

Improving on Popularity as a Proxy for Generality When Building Tag Hierarchies from Folksonomies

Fahad Almoqhim, David E. Millard, Nigel Shadbolt

Electronics and Computer Science, University of Southampton, Southampton, United Kingdom

{fibm1e09,dem,nrs}@ecs.soton.ac.uk

Abstract. Building taxonomies for Web content manually is costly and time-consuming. An alternative is to allow users to create folksonomies: collective social classifications. However, folksonomies have inconsistent structures and their use for searching and browsing is limited. Approaches have been proposed for acquiring implicit hierarchical structures from folksonomies, but these approaches suffer from the “generality-popularity” problem, in that they assume that popularity is a proxy for generality (that high level taxonomic terms will occur more often than low level ones). In this paper we test this assumption, and propose an improved approach (based on the Heymann-Benz algorithm) for tackling this problem by direction checking relations against a corpus of text. Our results show that popularity works as a proxy for generality in at most 77% of cases, but that this can be improved to 81% using our approach. This improvement will translate to higher quality tag hierarchy structures.

Keywords: Folksonomies, Taxonomies, Collective Intelligence, Social Information Processing, Social Metadata, Tag similarities.

1 Introduction

The transition from the Document Web, where content is produced mainly by the owners of websites, to the Social Web where users are not only information consumers but also content contributors, means that web content today is huge and constantly growing. Building and maintaining taxonomies for organizing such content manually by experts is costly and time-consuming. Consequently, an alternative approach is to allow users to contribute by tagging, this is a process that allows individuals to freely assign tags, descriptive metadata, to a web object or resource, producing a folksonomy (a set of user, tag, resource triples) as a result of that process [1].

Collaborative tagging is one of the most successful examples of the power of Collective Intelligence (CI) [2] for constructing and organizing knowledge in the Web. It has become a key part on most online portals, such as Delicious, Blogger, Flickr, Twitter and Facebook.

In recent years, folksonomies have emerged as an alternative to traditional classifications of organizing information [3,4]. They benefit from the power of collective

intelligence to offer an easier (in terms of time, effort and cognitive costs) approach to organizing web resources [5]. However, they share the inconsistent structure problem that is inherited from uncontrolled vocabularies, which causes many problems such as ambiguity, homonymy, synonymy, and basic level variation [6,7]. Consequently, many researchers have been working on approaches for acquiring latent hierarchical structures from folksonomies and constructing tag hierarchies [8,9,10]. Constructing tag hierarchies from folksonomies can be useful in different tasks, for example:

- **Improving Content Retrieval:** Although folksonomies have become a very popular method to describe web contents due to their simplicity of use [5], the lack of structure in folksonomies makes content retrieval tasks, like searching, subscription and exploration, limited [11,12]; they tend to have low recall performance and do not support efficient query refinement [13]. Tag hierarchies, therefore, can improve content retrieval tasks by making the relations between tags explicit [14,15]. In addition, Morrison found that searches conducted with tag hierarchies achieved better results than those conducted with search engines [16].
- **Building Lightweight Ontologies:** Ontology is the backbone of the semantic web [17], and an important knowledge structure for improving the organization, retrieval and management of heterogeneous content and widespread understanding of a specific domain. However, building and maintaining ontologies is so costly and time-consuming that it obstructs the progress of the Semantic Web development [18]. The large number of folksonomies offers a promising way to build tag hierarchies and then to construct lightweight ontologies. For instance, Mika provides a model of semantic and social networks for building lightweight ontologies from Delicious [19]. Also, Schmitz proposes subsumption-based model for constructing ontology from Flickr [13].
- **Enriching Knowledge Bases:** Since users constantly and freely tag new web contents, the tag hierarchies are up-to-date and hence can be used to update existing knowledge bases or enlarge their scope. For example, Kiu and Tsui present TaxoFolk, an algorithm that uses tag hierarchies for enriching existing taxonomies by unsupervised data mining techniques and augmented heuristics [20]. Furthermore, Zheng et al. propose an approach for enriching WordNet with tag hierarchies that extracted from Delicious [21]. Also, Van Damme et al. offer a comprehensive method for building and maintaining ontologies from tag hierarchies alongside some online resources [22].

However, current approaches to automatic tag hierarchy construction come with limitations [12] and [23], one of the most significant of which is the “generality-popularity” problem. This arises from the tendency of hierarchy construction algorithms to use popularity as a proxy for generality (this is explained further in Section 2.4). For example, if users tend to tag a picture of London attractions with “London” much more than “UK”, then “London” will have higher popularity and thus be placed in a more general position than “UK” despite the fact that the relation makes more sense semantically if “UK” is the more general term. In this research, we present an experiment to quantify the extent of the “generality-popularity” problem, and combine and extend prior research in tag hierarchy building and lexico-syntactic patterns

to propose an improved approach to building tag hierarchy that tackles this problem. Our approach works by correcting the taxonomic direction between popular and more general tags by using Hearst’s lexico-syntactic patterns [24] that are commonly used for acquiring taxonomic relations from large text corpora [25].

2 Related Work

2.1 Learning Concept Hierarchy from Text

The origins of automatic acquisition of latent hierarchical structures from unstructured content can be found in approaches to learning lexical relations from free text. These approaches can be seen in two directions: approaches that exploit clustering techniques based on Harris’ distributional hypothesis [26], e.g. [25] and [27]; or approaches that use lexico-syntactic patterns to acquire a certain semantic relation in texts, e.g. “is-a” or “such-as” relationship, e.g. [24] and [28]. Many of the latter direction of the approaches have focused on a key insight first expressed by Hearst in [24], that certain lexico-syntactic patterns (Table 1) can acquire a particular semantic relationship (hyponym/hypernym relationship) between terms in large text corpora [29].

Table 1. Hearst’s lexico-syntactic patterns for detecting hyponym/hypernym relations.

<i>No</i>	<i>Pattern</i>	<i>Example</i>
1	P such as $\{C_1, C_2 \dots, (\text{and} \mid \text{or})\} C_n$	European countries such as <i>England</i> and <i>Spain</i> .
2	Such P as $\{C_1, \dots\} * \{(\text{or} \mid \text{and})\} C_n$... works by such authors as <i>Herrick</i> , <i>Goldsmith</i> , and <i>Shakespeare</i> .
3	$C_1 \{, C_n\} * \{, \dots\} \{(\text{or} \mid \text{and})\}$ other P	... <i>apple</i> , <i>orange</i> , <i>banana</i> or other fruits .
4	P $\{, \dots\}$ including $\{C_1, \dots\} * \{(\text{or} \mid \text{and})\} C_n$... all common-law countries , including <i>Canada</i> and <i>England</i> .
5	P $\{, \dots\}$ especially $\{C_1, \dots\} * \{(\text{or} \mid \text{and})\} C_n$... most European countries , especially <i>England</i> , <i>Spain</i> , and <i>France</i> .

Lexico-syntactic patterns can capture different semantic relations, though hyponym/hypernym relationship seems to produce the most accurate results, even with no pre-encoded knowledge. Additionally, they occur frequently in texts and across their genre boundaries [24] and [30].

2.2 Learning Tag Hierarchy from Folksonomies

Recently there have been several promising approaches proposed for learning tag hierarchies from folksonomies. These approaches can be seen in three directions

based on using: clustering techniques, relevant knowledge resources or a hybrid of both to infer semantics from folksonomies.

Clustering Techniques based Approaches.

Clustering techniques are mostly based on agglomerative, bottom-up, approaches. First pair-wise tag similarities are computed and then divided into groups based on these similarities. After that, pair-wise group similarities are computed and then merged as one until all tags are in the same group [31].

Heymann and Garcia-Molinay [8] introduce an extensible greedy algorithm that automatically constructs tag hierarchies from folksonomies, extracted from Delicious and CiteULike. They use graph centrality [32] in the tag-tag co-occurrence network to identify the generality order of the tags. Their algorithm hypothesis is that the tag with the highest centrality is the most general tag thus it should be added to the tag hierarchy before others. Benz et al. [10] present an extension of Heymann's algorithm by applying tag co-occurrence as the similarity measure and the degree centrality as the generality measure. They tested their algorithm with the dataset gathered from Delicious and succeeded to produce clearer and more balanced tag hierarchies compared to the original algorithm.

C. Schmitz et al. [33] and P. Schmitz [13] used statistical models of tag subsumption for constructing tag hierarchies. C. Schmitz et al used the theory of association rule mining to analyze and structure folksonomies from Delicious. P. Schmitz adapted the work of [34] to introduce a subsumption-based model for building tag hierarchy from Flickr. Schwarzkopf et al. [35] extend the two algorithms in [8] and [33] by taking into account the tag context.

Mika [19] presents a graph-based model for constructing two tag hierarchies from folksonomies, extracted from Delicious, using statistical techniques. The first tag hierarchy is based on the overlapping set of user-tag networks, whereas the second is based on the overlapping set of object-tag networks. Hamasaki et al. [36] extended the work of Mika while considering the user-user relationship. In particular, the first tag hierarchy is modified by considering tagging information of the user's neighbors.

Solskinnsbakk and Gulla [9] constructed tag hierarchies from folksonomies extracted from Delicious by using morpho-syntactic and semantic similarity measures. Morpho-syntactic similarities are found by the Levenshtein distance, whereas the cosine similarity has been used to find the semantic similarity between tags. Plangprasopchok et al. [37] adapted affinity propagation proposed by Frey & Dueck [38] to build deeper and denser tag hierarchies from folksonomies. However, Strohmaier et al. [4] have proved that generality-based approaches to learning tag hierarchy, with degree centrality as generality measure and co-occurrence as similarity measure, e.g. [10] have a superior performance compared to probabilistic models, e.g. [37].

Knowledge Resources based Approaches.

Several existing knowledge resources, such as Wikipedia, WordNet and online ontologies, can be used to discover the meaning of tags and their relationships.

Laniado et al. [15] use WordNet to disambiguate and structure tags from Delicious. Angeletou et al. [39] present FLOR, an automatic approach for enriching folksonomies, extracted from Flickr, by linking them with related concepts in WordNet and online ontologies, using the Watson semantic search engine. Cantador et al. [40] introduce an approach that automatically maps tags, extracted from Delicious and Flickr, with Wikipedia concepts, and then associates those tags with domain ontologies. Similarly, Tesconi et al. [41] use Wikipedia as an intermediate representation between tags, extracted from Delicious, and some semantic resources, namely: YAGO and WordNet. Garcia et al. [42] propose an approach to automatically disambiguate polysemous, multiple related meanings, tags through linking them to DBpedia entries.

Hybrid Approaches.

Some approaches to learning tag hierarchies are based on the combination of both previous directions, clustering techniques and knowledge resources.

Specia and Motta [43] present a semi-automatic approach rely on clustering techniques and using WordNet and Google to structure tags, extracted from Delicious and Flickr. Giannakidou et al. [44] introduce a co-clustering approach for identifying the tag semantics by clustering tags, from Flickr, and relevant concepts from a semantic resource, WordNet. Lin et al. [45] propose an approach based on data mining techniques and WordNet concepts to discover the semantics in the tags and build tag hierarchies.

2.3 Limitations of Current Approaches

Although the **approaches that based on lexico-syntactic patterns** provide reasonable precision, their recall is low [46]. In addition, they are not appropriate to use them for acquiring semantic relations in tag collections since these collections tend to be much more inconsistent than text collections [47]. Moreover, Strohmaier et al., in their study of tag hierarchy building algorithms, show that the approaches tailored towards collaborative tagging systems outperform the approaches based on traditional hierarchical clustering techniques [4].

While several **approaches based on clustering techniques** have been offered solutions to structure folksonomies, they come with limitations [12] and [23]. These include the suffering from the “generality-popularity” problem. In practice a tag could be used more frequently not because it is more general, but because it is more popular among users. For instance, Plangprasopchok and Lerman [48] found, on Flickr, that the number of photos tagged with “car” are ten times as many as that tagged with “automobile”. By applying clustering techniques, the tag “car” is likely to have higher centrality, and thus it will be perceived as more general than “automobile”.

Knowledge resources based approaches are developed to partially solve the limitations of clustering techniques approaches. However, such resources are limited and they can only deal with standard terms [12]. This limitation is due to the tags nature in which they may contain spelling errors, abbreviations, idiosyncratic terms etc. Fur-

thermore, tags can be multi-lingual, which make these sources even harder to handle [23].

In this paper, we combine these approaches in order to benefit from the accuracy of lexico-syntactic patterns, while maintaining the flexibility and scalability of clustering techniques. We do this by using hyponym/hypernym patterns to check and correct the *direction* of taxonomic tag pairs in a tag hierarchy generated via clustering, thus addressing the “generality-popularity” problem.

3 Our Approach to Building High-Quality Tag Hierarchies

In previous work [49], we have shown that applying generality-based approaches to folksonomies constructed of user provided *tag pairs* results in a better quality hierarchy than those constructed of user provided *tags*. However, asking users to provide tag pairs rather than tags results in a poorer set of terms, and a less expressive hierarchy. This leads us to the insight of our new approach that if we could improve the *accuracy of directions* in relations constructed between tags by a generality-based approach, we would be able to improve the quality of the resulting tag hierarchy structure and semantics without sacrificing richness.

It has been shown that generality-based approaches of tag hierarchy construction show a superior performance compared to other approaches [4]. However, they suffer from the “generality-popularity” problem. To tackle this problem, our proposed approach extended a promising generality-based algorithm, based on [4], by using lexico-syntactic patterns applied to a large text corpus specifically the text of English Wikipedia. The patterns that our approach used are a combination of the well-known Hearst’s lexico-syntactic patterns (Table 1) and other two other direct patterns:

- “ *C* is a **P** ”
- “ *C* is an **P** ”

While lexico-syntactic patterns suffer from low recall [46], our approach leverages their reasonable precision to correct the taxonomic direction between popular and more general tags before using them to build the tag hierarchy. The algorithm we have used in our approach is an extension of Benz’s algorithm [10], which itself is an extension of Heymann’s algorithm [8]. Table 2 demonstrates the pseudo-code for the proposed algorithm.

The algorithm is affected by several parameters, including: occurrence threshold *occ* (the number of tag occurrences); similarity threshold *min_sim* (the number of tag co-occurrences with another tag); generality threshold *min_gen* (the number of tag co-occurrences with other tags); and patterns matching occurrences *p_occ1* and *p_occ2*. Empirical experiments were performed to optimize these parameters.

Table 2. Pseudo-code for the proposed algorithm.

Input: <i>user-generated terms (tags)</i>
Output: <i>tag hierarchy</i>

1. Filter the tags by an occurrence threshold *occ*.
2. Order the tags in descending order by generality (measured by degree centrality in the tag–tag co-occurrence network).
3. Starting from the most general tag, as the root node, add all tags *ti* subsequently to an evolving tag hierarchy:
 - (a) Calculate the similarities (using the co-occurrence weights as similarity measure) between the current tag *ti* and each tag currently present in the hierarchy, and append the current tag *ti* underneath its most similar tag *tag_sim*.
 - (b) If *ti* is very general (determined by a generality threshold *min_gen*) or no sufficiently similar tag exists (determined by a similarity threshold *min_sim*), append *ti* underneath the root node of the hierarchy.
 - (c) Check the taxonomic direction (*ti* → its suggested hypernym; i.e. *tag_sim* or the root) by using the proposed lexico-syntactic patterns, and calculate *p_occ1*; i.e. in total, how many (*ti* → its suggested hypernym), with using the proposed patterns, found in Wikipedia.
 - (d) Check the taxonomic direction (*ti* ← its suggested hypernym; i.e. *tag_sim* or the root) by using the proposed lexico-syntactic patterns, and calculate *p_occ2*; i.e. in total, how many (*ti* ← its suggested hypernym), with using the proposed patterns, found in Wikipedia.
 - (e) Correct the taxonomic direction if needed based on *p_occ1* and *p_occ2*.
4. Apply a post-processing to the resulting hierarchy by re-inserting orphaned tags underneath the root node in order to create a balanced representation. The re-insertion is done based on step 3.

4 Experimental Setup

To test the performance of our approach, we applied the original algorithm and our proposed algorithm, using five common tag similarity measures and with different similarity thresholds, to a large-scale folksonomy dataset collected from Delicious (see Section 4.2), yielding 20 different tag hierarchies. The five common similarity measures between *Tag 1* and *Tag 2* can be mathematically defined as follows:

$$\text{Matching} = |A \cap B| \quad (1)$$

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|} \quad (2)$$

$$\text{Jaccard} = \frac{|A \cap B|}{|A \cup B|} \quad (3)$$

$$Overlap = \frac{|A \cap B|}{\min(|A|, |B|)} \quad (4)$$

$$cosine = \frac{|A \cap B|}{\sqrt{|A| \times |B|}} \quad (5)$$

Where “A” is the set of the folksonomies that contains *Tag 1*, and “B” is the set of the folksonomies that contains *Tag 2*.

In this paper, we are focusing on checking and correcting the taxonomic tag pairs that we get from our proposed algorithm. Therefore, we evaluate all the taxonomic tag pairs from all the resulting 20 tag hierarchies against a gold-standard dataset, namely: WordNet. The detailed experimental setup is presented next.

4.1 Experimental Design

Fig. 1 summarized the process of the experimental design that we have used for performing our experiments detailed in this paper.

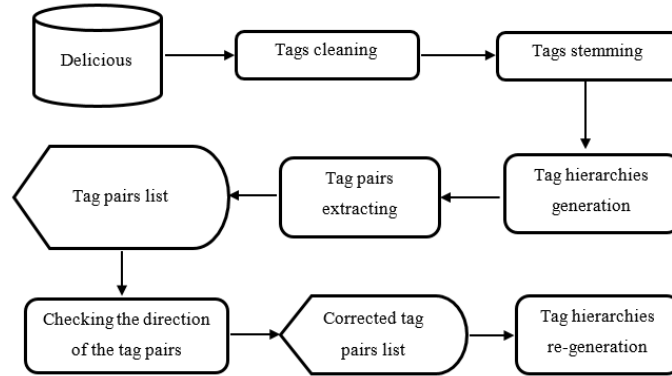


Fig. 1. The Process diagram of our experimental design

The above process consists of four main components, as follows:

- **Tags Normalising:** Before running the Tag Hierarchy Generation component, the tags are passed to the normalisation process that applies two steps: 1) **Tags Cleaning**, including: *Letters lower-case, symbol deleting and non-English letters deleting*. 2) **Tags stemming**, by using the will-known Porter Stemmer [50].
- **Tag Hierarchy Generation:** This component uses our proposed algorithm, except the steps (3.c – 3.e), to construct tag hierarchies from the tags.
- **Tag Pairs Direction Checking:** This is the most important component of our approach. It uses the steps (3.c – 3.e) of our proposed algorithm to check, and to correct if needed, the direction of the tag pairs that generated from the previous component. Note that since the produced tag pairs are stemmed, the Wikipedia and WordNet datasets are stemmed as well.

- **Tag Hierarchy Re-Generation:** It uses the Tag Hierarchy Generation to re-generate the tag hierarchy after correcting the direction of the taxonomic tag pairs.

4.2 Datasets

In our experiments, we have used two large datasets, as detailed follows:

- **Delicious Dataset:** To compare the performance of our proposed algorithms of building tag hierarchy compared to the original algorithm, we have used a large-scale folksonomy dataset from the PINTS experimental dataset¹ containing a systematic crawl of Delicious during 2006 and 2007. Table 3 summarized the statistics of the dataset.

Table 3. Statistics of the Delicious dataset.

<i>Dataset</i>	<i>Users</i>	<i>Tags</i>	<i>Resources</i>	<i>Tag assignments</i>
Delicious	532,924	2,481,698	17,262,480	140,126,586

- **Wikipedia Dataset:** To solve the “generality-popularity” tags problem by using the proposed lexico-syntactic patterns, we have chosen Wikipedia dataset. We selected to use Wikipedia since it is currently the largest knowledge repository available on the Web. The dataset that we have used contains 4,487,682 English Wikipedia articles².

4.3 Evaluation Methodology

To evaluate our proposed approach to building tag hierarchy against the original approach, we have chosen WordNet [51] dataset for two reasons:

- It is considered to be a gold-standard dataset for testing hyponym/hypernym relations building algorithms [29].
- And we avoided any dataset that was constructed automatically or based on Wikipedia since we have used it in our approach.

WordNet is a structured lexical database of the English language that build manually by experts. It contains 206,941 terms grouped into 117,659 synsets³. The synsets are connected by several lexical relations. The most important and frequently of these relations is the hyponym/hypernym relation. For our purpose we have extracted the taxonomic terms among synsets in WordNet.

¹ http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/DataSets/PINTSExperimentsDataSets/index_html

² As collected in March 2014.

³ <http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html>, as visited on June 2014.

5 Results and Analysis

In the first round of our experiment, we have applied our proposed algorithm and the original algorithm, using the five selected tag similarity measures, to the Delicious dataset, yielding 10 tag hierarchies. Then, we have rerun the experiment again but with a tag similarity threshold equal 0 to examine the effectiveness of using similarity threshold that suggested by the original algorithm. Finally, we have evaluated the direction correctness of all the taxonomic tag pairs from all the produced 20 tag hierarchies against WordNet. To give an impression of the results, Table 4 shows a few examples of the produced taxonomic tag pairs, using the five similarity measures under study.

Table 4. Examples of produced tag pairs for each of the selected similarity measures.

<i>Measure</i>	<i>Rank</i>	<i>Tag A</i>	<i>Tag B</i>	<i>Rank</i>	<i>Tag A</i>	<i>Tag B</i>
<i>Matching</i>	1	blog	design	1000	display	technology
<i>Dice</i>			design			lcd
<i>Jaccard</i>			design			lcd
<i>Overlap</i>			bloggerbeast			tft
<i>Cosine</i>			daily			lcd
<i>Matching</i>	100	daily	blog	5000	maple	php
<i>Dice</i>			news			willow
<i>Jaccard</i>			news			willow
<i>Overlap</i>			blog			willow
<i>Cosine</i>			news			willow
<i>Matching</i>	500	weather	news	10000	bridesmaid	dress
<i>Dice</i>			forecast			bridal
<i>Jaccard</i>			forecast			bridal
<i>Overlap</i>			noaa			dress
<i>Cosine</i>			forecast			bridal

And to get an overall view of how different each of the selected similarity measures is to others in terms of generating taxonomic tag pairs, Table 5 displays the overlap between the produced tag hierarchies based on these similarity measures.

Table 5. Overlap between tag hierarchies generated using selected similarity measures.

	<i>Matching</i>	<i>Cosine</i>	<i>Overlap</i>	<i>Jaccard</i>
Dice	0.15	0.71	0.16	0.57
Jaccard	0.09	0.40	0.10	
Overlap	0.71	0.24		
Cosine	0.22			

To give a comprehensive view of the evaluation against WordNet, we investigated the WordNet coverage of the investigated delicious dataset. Table 6 shows the Word-

Net coverage of the top delicious tags, whereas Table 7 illustrates the WordNet coverage of all the tags appeared in the produced tag hierarchies.

Table 6. WordNet coverage of tags in delicious dataset.

	<i>Top 10</i>	<i>Top 100</i>	<i>Top 500</i>	<i>Top 1000</i>
WordNet coverage	80.00%	77.00%	74.20%	71.10%

Table 7. WordNet coverage of tags in produced hierarchies.

	<i>With using similarity threshold</i>				
	<i>Matching</i>	<i>Dice</i>	<i>Jaccard</i>	<i>Overlap</i>	<i>Cosine</i>
WordNet coverage	39.50%	41.18%	41.91%	37.85%	40.61%
	<i>Without using similarity threshold</i>				
	<i>Matching</i>	<i>Dice</i>	<i>Jaccard</i>	<i>Overlap</i>	<i>Cosine</i>
WordNet coverage	38.89%	39.93%	31.30%	37.52%	39.90%

A number of factors limit the WordNet coverage of the tags and the taxonomic tag pairs. First, WordNet is a static knowledge resource, while the delicious dataset is an open-ended collection. Also, WordNet only covers the English language, whereas the delicious dataset contains multi-language tags. However, WordNet can be a reasonable reference for our purpose, i.e. tackling the “generality-popularity” problem, since a significant fraction of the popular tags in delicious is covered by WordNet; as shown in Table 6. Having established this the next step is to compare the tag pair directions produced by the original algorithm and our variation of the algorithm against the directions as defined in WordNet. This will give us a measure of how many times generality was a successful proxy for popularity in the original algorithm, and also the extent to which our approach improves on this.

Table 8 shows the results. For further improvement, we added a *min_p_occ* threshold in our proposed algorithm; to correct the generated taxonomic tag pairs, the occurrences number found in Wikipedia, by using the proposed lexico-syntactic patterns, need to be more than the *min_p_occ* threshold. The last column of Table 8 shows the improvement of using the *min_p_occ* threshold, which was more effective with the Matching, Dice and Jaccard similarity measures.

The first observation that can be drawn is that the original algorithm is moderately successful (as much as 76.96%), even though it blindly accepts popularity as a measure of generality. So while “generality-popularity” has been identified as a weakness of clustering approaches, using this assumption over three quarters of the generated relationships are in the right direction.

Table 8. Taxonomic tag pairs evaluation, using selected similarity measures and a similarity threshold for each measure, against WordNet

	<i>No of Tag Pairs found in WordNet</i>	<i>% Agreement with WordNet</i>		
		Original Algorithm	Our Algorithm	Our strict Algorithm
Matching	305	75.74%	77.38%	79.34%
Dice	130	47.22%	55.56%	61.11%
Jaccard	114	47.37%	64.91%	64.04%
Overlap	217	76.96%	81.11%	81.11%
Cosine	161	54.90%	64.71%	64.71%

The second observation that can be drawn is that there is a modest improvement achieved by our proposed algorithm compared to the original algorithm among all the selected tag similarity measures. This means, regardless of the similarity measure, our approach has succeeded in correcting the direction of taxonomic tag pairs that were generated in the wrong direction by the original algorithm. In the best case (Overlap) this leads to an accuracy of over 81%.

Table 9. Examples of wrong direction taxonomic tag pairs generated by original algorithm.

<i>Similarity Measure</i>	<i>Tag A</i>	<i>Tag B</i>	<i>Similarity Measure</i>	<i>Tag A</i>	<i>Tag B</i>
Matching	Faith	christian	Dice	Meat	Beef
	Footwear	shoes		primates	Monkey
	Society	culture		Road	Highway
	Wealth	money		Search	Google
	Poultry	chicken		Sweet	Candy
<i>Similarity Measure</i>	<i>Tag A</i>	<i>Tag B</i>	<i>Similarity Measure</i>	<i>Tag A</i>	<i>Tag B</i>
Jaccard	Coffee	espresso	Overlap	broadcast	Video
	Drink	alcohol		Canine	Dog
	Ireland	dublin		Footwear	shoes
	Pastry	tart		Poultry	chicken
	Puzzle	sudoku		Ride	Bike
<i>Similarity Measure</i>	<i>Tag A</i>	<i>Tag B</i>			
Cosine		bag	purses		
		sweet	candy		
		meat	beef		
		search	google		
		broadcast	radio		

Table 9 shows examples of these taxonomic tag pairs, which the original algorithm has generated them in the form of (*Tag A is-a Tag B*), where they have been found in WordNet as (*Tag B is-a Tag A*).

Given the large numbers of pairs generated by the algorithm and the moderate intersection of tags with WordNet (around 40%, as shown in Table 7) the low number of matched pairs is surprising. It may reflect the relatively small size of WordNet as compared to the delicious dataset, but it also may reflect the fact that our algorithm looks for *direct matches* in WordNet. One approach to increase the number of matches would be to use the transitivity of the generality relationship, this would match (and possibly correct the direction of) a tag pair, even if those tags were not directly linked in WordNet, but instead were part of a chain of generality relationships.

Another observation from Table 8 is that, among all the selected tag similarity measures, the Overlap measure yields the best performance of generating taxonomic tag pairs against WordNet, whereas Matching measure yields the biggest amount of generated tag pairs that found in WordNet regardless of the taxonomic direction.

Table 10. Taxonomic tag pairs evaluation, using selected similarity measures and without using a similarity threshold, against WordNet

	<i>No of Tag Pairs found in WordNet</i>	<i>% Agreement with WordNet</i>		
		Original algorithm	Our Algorithm	Our strict Algorithm
Matching	329	76.90%	77.81%	80.55%
Dice	150	51.33%	66.00%	66.67%
Jaccard	246	47.56%	66.26%	62.60%
Overlap	230	77.39%	80.00%	81.30%
Cosine	178	59.55%	67.42%	67.98%

Table 10 shows the results of rerunning the experiment but with a tag similarity threshold = 0. In addition to the previous observations on Table 8, Table 10 demonstrates that without using a similarity threshold, as suggested by the original algorithm, both the original algorithm and our variations can generate more taxonomic tag pairs that can be found in WordNet. Also, by using all selected tag similarity measures, both algorithms yield better taxonomic tag pairs structure and semantics.

6 Conclusion

Building and maintaining taxonomies for organizing Web content manually by experts is costly and time-consuming. Therefore, folksonomy has emerged as an alternative approach for organizing online resources. Yet, folksonomies are beset by many problems, due to the lack of consistent structure, such as ambiguity, homonyms, and synonymy. Thus many approaches have been proposed to resolve these problems by proposing mechanisms for acquiring latent hierarchical structures from folksonomies and constructing tag hierarchies. Among these approaches, it has been revealed that

generality-based approaches show a superior performance compared to other approaches. However, it has been argued that generality-based automatic tag hierarchy algorithms suffer from a “generality-popularity” problem, where they (sometimes inaccurately) assume that because a tag occurs more frequently it must be more general and thus appear higher in the hierarchy. Therefore, we have presented an experiment to measure this effect, and proposed an approach to reduce its impact. Our proposed approach extends a promising generality-based algorithm by using lexico-syntactic patterns for discovering hyponym/hypernym relations in order to distinguish between popular and general tags. For this purpose we have used Wikipedia as the text corpus, and for evaluation we have used WordNet as a gold-standard reference.

Our experiment reveals that generality acts as a successful proxy for popularity in 47% to 76% of cases (depending on the similarity measure used), and that the performance of our proposed algorithm outperforms the original algorithm, among all the selected tag similarity measures (correct in between 56% and 81% of cases). This means, regardless of the similarity measure, our approach has succeeded in correcting the direction of taxonomic tag pairs that were wrongly generated by the original algorithm. This improvement will result in building higher quality tag hierarchy structure and semantics.

In term of the comparison between the selected tag similarity measures, the Overlap measure yields the best performance of generating taxonomic tag pairs against WordNet. Finally, we have shown that removing the similarity threshold (in both the original algorithm and our variations) results in better taxonomic tag pairs, in terms of quantity and quality, irrespective of tag similarity measures.

For future work, we plan to investigate which lexico-syntactic patterns are most successful in correcting errors, and whether any introduce significant errors. This should give us a clear explanation of which patterns are more reliable in correcting the wrong direction of taxonomic tag pairs. Secondly, based on the results we achieved, we are planning to use a dynamic knowledge repository, such as a search engine, instead of a static knowledge resource, like Wikipedia. This should increase the coverage and occurrences of the tags in any tag collection. Finally, we intend to evaluate the tag hierarchies produced using our approach against more than one large reference taxonomies, this should give a measure of how the improvements in tag pair directions presented here translate into improved tag hierarchies.

Tagging has become an established method of crowd-sourcing structure on the Web, but folksonomies based on tags have serious weaknesses for both search and browsing, which is a primary use of structure on websites. Our hope is that our work will contribute towards the growing understanding of how more sophisticated hierarchical structure can be successfully derived from folksonomies, and that this will ultimately improve our interaction with the Social Web.

References

1. Vander Wal, T.: Folksonomy Coinage and Definition. (2007) Available at: <http://vanderwal.net/folksonomy.html>.
2. O'Reilly, T.: What is web 2.0: design patterns and business models for the next generation of software. [Online] Available at: <http://oreilly.com/web2/archive/what-is-web-20.html> (20 June 2013) (2005)
3. Gupta, M., Li, R., Yin, Z., Han, J.: An Overview of Social Tagging and Applications. In Aggarwal, C., ed. : Social Network Data Analytics. Springer, New York, NY (2011) 447-497
4. Strohmaier, M., Helic, D., Benz, D., Körner, C., Kern, R.: Evaluation of Folksonomy Induction Algorithms. ACM Transactions on Intelligent Systems and Technology 3(4) (2012) Article 74.
5. Mathes, A.: Folksonomies-cooperative classification and communication through shared metadata. Computer Mediated Communication 47(10) (2004) Available at: <http://adammathes.com/academic/computer-mediated-communication/folksonomies.pdf>.
6. Golder, S., Huberman, B.: Usage patterns of collaborative tagging systems. Journal of Information Science 32(2), 198–208 (2006)
7. Guy, M., Tonkin, E.: Tidying up tags. D-Lib Magazine 12(1) (January 2006) January, ISSN 1082-9873.
8. Heymann, P., Garcia-Molinay, H.: Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems. InfoLab Technical Report, Stanford (2006)
9. Solskinnsbakk, G., Gulla, J.: A Hybrid Approach to Constructing Tag Hierarchies. In : International conference on: On the move to meaningful internet systems: Part II, Hersonissos, Crete, Greece, pp.975-982 (2010)
10. Benz, D., Hotho, A., Stutzer, S.: Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. In : 2nd Web Science Conference (WebSci10), Raleigh, NC, USA (2010)
11. Begelman, G., Keller, P., Smadja, F.: Automated tag clustering: Improving search and exploration in the tag space. In : Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, pp.15-33 (2006)
12. Lin, H., Davis, J.: Computational and crowdsourcing methods for extracting ontological structure from folksonomy. In : 7th Extended Semantic Web Conference (ESWC'10), Heraklion, Greece, pp.472-477 (2010)
13. Schmitz, P.: Inducing ontology from flickr tags. In : Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland (2006)
14. Angeletou, S., Sabou, M., Specia, L., Motta, E.: Bridging the gap between folksonomies and the semantic web: An experience report. In : 4th European Semantic Web Conference (ESWC 2007), Innsbruck, Austria, pp.30-43 (2007)
15. Laniado, D., Eynard, D., Colombetti, M.: Using WordNet to turn a folksonomy into a hierarchy of concepts. In : 4th italian semantic web workshop: Semantic web application and perspectives, Bari, Italy, pp.192-201 (2007)

16. Morrison, P.: Tagging and searching: Search retrieval effectiveness of folksonomies on the world wide web. *Information Processing and Management* 44(4), 1562–1579 (2008)
17. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American* 284(5), 28-37 (2001)
18. Park, Y., Byrd, R., Boguraev, B.: Towards Ontologies On Demand. In : Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data (ISWC-03), Florida, USA (2003)
19. Mika, P.: Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web* 5(1), 5-15 (2007)
20. Kiu, C.-C., Tsui, E.: TaxoFolk: a hybrid taxonomy–folksonomy classification for enhanced knowledge navigation. *Knowledge Management Research & Practice* 8(1), 24-32 (2010)
21. Zheng, H., Wu, X., Yu, Y.: Enriching WordNet with Folksonomies. In : 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining (PAKDD'08), Osaka, Japan, pp.1075-1080 (2008)
22. Van Damme, C., Hepp, M., Siorpaes, K.: Folksonology: An integrated approach for turning folksonomies into ontologies. In : ESWC Workshop Bridging the Gap between Semantic Web and Web 2.0, Innsbruck, Austria, pp.57-70 (2007)
23. Solskinnsbakk, G., Gulla, J.: Mining tag similarity in folksonomies. In : 3rd international workshop on Search and mining user-generated contents (SMUC '11), Glasgow, Scotland, pp.53-60 (2011)
24. Hearst, M.: Automatic acquisition of hyponyms from large text corpora. In : 14th conference on Computational linguistics, Morristown, NJ, USA, pp.539-545 (1992)
25. Cimiano, P., Hotho, A., Staab, S.: Learning concept hierarchies from text corpora using formal concept analysis. *Journal of Artificial Intelligence Research* 24(1), 305-339 (2005)
26. Harris, Z.: *Mathematical structures of language*. John Wiley and Son (1968)
27. Faure, D., Nedellec, C.: A corpus-based conceptual clustering method for verb frames and ontology. In : the LREC Workshop on Adapting lexical and corpus resources to sublanguages and applications, pp.5-12 (1998)
28. Berland, M., Charniak, E.: Finding parts in very large corpora. In : the 37th Annual Meeting of the Association for Computational Linguistics (ACL), Stroudsburg, PA, USA, pp.57-64 (1999)
29. Snow, R., Jurafsky, D., Ng, A.: Learning syntactic patterns for automatic hypernym discovery. In : the Eighteenth Annual Conference on Neural Information Processing Systems (NIPS 2004), Vancouver, Canada, vol. 17 (2004)
30. Hearst, M.: Automated discovery of wordnet relations. In Fellbaum, C., ed. : *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA. (1998)
31. Wu, H., Zubair, M., Maly, K.: Harvesting social knowledge from folksonomies. In : 17th conference on Hypertext and hypermedia, Odense, Denmark, pp.111-114 (2006)
32. Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Semantic network analysis of ontologies. In : 3rd European Semantic Web Conference (ESWC2006), Budva, Montenegro, pp.514-529 (2006)

33. Schmitz, C., Hotho, A., Jäschke, R., Stumme, G.: Mining association rules in folksonomies. In : 10th IFCS Conference: Studies in Classification, Data Analysis and Knowledge Organization, Ljubljana, Slovenia, pp.261-270 (2006)
34. Sanderson, M., Croft, B.: Deriving concept hierarchies from text. In : 22nd ACM Conference of the Special Interest Group in Information Retrieval, Berkeley, California, USA, pp.206-213 (1999)
35. Schwarzkopf, E., Heckmann, D., Dengler, D., Kröner, A.: Mining the structure of tag spaces for user modeling. In : Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling, Corfu, Griechenland, pp.63-75 (2007)
36. Hamasaki, M., Matsuo, Y., Nishimura, T., Takeda, H.: Ontology extraction using social network. In : International workshop on the semantic web for collaborative knowledge acquisition, Hyderabad, India (2007)
37. Plangprasopchok, A., Lerman, K., Getoor, L.: From saplings to a tree: Integrating structured metadata via relational affinity propagation. In : In Proceedings of the AAAI workshop on Statistical Relational AI, Menlo Park, CA, USA (2010)
38. Frey, B., Dueck, D.: Clustering by passing messages between data points. *science* 315(5814), 972-976 (2007)
39. Angeletou, S., Sabou, M., Motta, E.: Semantically Enriching Folksonomies with FLOR. In : 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008), Tenerife, Spain (2008)
40. Cantador, I., Szomszor, M., Alani, H., Fernández, M., Castells, P.: Enriching ontological user profiles with tagging history for multi-domain recommendations. In : 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008), Tenerife, Spain (2008)
41. Tesconi, M., Ronzano, F., Marchetti, A., Minutoli, S.: Semantify del.icio.us: Automatically Turn your Tags into Senses. In : Social Data on the Web Workshop at the 7th International Semantic Web Conference, Karlsruhe, Germany (2008)
42. Garcia, A., Szomszor, M., Alani, H., Corcho, O.: Preliminary results in tag disambiguation using dbpedia. In : 1st International Workshop in Collective Knowledge Capturing and Representation, California, USA (2009)
43. Specia, L., Motta, E.: Integrating Folksonomies with the Semantic Web. In : 4th European conference on The Semantic Web: Research and Applications, Innsbruck, Austria, pp.624-639 (2007)
44. Giannakidou, E., Koutsonikola, V., Vakali, A., Kompatsiaris, Y.: Co-clustering tags and social data sources. In : 9th International Conference On Web-Age Information Management, Zhangjiajie, China, pp.317-324 (2008)
45. Lin, H., Davis, J., Zhou, Y.: An integrated approach to extracting ontological structures from folksonomies. In : 6th European Semantic Web Conference, Heraklion, Greece, pp.654-668 (2009)
46. Cimiano, P.: *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications* 27. Springer (2006)
47. Plangprasopchok, A., Lerman, K., Getoor, L.: Growing a Tree in the Forest: Constructing Folksonomies by Integrating Structured Metadata. In : 16th ACM SIGKDD international

conference on Knowledge discovery and data mining, Washington, DC, USA, pp.949-958 (2010)

48. Plangprasopchok, A., Lerman, K.: Constructing Folksonomies from User-Specified Relations on Flickr. In : 18th International World Wide Web conference, Madrid, Spain, pp.781-790 (2009)
49. Almoqhim, F., Millard, D., Shadbolt, N.: An Approach to Building High-Quality Tag Hierarchies from Crowdsourced Taxonomic Tag Pairs. In : the 5th International Conference on Social Informatics (SocInfo 2014), Kyoto, Japan, pp.25-27 (2013)
50. Porter, M.: An algorithm for suffix stripping. Program 14(3), 130-137 (1980)
51. Miller, G.: WordNet: a lexical database for English. Communications of the ACM 38(11), 39-41 (1995)