`, among others. Let us say the identified concept for `nyc` is `[New_York_city]`.

Once we have established the semantic concept underlying a social tag, and assuming the existence of taxonomic relations among concepts in the KB, we propose to exploit such relations expanding the concept towards its taxonomic ancestors until reaching a "reference" ancestor (stage 2a in Figure 1), which allows us to later categorise the concept as a content- or a context-based tag. In Section 4, we present and justify the considered "reference" concepts. Here, continuing with the example, we just mention that the concept `[New_York_city]` is an instance of the class `USA_cities`, and that expanding `USA_cities` we might find out that `[New_York_city]` also belongs to the classes `New_York_state_citites`, Cities and Locations. In Wikipedia, this kind of classification is given by its "Wiki categories"[7].

In the example, the semantic expansion is stopped at `Locations` class. This is a reference concept since it is uniquely associated to the `context-based` category (stage 3a in Figure 1).

Subjective and organisational categories

When a social tag is not a noun, we apply NLP techniques and a number of categorisation heuristics to determine whether it can be categorised as a subjective or an organisational tag.

The first step is to tokenise the tags, and determine the part of speech (PoS), i.e. noun, verb, adjective, adverb, preposition, etc., of each of the obtained tokens (stage 1b in Figure 1). For example, let us suppose that the input tag is `to_read`. This is tokenised as (`to`, `read`), and transformed into the tuple (`to <preposition>`, `read <verb>`).

---

[7] Wikipedia categories, http://en.wikipedia.org/wiki/Special/Categories

Next, we analyse the PoS tuple in order to find a subset of tokens that satisfies one of a set of patterns predefined for each category. In the example, the pattern `[<preposition> + <verb>]` may represent a task (stage 2b in Figure 1).

Finally, through a heuristic approach, we assign the found pattern to a category (stage 3b in Figure 1). Continuing with the example, a `task` is assigned to the `organisational` category. In the next section, we will describe our tag categorisation approach in more detail.

4 Categorisation of social tags

4.1 Tag categories

For each of the four purpose-based categories proposed in this work, we define a set of subcategories encompassing the types of entities that can be assigned to those categories. Table 2 shows these subcategories and examples of Flickr social tags automatically categorised by our approach.

In contents, we find physical and non-physical entities. As physical entities, we do have artefacts and living entities. Living entities can be split into animals and plants. Persons are considered as animals, and similarly, organisations are non-physical entities. Related to the context of an item, we find location and time entities. Within the subjective category, we distinguish between personal opinions and qualities. Finally, organisational entities are divided into self-references, tasks and actions.

Instead of directly assigning a social tag to a purpose-based category, we firstly identify the most suitable subcategory, and then obtain the corresponding main category. We detail this process in Sections 4.2 and 4.3.

[Table 2 about here]

4.2 Categorising content- and context-based social tags

The categorisation of content- and context-based tags is based on their mapping to semantic concepts described in an external KB. A major requirement imposed on the KB is that it has to provide a classification hierarchy (taxonomy) among its concepts. As described in Section 3, given a social tag, we should not only be able to map it to a semantic concept, but also to determine one of its taxonomic ancestors that would allow us to assign the tag to either the content-based category or the context-based category. The tag `nyc` (New York city) is not just a `city`, but also a `location`, which is a contextual entity.

In this context, WordNet [24], being a lexical database for the English language commonly exploited to support automatic text analysis and artificial intelligence applications, could be a good candidate for our categorisation goals. However, we are also interested in a repository flexible and capable of incorporating new concepts appearing in our daily lives (e.g., `web 2.0`, `podcast`, `diy`). Wikipedia, which is an online encyclopaedia updated collaboratively by the community, represents one of our best alternatives.

YAGO [34] is an ontological KB – supported by the W3C Linking Open Data (LOD) initiative[8] – which contains information harvested from WordNet and Wikipedia. As of September 2009, YAGO knows over 2 million entities such as people, organisations, cities, etc., and about 20 million facts about these entities. Among other semantic relations, it provides a multi-domain knowledge classification where the upper-level classes are the concepts existing in WordNet, and the lower-level classes correspond to Wikipedia categories. YAGO thus can be understood as a taxonomy divided into two parts: one invariant part which covers generic concepts that can be found in a dictionary, and other open part that extends the former with more specific classes about concepts not existing in a dictionary, but in an updated modern encyclopaedia.

The LOD project aims to extend the Web with a data commons by publishing open data sets as RDF[9] on the Web, and by setting RDF links between data items from different data sources. These RDF links can be for example followed by crawlers and Semantic Web search engines to provide sophisticated search and query capabilities over crawled data. Figure 2 shows that YAGO ontology forms part of the LOD repository, and directly links to WordNet and DBpedia [5], a huge knowledge base containing structured information extracted from Wikipedia.

[Figure 2 about here]

We propose to use YAGO in our social tag categorisation approach, as a way of bridging the gap between folksonomies (understood as non-interlinked social tag sets) and the Semantic Web (under the perspective of the LOD project). Figure 3 shows a portion of YAGO taxonomy. The coloured classes are reference concepts linked to content- and context-based (sub) categories, as explained in the following.

[Figure 3 about here]

To categorise a social tag, the first step is to map it to a YAGO class (stage 1a in Figure 1). In many cases, a direct mapping between the tag and the names of an ontology class is not possible. People add tags in singular and plural forms indistinctively, use morphological derivations (e.g., acronyms), make misspellings, and include compound nouns with different separation characters (e.g., `NewYork`, `New York`, `New_York`). For this reason, we perform a number of morphological transformations of the tags

8 Linking Open Data project, http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

9 Resource Description Framework (RDF), http://www.w3.org/RDF

obtaining a set of equivalent terms to be matched with YAGO class names. First, we apply stemming and stop word removal techniques to the tags. We also use an efficient algorithm that makes use of Google[10] "did you mean" functionality to split compound nouns with no spaces, and correct misspellings. All these tag processing and filtering techniques are presented in previous works [36][37]. Without going into details, we just give a couple of example transformations. The social tag newyork is transformed into the terms {`newyork`, `Newyork`, `NEWYORK`, `new_york`, `New_york`, `NEW_YORK`, `New_York`}. The social tag `barclona` is transformed into {`barclona`, `barcelona`, `Barcelona`, `BARCELONA`}. The generated tag derivations are then searched as names of the YAGO classes (see an example in Figure 4).

If a mapping is found, we proceed with the search of a reference YAGO class that uniquely identifies a content- or context-based subcategory (stage 2a in Figure 1). To do this, we recursively obtain the ancestor classes of the mapped concept, stopping when reaching a predefined reference class (one of those shown in Table 3). For example, let `Seagulls` be the social tag to categorise. After transforming it into different morphologic derivations, let us assume that we find a matching with the concept `seagull`, which belongs to the Wikipedia part of YAGO. Then, we extend that concept through its ancestors, and obtain concepts such as `larid`, `coastal diving bird`, `seabird`, …, `bird`, …, `animal`.

Finally, the linking between a semantic subcategory (reference YAGO class) and a purpose-oriented category is direct (stage 3a in Figure 1). The social tag `Seagulls` is an `animal`, and therefore has to be categorised as a content-based tag.

[Table 3 about here]

It is important to note that there is no overlap between reference classes, i.e. the hierarchy branches of content- and context-based are disjoint. Moreover, although the hierarchy branches of content-based subcategories intersect (e.g., a `person` is an `animal`, an `animal` is a `living entity`, and a `living entity` is a `physical entity`), our algorithm takes into account the level in which a mapping occurs. The lower-level reference classes are preferred to the upper-level ones (e.g., `Brad Pitt` is categorised as `person`, and not as `animal`, `living entity` or `physical entity`).

Of course, there are ambiguity problems when a tag-concept mapping is done, as shown in Figure 4 for the tag `java`.

[Figure 4 about here]

[10] Google web search engine, http://www.google.com

Terms have different meanings, and only one should be chosen for the corresponding social tag. Here, we obtain and categorise all the possible concepts mapped to a tag, and do not perform any disambiguation technique. We plan to do it in future work, as discussed in Section 8.

4.3 Categorising subjective and organisational social tags

When no mapping between a social tag and a YAGO concept is found, we analyse whether the tag can be categorised as subjective or organisational.

We identify PoS of each term forming the tag using NLP techniques [2]. After removing some stop-words (e.g., conjunctions), we consider the tag as a tuple of PoS [PoS_1, ..., PoS_k], and compare it with a set of PoS-tuple patterns defined for the subjective and organisational subcategories. Below, we list the PoS patterns defined for the subcategories, which have to be interpreted as regular expressions. An asterisk (*) means any element, and the 'OR' operator is used when a subcategory is defined by several PoS patterns.

- **Opinions**: [<adjective>] OR [<adverb>] OR [*<pronoun>*<adjective>*]
- **Qualities**: [*<adjective><noun>*]
- **Self-references**: [pronoun] OR [*<pronoun>*<noun>*] OR [*<verb>*<pronoun>*]
- **Tasks**: [*<preposition><verb>*]
- **Actions**: [<verb>]

If there is a match between the tag PoS-tuple and a subcategory PoS-pattern, then we categorise the tag with that subcategory. Otherwise, we do not categorise the tag, and discard it.

Some refinements may be done in these categorisation heuristics. For example, some tags categorised as qualities may be considered as content-based if we discard the adjectives: the tag big house could be transformed to the tag house.

Moreover, there may exist incorrect tag assignments within the subjective subcategories. For example, the tag bad hotel is categorised by our approach as a quality tag as it satisfies the [*<adjective><noun>*] regular expression, whereas it should be categorised as an opinion tag.

These issues have to be carefully studied in the future.

5 Recommendation algorithm

In general, recommender systems aim to provide personalised recommendations of items to users based on their previous behaviour as well as on other information gathered by item descriptions and user profiles. In particular, given the success of item recommendation in commercial websites, such as

Amazon.com and Netflix.com, it is considered worthwhile to exploit and evaluate our tag categorisation technique via the recommendation problem, using a graph-based algorithm, namely Random Walk with Restarts (RWR) [20].

Yildirim and Krishnamoorthy [43] propose a novel recommendation algorithm which performs Random Walks on a graph that denotes similarity measures between items. They evaluate their system using movie rating data from MovieLens. Liu et al. [19] propose a collaborative filtering approach that models user preferences derived from the ratings, by measuring the correlation between their rankings of the items rather than the rating values, using a Random Walk model. They evaluate their method on EachMovie and NetFlix data. Konstas et al. [17] perform RWR on a social graph using friendships and social tagging information, captured from the social network Last.fm. In these studies, Random Walk models are used based on their superiority to other collaborative filtering methods. As a result, we also opted to perform evaluation using an equivalent Random Walk model.

5.1 Random walk with restarts

A graph is a natural representation of data with some inherent relational structure. In a graph, objects and their relationships can be represented as nodes and weighted edges respectively, where weights denote the strength of the relationships. This abstraction allows us to integrate heterogeneous sources of data in a principled manner.

Measuring the relatedness of two nodes in the graph can be achieved using the Random Walks with Restarts (RWR) theory [20]. Starting from a node x, a RWR is performed by randomly following a link to another node at each step. Additionally, in every step, there is a probability α to restart at x. Let $\mathbf{p}^{(t)}$ be a column vector where $p_i^{(t)}$ denotes the probability that the random walk at step t is at node i. $\mathbf{q}$ is a column vector of zeros with the element corresponding to the starting node set to 1, i.e. $q_x = 1$. Also let $\mathbf{S}$ be the column normalised adjacency matrix of the graph. In other words, $\mathbf{S}$ is the transition probability table where its elements $S_{i,j}$ give the probability of j being the next state given that the current state is i. The stationary, or steady-state, probabilities for each node can be obtained by recursively applying (1) until convergence,

$$\mathbf{p}^{(t+1)} = (1-\alpha)\,\mathbf{S}\,\mathbf{p}^{(t)} + \alpha\,\mathbf{q} \tag{1}$$

where $\alpha \in [0,1]$.

The stationary probabilities give us the long term visit rate of each node given a bias towards a particular starting node. Therefore, $p_i^{(l)}$, where l is the state after convergence, can be considered as a measure of relatedness between nodes x and i.

5.2 Social graph

Random Walks with Restarts has recently attracted the interest of researchers in many different areas within Information Retrieval, starting from link analysis [25] to image annotation and retrieval [26][39], text classification [42], click-through data analysis [10] and collaborative recommendations [13].

In this study, we aim to evaluate the effect of tag categorisation in the context of folksonomy-based item recommendation using data crawled from the multi-topic social networking system of Flickr. RWR allows us to directly predict the preference of users to particular photos (to which we shall refer to as items from now on) from the data collection acquired, by taking into account not only their personal profiles in terms of item preferences (**UI**) but also their tagging behaviour, social network as well as similarly tagged items.

Specifically, we create our social graph by representing users, items and tags as nodes. User relationships (**UU**) are encoded using either uni- or bi-directional edges between the corresponding nodes. Similarly, we add edges between items and tags (**ITg**) as well as users and tags (**UTg**). More details are given in a later section.

6 Experiments

6.1 Data collection

For the purposes of automatic tag categorisation and folksonomy-based item recommendation, we collected live data from the Flickr social network. Flickr is a photograph and short-video oriented social network that allows users to upload and share their media with other users. In order to evaluate our tag categorisation strategy, we decided to use a dataset from Flickr because this is a multi-domain social system. Our approach could also be tested on a single-domain repository such as Last.fm (music) or MovieLens (movies). In doing so, however, our conclusions may be biased, as long as we would not be covering a significant part of the knowledge base (WordNet and Wikipedia) when categorising social tags. Nonetheless, a comparison of tag categorisation and recommendation results obtained with single- and multi-domain datasets constitutes an interesting study. We postpone it as future work. In this paper, we focus on the study of the feasibility of our automatic tag categorisation proposal in a generic framework with a wide range of topics and domains.

The main characteristics of the data collection acquired could be summarised as follows:

- Our main focus was on photos (items) rather than short videos, since the former are more widely accepted by the Flickr community at the moment.

- We consider two types of relationship between users and items: Users can either own photos or become fans of them, i.e. add them to their favourite items' list.
- Tags in Flickr are categorised as social (i.e. not specified by experts), meaning that only the owner of an item can annotate (tag) it.
- Items can be grouped by their owner into item-sets, based on his/her personal notion of similarity among them.
- Friendships are essentially uni-directional, established when a user decides to add another user into his contact list (similar to followers in micro-blogs like Twitter[11]). Along with uni-directional friendships, we also consider bi-directional bonds of friendship, defining them as mutual uni-directional links between two users and regard them in our analysis differently.

We extracted a representative portion of the Flickr social network comprising of 3223 users, 240648 items and 190105 tags, which are freely available[12] for other researchers.

This initial dataset was exposed to a three part filtering process. For the first part, our intention was to collect users who have intense activity in their profiles. This can be interpreted as the fact that they have popular items in their profiles, which means lots of contacts that might link to them both through friendship establishment and becoming fans of their items.

As a result, we collected the 500 items that were characterised as the most interesting on the 1st of January 2009. For these items, we extracted their owners' profiles, composed of their 150 top tags, the items tagged with these tags, and their contacts. Then, for each of the above items, we also obtained a list of 10 fans (the first 10 users provided by Flickr API). Similarly to the owners' profile obtaining step, for each fan we extracted the 150 top tags, the items tagged with these tags, and their contacts. In the previous steps, we discarded those users without tagged items.

For the second part, we aimed to maintain a dense graph on the user space. Firstly, we filtered all the gathered contacts, by discarding those contact relations which were not linking to owners and fans of most interesting items. To reduce the size of the item space, we discarded those items that were not favourite of at least two of the available users. Finally, we removed users with no tagged items, and tags not associated to an item existing in the collection at this point. Tags were also cut down if they were not used by at least two users to tag their items, and were exposed to a cleaning process (stemming, stop

[11] Twitter – Social messaging, http://www.twitter.com

[12] Flickr dataset, http://mir.dcs.gla.ac.uk/flickr

words removal, misspelling and compound noun processing, etc.), which is exemplified in Section 4.2, and explained in detail in previous works [36][37].

The third part dealt only with the tags collected so far. First of all, we machine translate them if necessary, using Google translate tool[13], which automatically detected the language of origin (if it was not English). Then, each tag was assigned one or more categories following the process described earlier. In case either of these two steps failed for a certain tag, then this was not included in the final set.

The outcome of this filtering process was a reduced data collection of 2022 users, 24263 items and 41742 tags (20055 content-based, 8300 context-based, 9013 subjective, and 4374 organisational[14]), with the derived sub-matrices' densities shown in Table 4, and which is as well made available at (blind for review).

6.2 Dataset

The combination of the $\mathbf{UU}$, $\mathbf{UI}$, plus $\mathbf{UTg}^{all}$, $\mathbf{ITg}^{all}$, $\mathbf{TgTg}^{all}$ (which from now on we shall refer to as the whole tag space) sub-matrices derived by the data collection method described above, resulted in the full social graph $\mathbf{S}$, shown in Figure 5.

[Figure 5 about here]

In our case, most of these sub-matrices are binary[15], either by definition (e.g., in the case of the **UI** sub-matrix, we only have evidence as of the ownership or preference of a user towards a certain item, which is essentially binary information; Same thing applies for the $\mathbf{ITg}^{all}$ sub-matrix; An item may or may not be associated with a single tag) or on purpose. Given the fact that the aforementioned sub-matrices were naturally binary, we opted to keep everything else bounded between 0 and 1.

More specifically, in the **UU** sub-matrix, we represent each uni-directional bond of friendship with the value of 0.5 and each bi-directional with 1. In this way, we tend to favour the latter category of friendships under the assumption that they are more concrete, similar to real life. What is more, bi-directional friendships are usually the norm in many other social networks such as Facebook[16] and

---

13 Google translate, http://translate.google.com

14 Note that due to ambiguity, a tag may belong to more than one category. See Section 6.5 for more details.

15 In Flickr system, a photo can only be tagged by one user (the owner), and thus the relation between a certain photo and each tag is binary. This is different to other tagging systems, where users can annotate any item, and thus a weight can be associated to an item-tag relation, in general based on the number of times the item has been annotated (by different users) with that particular tag.

16 Facebook – Social networking, http://www.facebook.com

Last.fm [17]. In the $\mathbf{UTg}^{all}$ sub-matrix, each edge is a normalised weight between 0 and 1, indicating the frequency of use of each tag for a specific user, extracted directly by the tag cloud in the profile of each user from the Flickr website. Finally, the $\mathbf{TgTg}^{all}$ sub-matrix is the item-based co-occurrence of tags, calculated using equation (2):

$$\mathbf{TgTg}^{all} = f\ ((\mathbf{ITg})^{\mathrm{T}} \cdot (\mathbf{ITg})), \tag{2}$$

where

$$f(a_{i,j}) = \begin{cases} a_{i,j} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}$$

with $a_{i,j}$ being the element of $\mathbf{TgTg}^{all}$ at the i^{th} and j^{th} position.

6.3 Evaluation protocol

We evaluated the tag categorisation technique under study using the RWR method in order to perform item recommendation, with either the whole tag space or parts of it, as will be clarified in the following section. In every experiment, we follow the same per user evaluation protocol adopted by [17] which we describe here briefly.

For each individual user in the dataset, we randomly select a list of 20% of the items he or she is only a fan of, and set zeros to the corresponding elements of **UI** and $\mathbf{UI}^{\mathrm{T}}$ sub-matrices. It should be noted, that it is important not to include owned items, since we do not want them to be recommended back to each user. What is more, as stated earlier it is the case in Flickr that only the owners can assign tags to their owned items. As a consequence, if we merely remove links in the **UI** sub-matrix then the recommendation would still be biased towards the owner's items unless we also remove the links in the **UTg** sub-matrix. However, this is no longer necessary because we are only including the items that each user is a fan of and thus there are no left links in the **UTg** sub-matrix.

We then create a query vector $\mathbf{q}$ so that $q_i = 1$ if $S_{u_a,i} > 0$, and $q_{u_a} = 1$, where $i = [1 \ldots N]$, $S_{u,i}$ is the *i*-th element of S corresponding to the *u*-th user, u_a is the current user under evaluation, and N is the total number of columns of **S**. Next, we normalise **q** so that $\|\mathbf{q}\|_1 = 1$.

Then, we perform a Random Walk on **S** which returns the stationary probability vector corresponding to u_a of all the items in the dataset. From this vector, we remove the remaining 80% of items user u_a was a fan of, so as to avoid recommending back items which were used while performing the RWR. We then re-order the remaining items in descending order, with the first element having been assigned the highest probability, denoting higher preference. This vector is pruned to contain the top 1000 items.

Note that in the case of the RWR method, we cannot directly compute the rating (or equivalently the predicted preference or not of a user for a certain item, as in our case with Flickr where ratings are considered binary) for each item as in standard collaborative filtering evaluation methodology. Instead we are only able to get a rank of the items in order of predicted preference. As a result, we are forced to evaluate the method as if it was being used in real-time purposes, i.e. judging both the predictive accuracy and precision of the system at the same time by using standard retrieval metrics such as Recall, MAP, and number of relevant retrieved items, similarly to [36].

6.4 Experimental setup

In this section, we illustrate the methodology for the individual recommendation experiments conducted using the RWR method. The outcome of each experiment was the aggregation of the lists containing the top 1000 items in descending order for each user, which were tested against the ground truth, i.e. the items owned or being indicated as favourite by a user.

As described in Section 5.1, there is a restart probability parameter α, which we set to the value of 0.8, so as to suppress the model to return to the initial query vector **q** more often and consequently perform random walks in the neighbouring elements of u_a, what was interpreted as stronger *personalisation* in [17]. After setting α, we then conducted two series of experiments, which aimed to validate our main hypothesis concerning tag categorisation.

The first series was preliminary using **UI**, the whole tag space and parts of **UU**, namely different versions of the latter sub-matrix that included only uni-directional friendships UU^{uni} (i.e. bi-directional bonds were considered of equal importance to uni-directional), only bi-directional friendships UU^{bi}, and both. The purpose of this series of experiments was to determine whether the inclusion of friendships were of benefit in the social graph, and thus be able to justify whether they should be included in the next series.

As will be shown and explained in the next sections, friendships were not found to be of significant importance and thus were not considered furthermore. As a result, the second series of experiments were conducted on a part of the social graph that contained **UI** and versions of $\mathbf{UTg}^{all}$, $\mathbf{ITg}^{all}$, $\mathbf{TgTg}^{all}$ with either content-based, context-based, subjective or organisational tags only. Four experiments were thus performed using all the tags in each category. However, taking into account that there was not enough balance as to the number of tags within each category (Section 6.1), we also performed another five experiments[17] with a reduced tag-set of 4374 (4K tags) in each category, i.e. the number of tags in the organisational category. The reduced tags were chosen randomly from the original tag-set. Finally, in

[17] One for each tag category, plus another being a subset of the whole tag space.

order to depict even more clearly that the performance of the recommendation is ascribed to the discriminative power of tags belonging to different categories, rather than the density of the graph, we performed a single experiment for the content-based tags only, that contained 8500 tags, i.e. of equal size to context-based and subjective tags.

6.5 Categorisation results

In this section, we present and discuss the results obtained by the categorisation of the social tags existing in the Flickr dataset described in Sections 6.1 and 6.2.

Table 5 shows the percentage of tags assigned to each purpose-oriented category before and after translating non-English tags. By translating and categorising these tags, we considerably reduced the set of tags that are not categorised from 40.8% to 32.4%. It is very important to highlight that the relative proportions of tags in each category are quite similar to those given in [7], where a non-noisy manual tag categorisation is carried out. This fact gives first insights about the validity of our proposal.

[Table 5 about here]

Going into more detail, Table 6 shows the percentage of tags assigned to each semantic subcategory.

[Table 6 about here]

The *entity* subcategory gathers those tags that were found in YAGO but were not classified in one of the content- and context-based subcategories. We realised that there are concepts that do have the entity concept as direct ancestor, so our approach could not categorise them correctly. Note that the real percentage of *physical entities* is 62.9% since artefact, living entity, animal, person and plant are subclasses of physical entity. An analogous situation happens with *living entities*, which represent the 22.1% of content-based tags. The small size of *time* subcategory is due to the fact that we discarded those tags containing numeric dates. Finally, we also notice that *non-physical entities* should have more subcategories. *Organisation* subcategory only represents the 4% of the total number of content-based tags.

Aiming to evaluate the accuracy of the proposed social tag categorisation mechanism, and provide preliminary quantitative insights about the influence of ambiguity on such mechanism, we conducted an empirical study based on manual evaluations of a significant number of tag assignments (categorisations). Specifically, 30 subjects, PhD students and research staff from our department, were recruited to evaluate the correctness of 3915 randomly selected tag assignments. This set of tag assignments represented 9.4% of the total number of tag assignments available in our dataset, and was built according to the categorisation percentages reported in Table 5. Each tag assignment was evaluated by 3 subjects, and was considered as *correct* if at least 2 of its evaluators judged it as correct. In general, there was a substantial agreement among subjects. Fleiss' Kappa coefficient [12] measuring evaluators' agreement was $\kappa = 0.78$ (a value $\kappa = 1$ means complete agreement). For each tag assignment evaluation, a subject was requested to state whether the tag was correctly assigned to 1) the corresponding subcategory, 2) to the ancestor category of such subcategory, or 3) at least to the main purposed-based category. To make such decisions, the evaluators were presented with the set of tags annotating the photo of each tag assignment to evaluate, so they were able to identify the semantic context (meaning) of the tag. Obtained accuracy results are shown in Table 7. In the following, we discuss these results for each subcategory.

[Table 7 about here]

Analysing wrong tag assignments to *entity* subcategory, we found out that there were many concepts in YAGO ontology that are directly linked to the generic WordNet "entity" class. Thus, for such concepts, our approach was not able to correctly infer their purpose-based categories, either from content- or context-based subcategories. This justifies the poor accuracy value (57.5%) obtained for that subcategory. A similar situation occurred for *physical entity* and *artefact* categories. We identified that many of the wrong tag assignments to these categories were due to the fact that YAGO ontology directly links specific concepts to the generic WordNet "physical entity" and "artefact" reference classes. In these cases, however, we saw that there were less wrong assignments between content- and context-based subcategories. Instead, there were many tags assigned to *physical entity* and *artefact* categories that should be categorised in one of their content-based subcategories. Hence, accuracy values for these subcategories were around 75%, much better than for *entity* subcategory. Moreover, we also found that most of the miscategorised physical entities corresponded to ambiguous tags whose categories may be difficult to discern between non-abstract and abstract classes: e.g., `nature`, `sky` and `beauty`. In many of these cases, evaluators stated that such tags were *entity* or *non-physical* instances. Our approach, on the contrary, identified different types of artefacts for those tags: books and magazines (e.g., `nature`), music bands, artists and songs (e.g., `sky`), and movies (e.g., `beauty`), to name a few.

Regarding social tags incorrectly categorised as *living entities*, evaluators highlighted two curious and unexpected issues. They discovered that some tags, such as `alabama`, `bomber`, and `ontario` were linked in some contexts to "living entity" class because these tags correspond to well known computer "viruses", while other tags were assumed to correspond to famous fictional characters: `alex` (the comic character), `sam` (the Olympic mascot), and `tails` (the video game character), among others. An accuracy of 67.8% was obtained for *living entity* subcategory. The accuracy values achieved for its descendant reference classes were diverse: 55.0% for *animals*, 75.0% for *people*, and 86.7% for *plants*. For *animal* subcategory, most of incorrect tag assignments were due to proper nouns of animals, e.g., `bubbles` (the chimpanzee), `liberty` (the dog), and `socks` (the cat). For *person* subcategory, we saw that there were many locations mapped to surnames. For example, `gorsky` is a popular Polish surname, but also corresponds to the name of several rural localities in Russia. Similarly, `shibuya` is a Japanese place name and surname. For *plant* subcategory, our approach was highly accurate. Nonetheless, it could not deal successfully with ambiguous tags such as `force` (wheat flake cereal vs. Physical quantity), `linden` (Tilia tree vs. place), and `victoria` (water lily flower vs. place).

Similarly to previous subcategories, the accuracy of *non-physical entities* (close to 80.0%) was affected by the ambiguity issue. For example, in some cases, `life` was categorised as *organisation* instead of *living entity* because of the pro-life organisation with that name; and `apartment` and `hole` were classified as *non-physical entities* because of the music bands with those names, instead of being categorised respectively as *location* and *physical entity*.

With respect to the assignments of *location* entities (accuracy of 75.6%), in general, the problem was clear. In many cases, location names correspond to nouns or adjectives, and depending on the context, they may be categorised incorrectly. Some examples are `black` (Alabama, USA), `white` (Georgia, USA), `sunset` (Utah, USA), and `wood` (Winsconsin, USA). *Time* entities, on the other hand, were easy to categorise, obtaining a very high accuracy (90.0%). In general, wrong time entities were associated to tags difficult to categorise depending on the context: e.g., `old`, `sunrise`, and `window`.

For *subjective entities*, we found out additional problems, which have to be addressed in the future. In our set of *opinion* tags (accuracy of 70.0%), we had a significant number of "objective" adjectives that do not necessarily express opinions of the taggers. Examples of these tags are: `long`, `wide`, `foggy`, `icy`, `rural`, and `urban`. Something similar happened for *quality* tags (accuracy of 82.5%). In certain contexts, there were adjectives that do not describe qualities of the corresponding nouns: e.g., `new year`, `good morning`, `pacific coast`, and `master photo`. Although we obtained an average categorisation accuracy of 90.0% for *subjective* tags, we plan to exploit SentiWordNet[18], a well known lexical resource for opinion mining, and investigate related works in Opinion Mining and Sentiment Analysis fields [27], to better identify the above tags, improve their categorisation, and evaluate their influence on item recommendation.

Regarding *organisational entities*, we encountered the following limitations. *Self-reference* tags were categorised correctly with a high accuracy of 85.7%. Nonetheless, evaluators found some interesting cases, such as `The Curious Case of Benjamin Button Get It` (related to the drama movie), and `The World Around Us` (related to the documentary television series), which are not self-references,

[18] SentiWordNet, http://sentiwordnet.isti.cnr.it

and may be useful in a recommendation application. A number of *task* tags were categorised incorrectly because of certain prepositions, e.g., `Dog at Play`, `Festival of Lights`, `Places to Relax`, and `Playing with Zoom`. It can be seen that some of these tags are not *tasks* but *actions*, which could explain why the categorisation accuracy of *task* tags increases from 60.0% to 76.7% for the *organisational* main purpose-based category. Finally, *action* tags (accuracy of 75.0%) showed a more difficult problem to deal with. In some cases, there were nouns that were identified as verbs, and thus they were incorrectly categorised: e.g., `comment`, `doodle`, `dress`, `email`, `gyp`, `ruffle`, and `scrawl`.

In summary, based on our empirical study, our social tag categorisation approach seems to work quite well, achieving an average accuracy of 80.8%. However, it has to be improved by addressing the ambiguity of the tags, and by dealing with other semantic aspects described above. Potential research lines in these directions are depicted in Section 8.

6.6 Recommendation results

Next, we discuss the results of the evaluation of our social tag categorisation technique using folksonomy-based recommendation. The baseline upon which all comparisons are made was considered the performance of the recommender using only the **UI** sub-matrix.

As explained in Section 6.4, the purpose of the first set of experiments was to determine whether or not to include friendships in the social graph. As shown in Table 8, the incorporation of **UU** or parts of it did not actually add to the performance of the system; in fact it scored lower precision in higher ranks than the baseline. This is an interesting and quite surprising finding since friendship might help collaborative filtering in finding neighbours of the active user. We believe friendships could be more useful for recommendation in other types of social systems, such as Facebook, which is oriented to manage social contacts. Flickr, in contrast, is focused on photo sharing. This may be the reason of our results.

[Table 8 about here]

The second set of experiments validated our hypothesis that there exist certain categories of tags that work better in the context of item recommendation. In Figure 6, it is clearly shown that both organisational and subjective tags are statistically significantly outperformed by content- and context-based tags. What is more, organisational tags exhibit a performance equivalent to the baseline, strengthening our idea that they are particularly noisy. Using either content- or context-based tags, it can be seen in Table 8 that they achieve MAP and number of relevant retrieved items equivalently as in the case of using the whole tag space. There is even a significant trend in the case of content-based tags to outperform the case of using the whole tag space.

It can of course be argued that the aforementioned results are a consequence of matrix density, since organisational tags are fewer than context-based and subjective tags, and considerably fewer than content-based tags. Figure 7 however shows that the picture is on average the same even if we use the same amount of 4K tags in each category. The sparsity naturally plays a role, in the sense that MAP has globally decreased, nevertheless content- and context-based tags are still superior to the rest two categories and have an equivalent performance to using the whole tag space. Another argument to explain the obtained performance results may be the fact that controlling the tag frequencies does not take into account the number of items evaluated for each tag category. This is not the case in our experiment, since the number of relevant items retrieved is almost the same for each 4K tag set, as shown in Table 8. By computing precision for different recall cut-off values we assure that the obtained performance differences are valid.

Finally, there is an inconsistency with content-based tags, compared to context-based and the whole tag space, between Figures 6 and 7. The reason is that we effectively remove a large subset of useful tags when going from the original tag space to the reduced 4K. This is validated when increasing the amount of content-based tags to 8500 (of same size to context-based and subjective), where the overall performance is increased (Table 8).

[Figure 6 about here]

[Figure 7 about here]

7 Discussion

Social tags not only describe content and context information of the annotated items, but also subjective and organisational aspects of the users. Based on this fact, we hypothesised that not all tags are useful for folksonomy-based recommendation, and claimed that a categorisation of tags based on the users' tagging purposes can help to discard non-relevant tags, and thus improve content retrieval processes. Aiming to validate the above hypothesis, we addressed three research questions.

First, in order to categorise social tags, their underlying concepts have to be understood. Is it possible to identify such concepts in a simple and generic way? We present an automatic mapping between social tags and semantic entities described in YAGO [34], an ontology that links WordNet and Wikipedia. We thus cover multiple domains as is the case of Flickr system, and cope with new vocabulary appearing in our daily lives. With the proposed technique, we were able to categorise 67.6% of the collected tags, which represents a considerable proportion of significant tags if we take into consideration the noisy and multi-linguistic nature of the dataset.

Second, once the social tags are mapped to semantic concepts, is it possible to assign these concepts to purpose-oriented categories? We propose an automatic mechanism that exploits the semantic concept relations given by YAGO to transform the tag concepts into semantic entities, which can uniquely be assigned to content- and context-based categories. For a representative dataset from Flickr system, the proportion of tags assigned to those categories (32.9% and 13.0% respectively, as shown in Table 5) is similar to those given in [7], where a non-noisy manual tag categorisation is presented. This supports the correctness of our automatic tag categorisation approach.

Third and finally, does the obtained categorisation really improve folksonomy-based recommendation? The analysis of the results show that the incorporation of content- and context-based tags instead of subjective and organisational improved the performance of the system. This can be accounted to the fact that the latter tend to be more susceptible to noise than the former. Essentially, tags in these categories may help users for self-organisation purposes, especially in the case of hundreds of items in one's account, or may express very personal opinions that do not apply to the rest users. As a result, it is expected that they might not be of help as in our case of collaborative recommendation; tags characterising individuality simply cannot extrapolate in the multi-user community scenario. On the other hand, content- and context-based tags could be considered as of more global usage than the rest and consequently of additional re-usability by more than one users. Adding to this their similar performance compared to the usage of the whole tag space, it can be argued that their use alone in the social graph can serve the purposes of folksonomy-based recommendation sufficiently.

Another interesting point to raise is the fact that content- and context-based tags are actually the most popular in our data collection. Therefore, this study actually shows that even performing a stemming of the long tail of the tag distribution which follows a power law may actually perform equivalently to using the whole tag space. In other words, keeping the most frequent tags, disregarding their semantic content, can be justified since their majority shall contain content- and context-based tags.

Of course it can be argued that tag categorisation could have been exploited more, i.e. we could have integrated this idea directly into the social graph. An idea would be to apply edges to tags belonging to the same category in the $\mathbf{TgTg}^{all}$ (instead of calculating co-occurrence), and thus be taken into account by the RWR during the recommendation process. Even though this idea could provide with a useful outcome, still our approach aims and proves that we can effectively reduce the tag space in a similar to filtering process, using either the content- or context-based tags in isolation, which has the additional benefit of efficiency.

8 Conclusions and future work

In this paper, we presented an approach to automatically categorise social tags based on the users' intentions, namely content-based, context-based, subjective and organisational. This technique maps tags to semantic concepts existing in the multi-domain YAGO ontology [34], which is a Semantic Web knowledge base with structured information extracted from WordNet and Wikipedia, and uniquely assigns them to content- and context-based categories. The identification of subjective and organisational tags is based on NLP and regular expression heuristics.

Our main goal was to study whether the distinction of tags in these categories is of benefit to folksonomy-based recommender systems. We executed a RWR recommendation algorithm [17] using sets of tags from different categories, and were able to show that content- and context-based tags are superior to subjective and organisational tags, achieving equivalent performance to that obtained using the whole tag space.

As a proof of concept of our tag categorization framework, we decided to use YAGO ontology as the KB whose semantic concepts are linked to multi-domain social tags. Since this KB covers WordNet and, more importantly, a significant part of DBpedia (Wikipedia), using it, we are able to map proper nouns (e.g., people, locations, organizations, events, etc.) and "contemporary" terms. Nonetheless, the utilization of a larger number of external knowledge bases would help missing less tag-concept mappings, and could be exploited for disambiguation purposes.

Semantic ambiguity of tags would be a consideration to deal with in the future. Recent works have addressed this problem. Au Yeung et al. [6] propose to cluster the items tagged by the users. Based on the obtained clusters, the authors find relations between tags, potentially useful for disambiguation purposes. Weinberger et al. [40] compute the Kullback-Leibler divergence [18] to measure the ambiguity between pairs of tags. In this work, we have not addressed this issue when categorising social tags based on their intention. We plan to study disambiguation strategies that take into account the "context" of a social tag within a user or item profile [21][31]. For example, let us assume that we retrieve the tag "java" from a user/item profile, and we have to decide whether it refers to the well known programming language or to the Indonesian island. Let us also assume that profile contains tags such as "computer", "technology", etc. Since "java" co-occurs with these tags when it refers to the programming language much more frequently than when it refers to the Indonesian island, we could state with a certain confidence that, in this case, the tag meaning correspond to the programming language.

Analysing our categorisation results, we found that, in most of the cases, ambiguities occurred with social tags classified into both content and context categories, especially in those cases where the social tags corresponded to locations. Thus, although it would be convenient to correctly disambiguate and classify such tags, the results obtained with our recommendation model are still valid as its most accurate recommendations were obtained exploiting content- and context-based tags. Ambiguities in subjective and organisational tags may occur but their influence in the recommendations is relatively much lower. Nonetheless, for recommendation purposes, we find very interesting the possibility of exploring sentiment analysis approaches to enhance our subjective and organisational tag categorisation strategy based on regular expressions. As discussed in the paper, there may exist incorrect tag assignments to subjective subcategories. For example, the tag `bad hotel` is categorised by our approach as a "quality" tag as it satisfies the `[*<adjective><noun>*]` regular expression, whereas it should be categorised as an "opinion" tag.

In a folksonomy-based item recommender system, a potential future research line is the incorporation and exploitation of relations between social tags that go beyond co-occurrence based similarities. The transformation of tags into ontology concepts allows us to infer and use semantic relations between these concepts for recommendation purposes. Synonym (e.g., `android` and `humanoid`, `funicular` and `cable railway`) and morphological similarities (e.g., `blog`, `blogs`, `blogging`) between concepts could be very useful to better identify related annotated items. Preliminary work in this direction has already been done [9].

Acknowledgements

This research was supported by the European Commission (SALERO, FP6-027122), by the Spanish Ministry of Science and Education (RIM3, TIN2008-06566-C04-02), and by the Spanish Ministry of Industry (i3media, CENIT-2007-1012).

References

[1] 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 Transactions on Knowledge and Data Engineering* 17 (6), pp. 734-749.

[2] Alfonseca, E., Moreno-Sandoval, A., Guirao, J. M., Ruiz-Casado, M. 2006. The Wraetlic NLP Suite. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC'06)*.

[3] Ames, M., Naaman, M. 2007. Why We Tag: Motivations for Annotation in Mobile and Online Media. In *Proceedings of the 25th ACM Conference on Human Factors in Computing Systems (CHI'07)*, pp. 971-980.

[4] Angeletou, S. 2008. Semantic Enrichment of Folksonomy Tag-spaces. In *Proceedings of the 7th International Semantic Web Conference (ISWC'08)*, pp. 889-894.

[5] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z. 2007. DBpedia: A Nucleus for a Web of Open Data. In *Proceedings of the 6th International Semantic Web Conference (ISWC'07)*, pp. 722-735.

[6] Au Yeung, C. M., Gibbins, N., Shadbolt, N. 2009. Contextualising Tags in Collaborative Tagging Systems. In *Proceedings of the 20th ACM Conference on Hypertext and Hypermedia (Hypertext'09)*, pp. 251-260.

[7] Bischoff, K., Firan, C. S., Nejdl, W., Paiu, R. 2008. Can All Tags Be Used for Search? In *Proceeding of the 17th ACM Conference on Information and Knowledge Management (CIKM'08)*, pp. 203-212.

[8] Cantador, I., Bellogín, A., Castells, P. 2008. A Multilayer Ontology-based Hybrid Recommendation Model. *AI Communications* 21(2-3), pp. 203-210.

[9] Cantador, I., Szomszor, M., Alani, H., Fernández, M., Castells, P. 2008. Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations. In *Proceedings of the 1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb'08)*, pp. 5-19,

[10] Craswell, N., Szummer, M. 2007. Random Walks on the Click Graph. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'07*), pp. 239-246.

[11] De Gemmis, M., Lops, P., Semeraro, G., Basile, P. 2008 Integrating Tags in a Semantic Content-based Recommender. *In Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys'08)*, pp. 163-170.

[12] Fleiss, J. L., Cohen, J. 1973. The Equivalence of Weighted Kappa and the Intraclass Correlation Coefficient as Metrics of Reliability. *Educational and Psychological Measurements* 33, pp. 613-619.

[13] Fouss, F., Pirotte, A., Renders, J. M., Saerens, M. 2007. Random-walk Computation of Similarities between Nodes of a Graph with Application to Collaborative Recommendation. *IEEE Transactions on Knowledge and Data Engineering* 19(3), pp. 355-369.

[14] Gemmell, J., Shepitsen, A., Mobasher, M., Burke, R. 2008. Personalization in Folksonomies Based on Tag Clustering. In *Proceedings of the 6th Workshop on Intelligent Techniques for Web Personalization and Recommender Systems*.

[15] Golder, S. A., Huberman, B. A. 2006. Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science* 32(2), pp. 198-208.

[16] Halpin, H., Robu, V., Shepherd, H. 2007. The Complex Dynamics of Collaborative Tagging. In *Proceedings of the 16th International Conference on World Wide Web (WWW'07)*, pp. 211-220.

[17] Konstas, I., Stathopoulos, V., Jose, J. M. 2009. On Social Networks and Collaborative Recommendation. In *Proceeding of the 32nd Annual ACM SIGIR Conference (SIGIR'09)*, pp. 195-202.

[18] Kullback, S., Leibler, R. A. 1951. On Information and Sufficiency. *The Annals of Mathematical Statistics* 22(1), pp. 79-86.

[19] Liu, N. N., Yang, Q. 2008. Eigenrank: a Ranking-oriented Approach to Collaborative Filtering. In *Proceeding of the 31st Annual ACM SIGIR Conference (SIGIR'08)*, pp. 83-90.

[20] Lovasz, L. 1996. Random Walks on Graphs: A Survey. *Combinatronics* 2, pp. 1-46.

[21] Markines, B., Cattuto, C., Menczer, F., Benz, D., Hotho, A., Stumme, G. 2009. Evaluating Similarity Measures for Emergent Semantics of Social Tagging. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*, pp. 641-650.

[22] Marlow, C., Naaman, M., Boyd, D., Davis, M. 2006. HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, ToRead. In *Proceedings of the 17th ACM Conference on Hypertext and Hypermedia (Hypertext'06)*, pp. 31-40.

[23] Mika, P. 2005. Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. *Journal of Web Semantics* 3(2-3), pp. 211-223.

[24] Miller, G. 1995. WordNet: A Lexical Database for English George. *Communications of the ACM* 38(11), pp. 39-41.

[25] Page, L., Brin, S., Motwani, R., Winograd, T. 1998. The PageRank Citation Ranking: Bringing Order to the Web. *Technical report, Stanford InfoLab*.

[26] Pan, J. Y., Yang, H. J., Faloutsos, C., Duygulu, P. 2004. GCap: Graph-based Automatic Image Captioning. In *Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, pp. 146-154.

[27] Pang, B., Lee, L. 2008. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval* 2(1-2), pp. 1-135.

[28] Plangprasopchok, A., Lerman, K. 2009. Constructing Folksonomies from User-Specificed Relations on Flickr. In *Proceedings of the 18th International Conference on World Wide Web (WWW'09)*, pp. 781-790.

[29] Quintarelli, E., Rosati, L., Resmini, A. 2007. FaceTag: Integrating Bottom-up and Top-down Classification in a Social Tagging System. In *Proceedings of the 8th Annual IA Summit*.

[30] Sen, S., Lam, S. K., Rashid, A. M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, M. F., Riedl, J. 2006. Tagging, Communities, Vocabulary, Evolution. In *Proceedings of the 20th ACM Conference on Computer Supported Cooperative Work (CSCW'06)*, pp. 181-190.

[31] Shepitsen, A., Gemmell, J., Mobasher, B., Burke, R. 2008. Personalized Recommendation in Social Tagging Systems using Hierarchical Clustering. In *Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys'08)*, pp. 259-266.

[32] Smith, G. 2004. Folksonomy: Social Classification. In *http://atomiq.org/archives/2004/08/folksonomy_social_classification.html*

[33] Specia, L., Motta, E. 2007. Integrating Folksonomies with the Semantic Web. In *Proceedings of the 4th European Semantic Web Conference (ESWC'07)*, pp. 624-639.

[34] Suchanek, F. M., Kasneci, G., Weikum, G. 2008. YAGO: A Large Ontology from Wikipedia and WordNet. *Journal of Web Semantics* 6(3), pp. 203-217.

[35] Suchanek, F. M., Vojnovic, M., Gunawardena, D. 2008. Social Tags: Meaning and Suggestions. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management (CIKM'08)*, pp. 223-232.

[36] Szomszor, M., Alani, H., Cantador, I., O'Hara, K., Shadbolt, N. R. 2008. Semantic Modelling of User Interests based on Cross-Folksonomy Analysis. In *Proceedings of the 7th International Semantic Web Conference (ISWC'08)*, pp. 632-648.

[37] Szomszor, M., Cantador, I., Alani. H. 2008. Correlating User Profiles from Multiple Folksonomies. In *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia (Hypertext'08)*, 33-42.

[38] Tso-Sutter, K. H., Marinho, L. B., Schmidt-Thieme, L. 2008. Tag-aware Recommender Systems by Fusion of Collaborative Filtering Algorithms. In *Proceedings of the 2008 ACM Symposium on Applied Computing (SAC'08)*, pp. 1995-1999.

[39] Urban, J., Jose, J. M. 2006. Adaptive Image Retrieval using a Graph Model for Semantic Feature Integration. In *Proceedings of the 8th ACM international Workshop on Multimedia Information Retrieval (MIR'06)*, pp. 117-126.

[40] Weinberger, K. Q., Slaney, M., Van Zwol, R. 2008. Resolving Tag Ambiguity. In *Proceedings of the 16th ACM International Conference on Multimedia (MM'08)*, pp. 111-120.

[41] Xu, Z., Fu, Y., Mao, J., Su, D. 2006. Towards the Semantic Web: Collaborative Tag Suggestions. In *Proceedings of the WWW'06 Collaborative Web Tagging Workshop*.

[42] Xu, Y., Yi, X., Zhang, C. 2006. A Random Walks Method for Text Classification. In *Proceedings of the 2006 SIAM International Conference on Data Mining (SDM'06)*, pp. 340-347.

[43] Yildirim, H., Krishnamoorthy, M. S. 2008. A Random Walk Method for Alleviating the Sparsity Problem in Collaborative Filtering. In *Proceedings of the 2nd ACM Conference on Recommender Systems (RecSys'08)*, pp. 131-138.

[44] Zhen, Y., Li, W., Yeung, D. 2009. TagiCoFi: Tag Informed Collaborative Filtering. In *Proceedings of the 3rd ACM Conference on Recommender Systems (RecSys'09)*, pp. 69-76.

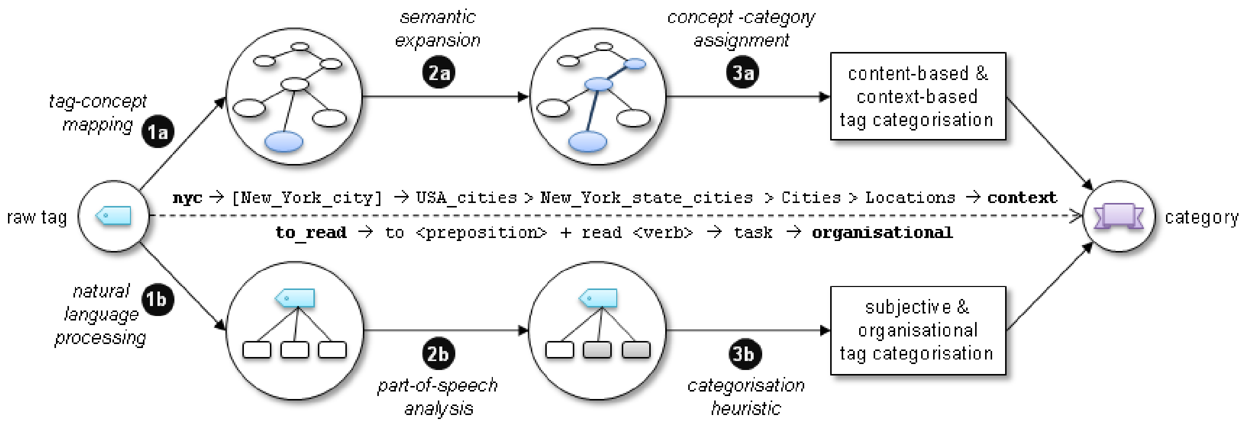

Fig. 1: Purpose-oriented categorisation of social tags

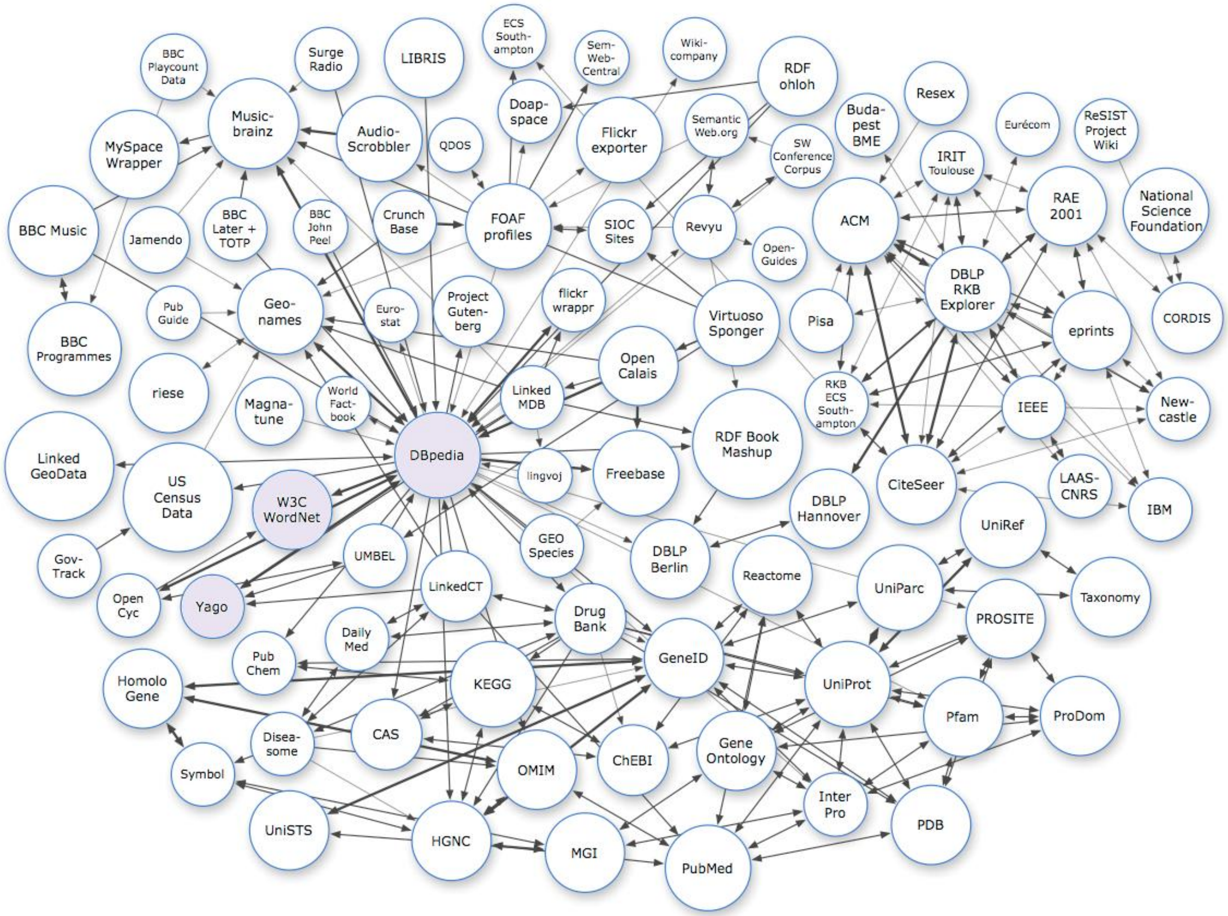

Fig. 2: Data sets published and interlinked by The W3C Linking Open Data project (July 2009). Coloured data sets represent the main semantic data sources exploited by our social tag categorisation proposal

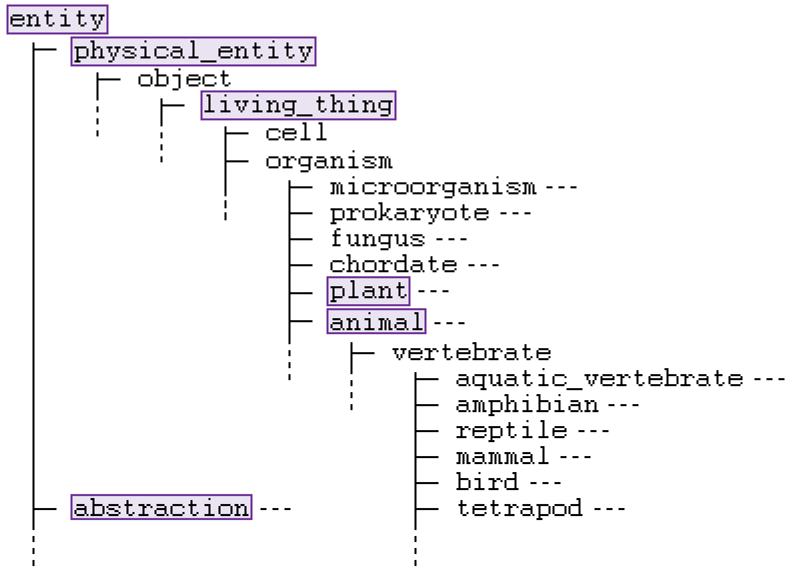

Fig. 3: A part of YAGO taxonomy. Coloured classes are reference concepts mapped to our purpose-oriented categories

Subcategory: animal

Java_(chicken)

wikicategory_Chicken_breeds

Subcategory: location

Java,_Georgia

wikicategory_Cities,_towns_and_villages_in_Georgia

Java,_South_Dakota

wikicategory_Towns_in_South_Dakota

Java,_New_York

wikicategory_Towns_in_New_York

Java,_(island)

wikicategory_Islands_of_Indonesia

Subcategory: non-physical

Java_(band)

wikicategory_French_hip_hop_groups

Java_(board_game)

wikicategory_Economic_simulation_board_games

Java_(programming_language)

wikicategory_Java_specification_requests

Subcategory: person

Java_(actor)

wikicategory_Film_actors

Fig. 4: Semantic subcategories, concepts and YAGO classes associated to the social tag `java`

	Users	Items	Tags
Users	UU	UI	UTg^{all}
Items	UI^{T}	0	ITg^{all}
Tags	$(\mathrm{UTg}^{all})^{\mathrm{T}}$	$(\mathrm{ITg}^{all})^{\mathrm{T}}$	TgTg^{all}

Fig. 5: Social graph **S** and the comprising sub-matrices

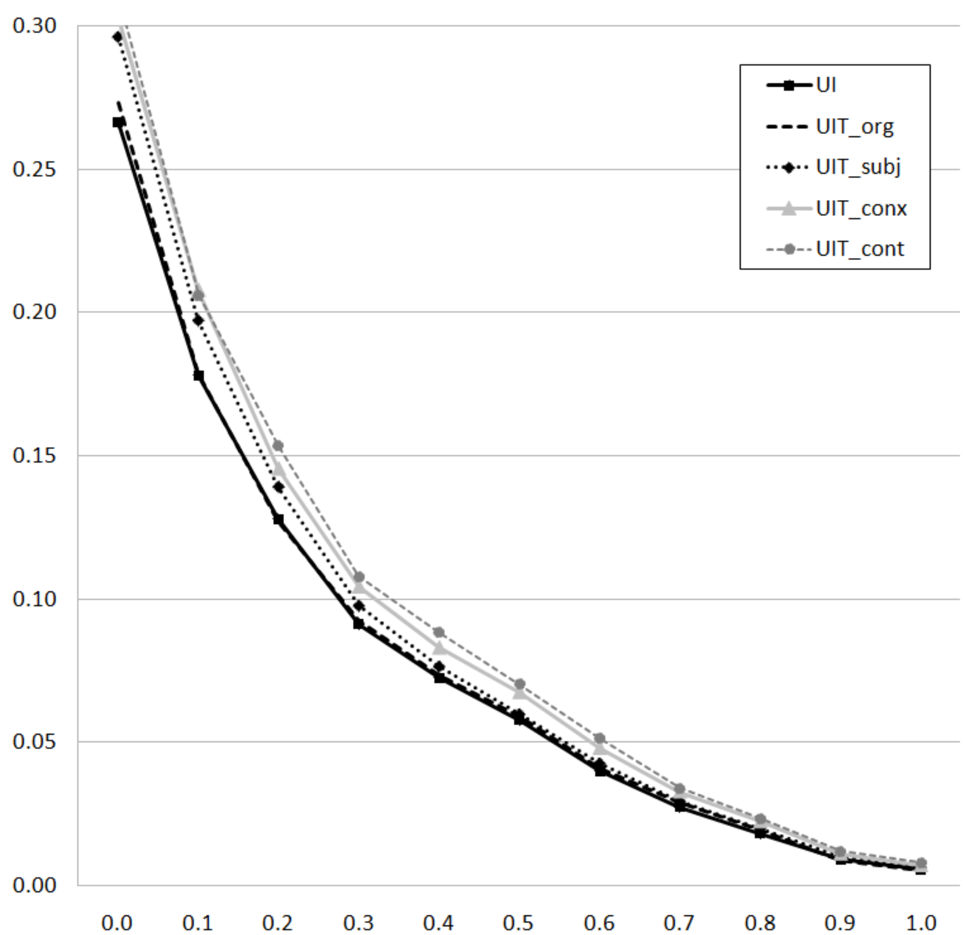

Fig. 6: Interpolated Recall-Precision Curves for experiments using either the whole or part of the tag space

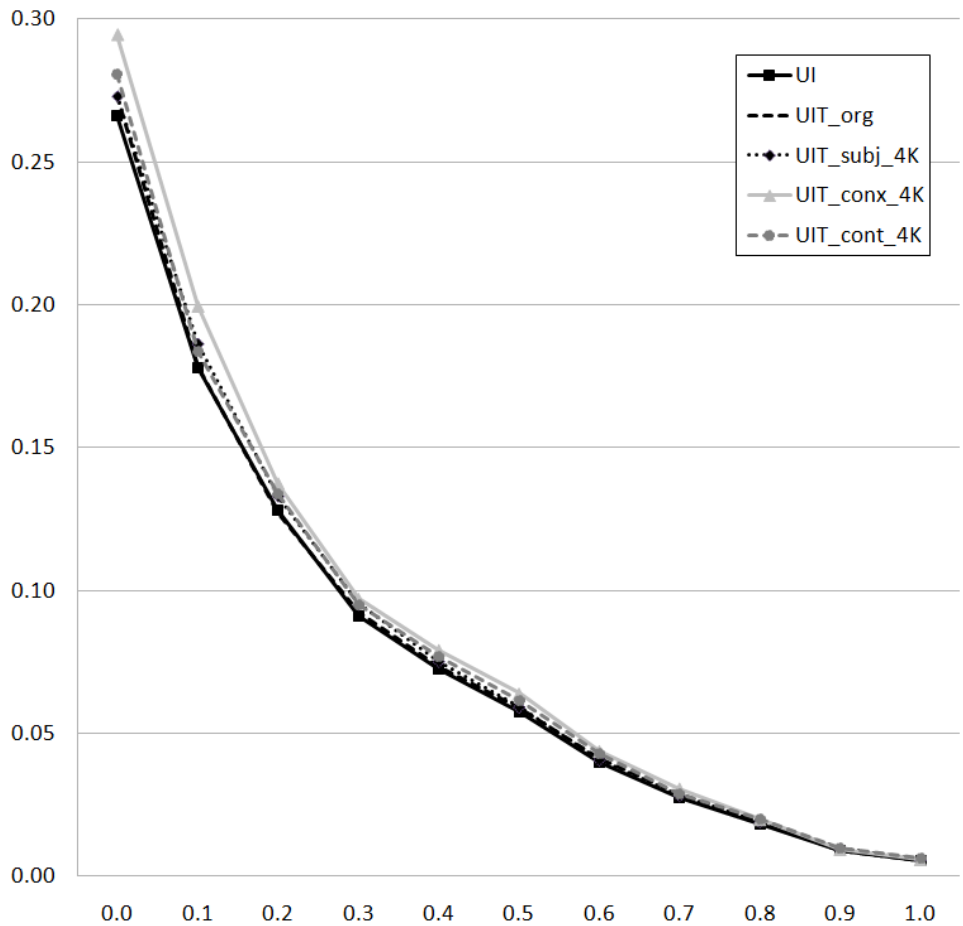

Fig. 7: Interpolated Recall-Precision Curves for experiments using the reduced 4K tag space

Our categories	Xu et al. [41]	Sen et al. [30]	Golder et al. [15]	Bischoff et al. [7]
Content-based	Content-based	Factual	What or who is about	Topic
	Attribute		What it is	Type
			Who owns it	Author/owner
Context-based	Context-based		Refining other categories	Time
				Location
Subjective	Subjective	Subjective	Qualities/ characteristics	Opinions/qualities
Organisational	Organisational	Personal	Task organisation	Usage context
			Self reference	Self reference

Table 1: Comparison of purpose-based categorisation of social tags, adapted from [7]

Category	Subcategory	Flickr tag examples
Content-based	*Physical entity*	`food, glue, heart, ice`
	└─ *Artefact*	`comb, finger, helicopter, table`
	└─ *Living entity*	`cell, clone, life, mushroom`
	└─ *Animal*	`caterpillar, frog, pigeon, pet`
	└─ *Person*	`boy, daniel, friend, sister`
	└─ *Plant*	`cactus, cereal flower, tree`
	Non-physical entity	`cloud, feminism, noise, tennis`
	└─ *Organisation*	`bmw, ibm, religion, rolling stones`
Context-based	*Location*	`california, rome, spain, wedding`
	Time	`halloween, march, sixties, winter`
Subjective	*Opinion*	`oh damn, so cute, unforgettable`
	Quality	`golden picture, geometric elegance`
Organisational	*Self-reference*	`i love you, her, missing you`
	Task	`time for change,  do not want to know`
	Action	`avoid, hiking, explore page, sit`

Table 2: Proposed purpose-oriented categories and semantic subcategories, with examples of real Flickr tags categorisations

Category	Subcategory	YAGO reference classes
Content-based	*Physical entity*	physical_entity
	└─ *Artefact*	artifact
	└─ *Living entity*	living_thing, life_form, live_body
	└─ *Animal*	animal
	└─ *Person*	person, human_body, kin
	└─ *Plant*	plant, plant_part
	Non-physical entity	abstraction
	└─ *Organisation*	organization
Context-based	*Location*	location, land, geological_formation, social_group
	Time	time, time_interval, time_period, time_unit

Table 3: YAGO reference classes associated to the considered content- and context-based subcategories

	All tags	4K tags
User-User (**UU**)	$6.1 \cdot 10^{-3}$	
User-Item (**UI**)	$2.2 \cdot 10^{-4}$	
User-Tag (**UTg^{all}**)	$5.7 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$
Item-Tag (**ITg^{all}**)	$6.5 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$
Tag-Tag (**TgTg^{all}**)	$2.2 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$
Social graph (**S**)	$2.2 \cdot 10^{-4}$	$2.2 \cdot 10^{-4}$

Table 4: Densities of the sub-matrices that comprise the social graph S, in the case of the full tag space, and in the case of random selection of 4K tags (see Section 6.6)

	Without translation	**With translation**
Content-based	27.8	32.9
Context-based	11.3	13.0
Subjective	13.4	14.6
Organisational	6.7	7.2
Unknown	40.8	32.4

Table 5: Percentages of social tags assigned to each purpose-oriented category (before and after non-English term translations). “Unknown” gathers those social tags that were not assigned to any category

Content-based		Context-based	
Physical entity	14.5	*Location*	92.7
└─ *Artefact*	26.3	*Time*	7.3
└─ *Living entity*	6.1	**Subjective**	
└─ *Animal*	2.4	*Opinion*	53.0
└─ *Person*	3.6	*Quality*	47.0
└─ *Plant*	9.9	**Organisational**	
Non-physical entity	31.5	*Self-reference*	79.8
└─ *Organisation*	0.4	*Task*	4.0
Entity	5.3	*Action*	16.2

Table 6: Percentages of social tags assigned to each semantic subcategory (with respect to the total number of tags in the corresponding purpose-based category)

Main category	***Subcategory***	***Category***	***#evaluated assignments***	***Accuracy (subcategory)***	***Accuracy (category)***	***Accuracy (main category)***
Content-based	*Physical entity*	-	296	55.0%	-	77.5%
	Artefact	*Physical entity*	537	70.0%	72.5%	72.5%
	Living entity	*Physical entity*	124	67.8%	75.0%	82.1%
	Animal	*Living entity*	49	55.0%	55.0%	65.0%
	Person	*Living entity*	73	75.0%	75.0%	83.3%
	Plant	*Living entity*	202	86.7%	88.4%	96.0%
	Non-physical entity	-	643	63.1%	-	84.4%
	Organisation	*Non-physical entity*	8	85.0%	95.0%	96.5%
	Entity	-	108	57.5%	-	57.5%
Context-based	*Location*	-	723	75.6%	-	75.6%
	Time	-	57	90.0%	-	90.0%
Subjective	*Opinion*	-	461	70.0%	-	97.5%
	Quality	-	409	82.5%	-	82.5%
Organisational	*Self-reference*	-	180	85.7%	-	91.4%
	Task	-	9	60.0%	-	66.7%
	Action	-	36	75.0%	-	75.0%
				72.1%	**76.8%**	**80.8%**

Table 7: Average accuracy results of the proposed social tag categorisation strategy

	UI	**UIT org**	**UIT subj**	**UIT conx**	**UIT cont**	**UIT cont 8K**	**UIT all**
MAP	0.0681	0.0692	**0.0745***	**0.0797**†	**0.0833***	**0.0774**	**0.0837**
#relevant	11680	**11793***	**12062***	**12110**	**12391***	**12124**	**12356**

	UIT subj 4K	**UIT conx 4K**	**UIT cont 4K**	**UIT all 4K**	**UIF**	**UIF uni**	**UIF bi**
MAP	0.0714	**0.0756**†	**0.0720***	**0.0738***	0.0478	0.0485	0.0493
#relevant	*11884*†	**11997**	*11880*†	**12006**†	11298	11294	11305

Table 8: Results of the experiments. The 4K tag sets are compared to UIT org. Bold typeset indicates statistical significance at p < 0.001 compared to the baseline UI model. Italic typeset indicates statistical significance at p < 0.05 compared to the baseline UIT org model. * and † indicate statistical significance at p < 0.001 and p < 0.05 respectively between consecutive experiments. In all cases, a two-sample, two-tailed t-test was used

Figure01

Click here to download high resolution image

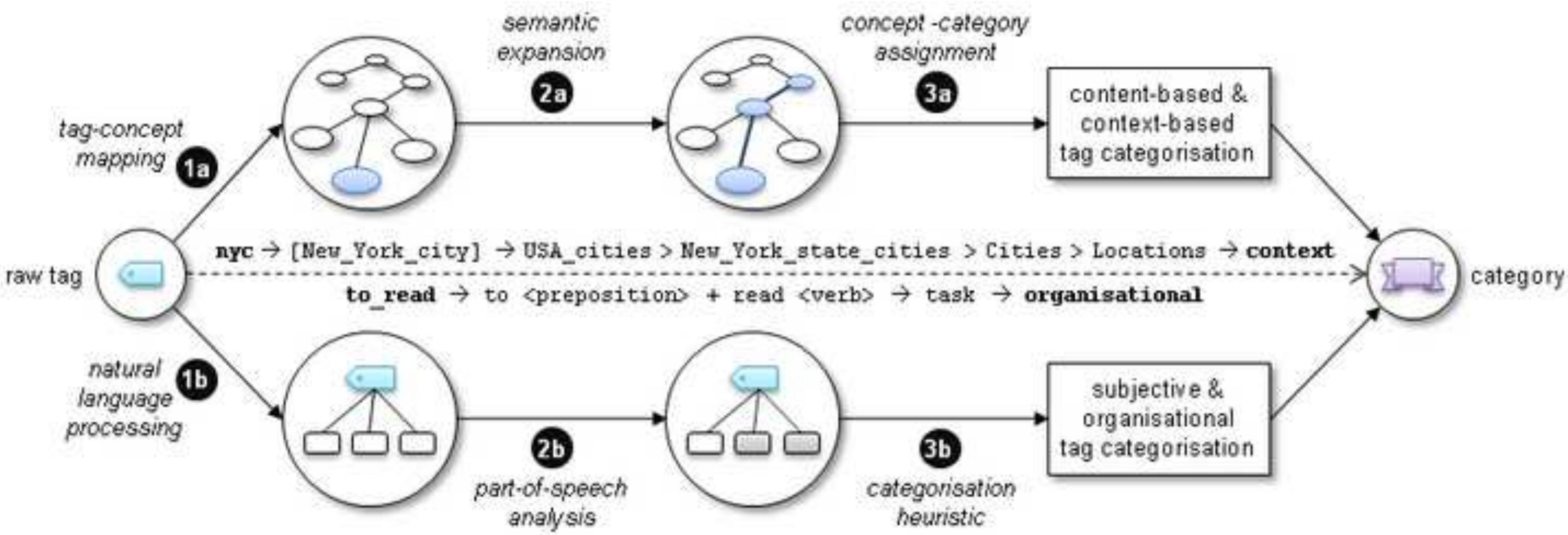

Figure02

Click here to download high resolution image

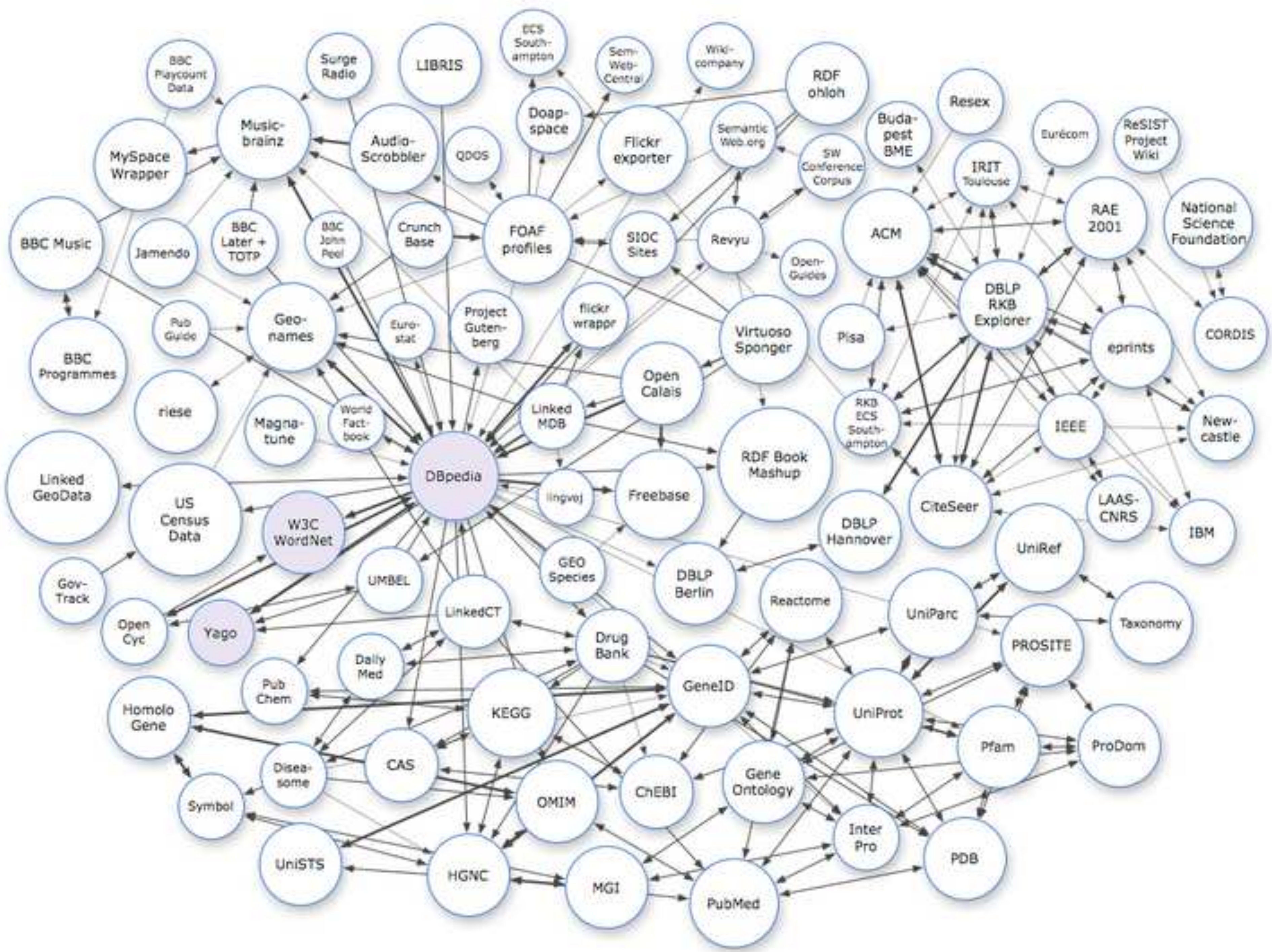

Figure03

Click here to download high resolution image

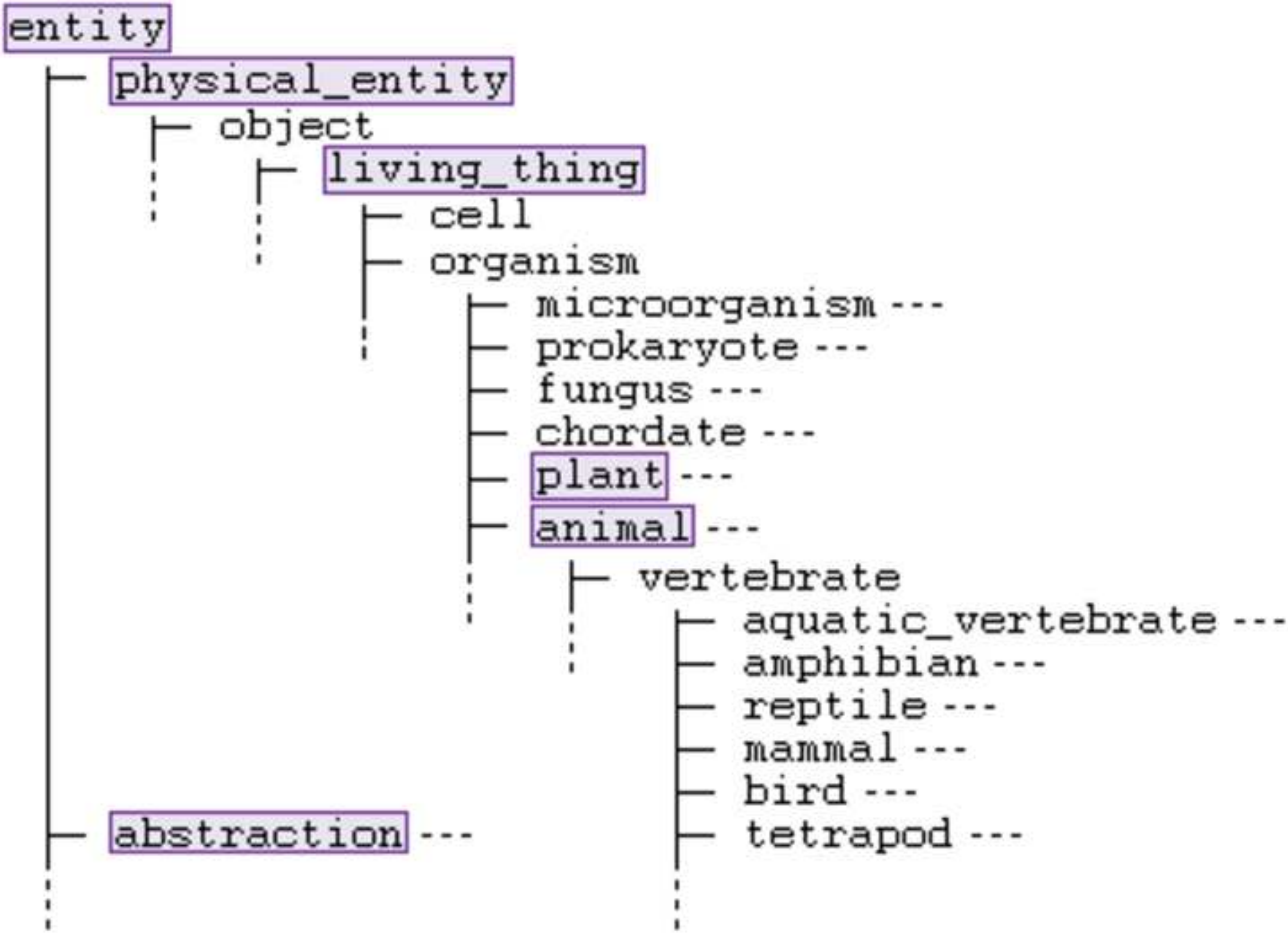

Figure05
Click here to download high resolution image

	Users	Items	Tags
Users	UU	UI	UTg^{all}
Items	UI^{T}	0	ITg^{all}
Tags	$(UTg^{all})^{T}$	$(ITg^{all})^{T}$	$TgTg^{all}$

Figure06
Click here to download high resolution image

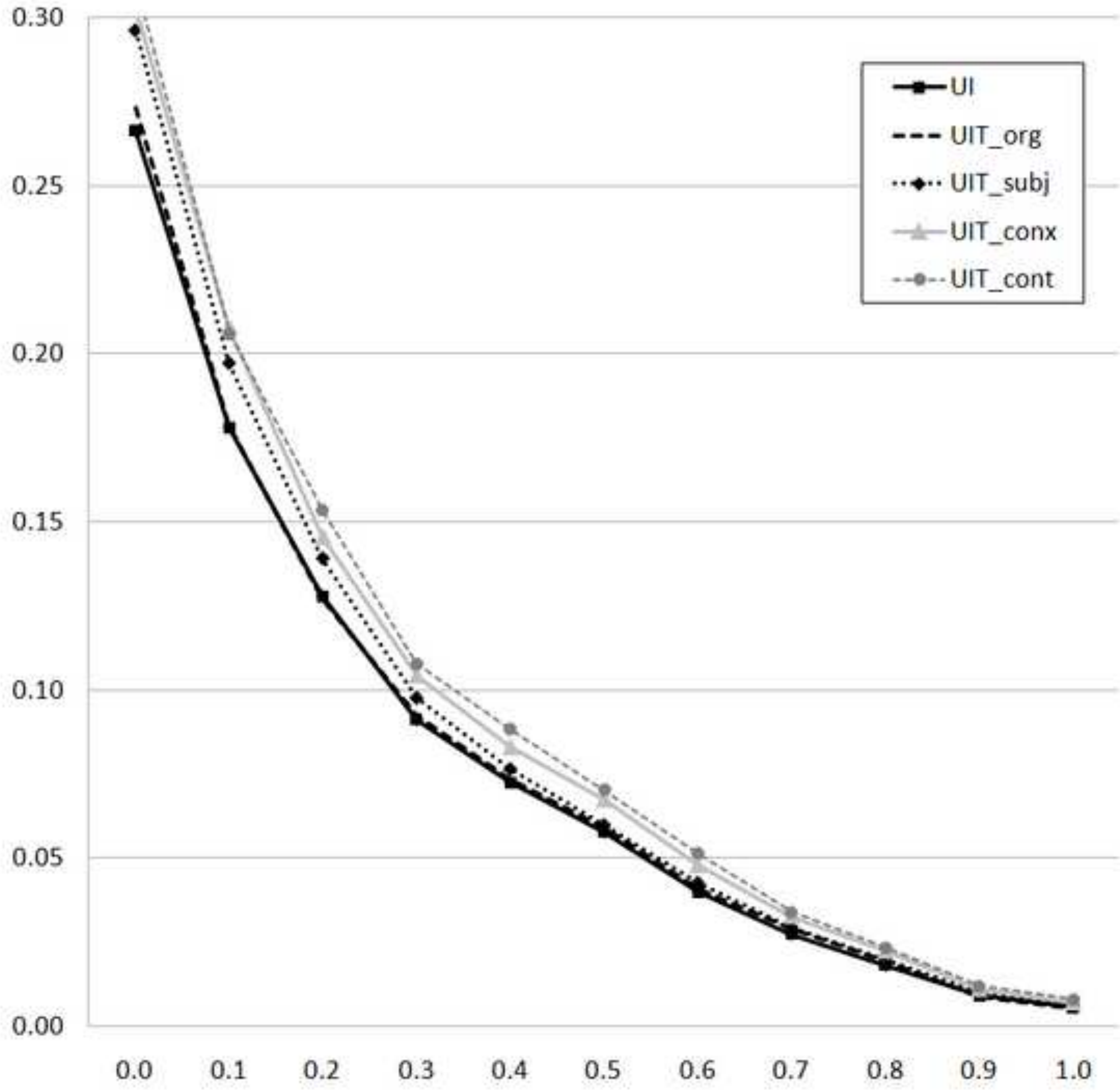

Figure07
Click here to download high resolution image