UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Electronics and Computer Science

Building Tag Hierarchies Based on Co-occurrences and Lexico-Syntactic Patterns

by

Fahad Ibrahim Bin Moqhim

Thesis for the degree of Doctor of Philosophy

June 2016

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Electronics and Computer Science

Thesis for the degree of Doctor of Philosophy

BUILDING TAG HIERARCHIES BASED ON CO-OCCURRENCES AND LEXICO-SYNTACTIC PATTERNS

Fahad Ibrahim Bin Moqhim

Knowledge structures, such as taxonomies, are key to the organization and management of Web content, but are expensive to build manually. In this thesis we explore the issues around automatically building effective tag hierarchies from folksonomies (collective social classifications), and propose changes to the state-of-the-art methods that improve their performance. These changes aim to tackle the "generality-popularity" tags problem, in that popularity is assumed (sometimes inaccurately) to be a proxy for generality, i.e. high-level taxonomic terms will occur more often than low-level ones.

The effectiveness of this research is demonstrated in four experiments. The first experiment explores whether taxonomic tag pairs captured directly from users change the quality of constructed tag hierarchies. The second experiment examines the possibility of using personal tag relationships constructed by users to improve the accuracy of learned taxonomic tags. The third experiment demonstrates the potential of using lexico-syntactic patterns applied to a closed text corpus to improve the direction of automatically derived tag pairs in order to build higher quality tag hierarchies. The last experiment investigates the possibility of using an open knowledge repository instead of a closed knowledge resource to increase the tags coverage in any tag collection, and consequently the quality of learned tag hierarchies.

The results of our experiments show that collecting taxonomic tag pairs increases the semantic quality of the tag hierarchy, but at the expense of expressivity, and with some degradation of user experience. Secondly, personal tag relationships can be used to improve the accuracy of constructed taxonomic tags, but with limited success if the personal tag relationships and the learned taxonomic tags are not extracted from the same tagging system. Finally, lexico-syntactic patterns applied to a closed large text corpus (e.g. Wikipedia) can be used to improve the accuracy of directions in relations constructed between tags by a generality-based approach to tag hierarchy construction, and this would be improved further if an open corpus (e.g. the Web) is used instead of a closed one, which consequently improves the quality of the learned tag hierarchies in terms of structure and semantics.

Table of Contents

List of Tables

List of Figures

Declaration of Authorship

I, Fahad Bin Moqhim, declare that this thesis entitled "*Building Tag Hierarchies Based on Co-occurrences and Lexico-Syntactic Patterns*" and the work presented in it are my own and have been generated by me as the result of my own original research. I confirm that:

1. This work was done wholly or mainly while a candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2013. 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, 2013. Springer.

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2014. Improving on Popularity as a Proxy for Generality When Building Tag Hierarchies from Folksonomies. In the 5th International Conference on Social Informatics (SocInfo 2014). Barcelona, Spain, 2014. Springer.

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2015. The horse before the cart: improving the accuracy of taxonomic directions when building tag hierarchies. In the 8th Saudi Students Conference in the UK. London, 2015. Imperial College Press.

Acknowledgements

This thesis would not have been possible without the immense support that I received during a candidature for a research degree at the University.

First of all, I would like to greatly thank my first supervisor, Dr David Millard, for his outstanding support, guidance and inspiration during the completion of my PhD study. I would like also to take this opportunity to express my appreciation to my second supervisor, Professor Sir Nigel Shadbolt, for his time and valuable suggestions to improve my work.

I wish also to extend my sincere gratitude to my family: my parents for their endless support and praying for me, my brothers and sisters for all their support and wishes. They have always encouraged me to follow my dreams.

I would like to express my gratefulness to my friends and colleagues who helped and supported me during the research. I would also like to thank all people who participated in this study.

Finally, I acknowledge and thank the government of Saudi Arabia, through the National Centre for Assessment (Qiyas), for this scholarship enabling me to pursue my PhD study.

Definitions and Abbreviations

AI	Artificial Intelligence
AUT	Area Under Tree; a measure for structural evaluation of hierarchies
CI	Collective Intelligence
KR	Knowledge Representation
PT	Produced (Learned) Taxonomy
RT	Reference Taxonomy
TF	Taxonomic F-measure
TP	Taxonomic Precision
TR	Taxonomic Recall

Chapter One

Introduction

Knowledge structures, such as taxonomies or ontologies, are key to the organization and management of information, but are expensive to build manually. In this thesis we explore the issues around automatically building effective knowledge structures based on collective intelligence, and propose changes to the state-of-the-art methods that improve their performance. Building knowledge structures for organising web resources helps in promoting browsing, searching and retrieval (McGuinness, 2003; Garshol, 2004; Giunchiglia & Zaihrayeu, 2009). Knowledge structures are also fundamental in constructing lexical resources, which play a vital role in preparing, processing and organizing the information and knowledge required by machines and humans (Bloehdorn, 2008).

A knowledge structure that formally defines concepts, like mathematical theories, has many structural properties, while a knowledge structure that defines concepts very loosely, like document and hyperlink, has few structural properties (Gruninger et al., 2008). Figure 1-1 shows the formal structural complexity of the main existing knowledge structures that have been described in Section 2.3, whereas Figure 1-2 shows the distinctions between them in terms of the level of structures.

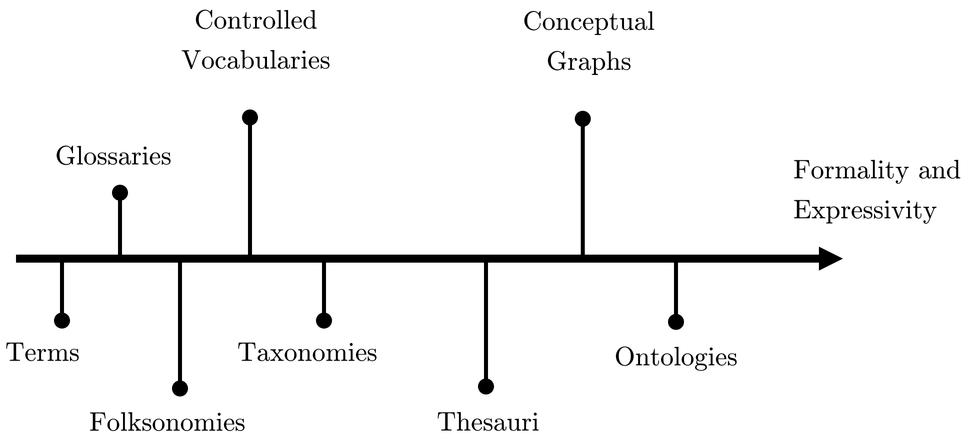


Figure 1-1: Structure level of knowledge structures, based on (Section 2.3), (Uschold & Gruninger, 2004) and (Giunchiglia & Zaihrayeu, 2009)

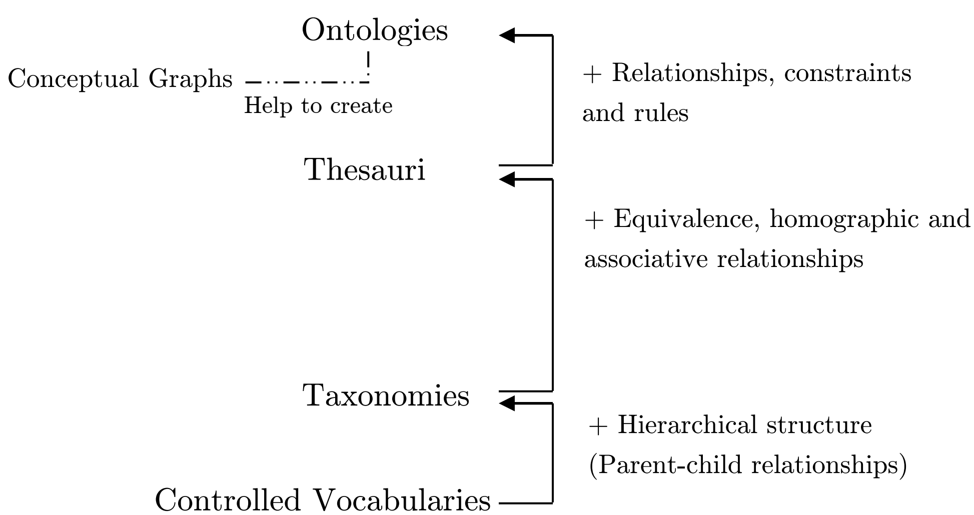


Figure 1-2: Structural distinctions between knowledge structures, based on (Section 2.3) and (Cardoso, 2007)

For lots of information systems, the power of the computational operations provided relies on the level of structure and completeness of the data (Gruber, 2008). Consequently, the more formal structural complexity a knowledge structure has, the richer the semantics that can be obtained, however, the cost is consequently much greater (Figure 1-3). In order to choose which one fits the system requirements with a minimum cost, it is necessary to study the key existing knowledge structures (Section 2.3) and their usages. A catalogue system, like Phonebook Yellow Page listings, can work properly with a controlled vocabulary, whereas a taxonomy is needed for a website organization and navigation support. And if the system, for example, needs to define properties, relations and rules on concepts for the purpose of machine-interpretability, creating an ontology may worth the extra cost. It should be noted that, although a knowledge structure has a high level of formality, it may cause a tension in the semantics obtained from it. This tension has been defined as the Semantic Gap, which is caused by the difference of semantics as expressed by human and as expressed by machines (Millard et al., 2005).

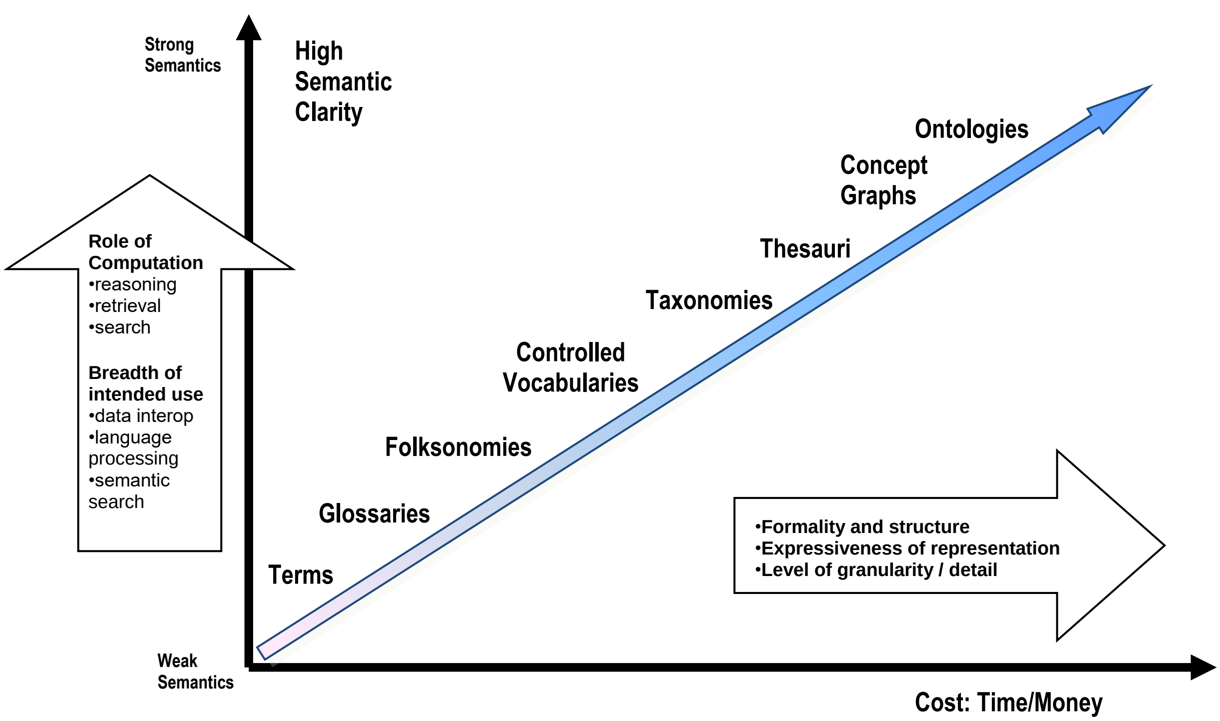


Figure 1-3: Semantic level of knowledge structures vs. cost; adapted from (Obrst, 2003), (McGuinness, 2003) and (Gruber, 2008)

A common knowledge structure for organizing online resources is taxonomy (Bloehdorn, 2008). Moreover, taxonomy is a fundamental part of building an ontology (Guarino, 1998; Nicola & Missikoff, 2016), which is seen as the backbone of the semantic web (Berners-Lee et al., 2001). As web content today is huge and constantly growing, building and maintaining taxonomies for such content manually is costly and time-consuming. Folksonomy has emerged to deal with this issue by providing a collective approach based on social classification and harnessing the power of collective intelligence (Quintarelli, 2005).

Folksonomy is “the result of tagging of information objects by a user freely using keywords relevant to the object being tagged” (Vander Wal, 2007). In (Hotho et al., 2006), Folksonomy is defined formally as follows:

Definition 1-1 Folksonomy is a tuple F := (U, T, R, Y) where:

— U, T, and R are finite sets, whose elements are called users, tags and resources, respectively, and

— Y is a ternary relation between them, i. e., $Y \subseteq U \times T \times R$, whose elements are called tag assignments.

Figure 1-4: An example of Folksonomy

Although folksonomies became popular as part of collaborative tagging systems, they are beset by many problems, due to the lack of consistent structure, such as homonym, synonym and basic level variation (Golder & Huberman, 2006; Tommasel & Godoy, 2015).

Figure 1-4 shows an example of folksonomy where three resources are tagged by three users with a number of tags. Due to the lack of consistent structure in folksonomies the following problems could hamper the process of searching:

> **Tag homonym** that occurs among tags having the same spelling but with different meanings. For example, searching for *apple* in the above folksonomy will return both *R1* and *R3* resources, regardless of the meaning of apple in the context of the query, a company or fruit.
>
> **Tag synonym** that occurs among different tags expressing the same meaning. Synonym can cause exclusion of the searching results. For example, searching for *mobile phone* will return only *U2*, and not *U1*.
>
> **Tag basic level variation** that arises when tags with different levels of specificity are used to tag relevant resources; i.e. they relate to the same concept. For example, searching for *smartphone* will return only *U2*, and not *U1*, as the lack of structure in folksonomies does not reveal the fact "iPhone is a smartphone". Furthermore, this problem obstruct the result diversity (Zwol et al., 2008), where similar results in distinctive sets are grouped.

To overcome these problems, many studies such as (Heymann & Garcia-Molina, 2006; Solskinnsbakk & Gulla, 2010; Benz et al., 2010) have been conducted to acquire the latent hierarchical structures in social tagging and building tag hierarchies (common taxonomies). However, these approaches come with limitations (Plangprasopchok & Lerman, 2009; Lin & Davis, 2010; Solskinnsbakk & Gulla, 2011).

One of the most significant of these limitations is the "popularity-generality" tag problem. This arises from the tendency of hierarchy construction algorithms to use popularity as a proxy for generality (this is explained further in Section 3.4). 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. We have applied a generality-based approach on Delicious dataset and found that many general tags wrongly become hyponyms of less general ones as they are more popular. Table 1-1 shows few examples of these wrong direction taxonomic tag pairs.

Table 1-1 Examples of wrong direction taxonomic tag pairs occurred by the "generality-popularity" tags problem

Hyponym	*hypernym*
Broadcast	Video
Canine	Dog
Footwear	Shoes
Poultry	Chicken
Sweet	Candy

In this research we explored three approaches to overcome this problem. In the first we propose a change to the current tagging approach for making a big change to the type of knowledge structure that can be built. The new tagging approach (Figure 1-5) takes the form of "is-a" relationship, where users should type two related tags; i.e. Tag t1 is a tag for the resource and Tag g1 is a generalization of Tag t1. And if Tag t1 is more popular than Tag g1, it will still be a subclass of Tag g1 by the new tagging approach. This simple relationship (Tag t1 is-a Tag g1) will not only help in tacking the "generality-popularity" tags problem, but also will provide tag pairs that are more expressive than tags alone for constructing high quality tag hierarchies in terms of semantics and structure. In this thesis, by "the quality of tag hierarchy semantics" we mean the accuracy of the hyponym/hypernym relationship between tags within the hierarchy against a reference taxonomy, and by "the quality of tag hierarchy structure" we mean the expression of the tag hierarchy (hierarchy width and depth).

We have examined the performance of our new tagging approach and 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 *individual tags*. However, asking users to provide tag pairs rather than tags results in a poorer set of terms, and a less expressive hierarchy (Chapter 4). This leads us to the insight of our second approach to tackle the "generality-popularity" tags problem 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.

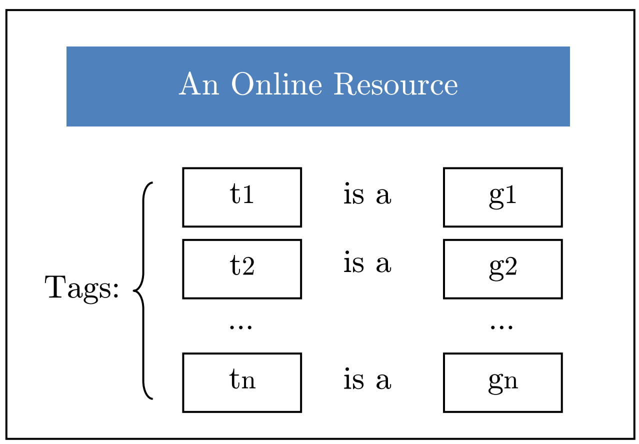


Figure 1-5: The new tagging approach

Our second approach proposes to use the relationships between tags that expressed by users in a collaborative tagging system for improving the accuracy of the taxonomic tag directions built from individual tags in that collaborative tagging system. We tested this approach by using Delicious Bundles (users in Delicious can group relevant tags into bundles) and tag hierarchies constructed from a Delicious dataset, as well as tag hierarchies constructed from another folksonomy dataset (e.g. a Flickr dataset). While some works, e.g. (Plangprasopchok et al., 2010b), suggest to use the relationships between tags that explicitly created by users to create a common tag hierarchy, we propose to use these relationships to improve the accuracy of the taxonomic tag directions that constructed from individual tags. Our results show that the coverage of the examined taxonomic tag pairs generated from a Flickr dataset found in Delicious Bundles is much less than the ones generated from the Delicious dataset (Section 5.5). This indicates that is better to use personal tag relationships from the same collaborative tagging system that used to construct tag hierarchies. However, not every tagging system allows users to organize content hierarchically.

Our third approach combines and extends prior research in tag hierarchy construction and lexico-syntactic patterns to propose an improved approach to building tag hierarchies. The approach works by correcting the taxonomic direction between popular and more general tags by using Hearst' s lexico-syntactic patterns (Hearst, 1992) that are commonly used for acquiring taxonomic relations from large text corpora (Cimiano et al., 2005). We will test the effectiveness of this proposed approach by using two different types of text corpora: 1) a closed text corpus (e.g. a download of English Wikipedia), and 2) an open text corpus (e.g. the Web via Bing). This also will allow us to make sure of the results accuracy that we will get by making a comparison evaluation between the two corpora, and increase the coverage and occurrences of the tags in any tag collection.

1.1 Motivation for the Research

Tag hierarchies have emerged to combine the features of folksonomies and taxonomies since they are constructed from folksonomies and considered as a communal taxonomy (Heymann & Garcia-Molina, 2006; Plangprasopchok, 2010). There are a number of motivations for creating tag hierarchies:

Improving Content Retrieval: Although folksonomies have become a very popular method to describe web contents due to their simplicity of use (Mathes, 2004), their lack of structure limits content retrieval tasks, like searching, subscription and exploration (Begelman et al., 2006; Limpens et al., 2008; Angeletou et al., 2008; Lin & Davis, 2010); they tend to have low recall performance and do not support efficient query refinement (Schmitz, 2006). In contrast, tag hierarchies can improve content retrieval tasks by making the relations between tags explicit (Angeletou et al., 2007; Laniado et al., 2007; Liu et al., 2010; Yao et al., 2012; Dong et al., 2015; Jabeen et al., 2016). In addition, Morrison found that searches conducted with tag hierarchies achieved better results than those conducted with search engines (Morrison, 2008). Zhuhadar et al. use tag hierarchies to continuously enrich taxonomies in a real-world application (i.e. Massive Open Online Courses (MOOCs) platform at Western Kentucky University) and found that tag hierarchies can be used for more efficient information discovery in MOOCs' platforms (Zhuhadar et al., 2015).

Building lightweight ontologies: Ontology is the backbone of the semantic web (Berners-Lee et al., 2001), 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 costly and time-consuming and obstructs the progress of the Semantic Web development (Section 2.3). The popularity of folksonomies offers a promising way to build tag hierarchies and then to construct lightweight ontologies (ontologies with few formal axioms and constraints referred to as *lightweight ontologies* (McGuinness, 2003; Giunchiglia & Zaihrayeu, 2009)). For instance, Mika provides a model of semantic and social networks for building lightweight ontologies from Delicious[1] (Mika, 2007). Also, Schmitz proposes subsumption-based model for constructing ontology from Flickr[2] (Schmitz, 2006).

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 (Plangprasopchok, 2010; Wang et al., 2015). For example, Kiu and Tsui present an algorithm (TaxoFolk) that uses tag hierarchies to enrich existing taxonomies by unsupervised data mining techniques and augmented heuristics (Kiu & Tsui, 2010). Furthermore, Zheng et al. propose an approach for enriching WordNet[3] with tag hierarchies that are

[1] http://delicious.com

[2] http://www.flickr.com

[3] http://wordnet.princeton.edu

extracted from Delicious (Zheng et al., 2008). Also, Van Damme et al. offer a comprehensive method for building and maintaining ontologies from tag hierarchies alongside some online resources (Van Damme et al., 2007).

1.2 Research Hypothesis and Questions

The research in this thesis examines the following hypothesis:

Lexico-syntactic patterns applied to a large text corpus can be used to improve the accuracy of directions in relations constructed between tags by an approach to tag hierarchy construction, and to improve the quality of the resulting tag hierarchy structure and semantics.

The following three main research questions have been identified:

1. To what extent do high quality tag pairs captured directly from users change the quality of constructed tag hierarchies?
2. Can lexico-syntactic patterns applied to a closed text corpus improve the direction of automatically derived tag pairs, and how is this affected when the lexico-syntactic patterns are applied to an open text corpus, such as the open web?
3. Will the improvement of the accuracy of taxonomic tag directions translate to higher quality tag hierarchy structure and semantics?

Question one explores the impact of two things. First, the impact of gathering taxonomic tag pairs from users rather than individual tags in the quality of the learned tag hierarchies. Second, the impact of using personal tag relationships created by users on improving the accuracy of taxonomic tag directions constructed from individual tags by a generality-based approach.

For the first exploration, we propose a new tagging approach that takes the form of "is-a" relationship, where users should type two related tags; i.e. Tag t1 is a tag for the resource and Tag g1 is a generalization of Tag t1. This simple relationship

(Tag t1 is-a Tag g2) will not only help in tackling the "generality-popularity" tags problem, but will also provide more expressive of the tags than tags alone for constructing high-quality tag hierarchies. To test the proposed tagging approach, we need to perform an experiment for collecting taxonomic tag pairs and individual tags annotated to the same resources, and then building tag hierarchies from these datasets. This will allow us compare the quality of these tag hierarchies in terms of structure and semantics, as well as the usability cost (i.e. cognitive effort) of using the proposed tagging approach compared to the normal one (individual tags).

For the second exploration, we need to perform an experiment for using personal tag relationships created by users in a collaborative tagging system to improve the accuracy of the taxonomic tag directions built from individual tags, and then comparing the accuracy of the improved taxonomic tags to the original ones.

Question two focuses on how the accuracy of some taxonomic tag directions could be improved by applying lexico-syntactic patterns to a large text corpus. Using lexico-syntactic patterns to check the directions of taxonomic tags might help in tackling the "generality-popularity" tags problem. By "the accuracy of taxonomic tag directions" we mean the validation of the direction between hyponym and hypernym tags; i.e. Tag t1 is-a Tag g2 or vice versa. In order to answer this question, we need to perform an experiment for generating taxonomic tags, correcting the directions of those taxonomic tags, and comparing the accuracy of the corrected taxonomic tags to the original ones. Finally we need to examine the impact of using an open knowledge repository (e.g. the Web) instead of a closed text corpus (e.g. a download of English Wikipedia).

Question three concerns whether the improvement of the accuracy of directions in relations constructed between tags by a generality-based approach would improve the quality of the resulting tag hierarchy structure and semantics or not. By "tag hierarchy structure" we mean the expression of the tag hierarchy (hierarchy width and depth), and by "tag hierarchy semantics" we mean the quality of the hyponym/hypernym relationship between tags within the hierarchy. In order to answer this question, we need to perform an experiment for building tag hierarchies by our improved algorithm, as well as the original algorithm, and compare the quality of these tag hierarchies in terms of structure and semantics.

1.3 Research Contributions

This thesis investigates the current generality-based approaches for tag hierarchy construction to discuss their strengths and weakness, and to provide a solution to reduce their limitations. Through the methodology used in this research and its findings, the key contributions are as follows:

- **A Methodology for the Evaluation of Tag Hierarchies**. Evaluating an approach to tag hierarchy construction is a major challenge since there is not yet a golden evaluation dataset or a proper evaluation methodology of hierarchical structures. This thesis presents a broad evaluation process that involves mix of objective and subjective metrics to evaluate three aspects: the semantics of tag hierarchies, the expression of tag hierarchies and the usability of the tagging approach that have been used to collect the tags, in terms of efficiency, effectiveness and satisfaction (explained in Chapter 3).

- **Tag Hierarchies Construction Approach based on Crowdsourced Taxonomic Tag Pairs**. The thesis introduces a new tagging approach for moving from collective folksonomies to collective taxonomies. In other words, we propose making small changes to the current tagging approach, by asking participants to tag in the form of "is-a" relationship, in order to make a big change to the type of knowledge structure that can be built. These small changes will cope with the lack of a consistent structure in folksonomies, raise their semantic and keep the interaction cost of the process down. To test the proposed tagging approach and collect data for executing the experiment, the TagTree System is introduced. Also, an experiment was conducted to ascertain whether the new tagging approach has a genuine impact on the semantic of the learned taxonomic tags and whether this in turn has an impact on the tag hierarchy as a whole. This contribution comprises all of the proposed tagging approach and algorithm, the data from this experiment, the implementation of the TagTree System, and the analysis of the results (explained in Chapter 4).

- **Tag Hierarchies Construction Approach based on Personal Tag Relationships and Individual Tags**. The thesis proposes an approach that extended a promising generality-based approach by using personal tag relationships created by users (e.g. Delicious Bundles). Contrary to previous work done on creating taxonomic relations based on personal tag relationships, our proposed tagging approach uses personal tag relationships to check, and not to create, the taxonomic tag directions that are built from individual tags. An experiment was conducted to evaluate our proposed approach and algorithm to building tag hierarchy against the original approach. To evaluate this we have chosen WordNet, as a reference taxonomy, to find which one of the two approaches produces more accurate taxonomic tags (explained in Chapter 5).

- **Tag Hierarchies Construction Approach based on Lexico-Syntactic Patterns Applied to a Text Corpus**. The thesis proposes an approach to building tag hierarchy that extended a promising generality-based approach by using lexico-syntactic patterns applied to a large text corpus specifically the text of English Wikipedia. The patterns that our approach uses are a combination of the well-known Hearst's lexico-syntactic patterns (Table 3-1), and another direct pattern: "*Tag t1* is a/an *Tag g1*". Contrary to previous work done on creating taxonomic relations based on lexico-syntactic patterns, our proposed tagging approach is novel because it uses the proposed lexico-syntactic patterns to check, and not to create, the taxonomic tag directions that are built from a tag collection. While lexico-syntactic patterns tend to achieve a very high level of precision, but low recall, 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. An experiment was conducted to evaluate our proposed approach and algorithm to building tag hierarchy against the original approach. To evaluate this we have chosen WordNet, as a reference taxonomy, to find which one of the two approaches produces more accurate taxonomic tags. Also, another experiment was conducted to ascertain whether an open text corpus (e.g. the Web) can be used, instead of a closed text corpus (e.g. English Wikipedia), in our proposed tag hierarchy construction to improve the semantic of the learned taxonomic tags, and whether this in turn has an impact on the tag hierarchy as a whole. This contribution comprises the proposed tag hierarchy construction approach, the implementation of the proposed algorithm, and the analysis of the results (explained in Chapter 5).

1.4 Publications

Throughout the process of this thesis, a number of publications have been published and presented in different conferences. A summary of these publications and achievements is presented as follows:

- **An approach to building high-quality tag hierarchies from crowdsourced taxonomic tag pairs** (Almoqhim et al., 2013). This paper was presented at the 5th International Conference on Social Informatics 2013 in Kyoto, Japan, and established my initial approach to building tag hierarchies. The proposed tagging approach, the proposed algorithm to building tag hierarchy and the proposed the evaluation process to evaluate tag hierarchies were introduced, along with the results and analysis of the resulting tag hierarchies as well as the motivation for the research.

- **Improving on popularity as a proxy for generality when building tag hierarchies from Folksonomies** (Almoqhim et al., 2014). This paper was presented at the 6th International Conference on Social Informatics 2014 in Barcelona, Spain. Based on the results we achieved in the previous paper, this paper introduced our new proposed approach to building tag hierarchy based on lexico-syntactic patterns applied to a large text corpus; i.e. English Wikipedia corpus. The proposed approach and algorithm were introduced along with the results and analysis of the resulting taxonomic tags.

- **The horse before the cart: improving the accuracy of taxonomic directions when building tag hierarchies** (Almoqhim et al., 2015). This paper was presented at the 8th SSC 2015 in London, United Kingdom. In this paper, we extend the work presented in our previous paper and show further improvement in building high-quality tag hierarchy. By extracting all transitive hyponym/hypernym relations in WordNet we were able to evaluate our approach with a more reasonable size of taxonomy reference. Whereas in the previous paper we had extracted 364,135 direct taxonomic terms among synsets in WordNet, we have, for this paper, extracted 2,153,520 direct and inherited taxonomic terms among synsets in WordNet.

Based on the results we achieved in the above publications, we are planning to submit a significant journal article that will summarise our work in this thesis along with showing further improvement to our approach to building tag hierarchy construction by using an open knowledge repository, i.e. the Web via Bing, instead of a closed knowledge resource, i.e. a download of English Wikipedia.

1.5 Outline of the Thesis

This thesis is divided into seven chapters and this section provides a summary of the content of each. This chapter has started with an overview of context and motivation, the hypothesis for this research including the research questions to be answered, the outlines the key contributions, and the publications that have been published during the research. The remainder of this thesis is organised as follows:

Chapter 2 gives an overview of the field of knowledge representation including the benefits and the influences of building knowledge structures. The chapter also introduces a framework for describing and comparing knowledge structures to distinguish between them in terms of cost, structure and semantic level.

Chapter 3 focuses on related work that has been undertaken on the one of the key existing knowledge structures, i.e. folksonomy, and how to improve it by acquiring latent hierarchical structures from it and constructing common tag hierarchies. A comprehensive review of the current approaches to constructing tag hierarchies from collaborative tagging and their limitations are discussed. In addition, a broad evaluation process for automated hierarchical structures construction is introduced.

Chapter 4 introduces a new tagging approach that proposes a change to the current tagging approach to cope with the lack of a consistent structure in folksonomies and build advanced knowledge structures; moving from collective folksonomies to collective taxonomies. The chapter also introduces the TagTree System and describes the experiment that has been conducted to investigate the impact of the new tagging approach on building high-quality tag hierarchies. The improvements that are achieved by this proposed tagging approach will be highlighted.

Chapter 5 proposes three approaches to improve the accuracy of directions in relations constructed between tags by a generality-based approach. The key aim of these approaches is tackling the " generality-popularity" tags problem, in that the current generality-based approaches assume that popularity is a proxy for generality, by correcting the directions of taxonomic tags against a knowledge resource. The proposed approaches and knowledge resources will be described, and the analysis of the results will be discussed.

Chapter 6 examines whether the improvement of the accuracy of directions in relations constructed between tags by a generality-based approach would improve the quality of the resulting tag hierarchy structure and semantics or not. An extensive evaluation to assess the tag hierarchies produced using our improved approach described in the previous chapter will be discussed.

Chapter 7 closes the thesis with the conclusions drawn from it and links them to the findings achieved. Also, some the possible future work of the research will be presented in this chapter.

Chapter Two

Knowledge Representation on the Web

Originally, knowledge representation (KR) was a subfield of artificial intelligence (AI), and it dates back to the 1950s when John McCarthy introduced an important paper called "Programs with Common Sense", which was reprinted in [Minsky, 1968] (Buchanan, 2005). From McCarthy's work, as well others, researchers argued that AI could be seen as automatic reasoning with declarative knowledge representations and the key research challenge is to discover how to represent human knowledge in machines and use it computationally to infer new knowledge as well as to solve problems (Harmelen et al., 2008).

KR is a research area formed by a range of disciplines, theories and techniques. According to (Sowa, 2000), KR is a multidisciplinary field that builds upon three other fields: logic, ontology, and computation. Logic is for the formal structure and the rules of reasoning; ontology is a knowledge structure for defining and declaring a set of concepts and their relationships (as without ontology the concepts are poorly defined and ambiguous); and computation is for implementing the logic and ontology in computer applications. Since the aim of this research is to build high level knowledge structures based on collective intelligence, the following sections will focus on the second aspect of KR: knowledge structures.

2.1 Knowledge Structures based on Collective Intelligence

Harnessing Collective Intelligence (CI) is one of the great challenges of our times (Lykourentzou et al., 2011) and for many years it has been an active research area of various fields such as biology, social sciences, computer science and engineering (Leimeister, 2010).

Malone et al. define collective intelligence very broadly as *"groups of individuals doing things collectively that seem intelligent"* (Malone et al., 2009). By this definition, collective intelligence is not a recent phenomenon and has actually existed for a very long time. For instance, families, organisations and countries are all groups of individuals that have the potential to collectively do things intelligently. Moreover, ant and beehives colonies are groups of individuals (insects), engaged in activities, such as finding food, that seem intelligent. Even a single human brain can be seen as a group of individual neurons that collectively act in an intelligent way (Malone, 2008).

The spread of Web 2.0 applications and simple technologies has led to the emergence of new meaning and forms of collective intelligence by encouraging people to engage more significantly and more effectively in building the content of the Web (Leimeister, 2010). Several collaborative knowledge construction applications on the Web, such as wikis and collaborative tagging, are successful at motivating individuals to express their knowledge, which leads to the surprising revolution of novel knowledge (Maleewong et al., 2008). The most popular successful example of these collective intelligence applications is Wikipedia (Tapscott & Williams, 2008), which is the largest multilingual online encyclopaedia in the world. It provides around 40 million articles in 292 languages, and over 5 million in English, which have been written collaboratively by thousands of volunteers worldwide (Wikipedia, 2016).

Collaborative tagging, also known as social tagging or social indexing, is another popular successful example of the power of collective intelligence for creating and organising knowledge. It is recognised as one of the best approaches for assigning metadata to web resources. Moreover, at present, collaborative tagging has become a key part on most online portals, such as Delicious, Flickr, Blogger and Facebook (Gupta et al., 2011; Kalboussi et al., 2015). Tagging is a process that allows individuals to assign tags to a web object or resource. These tags can facilitate the classification and categorisation of the Web content and can be considered as knowledge themselves.

The knowledge structure derived from the practice of collaboratively tagging resources by individuals is often referred to as a folksonomy. Compared to manual metadata creation, social tagging and folksonomies use the power of collective intelligence to offer a simpler, cheaper and a more natural approach to organising web resources (Macgregor & McCulloch, 2006). Folksonomies, however, share the inconsistent structure problem that is inherited from uncontrolled vocabularies, such as homonym, synonym and basic level variation. Consequently, many researchers have been working on approaches for acquiring latent hierarchical structures from folksonomies and constructing tag hierarchies. In this thesis, we will investigate these approaches and identify their limitations, and look at how it is possible to use the power of collective intelligence to build and improve tag hierarchies.

2.2 Motivation of Knowledge Structures

In recent years researchers have shown an increased interest in knowledge structures (Staab & Studer, 2009) which are seen as the building blocks of the Semantic Web (Berners-Lee et al., 2001; Shadbolt et al., 2006). Knowledge structures can be used in various domains, such as classification of resources, lexical resources, metadata descriptions, data integration, and queries and deductive reasoning. These application domains will be briefly described as follows:

> **Classification of Resources**: In knowledge organization and library science there is a need to organize knowledge resources, e.g. documents, in order to facilitate browsing, searching and retrieval. A common technique for organizing resources is to use a formal classification structure, i.e. taxonomy. This structure is useful for organizing and grouping similar

resources according to important characteristics in a hierarchical manner with some relationship (Bloehdorn, 2008). In the past, resources classification has been the domain of conventional library and archiving. But more recently a lot of interest has also focused on the organisation of the information existing on the web such as the Open Directory Project[4], or search engines such as Google[5] and Yahoo![6].

Lexical Resources: Lexical resources, of a given language, aim at structuring the words along with various forms, e.g. nouns, verbs etc., and linking them through some relations, such as synonym (equivalent in meaning), antonym (opposite in meaning) or meronym (a part-whole relationship). Lexical resources are broadly used for the study of natural language and knowledge engineering. They play a vital role in preparing, processing and organizing the information and knowledge required by machines and humans (Miller, 1995).

Metadata Descriptions: Metadata is often seen as "data about data". Besides classification of resources, metadata descriptions are used for a detailed specification of an object, e.g. a document, and for both its data and structure. The main purpose of metadata is to improve and facilitate the retrieval of relevant information (Mathes, 2004).

Data Integration: It is the process of combining data stored in different sources and providing a unified view of this data to the user (Lenzerini, 2002). Data sources tend to be represented in rooted trees (i.e. knowledge structures), whose nodes are assigned with natural language labels so data integration can be achieved by discovering semantic relations that exist between the nodes (Giunchiglia & Zaihrayeu, 2009).

Queries and Deductive Reasoning: Explicit domain knowledge tends to achieve the aim of answering queries from the represented knowledge.

[4] http://www.dmoz.org

[5] http://www.google.com

[6] http://www.yahoo.com

Thus the system should also return tacit knowledge that can be inferred based on the represented knowledge. In order to achieve the desired reasoning, knowledge needs to be perfectly represented with a formal semantics (Bloehdorn, 2008).

2.3 Existing Knowledge Structures

For various purposes many knowledge structures are used in different communities to represent knowledge. There are some attempts to categorise these knowledge structures into different types, according to the degree of formality, complexity of the structure, and expressivity of the language used to define them (Giunchiglia et al., 2006).

For example, (Giunchiglia & Zaihrayeu, 2009) considered all knowledge structures that make few or no use of formal axioms and constraints as *lightweight ontologies,* and they divided them into two main types based on the degree of formality and expressivity: *informal lightweight ontologies* (also known as terminological ontologies (Gamper et al., 1999)), such as glossaries, and *formal lightweight ontologies*, such as formal ontologies. They also divided them into two main types based on their usage: *descriptive lightweight ontologies* (for defining the terms and the domain nature, such as thesauri) and *classification lightweight ontologies* (for classifying and accessing huge collections of data, such as faceted classifications).

McGuinness divides knowledge structures into two main types based on the ability of a knowledge structure to clearly identify concepts: *simple ontologies* and *structured ontologies*. Simple ontologies use natural language to define terms and they are mainly developed for human use, while structured ontologies are considered for both human and machine interaction by adding the ability to reason (McGuinness, 2003).

A wide range of formalizations or structures have been proposed to represent knowledge, which serves various fields, such as information science, AI and the semantic Web. However, dealing with the organization of the digital knowledge space by many fields creates confusion in the terminology used in knowledge structures (Gilchrist, 2003). Consequently the sheer range of various works in knowledge structures raises the probability that knowledge structures have being introduced with no common understanding of their definition, functions and implementation (Gruninger et al., 2008).

This subsection presents a systematic overview of common knowledge structures using a simple framework. This framework helps to understand and describe the knowledge structures, and thus to distinguish between them. The framework (Table 2-1) consists of four components, including: Definition, Intended Purpose, Main Drawbacks and Example of Use.

Table 2-1: Framework of knowledge structures description

Component	*Description*
Definition	to define the knowledge structure in terms of its main elements and relationships
Intended Purpose	to identify the main purpose of a knowledge structure and list its intended purpose
Main Drawbacks	to point out the main disadvantages in terms of semantics, structure and cost.
Example of Use	to mention some examples of usage in practice

A comprehensive description of the key existing knowledge structures based on the framework above will be presented in the following subsections.

2.3.1.1 Glossaries

Definition: A glossary is an alphabetical list of terms in a specific domain with their definitions. Similar to an index, a glossary may also have other fields, including: *see* and *see also* references to related topics, and extra details relevant to the term itself like its language or pronunciation (Pepper, 2000).

Intended Purpose: A first step in constructing a domain of knowledge is to gather a list of relevant terms to the domain with their definitions; i.e. glossary. This first step aids to constitute the linguistic surface manifestation of domain concepts (Velardi et al., 2008).

Main Drawbacks: Like traditional classifications, glossaries share the issue of a high cost of development as building glossaries is time-consuming and costly, as well as being an incremental procedure that must be continuously maintained (Velardi et al., 2008).

Examples of Use: Traditionally, a glossary appears at the end of a book (sometimes at the beginning) and contains a list of terms that recently introduced, specialized or uncommon within that book. Large list of glossaries can be found at Glossarist[7].

2.3.1.2 Folksonomies

Definition: The term "Folksonomy" was first coined by the information architect Thomas Vander Wal through the AIfIA mailing list, to mean the widespread practice of tagging using freely chosen terms by people (Vander Wal, 2007). It is a combination of the words "folks" (people) and "taxonomy"; and taxonomy is a blend of "taxis" (classification) and "nomos" (management). (Gupta et al., 2011).

Note that various definitions for the term "Folksonomy" exist in the literature (Spiteri, 2007). However, Vander Wal, who coined the term, states that folksonomy is "the result of tagging of information objects by a user freely using keywords relevant to the object being tagged" (Vander Wal, 2007). As a result, folksonomy consists of the tag, the object and the user, without doing any further process. In this thesis, we keep the original meaning of folksonomy as it is proposed by Vander Wal and, for differentiation matters, we use "tag

[7] http://glossarist.com

hierarchies" to mean the result of acquiring latent hierarchical structures from folksonomies.

Intended Purpose: An attractive feature of folksonomies is its inclusiveness, as they reflect the users' ways of thinking in order to meet their needs. Furthermore, folksonomies give a great opportunity to study and analysis behaviour of users by observing how they annotate their own resources, e.g. (Farooq et al., 2007), (Lee et al., 2009) and (Golbeck et al., 2011). Although folksonomies was seen in its early stage as an approach that incurs a relatively high interaction cost for the general user (Hong et al., 2008), it has since been recognised as a simple approach for assigning metadata to web content compared to traditional indexing, such as controlled vocabularies. Folksonomies do not need professional indexers as they allow anyone to freely type tags or terms to a resource. In this way they can accommodate new concepts easily. Folksonomies can therefore make a considerable contribution to public catalogues, such as library catalogues, by allowing users to create and organize their own personal information in that catalogue (Spiteri, 2007).

Main Drawbacks: Folksonomies share the inconsistent structure problem that is inherited from uncontrolled vocabularies, which causes many problems such as homonymy (same spelling with different meanings), synonymy, and basic level variation (Mathes, 2004; Golder & Huberman, 2006; Guy & Tonkin, 2006; Andrews & Pane, 2013; Chen et al., 2014).

Examples of Use: Today there are many tools and applications that use a folksonomy approach to ask people assigning descriptions to the resources. For example, social bookmarking tools like Delicious[8], photo-sharing sites like Flickr[9], blogging sites like Blogger[10], social networking sites like Facebook[11],

[8] http://delicious.com

[9] http://www.flickr.com

[10] http://www.blogger.com

[11] http://www.Facebook.com

cataloguing sites like LibraryThing[12], Social news sites like Digg[13] and education sites like Southampton EdShare[14]. Moreover, many researchers propose to use folksonomies in creating other knowledge structures, like taxonomies (Kiu & Tsui, 2010) and ontologies (Fang et al., 2016), and also in building recommender systems (Godoy & Corbellini, 2016).

2.3.1.3 Controlled Vocabularies

Definition: A controlled vocabulary is an organized and finite list of terms (words, phrases or notations) that can be used for classification. It can be seen as a type of metadata to firstly tag resources, and then to effectively find them by browsing or searching (Lee–Smeltzer, 2000).

Intended Purpose: Controlled vocabularies assist authors to designate the appropriate terms to documents, which helps them retrieve information afterwards. Also, using controlled vocabularies avoids the possibility of inconsistency structures that may be caused, for example, by misspellings, ambiguity, synonyms or homonyms (Macgregor & McCulloch, 2006).

Main Drawbacks: A key drawback of controlled vocabularies is the high cost of development and maintenance. A controlled vocabulary method requires experts to develop it, an agreed view of the domain, and skilled users (Quintarelli, 2005).

Examples of Use: A controlled vocabulary can be a universal scheme like Phonebook Yellow Page listings, to a specific discipline like the National Library of Medicine classification and subject headings (MeSh), or a custom scheme for a specific system like Amazon[15].

---

[12] http://www.librarything.com

[13] http://digg.com

[14] http://www.edshare.soton.ac.uk

[15] http://www.amazon.com

2.3.1.4 Taxonomies

Definition: Taxonomy is typically a controlled vocabulary with a hierarchical structure between the terms. This hierarchical structure describes a term by making its parent-child relationships with other terms explicit (Garshol, 2004). A formal taxonomy needs a complete structure of both categories and perspectives (Parunak, 1993). The nodes of taxonomy can be grouped into categories, based on their characteristics, and these categories can also be formed in different views.

Intended Purpose: Taxonomies allow things to be categorised using a flexible level of detail and clarify the relationship between them. This relationship enriches the semantics of controlled vocabularies so the semantics of a term can be captured by humans and mechanises through analysing the relationship between the term and the terms around it in the hierarchy (Cardoso, 2006). Taxonomy is the backbone structure of an ontology in which the relations are "is-a", while the rest of the ontology structure provides supplementary information about the related domain and may involve other relations such as "part-of", "located-in" and "is-parent-of" (Guarino, 1998).

Main Drawbacks: Similar to controlled vocabularies, taxonomies are costly in terms of time and effort, and require professional and skilled users. Furthermore, since taxonomies have a relationship between terms, some expertise in information structures is necessary. In addition, a taxonomy works best when the categories are pre-identified for a given domain and do not change over time; i.e. these taxonomies that are rigid and conservative (Quintarelli, 2005).

Examples of Use: Historically, taxonomies are used for classifying animals or plants by biologists based on a set of natural relationships. Another example is Dmoz[16], which is a directory of Web pages.

[16] http://www.dmoz.org

2.3.1.5 Thesauri

Definition: A thesaurus is a networked collection of controlled vocabularies. It is an extension of a controlled vocabulary by adding other forms relating to the terms and highlighting the semantic relationships between them (Garshol, 2004). According to (ANSI/NISO, 2005), the relationships used in thesauri are four different kinds: *equivalence* (synonymy: two terms have the same meaning), *homographic* (homonyms: two terms have the same spelling with different meanings), *hierarchical* (parent-child relationships), and *associative* (is-related-to relationships).

Intended Purpose: Thesauri help in enhancing the information retrieval tasks by indicating the semantic relationship between terms. In addition, thesauri promote consistency while using terms to annotate resources. Also, thesauri offer a way to convert the natural language of indexers, authors and users into controlled vocabularies used for classifying and retrieval (ANSI/NISO, 2005).

Main Drawbacks: Since a thesaurus is a networked collection of controlled vocabularies and an extension of a taxonomy, it inherits the same cost issues, for both time and effort. It requires professionals to accurately develop it, with predefined categories, homogeneous and stable items, and skilled users to perfectly use it. Although thesauri provide broader/narrower term specifications, which allows for simple hierarchy to be deduced, they do not usually supply explicit hierarchies (McGuinness, 2003).

Examples of Use: There are general thesauri such as: Roget's Thesaurus[17] and UNESCO Thesaurus[18]; and specialist thesauri such as: European Thesaurus on International Relations and Area Studies[19], Thesaurus for Graphic Materials[20] and British Education Index Thesaurus[21].

[17] http://machaut.uchicago.edu/rogets

[18] http://www2.ulcc.ac.uk/unesco

[19] http://www.einiras.org/services/eurothesaurus.cfm

[20] http://www.loc.gov/pictures/collection/tgm

[21] http://www.leeds.ac.uk/educol/BEID.html

2.3.1.6 Conceptual Graphs

Definition: Conceptual graph is a logic-based knowledge representation introduced by John F. Sowa in (Sowa, 1976), based on existential graphs invented by Charles Sanders Peirce, to represent the conceptual schemas used in database systems. Sowa extended his work in (Sowa, 1984) to apply it to a wide range of fields, including: AI, computer science and information science. Conceptual Graphs can be considered as a set of formal languages whose objects are graphs and whose deductions can be computed through graph-based operations (Mugnier, 1992). Conceptual graphs are based upon the following general form:

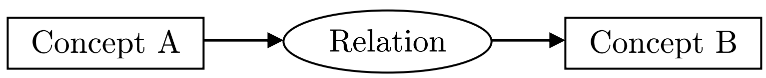


Figure 2-1: The general form of conceptual graphs

This could be read as: "The relation of a Concept A is a Concept B". The direction of the arrows determines the reading direction (Polovina, 2007). A well-known and broadly used example of conceptual graphs is illustrated in Figure 2-2.

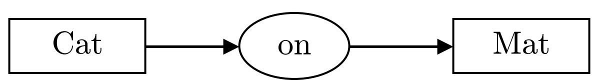


Figure 2-2: "Cat on Mat" example

A conceptual graph is a bipartite graph; i.e. all arrows either go from a concept to a relation or vice versa; from a relation to a concept.

Intended Purpose: Although conceptual graphs has been introduced to represent the conceptual schemas used in database systems, they also can be used for knowledge representation, reasoning and natural language processing (Sowa, 2008).

Main Drawbacks: Building and maintaining a conceptual graph is costly in terms of time and effort, and its rules are not only hard to create but also hard to use them in different domains (Crampes & Ranwez, 2000; Zhang & Yu, 2001).

Examples of Use: The Cogitant library[22] is a set of C++ classes to easily develop applications based on conceptual graphs. Another example is CoGui[23], which is a free graph-based visual tool for constructing conceptual graph knowledge structures represented, and compatible with the Cogitant. Also, conceptual graphs can be used for the process of creating, comparing and merging ontologies (Corbett, 2004).

2.3.1.7 Ontologies

Definition: In (Gruber, 1993), the notion of an ontology is originally defined as an "*explicit specification of a conceptualization*", and in (Borst., 1997), it is defined as a "*formal specification of a shared conceptualization*". Borst's definition emphasises that the conceptualization should be expressed in a formal machine readable way and share a common view between all parties that deal with that ontology. Typically, an ontology consists of some representational primitives, including: classes (sets), attributes (properties), relationships and information about the meaning of these representational primitives and their constraints (Gruber, 2009). Ontologies are considered as taxonomies but with richer semantic relationships between terms and attributes, and also solid rules to identify terms and relationships (Guarino, 1998).

Intended Purpose: Ontologies offer a universal knowledge representation of heterogeneous content and widespread understanding of a specific domain. They enable the domain to be communicated between human beings and machines (d'Aquin & Noy, 2012). Ontological analysis identifies the structure

---

[22] http://cogitant.sourceforge.net

[23] http://www2.lirmm.fr/cogui

of knowledge for a given domain, and thus shapes the heart of a knowledge representation system for that domain (Chandrasekaran et al., 1999).

Main Drawbacks: Although ontologies can be considered as the backbone of the semantic web, building and maintaining ontologies is so costly and time-consuming that it obstructs the progress of the Semantic Web development (Horrocks, 2013; Petrucci, 2015).

Examples of Use: There are many published ontologies in the Web such as: Good Relations[24] (an ontology for describing online products), Friend of a Friend[25] (FOAF; an ontology for describing people, their activities and their relations) and Gellish English dictionary[26].

2.3.1.8 Summary of Existing Knowledge Structures

Table 2-2 summarises the Key of Existing Knowledge Structures, which have been presented in (Section 2.3).

[24] http://purl.org/goodrelations

[25] http://www.foaf-project.org

[26] http://www.gellish.net

Table 2-2: Summary of the key of existing knowledge structures

Knowledge Structure	Definition	Intended Purpose	Main Drawbacks	Example of Use
Glossary	List of terms in a specific domain with their definitions	A first step in constructing a domain of knowledge	Time-consuming and costly development and maintenance	Glossary appears at the end of a book and glossarist.com
Folksonomy	A user-generated categorization by collaboratively tagging online resources	Cheap and inclusive; they reflect users' ways and meet their needs	Inconsistent structure problem	Most Web 2.0 applications, like Delicious and Flickr
Controlled vocabulary	An organized and finite list of terms (words, phrases or notations) that can be used for classification	Assisting authors; Avoiding inconsistency structures.	Time-consuming and costly development and maintenance	DDC (universal), MeSh (domain) and Amazon (system)
Taxonomy	Controlled vocabulary with a hierarchical structure (parent-child relationships), Formal taxonomy needs a complete structure of both categories and perspectives.	Enriching the semantics of controlled vocabularies, and it is the backbone structure of any ontology. Formal taxonomy is more productive.	Time-consuming and costly development and maintenance	animals/plants classifications and Dmoz.org
Thesaurus	An extension of controlled vocabulary by arranging the terms with more semantic relationships	Enhancing the information retrieval function and promoting consistency	Time-consuming and costly development and maintenance	WordNet (general) and British Education Index (specific)
Conceptual graph	A logic-based knowledge representation based on existential graphs	Based on first-order logic, humanly readable and computationally tractable	Time-consuming and costly development and maintenance, and its rules are hard to create and to use them in different domains	The Cogitant library and CoGui tool
Ontology	A formal specification (classes, properties, relationships, constraints and rules) of a shared conceptualization	A universal KR of heterogeneous content of a specific domain, and human beings & machines readable	Time-consuming and costly development and maintenance	Dublin Core, FOAF and Gellish English dictionary

2.4 Chapter Summary

This chapter has overviewed knowledge structures as a main part of KR and a key player in various fields, such as information science, AI and the semantic Web. An appropriate structure is required for expressing knowledge by declaring a set of concepts of that knowledge and their relationships. The motivations of knowledge structures are expressed and a framework for describing knowledge structures is introduced. This framework is used to describe and compare the difference between the key existing knowledge structures. Also, the comparison helps to determine the position of a new knowledge structure in terms of cost, structure and semantic level compared to the key existing knowledge structures.

The following chapter will highlight Folksonomy, as one of the key existing knowledge structures, and look at how it is possible to use the power of collective intelligence to build and improve them.

Chapter Three

Tag Hierarchies Construction and Evaluation

In recent years, folksonomies (Section 2.3) have emerged as an alternative approach to traditional classifications, e.g. taxonomies, of organising information (Kiu & Tsui, 2010; Gupta et al., 2011; Strohmaier et al., 2012; Zahia & Mohamed, 2013). They benefit from the power of collective intelligence to offer a simpler, cheaper and more flexible approach to organising web resources. However, they share the inconsistent structure problem that is inherited from uncontrolled vocabularies (Section 2.3). As a result, much research work has been done on resolving this problem by acquiring latent hierarchical structures from folksonomies and constructing common tag hierarchies like (Heymann & Garcia-Molina, 2006; Solskinnsbakk & Gulla, 2010; Benz et al., 2010; Oramas, 2014; Almoqhim et al., 2015; Fang et al., 2016).

3.1 Tagging Motivations

Many studies on collaborative tagging systems discovered that users have different tagging motivations across and even within tagging systems, which has an impact on the properties of resulting folksonomies (Nov & Ye, 2010; Strohmaier et al., 2012; Wu et al., 2016). (Sen et al., 2006) suggest three factors that may affect the selection of tags: user's personal tendency, community influence and the tag recommendation algorithm that suggests tags to users while they are tagging. Furthermore, (Marlow et al., 2006) found that some of tagging motivations are influenced by the system design and the way by which users are exposed to inherent tagging practices. For example, some researchers found that the key motivation for Flickr users is social factors (Marlow et al., 2006; Ames & Naaman, 2007), whereas others believed that Delicious users are motivated mostly by personal interests (Hammond et al., 2005; Golder & Huberman, 2006).

(Marlow et al., 2006) claim that the motivations of tagging can be divided into two general categories: organisational and social practices. The first expresses the use of tagging as an alternative approach to traditional classifications; whereas the latter arises from the social interaction nature with others, where users add some additional information, and express themselves and their opinions about the tagged resources. Similarly, (Hammond et al., 2005) call these two practices as altruistic and selfish. (Ames & Naaman, 2007) extend the work of (Marlow et al., 2006) by suggesting a classification of tagging motivations in two dimensions: function (organisation and communication) and sociality (self, friends/family and public). They conducted in-depth interviews with Flickr users and found that most of the users used the tags to organise resources for both themselves and the public, as well as to communicate with others.

(Gupta et al., 2011) list ten different types of tagging motivations, as follows: future retrieval, contribution and sharing, attract attention, play and competition, self-presentation, opinion expression, task organization, social signalling, money and technological ease. (Körner et al., 2010) distinguish between two main kinds of tagging users: categorisers (who's motivation is to categorise resources) and describers (who's motivation is to provide details of the resources).

3.2 Tagging Content

The early works on collaborative tagging were focusing on tagging content analysis, for instance: exploring the types of tags, their grammatical forms, and their usage (Golder & Huberman, 2006; Marlow et al., 2006; Spiteri, 2007).

(Gupta et al., 2011) list several types of tags, based on other works, such as: Content-Based Tags (to identify the actual content of the resource), Context-Based Tags (to provide the context of a resource, e.g., locations or time), Attribute Tags (to provide inherent attributes that cannot be derived from the content directly, e.g., author of a piece of content), Ownership Tags (to identify who owns the resource), Subjective Tags (to express user's opinion and emotion), Organisational Tags (to identify personal stuff, e.g., mywork; or to serve as a reminder of tasks, e.g., to-read), Purpose Tags (to support the information seeking task of other users, e.g., learn about LaTeX). Based on analysis of folksonomies from Delicious, (Munk & Mork, 2007) found that only a few terms dominate the set of tags assigned to a resource. These terms are basically consist of a number of broad and general content categories, which are common to all users.

(Spiteri, 2007) analysed tags extracted from three different folksonomies and observed that users use *nouns* for tagging (94% - 97% of tags, based on the selected folksonomy) more than other grammatical forms, such as adjectives. Similarly, (Guy & Tonkin, 2006) found that the majority (≈ 90%) of investigated tags in their samples from both Delicious and Flickr are nouns. (Spiteri, 2007) found also that users tagged *things* (76% - 90% of tags, based on the selected folksonomy) more than other types of concepts, such as activities or events.

3.3 Approaches for Tagging

While current tagging systems mostly support one type of tagging approaches, i.e, individual tags (Gupta et al., 2011), in some cases they allow the definition of a shallow hierarchy of user tags. For example, users in Delicious can group relevant tags into bundles. For instance, tags like "Java", "C#" and "Python" can be combined into a bundle that called "Programming Languages". Similarly, users in Flickr can group relevant photos into sets and also group relevant sets into collections. Also, GroupMe! allows users to tag and group relevant web resources into collections (Abel et al., 2008). While these systems allow users to group relevant tags together, other systems state the type of the relationship. For example, BibSonomy[27] allows users to make subtag/supertag relations between tags (Benz et al., 2010). Another example is Semdrops (Torres et al., 2011), which is a Firefox plugin that allows users to annotate web resources with different kinds of semantic tags including: category, property and attribute tags. As these systems allow users to create these relations by freely using tags, tag hierarchy learning from these shallow hierarchies still face some of the inconsistent structure challenges (Plangprasopchok et al., 2010b).

Commonly, no overall hierarchical representation of the tags is given in social tagging systems; therefore researchers have focused on developing approaches to extract hierarchical structures from individual tags that can be extracted from popular tagging systems, such as Delicious, Flickr, BibSonomy and CiteULike[28].

3.4 Approaches for Tag Hierarchies Construction

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 (Harris, 1968), e.g. (Cimiano et al., 2005) and (Faure & Nedellec, 1998); 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. (Hearst, 1992) and (Berland & Charniak, 1999). Many of the latter direction of the approaches have focused on the insight expressed by Hearst in (Hearst, 1992), that certain lexico-syntactic patterns (Table 3-1) can acquire a particular semantic relationship (hyponym/hypernym relationship) between terms in large text corpora (Snow et al., 2004).

Lexico-syntactic patterns can capture different semantic relations, though the hyponym/hypernym relationship seems to produce the most accurate results,

[27] http://www.bibsonomy.org

[28] http://www.citeulike.org

even with no pre-encoded knowledge. Additionally, they occur frequently in texts and across their genre boundaries (Hearst, 1992) and (Hearst, 1998).

Table 3-1 Hearst's lexico-syntactic patterns for detecting hyponym/hypernym relations.

No	*Pattern*	*Example*
1	**P** such as –C_1, C_2 ... , (and — or)″ C_n	**European countries** such as *England* and *Spain*.
2	Such **P** as –C_1 ,″ * –(or — and)″ C_n	…works by such **authors** as *Herrick*, *Goldsmith*, and *Shakespeare*.
3	C_1 –, C_n″ * –,″ –(or — and)″ other **P**	…*apple*, *orange*, *banana* or other **fruits**.
4	**P** –,″ including –C_1,″ * –or — and″ C_n	…all **common-law countries**, including *Canada* and *England*.
5	**P** –,″ especially –C_1,″ * –or — and″ C_n	…most **European countries**, especially *England*, *Spain*, and *France*.

The approaches based on linguistic patterns, however, are not appropriate to build semantic relationship between terms in the tags collections since these tend to be ungrammatical and are more inconsistent than text collections (Plangprasopchok et al., 2010b). Moreover, Strohmaier et al., in their comprehensive study of tag hierarchy construction algorithms, show that the algorithms tailored towards social tagging systems outperform the algorithms based on traditional hierarchical clustering techniques (Strohmaier et al., 2012).
Recently there have been several promising approaches proposed for constructing 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.

3.4.1 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 (Wu et al., 2006).

(Heymann & Garcia-Molina, 2006) propose an extensible greedy algorithm that automatically constructs tag hierarchies from folksonomies, extracted from Delicious and CiteULike. They use graph centrality (Wasserman & Faust, 1994) in the tag-tag co-occurrence network to identify the generality order of the tags. Their claim is that the tag with the highest centrality is the most general tag thus it should be merged with the hierarchy before others. (Benz et al., 2010) 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 data set gathered from Delicious, and showed that the performance of their extended algorithm outperforms the original algorithm. (Benz et al., 2011) have studied different measures of tag generality and found that the "popularity" of a tag seems to be a good proxy for "generality", and the degree centrality measure can differentiate well between abstract and concrete tags, whereas (Cattuto et al., 2008) found that the tag co-occurrence is a good measure to extract taxonomic relationships between tags.

(Schmitz et al., 2006) and (Schmitz, 2006) use statistical models of tag subsumption for constructing tag hierarchies. C. Schmitz et al adopted the theory of association rule mining to analyse and structure folksonomies from Delicious. P. Schmitz adapted the work of (Sanderson & Croft, 1999) to propose a subsumption-based model for constructing tag hierarchical relations from Flickr. (Schwarzkopf et al., 2007) extend the two algorithms in (Heymann & Garcia-Molina, 2006; Schmitz et al., 2006) by taking into account the tag context .

(Mika, 2007) 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.,

2007) extended the work of Mika while considering the user-user relationship. In particular, the first tag hierarchy in Mika's work is modified by considering tagging information of the user neighbours.

(Solskinnsbakk & Gulla, 2010) constructed tag hierarchies from folksonomies extracted from Delicious by using a combination of 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., 2010a) adapted affinity propagation introduced by (Frey & Dueck, 2007) to construct deeper and denser tag hierarchies from shallow personal hierarchies in Flickr. Yet (Strohmaier et al., 2012) have shown that generality-based approaches of tag hierarchy, with degree centrality as generality measure and co-occurrence as similarity measure, e.g. (Benz et al., 2010), have a superior performance compared to probabilistic models, e.g. (Plangprasopchok et al., 2010a).

(Rêgo et al., 2015) propose a binary classification approach to detect subsumption relations between tags in folksonomies, extracted from Bibsonomy. They claim that their approach also deal with class imbalance problem, i.e. the classifier is severely biased towards predicting the majority class. (Cai et al., 2016) propose an approach for building ontologies from folksonomies, extracted from Delicious, based on context-aware basic level concepts detection.

3.4.2 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., 2007) use WordNet to disambiguate and structure tags from Delicious. (Angeletou et al., 2008) 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[29] semantic search

[29] http://watson.kmi.open.ac.uk

engine. (Cantador et al., 2008) 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., 2008) use Wikipedia as an intermediate representation between tags, extracted from Delicious, and some semantic resources, namely: YAGO[30] and WordNet. (Garcia et al., 2009) propose an approach to automatically disambiguate polysemous, multiple related meanings, tags through linking them to DBpedia[31] entries. Likewise, (García-Silva et al., 2015) use DBpedia, as well as other knowledge resources: OpenCyc[32] and UMBEL[33], to construct domain ontologies from folksonomies, extracted from Delicious.

3.4.3 Hybrid Approaches

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

(Specia & Motta, 2007) present a semi-automatic approach that relies on clustering techniques and using WordNet and Google to structure tags, extracted from Delicious and Flickr. (Giannakidou et al., 2008) 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., 2009) propose an approach based on data mining techniques and WordNet concepts to discover the semantics in the tags and build tag hierarchies.

(Gu et al., 2015) present a supervised approach for tag hierarchy construction in open source communities by using co-occurrence tag networks and categories extracted from SourceForge[34]. (Joorabchi et al., 2015) use machine learning

[30] http://www.mpi-inf.mpg.de/yago-naga/yago

[31] http://dbpedia.org

[32] http://sw.opencyc.org

[33] http://www.umbel.org

[34] https://sourceforge.net

techniques and Wikipedia concepts to structure folksonomies extracted from a Q&A website, i.e. StackOverflow[35]. (Fang et al., 2016) present a framework consisting of three stages: concept discovery, concept relationship extraction, and concept hierarchy construction that uses clustering techniques and Wikipedia to build visual ontologies from annotated images extracted form Flickr.

Table 3-2 summarises the approaches for constructing tag hierarchies from folksonomies that are reviewed in our work.

[35] https://stackoverflow.com

Table 3-2: Summary of the main reviewed learning tag hierarchy approaches

Approach	Class	Data Source	Brief description
(Heymann & Garcia-Molina, 2006)	Clustering Techniques based Approaches	Delicious & CiteULike	They use graph centrality in the tag-tag co-occurrence network to identify the generality order of the tags; i.e. the tag with the highest centrality is the most general tag thus it should be added to the tag hierarchy before others.
(Schmitz et al., 2006)		Delicious	They used the theory of association rule mining to analyse and structure folksonomies.
(Schmitz, 2006)		Flickr	They adapted the work of (Sanderson & Croft, 1999) to introduce a subsumption-based model for building tag hierarchy.
(Schwarzkopf et al., 2007)		Delicious	They extend the two algorithms in (Heymann & Garcia-Molina, 2006) and (Plangprasopchok et al., 2010b) by taking into account the tag context.
(Mika, 2007)		Delicious	They present a graph-based model for constructing two tag hierarchies from folksonomies. 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., 2007)		Polyphonet	They extended the work of (Mika, 2007) while considering the user-user relationship. In particular, the first tag hierarchy is modified by considering tagging information of the user's neighbours.
(Solskinnsbakk & Gulla, 2011)		Delicious	They constructed tag hierarchies from folksonomies 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.
(Benz et al., 2010)		Delicious	They present an extension of (Heymann & Garcia-Molina, 2006) algorithm by applying tag co-occurrence as the similarity measure and the degree centrality as the generality measure. They succeed to produce clearer and more balanced tag hierarchies compared to the original algorithm.
(Plangprasopchok et al., 2010b)		Flickr	They adapted affinity propagation proposed by Frey & Dueck (Frey & Dueck, 2007) to build deeper and denser tag hierarchies from folksonomies.
(Laniado et al., 2007)	Knowledge Resources based Approaches	Delicious	They use WordNet to disambiguate and structure the tags.
(Angeletou et al., 2008)		Flickr	They present FLOR, an automatic approach for enriching folksonomies by linking them with related concepts in WordNet and online ontologies, using the Watson semantic search engine.
(Cantador et al., 2008)		Delicious & Flickr	They introduce an approach that maps the tags with Wikipedia concepts, and then associates those tags with domain ontologies.
(Tesconi et al., 2008)		Delicious	They use Wikipedia as an intermediate representation between the tags and some semantic resources (YAGO & WordNet)
(Garcia et al., 2009)		Flickr	They propose an approach to disambiguate homonym tags through linking them to DBpedia entries.
(Specia & Motta, 2007)	Hybrid Approaches	Delicious & Flickr	They present a semi-automatic approach rely on clustering techniques and using WordNet and Google to structure tags.
(Giannakidou et al., 2008)		Flickr	They introduce a co-clustering approach for identifying the tag semantics by clustering tags, and relevant concepts from WordNet.
(Lin et al., 2009)		CiteULike & Flickr	They propose an approach based on data mining techniques and WordNet concepts to discover the semantics in the tags.

3.4.4 Limitations of the Approaches

Although several approaches based on clustering techniques have been tried to structure folksonomies (Section 3.4.1), they come with limitation, which include the suffering from the "generality-popularity" tags 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 example, Plangprasopchok and Lerman found, on Flickr, that the number of photos tagged with "car" are ten times as many as those tagged with "automobile". By applying clustering techniques, "car" is likely to have higher centrality, and thus it will be more general than "automobile". Therefore, while tag statistics are an important source for constructing tag hierarchies, they are not enough evidence to discover concept hierarchies (Plangprasopchok & Lerman, 2009).

Knowledge resources based approaches (Section 3.4.2) have been developed to partially solve the limitations of clustering techniques approaches. However, the resources these approaches use are limited and typically can only deal with the standard terms (Lin & Davis, 2010). This limitation is due to the tags nature in which they may contain spelling errors, abbreviations, idiosyncratic terms etc. Furthermore, tags can be multi-lingual, which make these sources even harder to handle (Solskinnsbakk & Gulla, 2011). In fact, some researchers found collaborative tagging proposes an alternative solution to ontologies for creating and organising online knowledge (Shirky, 2005; Dix et al., 2006). Also other researchers found tag hierarchies are a great source for building and enriching ontologies, but not vice versa (Section 1.1).

3.5 Evaluation of Building Tag Hierarchies Approaches

Evaluating an approach to taxonomies construction is a major challenge since there is not, as yet, a golden evaluation dataset (Garcia-Silva et al., 2012; Strohmaier et al., 2012), nor a common evaluation methodology of hierarchical structures (Zheng et al., 2008; Yang & Callan, 2009; Andrews & Pane, 2013).

Thus the judgement of the existing approaches was often based on personal investigation of relevant experts by evaluating a portion of the produced taxonomies (Lin et al., 2009; Zhang et al., 2014). However, conducting precise evaluation of large taxonomies is enormously time-consuming, and sometimes almost impossible in reality. This thesis proposes a broad evaluation process of tag hierarchy construction approaches that adopted from relevant evaluation metrics and aims to assess three aspects:

- The quality of the learned tag hierarchies in terms of semantics.
- The expression (depth and width) of the learned tag hierarchies.
- The usability of the used tagging approach, in terms of efficiency, effectiveness and satisfaction.

These aspects can be evaluated by using four evaluation metrics. The details of these metrics are as follows:

3.5.1 Semantic Evaluation

Taxonomies are typically created towards a specific domain or application, and not only for representing knowledge. This makes different aspects of the taxonomy are more or less important, based on that domain or application. Thus, designing a common evaluation methodology for assessing the overall semantic quality of a taxonomy is a difficult task. However, (Dellschaft & Staab, 2006) state that using reference-based evaluation metrics is practically feasible for large scale taxonomies. And as an additional check for the validity of this metrics, human-based metrics is suggested to use, where human subjects are asked to judge the semantic quality of a subset of the learned taxonomic tag pairs.

3.5.1.1 Evaluation against Reference Taxonomy

This type of evaluation is performed by comparing how similar a produced taxonomy is to a related reference taxonomy. To perform the comparison between a produced taxonomy (PT) and a reference taxonomy (RT), a number of valuable measures have been proposed in the related literature review. Dellschaft and Staab propose two measures: taxonomic precision (tp) and taxonomic recall (*tr*) for comparing concept hierarchies (Dellschaft & Staab, 2006). The main idea is to compare the positions of two common concepts (*c*) in both hierarchies (local measure), and then to compare the two whole hierarchies (global measure). First, to compute the local measure, a concept that is present in both hierarchies is identified, and then characteristic excerpts (*ce*) of the concept are extracted. These excerpts contain the ancestors (super-concepts) and descendants (sub-concepts) of the concept that are present in both hierarchies. The position of the concept in both hierarchies will be similar if both excerpts are similar. Finally, to compute the global measure, all the local values are summed up over the common concepts in both hierarchies.

The local measure of taxonomic precision (*tp*) and taxonomic recall (*tr*) are mathematically defined, respectively, as follows:

$$tp(c, PT, RT) = \frac{|\mathrm{ce}(\mathrm{c}, \mathrm{PT}) \cap \mathrm{ce}(\mathrm{c}, \mathrm{RT})|}{|\mathrm{ce}(\mathrm{c}, \mathrm{PT})|} \qquad \textbf{Equation 3-1}$$

$$tr(c, PT, RT) = \frac{|\mathrm{ce}(\mathrm{c}, \mathrm{PT}) \cap \mathrm{ce}(\mathrm{c}, \mathrm{RT})|}{|\mathrm{ce}(\mathrm{c}, \mathrm{RT})|} \qquad \textbf{Equation 3-2}$$

Note that *tp* and *tr* are, in fact, the inverse of each other:

$$tp(c, PT, RT) = \frac{|\mathrm{ce}(\mathrm{c}, \mathrm{PT}) \cap \mathrm{ce}(\mathrm{c}, \mathrm{RT})|}{|\mathrm{ce}(\mathrm{c}, \mathrm{PT})|} = tr(c, RT, PT) \qquad \textbf{Equation 3-3}$$

The global measure of taxonomic precision (*TP*) is mathematically defined, as follows:

$$TP(PT, RT) = \frac{1}{|\mathrm{Cp} \cap \mathrm{Cr}|} \sum_{c \in Cp \cap Cr} tp(c, PT, RT) \qquad \textbf{Equation 3-4}$$

Where Cp is the set of the concepts in the produced taxonomy, and Cr is the set of the concepts of the reference taxonomy. The global measure of taxonomic recall (TR) is computed analogously. To give an overall overview and balance the values of TP and TR, taxonomic F-measure (TF) is computed as the harmonic mean of taxonomic precision and recall as follows:

$$TF(RT, PT) = \frac{2.\mathrm{TP}(\mathrm{RT}, \mathrm{PT}) \times \mathrm{TR}(\mathrm{RT}, \mathrm{PT})}{\mathrm{TP}(\mathrm{RT}, \mathrm{PT}) + \mathrm{TR}(\mathrm{RT}, \mathrm{PT})} \qquad \textbf{Equation 3-5}$$

3.5.1.2 Evaluation by Human Assessment

Evaluating the overall semantic quality of a tag hierarchy is a challenge even for skilled human subjects. For human-based evaluation, we adopt a simpler but effective approach that used by (Strohmaier et al., 2012) in the scope of this thesis. This approach should be feasible as a further check for the validity of the reference-based evaluation.

To use this metrics, a manageable subset of direct taxonomic pairs (t1, t2) from the learned tag hierarchy will be extracted to be manually judged as to whether they are related, and if they are, then how. The relation between each term pair can be one of the following options:

1. t1 is the same as t2.
2. t1 is a (kind of/part of) t2.
3. t1 is somehow related to t2.
4. t1 is not related to t2.
5. Unclear; because the meaning of t1 or t2 is not clear.

The insight behind this approach is that a better tag hierarchy will have a higher percentage of pairs being judged as “kind of” or “part of”, and a lower percentage of pairs being judged as “not related”.

3.5.2 Structural Evaluation

Although the measures mentioned above assess how consistent a produced tag hierarchy is, they do not measure the expression of the tag hierarchy – for example, a long chain of parent/child concepts may be semantically valid, but lacks the broad and symmetrical shape that makes a hierarchy useful for search or browsing. Therefore, this structural evaluation is considered as complementing the semantic evaluation presented in the previous section. The structural evaluation considers that the better tag hierarchy is a bushier and deeper hierarchy. In other words, the concepts in such hierarchies are broadly listed (hierarchy width), while each concept is branched in adequate detail (hierarchy depth).

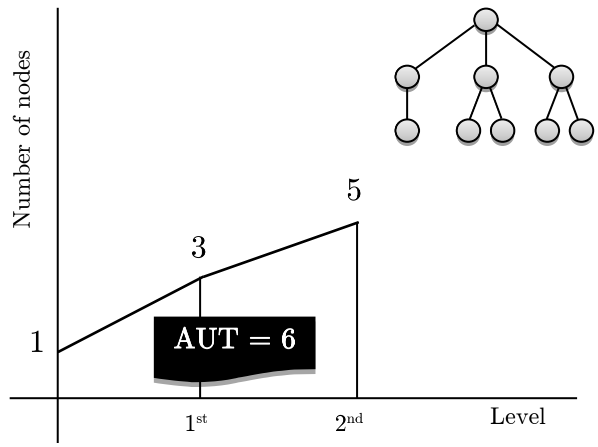


Figure 3-1: Example of calculating AUT metrics

To perform this evaluation, (Plangprasopchok et al., 2010b) introduces a simple measure known as Area Under Tree (AUT), which considers both the width and depth of taxonomy. To compute AUT for a hierarchy, the distribution of nodes in each level is computed first, and then the area under the distribution is calculated. For example, suppose that there is a tag hierarchy with the root, 3 nodes at the first level and 5 nodes at the second level. With the scale of hierarchy depth set to 1.0, AUT of this tag hierarchy would be: $0.5 \times (1 + 3) + 0.5 \times (3 + 5) = 6$ (a sum of trapezoids; as shown in Figure 3-1).

3.5.3 Usability Evaluation

Although most of collaborative tagging systems support one way of tagging (individual tags), in some cases they provide different tagging approaches (Section 3.3). The tagging approach in any social tagging system is a part of the process of building knowledge structures from that system. Thus, where a new interface is involved, the usability of the used tagging approach should be considered in evaluating an approach for tag hierarchy construction. For example, this thesis propose a new tagging approach (Section 4.1) that requires users to provide tag pairs in the form of "is-a" relationship, rather than individual tags, therefore, the usability of using the new approach should be measured. In other words, cognitive effort required from users to create tag pairs using the new approach should be measured compared to the normal approach (i.e. individual tags). This will give an indication whether the usability cost is acceptable for such new tagging approach or not.

For this evaluation measure, we will adopt the System Usability Scale (SUS), which is proposed by (Brooke, 1996), to use it as usability evaluation metrics in the scope of this thesis (Table 3-3) SUS is a Likert scale questionnaire, consisting of 10 items that is seen as a common standardized tool and has been used and verified in many domains (Greene et al., 2006). SUS is a simple and low-cost, but effective and reliable tool for evaluating usability in terms of efficiency, effectiveness and satisfaction (Ravendran et al., 2012), which ISO 9241-11 suggests to cover by any measure of usability (ISO, 1998). Efficiency is the level of the resources consumed by users, e.g. time, to perform a given task, whereas effectiveness is the ability of users to complete that task. Satisfaction is measured by users' subjective reaction to using the system.

According to Tullis and Stetson, who assessed the usability of two Web sites by using five usability surveys, SUS yields the most reliable results across a wide range of sample sizes, including small ones (Tullis & Stetson, 2004). Also, Bangor et al. analysed the results of many SUS surveys collected from difference usability evaluations for a ten year period and found that SUS was highly reliable and valuable over various interface types (Bangor et al., 2008). However, besides the usability of the product being assessed, SUS ratings are affected by the user experience, which can dramatically influence overall SUS scores by 15-16% between users who have "never" and "extensive" experience of that product (McLellan et al., 2012).

Table 3-3 Usability evaluation of tagging approach; adapted from (Brooke, 1996)

1. I think that I would like to use this approach frequently.
2. I found this approach unnecessarily complex.
3. I thought this approach was easy to use.
4. I think that I would need the support of a technical person to be able to use this approach.
5. I found the various functions in this approach were well integrated.
6. I thought there was too much inconsistency in this approach.
7. I would imagine that most people would learn to use this approach very quickly.
8. I found this approach very cumbersome to use.
9. I felt very confident using this approach.
10. I needed to learn a lot of things before I could get going with this approach.

Each statement in the survey has to be rated on a five-point scale of "Strongly Disagree" to "Strongly Agree". The survey yields a single score, from 0 to 100, representing the overall usability of the tagging approach being evaluated, whereas scores for individual statements are not meaningful on their own (Brooke, 1996). Note that Brooke did not determine when a SUS score is acceptable. Some researchers pointed out that it is more difficult to show a product is acceptable in terms of usability than if it is not. Bangor et al. found that a product with SUS scores below 50 will mostly have usability difficulties, whereas scores between 70 and 89, though promising, do not assure high acceptance of usability (Bangor et al., 2008).

3.6 Chapter Summary

This chapter and the previous chapter have covered the related literature concerning building knowledge structures based on collective intelligence. The knowledge structure that this research will focus on is Tag Hierarchy; since it is a result of a successful application of harnessing the power of collective intelligence.

This chapter has highlighted the aspects of collaborative tagging and tag hierarchies, which are acquired from folksonomies. In recent years many approaches have been offered for constructing tag hierarchies from collaborative tagging. These approaches can be seen in three directions based on using: clustering techniques, relevant knowledge resources or a hybrid of both. Generality-based approaches to building tag hierarchy outperform probabilistic models, and approaches should not rely on static knowledge resources due to their limitations. All these approaches, however, come with limitations and there is a need to improve the current work in order to gain high-quality tag hierarchies.

One of the most significant of these limitations is the "popularity-generality" tags problem, where generality-based approaches (sometimes inaccurately) assume that because a tag occurs more frequently it must be more general and thus appear higher in the hierarchy. To overcome this problem we propose a new tagging approach and algorithm that will be explained and tested in the following chapter.

Chapter Four

Building Tag Hierarchies from Crowdsourced Taxonomic Tag Pairs

This chapter describes a pilot study to explore the impact of gathering taxonomic tag pairs rather than individual tags in tackling the "popularity-generality" problem (Section 3.3). In other words, we propose a new tagging approach that takes the form of "is-a" relationship, where users should type two related tags; i.e. Tag t1 is a tag for the resource and Tag g1 is a generalization of Tag t1. And if Tag t1 is more popular than Tag g1, it will still be a subclass of Tag g1 by this new tagging approach. This simple relationship (Tag t1 is-a Tag g1) will not only help in tackling the "generality-popularity" tags problem, but will also provide more expressive of the tags than tags alone for constructing high-quality tag hierarchies.

In this tagging approach, instead of Folksonomy, the knowledge structure derived from the practice of the new approach of collaboratively tagging online resources by the crowd will be known as "TagTree". Since our tagging approach is new, a web-based prototype, the TagTree System, will be introduced to test it as well as to collect data for executing the experiment.

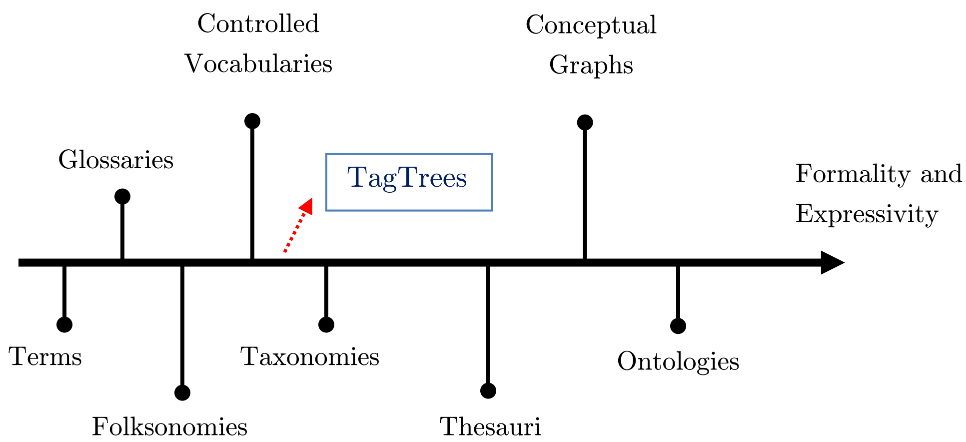


Figure 4-1: The expected structural level of TagTrees in Figure 1-1

In this new tagging approach, we propose making a change (tagging in the form of is-a relationship) to the current tagging approach (i.e, individual tags; Section 3.3) in order to make a big change to the type of knowledge structure that can be built (i.e. from *folksonomies* to *TagTrees*). This change will cope with the lack of a consistent structure in folksonomies (Section 2.3) by raising their structure (Figure 4-1) and then raising their semantic value, while keeping the interaction cost of the process down (Figure 4-2).

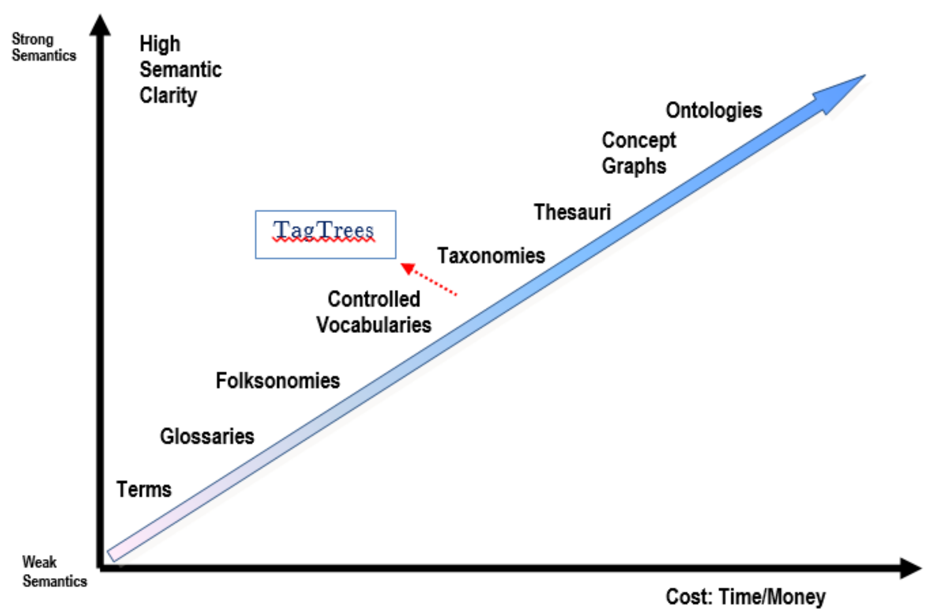


Figure 4-2: The expected semantic level of TagTress in Figure 1-3

This chapter will describe the proposed social tagging approach and then highlight the experimental design and implementation of the proposed system. Also, the chapter will identify the dataset and evaluation metrics we propose to use to evaluate and make a comparison between the tag hierarchies produced from the proposed approach (tag pairs) and the normal approach (individual tags) of social tagging.

4.1 Proposed Social Tagging Approach

The aims of the proposed tagging approach is to cope with the lack of a consistent structure in folksonomies by raising their semantic value, while keeping the good features of social tagging and folksonomies, as mentioned earlier (Section 2.3).

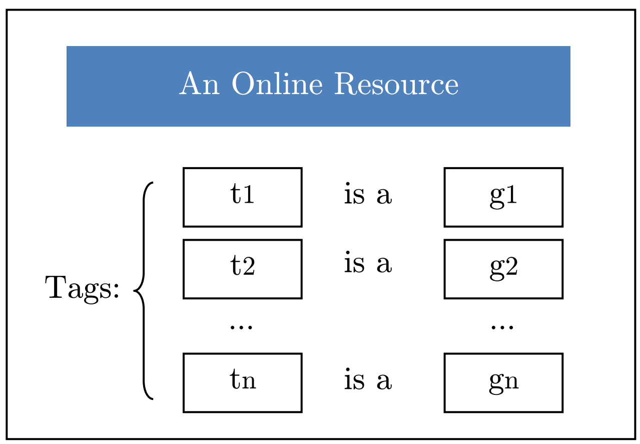


Figure 4-3: The new tagging approach

In the new tagging approach (Figure 4-3), the benefit of the power of collective intelligence will be extended by involving human knowledge in building the desired knowledge structure, TagTree. The user is required to tag the resource in the form of "is-a" relationship, where t1 (the left side) is a tag for the resource and g1 (the right side) is a generalization of t1, and could be a related tag to the resource as well. For example, a picture of The Tower of London can be tagged as follows: "Tower of London" is a "tower", "Tower of London" is a "London attraction", or "tower" is a "building". Also, the users can tag as much as they want for each resource in this way.

It should be noted that the new tagging approach has some limitations. One of these limitations is the lack of supporting some types of tags, such as "Organisational Tags", where users use some tags to identify personal stuff, e.g., mywork; or to serve as a reminder of tasks, e.g., to-read. Moreover, users may struggle to use non-noun tags, such as adjective tags, with the new tagging approach, though some researchers found that the majority of the used tags in collaborative tagging systems are nouns (Section 3.2). While the main aim of the proposed tagging approach is to cope with the lack of a consistent structure in folksonomies, it should not prevent users from freely choosing tags to meet their needs. One of the solutions that can be considered in the future work is to allow users to use both "*tag pairs*" as well as "*individual tags*" to tag the same resources. This will give the users more flexibility to reflect their ways of thinking and express their opinions, besides providing semantically rich tag pairs for constructing high-quality tag hierarchies. While the new tagging approach will share some of the issues of folksonomies, such as abbreviations and spelling variations and errors, it also raises the accuracy of the semantic relationship between tags and should help the related computational techniques (Section 3.4) in improving the quality of tag hierarchies.

4.2 Proposed Algorithm

As the key aim of this chapter is to explore the impact of gathering taxonomic tag pairs rather than individual tags in coping with the "popularity-generality" problem, we will take advantage of an existing algorithm to extend it for our purpose.

Based on the literature review of tag hierarchy learning (Section 3.4), the selected algorithm should meet the following criteria:

- It should be tailored towards the characteristics of social tagging systems.
- It should be based on generality-based approaches, with degree centrality as generality measure and co-occurrence as similarity measure, since they outperform others.
- It should not rely on static knowledge resources due to their limitations.

Algorithm 4-1: Pseudo-code for the proposed algorithm for building tag hierarchies from crowdsourced taxonomic tag pairs, adopted from (Heymann & Garcia-Molina, 2006) and (Benz et al., 2010) to deal with “tag pairs”, instead of “individual tags” (Lines 1-4, 6 and 19).

***Input**: Paris of user-generated tags in the form of “t is-a g”*

***Output**: tag hierarchy*

Functions: Several functions are assumed: sim(t*i*,t*j*): Calculate the similariy (using the co-occurrence weights as similarity measure) between t*i* and t*j*. GetNodes(*TagHierarchy*): returns all nodes in the given hierarchy, *TagHierarchy*. AppendToNode(*TagHierarchy*,t*i*,t*j*): append t*j* underneath t*i* in the given hierarchy, *TagHierarchy*. GetParent(t,g): return tag g. GetChild(t,g): return tag t.

Parameters: Several parameters are required to be set: tag occurrences threshold (*occ)*, tag-tag similarity threshold (*min˙sim)*, tag generality threshold (*min˙gen)*.

```
1.  TagPairs = Filter the tag pairs “t, g” by an occurrence threshold occ.
2.  TagPairs = Order the tag pairs in descending order by generality (measured
    by degree centrality in the g–g co-occurrence network).
3.  TagHierarchy = {∅, root= GetParent(t,g)}
4.  AppendToNode(TagHierarchy, GetChild(t,g), root)
5.  for i = 1 … |TagPairs| - 1 do
6.      ti = GetParent(TagPairs [i])
7.      MostSimilarVal = 0.
8.      for all tj ∋ GetNodes(TagHierarchy) do
9.          if sim(ti,tj) > MostSimilarVal then
10.             MostSimilarVal = sim(ti,tj)
11.             MostSimilar = tj
12.         end if
13.     end for
14.     if MostSimilarVal > min˙sim and MostSimilarVal < min˙gen then
15.         AppendToNode(TagHierarchy, ti, MostSimilar)
16.     else
17.         AppendToNode(TagHierarchy, ti,root)
18.     end if
19.     AppendToNode(TagHierarchy, GetChild(t,g), ti)
20. end for
```

Benz's algorithm has a superior performance compared to other state-of-the-art tag hierarchy induction algorithms based on a comparative study introduced by Strohmaier et al. (Section 3.4.1). Consequently, the proposed algorithm will be an extension of Benz's algorithm, which itself is an extension of Heymann's algorithm. Algorithm 4-1 demonstrates the pseudo-code for the proposed algorithm.

4.2.1 Description of the Algorithm

The algorithm starts by filtering the tag pairs (extracted from the folksonomy dataset) by an occurrence threshold *occ* (Line 1). Then, it orders the tag pairs in descending order by generality that measured by degree centrality in the (g–g) co-occurrence network (Line 2). After that, the algorithm starts with the most general tag pair and consider the hypernym of this pair (g) as the root node of the hierarchy, and then append the hyponym of the pair (t) underneath the root (Line 3-4). Then, it adds each tag t*i* (g) in the tag pair list subsequently to an evolving tag hierarchy (Lines 5-7). It decides where to add each tag t*i* (g) by calculating its similarity (using the co-occurrence weights as a similarity measure) to each tag currently present in the hierarchy t*j*, and appends the current tag t*i* (g) underneath its most similar tag *MostSimilar*. If t*i* (g) is very general (determined by a generality threshold *min'gen*) or no sufficiently similar tag exists (determined by a similarity threshold *min'sim*), the algorithm appends t*i* (g) underneath the root node of the hierarchy. Then it appends the hyponym of the current tag pair (t) underneath t*i* (g) (Lines 8-19). Finally, the algorithm applies 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 process is done by the steps in (Lines 5-20).

Compared to the original algorithm (Heymann-Benz algorithm), the proposed algorithm extends the benefit of the power of collective intelligence by involving human knowledge in learning higher quality tag hierarchies. Nearly a half of the resulting tag hierarchy is created by the agreement of the crowd since the adding iteration step (Lines 4 and 19) is based on a pair of tags with an explicit semantic relation, which should increases the performance of the resulting hierarchies. Although the comparative study of Strohmaier et al. skipped the pre-processing step, which is used in Heymann-Benz algorithm to deal with synonym or ambiguous tags, the proposed algorithm can resolve these problems without extensive pre-processing, as will be shown in the following section.

4.2.2 Settings of the Algorithms

The original and our algorithms are affected by several parameters, and here is a brief description of the settings used to run them:

— **Tag Occurrences Threshold *occ***: In general, the bigger of the number of tag pair/tag occurrences, the stronger agreement between users on a proper view of the shared context. As the size of the sample study in this pilot experiment is small, we have chosen to include tags occurring more than 2 times for the "tag pairs" dataset, and 4 times for the "individual tags" dataset. Increasing these thresholds led to an inadequate number of tag pairs/tags for building tag hierarchies.

— **Tag-Tag Similarity Threshold *min_sim***: Each candidate tag from the folksonomy is appended as a child of the most similar node in the hierarchy if its similarity to that node is greater than a similarity threshold, otherwise it is appended to the root of the hierarchy. The best value for this threshold were 0.011 for the "tag pairs" dataset, and 0.013 for the "individual tags" dataset. Lowering the thresholds led to relatively unrelated tags, while made it higher resulted in an unbalanced hierarchy as many tags were appended to the root of the hierarchy.

— **Tag Generality Threshold *min_gen***: Generality in the original and our algorithms is measured by degree centrality in the tag–tag co-occurrence network. The best value for this threshold were 0.1 for the "tag pairs" dataset, and 0.2 for the "individual tags" dataset. A child of a node is also a child of the root of the hierarchy, however, lowering the threshold resulted in an unbalanced hierarchy as many tags were appended to the root of the hierarchy.

4.3 TagTree System

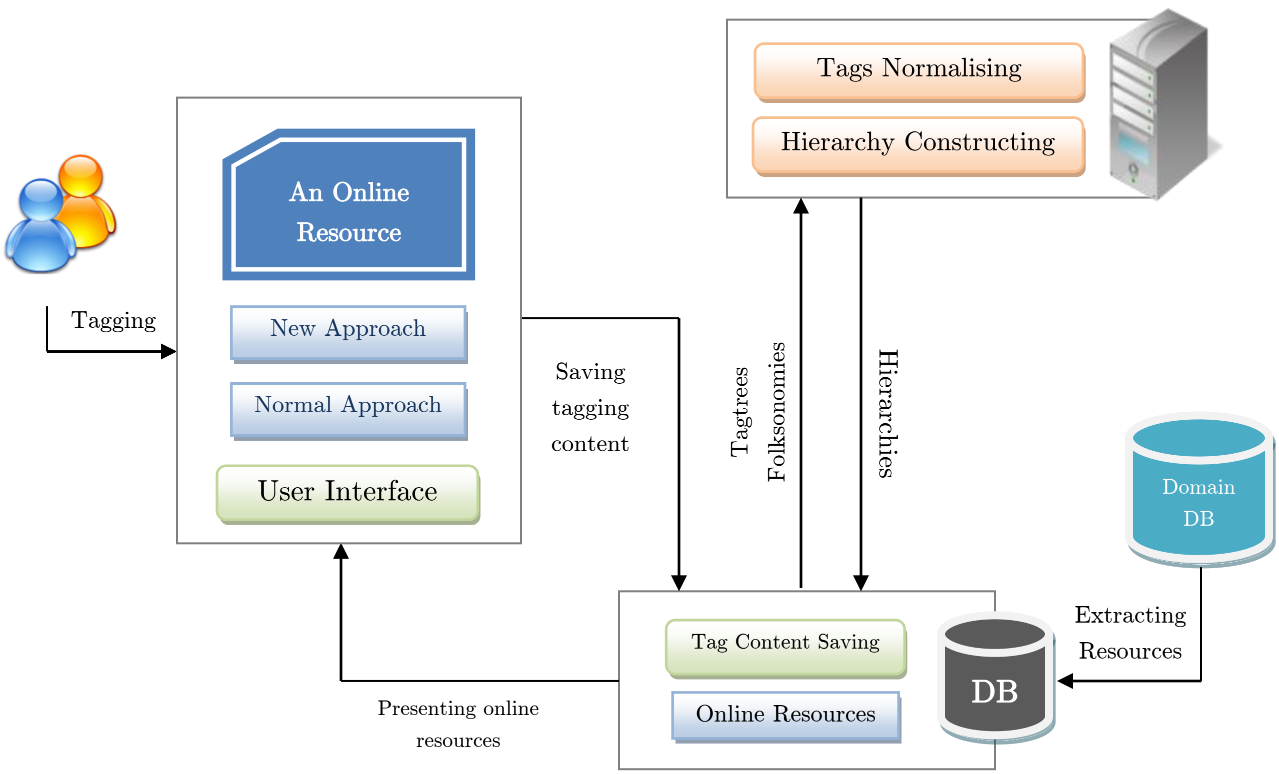


Figure 4-4: The architecture of the TagTree system

To test the proposed tagging approach and collect data for executing the experiment, the TagTree System was created. It is a web-based prototype which aims to build a high level knowledge structure (TagTree) through allowing the participants to tag some online resources by using the new approach of social tagging, as well as the normal approach; i.e. individual tags. Figure 4-4 illustrates the architecture of the TagTree system.

To help us decide how to design the page of the resources and present them, a focus group with 10 participants was conducted. Three options were suggested for how to present the resource, including: 1) present only image of the attraction, 2) present image and the name of the attraction, or 3) present image, the name and the description of the attraction. After that, the participants were divided into two small groups to evaluate and discuss the three options. Based on this, the last option (Figure 4-5) was unanimously nominated by the participants to be the presentation of the resources, as it is the best in terms of clearness and generating a bigger number of related tags to the resources.

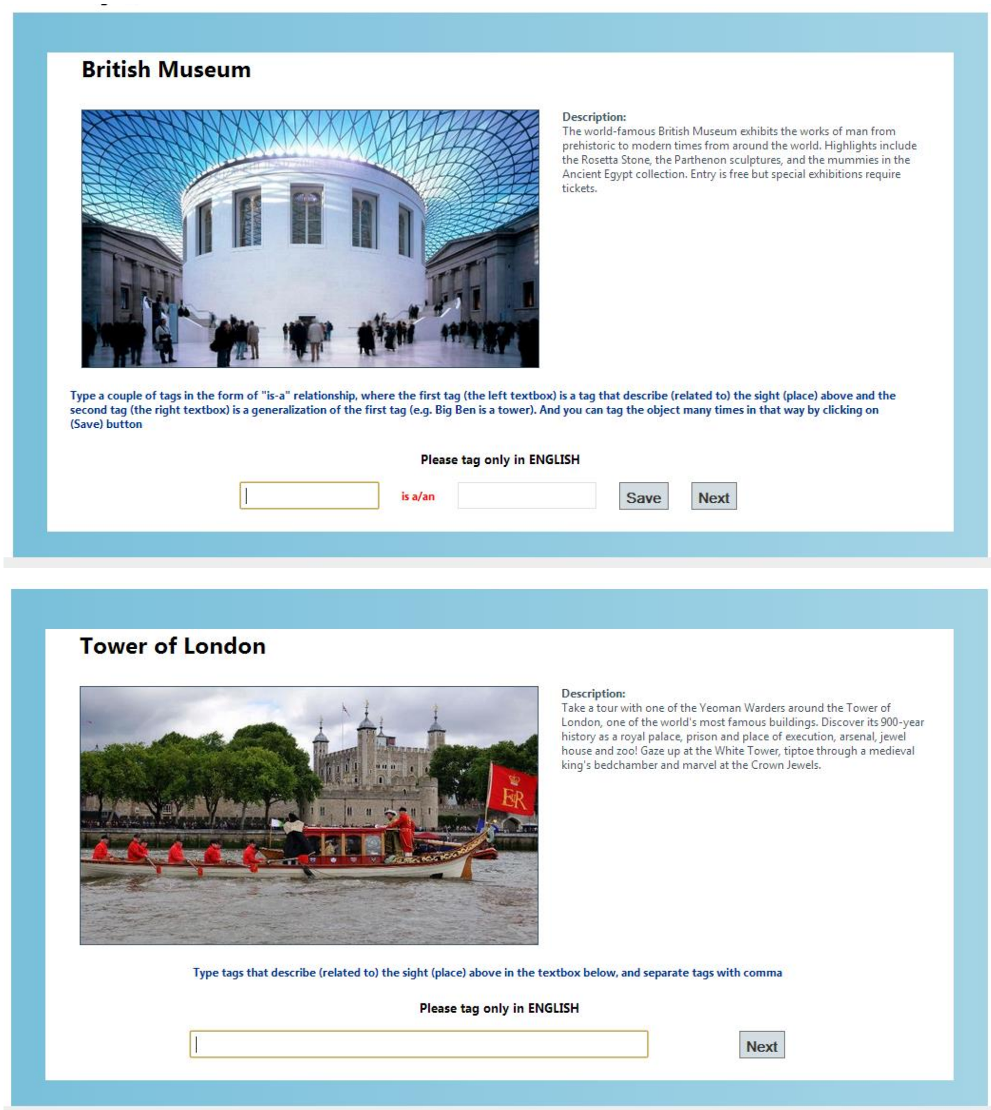


Figure 4-5: The new (top) and the normal (bottom) tagging approaches

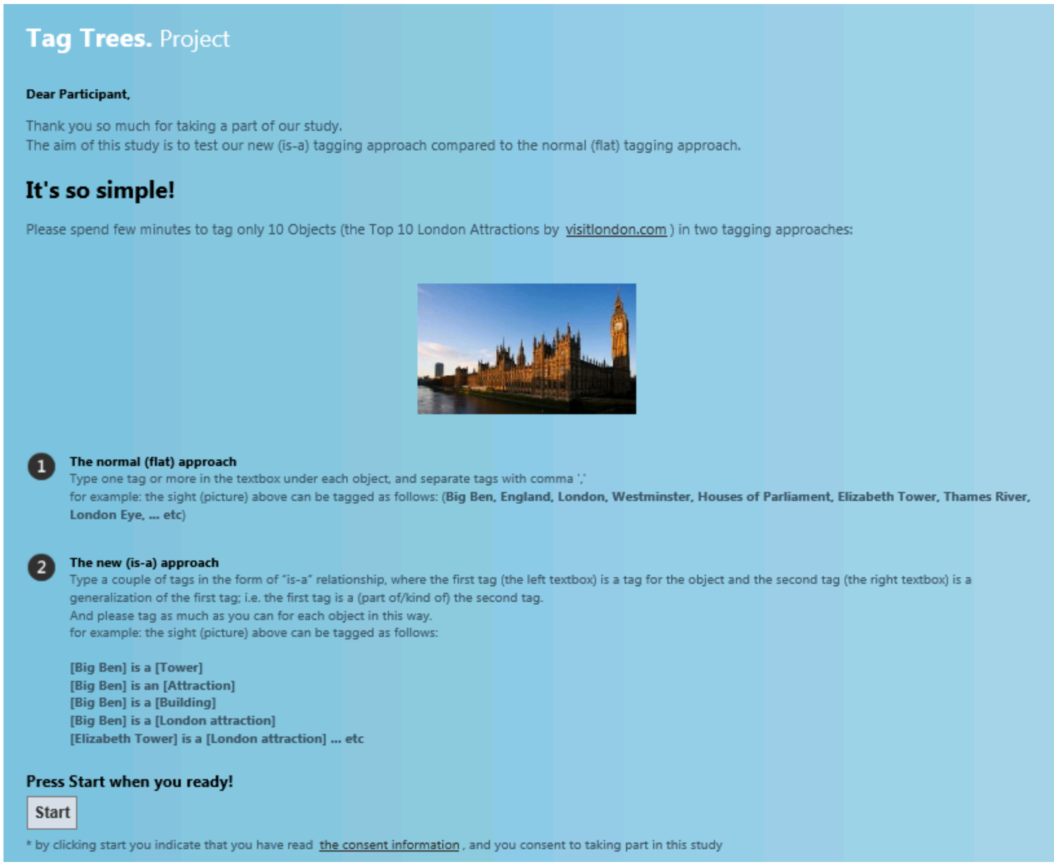


Figure 4-6: The main page of the TagTree system

The TagTree System consists of four main parts: User Interface, Tag Content Saving, Tags Normalising, and Hierarchy Constructing. Each part is described below:

User Interface: Figure 4-6 shows the main page of the TagTree system. It describes to users how to use the new tagging approach, with an example, and allows them to tag some resources (for more details see Section 4.4) by using the new tagging approach, "tag pairs", and the normal tagging approach, "individual tags". By clicking on "Start" button users can see and tag the resources. Each resource is presented to the users in a separate page that includes an image, the name and the description of the attraction. (Figure 4-5).

Tag Content Saving: This component receives the tag content from users and saves it into the system database. The tag content includes: user ID (user session), tags and time spent for each tagging action by the user.

Tags Normalising: Before running the Hierarchy Constructing component, the tags are passed on to the normalisation process that we implemented and several filters for cleaning the tags are applied, including: Letters Lower-case, Non-English Deleting and Stop Words[36] Removing.

Hierarchy Constructing: This component uses the adopted and original algorithms to build the tag hierarchies from tagtrees and folksonomies.

4.4 Datasets

Since the research proposes a new tagging approach, the TagTree system is implemented to gather the data collection of the experiment. The Top 10 London Attractions[37], elected by visitlondon.com, is selected to be the resources used in the TagTree system for the following reasons:

1. "London Attractions" is a general and popular domain. The participants can freely use different types of tags (Section 3.2) to describe them.
2. The Top 10 London Attractions are good resources since they are elected based on visitor numbers, which means they are well known, so the target participants, who are mainly researchers and students at UK universities, are highly likely to effectively participate in the experiment.
3. Choosing a small size of related resources and using them for both tagging approaches, the new and the normal, helps in finding shared concepts in order to construct tag hierarchies from both approaches.

Table 4-1 shows a descriptive statistics of the data collection, after performing the normalisation process.

[36] Words with little meaning such as: "a", "am", "is", "and", "the" …etc

[37] http://www.visitlondon.com/things-to-do/sightseeing/tourist-attraction/top-ten-attractions

Table 4-1 Descriptive statistics of the data collection by the new (tag pairs) and normal (individuals tags) tagging approaches.

Approach	*Users*	*Tags*	*Resources*	*Tag assignments*
New Approach	215	275 tag pairs (from 235 tags)	10	333 tag pairs
Normal Approach	215	320 tags	10	550 tags

4.5 Evaluation Methodology

The link to the TagTree system, including the usability evaluation survey (Section 3.5.3), was disseminated via emails and social media, and was live for a period of five weeks. The Ethics form of the experiment has been approved by FPAS Ethics Committee at the University of Southampton (Reference No. 4864, on 5/12/2012; for more details see the Appendix).

A total of 215 participants, who are mainly researchers and students at UK universities, took part in this experiment, as a voluntary contribution to the study. Each participant was required to tag ten resources, Top 10 London Attractions (Section 4.4), by using the new tagging approach and the normal tagging approach; i.e. each approach will be used for tagging five resources. The order in which the participants use the tagging approaches is rotated, so that half the participants use the tag pairs approach first, and half use the normal approach first. Finally, the participant was asked to complete an online survey (SUS; Table 3-3) in order to evaluate each tagging approach in terms of usability.

For the scope of the research, two SUS statements have been removed since they are not appropriate to our evaluation, including:

- I found the various functions in this approach were well integrated.
- I thought there was too much inconsistency in this approach.

And have been replaced with the following two statements:

- I could express the ideas that I want to by using this approach.
- I was satisfied with the quality of what I wrote.

Two data sets were extracted from the TagTree system. The first one was collected by the new tagging approach and consists of tagged resources, tags pairs, users and tagging time spent durations. The second one was collected by the normal tagging approach and consists of tagged resources, individual tags, users and tagging time spent durations. In the experiment, three tag hierarchies are produced and in order to differentiate between them we gave each one of them a different name as follows:

1. **TagTree**: By using our extended algorithm and the first data set.
2. **Tag Hierarchy A**: By using the Heymann-Benz algorithm and the second data set.
3. **Tag Hierarchy B**: By using the Heymann-Benz algorithm and using the first data set in which {tag1 is-a tag2} is considered as individual tags, i.e. ignoring the "is-a" relations.

These three tag hierarchies are evaluated by the proposed evaluation metrics in section 3.5. This evaluation is needed to answer three questions:

1. **Which one of the two tagging approaches produces the highest semantics quality of tag hierarchies?** (evaluation metrics: 3.5.1.1 and 3.5.1.2)
2. **Which one of the two tagging approaches produces the most expressive tag hierarchies (hierarchy width and depth)?** (evaluation metrics: 3.5.2)
3. **Which one of the two approaches achieves the best usability evaluation, in terms of efficiency, effectiveness and satisfaction?** (evaluation metrics: 3.5.3)

To answer the first question we need to perform an evaluation against a reference taxonomy (evaluation metrics: 3.5.1.1). Unfortunately no suitable reference taxonomy for the purpose of the experiment domain was found. Therefore, two researchers from the WAIS group at the University of Southampton, who had no special knowledge of the domain (London Attractions) but did have expertise in knowledge structures (taxonomies and ontologies), were asked to create appropriate reference taxonomy of the experiment domain.

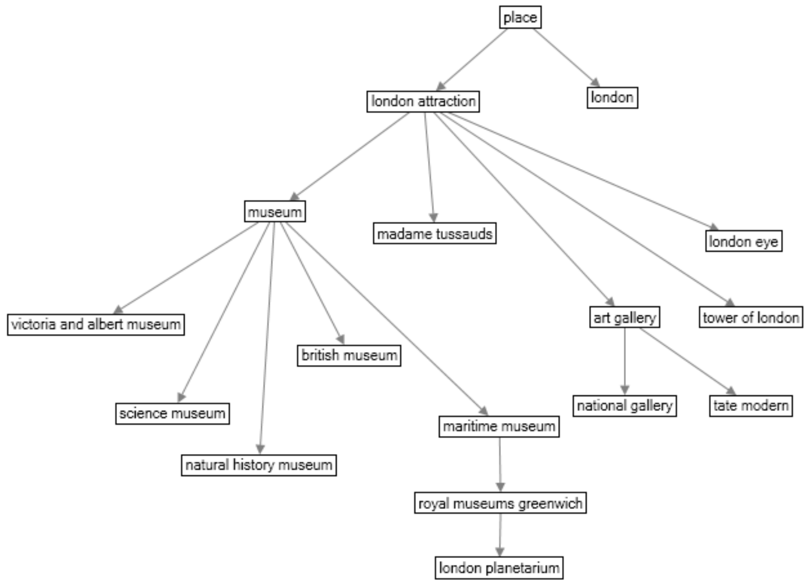


Figure 4-7: Reference taxonomy of the experiment domain; the arrows directions indicate the direction of the "is-a" relationship between tags, where tag1 → tag2 reads as tag1 is a hypernym of tag2.

They were asked to sit together and provided with a list that contains all unique tags from the produced tag hierarchies without relations, and they had the option to use them or not. They were not told the tags where extracted from and did not see the produced tag hierarchies. Based on this, they created a taxonomy that is shown in (Figure 4-7). While this taxonomy is not exhaustive, its nodes are relevant to the experiment domain and on inspection the relations between them can be seen to be semantically accurate. For the purpose of the experiment evaluation, with the limitation noted above, this taxonomy will be considered as a reference taxonomy to perform the evaluation against a reference taxonomy metrics (Section 3.5.1.1). And as an additional check for the validity of this, a human assessment of the semantic quality of the learned tag hierarchies will be performed.

Since the size of the sample study in this pilot experiment is small, we have asked three researchers from the WAIS group at the University of Southampton, who did have expertise in knowledge structures, to take part in this study. Their job was to judge to how two terms are related, and if they are, then how (Section 3.5.1.2).

4.6 Results and Analysis

In order to give a visual impression of the results, Figure 4-8, Figure 4-9 and Figure 4-10 depict the three produced tag hierarchies. It should be noted that these produced tag hierarchies obviously are not exhaustive and do not represent the whole domain of London Attractions, however, they should represent shared conceptualizations that are hidden in the folksonomies. Thus using these tag hierarchies, for example, to support searching for content will be useful if the user enters a query with keywords that are already part of that shared conceptualizations.

The main aim of this section is to evaluate the quality of the tag hierarchy semantics and structure that produced from tagtrees by using the proposed tagging approach (tag pairs), compared to the one that produced from folksonomies by using the normal tagging approach (individual tags). The following sections will show the results and discussions of this evaluation.

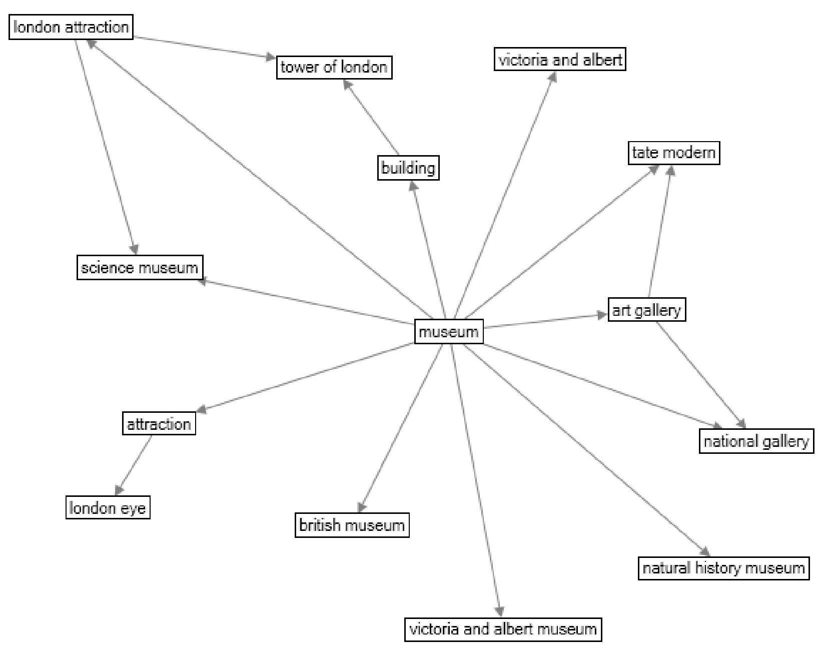


Figure 4-8: TagTree; the arrows directions indicate the direction of the "is-a" relationship between tags, where tag1 → tag2 reads as tag1 is a hypernym of tag2.

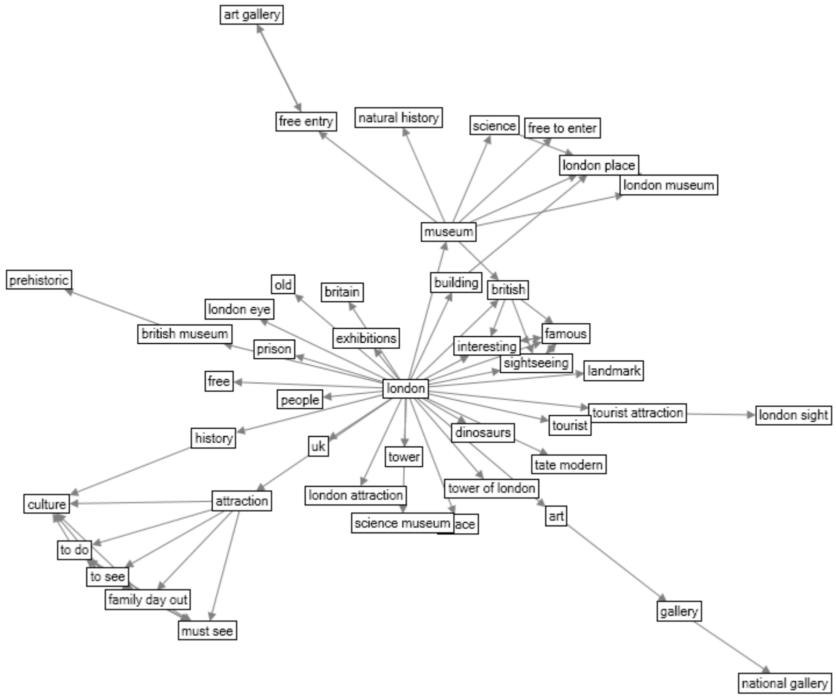


Figure 4-9: Tag Hierarchy A; the arrows directions indicate the direction of the "is-a" relationship between tags, where tag1 → tag2 reads as tag1 is a hypernym of tag2.

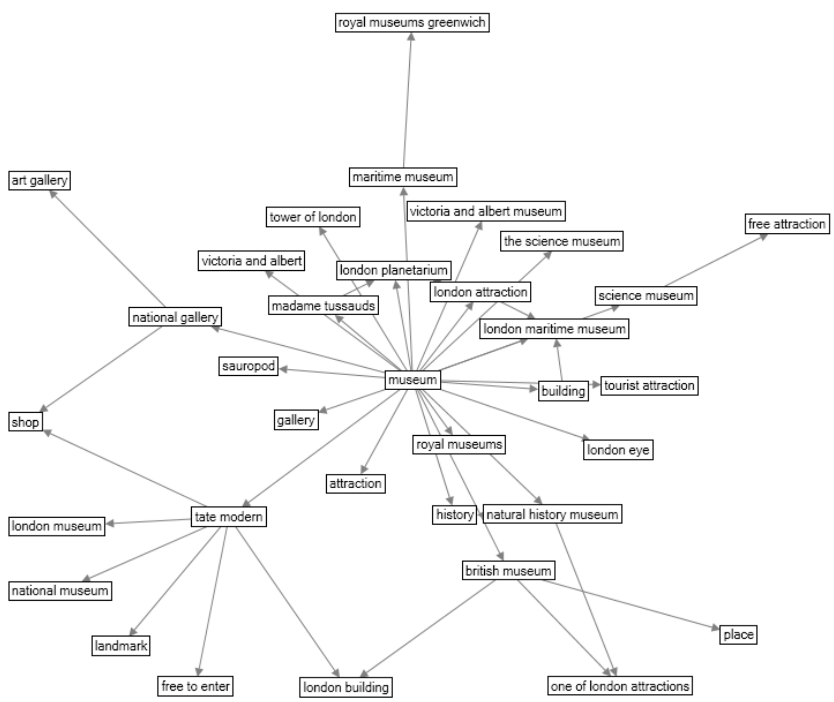


Figure 4-10: Tag hierarchy B; the arrows directions indicate the direction of the "is-a" relationship between tags, where tag1 → tag2 reads as tag1 is a hypernym of tag2.

4.6.1 Results of Semantic Evaluation

The semantic evaluation was undertaken in two steps. Firstly by comparing how similar a produced taxonomy is to a related expert-crafted taxonomy (as a reference taxonomy), and secondly by comparing a sample of relationships in the hierarchy to human judgment.

4.6.1.1 Reference-based Evaluation

Figure 4-11 shows the results of the semantic evaluation against the reference taxonomy. The y-axis illustrates the similarity between each tag hierarchy and the reference taxonomy. The similarity is measured by using several measures that are explained in Section 3.5.1.1, including: taxonomic precision (TP), taxonomic recall (TR) and taxonomic F-measure (TF). Obviously, more similarity between a tag hierarchy and the reference taxonomy indicates that tag hierarchy has a higher quality.

The first observation that can be drawn from these empirical results is that there is a remarkable difference between the tag hierarchy built from the new knowledge structure we proposed, TagTree, and the tag hierarchy built from individual tags; i.e. Tag Hierarchy A. Our proposed extended algorithm yields a tag hierarchy from our proposed tagging approach that is more similar to the reference taxonomy with taxonomic F-measure (TF) equal to 70.16%.

Another important observation is that the semantic quality of Tag Hierarchy B is much better than the semantic quality of Tag Hierarchy A, although both have been constructed by the same process (Heymann-Benz algorithm and individual tags). However, Tag hierarchy B is built from tags originally collected from the new tagging approach. This all confirms our expectation, which is that making a change to the current tagging approach in order to make a big change to the quality of knowledge structure that can be built. To further check the validity of this, a human assessment of the learned tag hierarchies was performed, whose results will be discussed next.

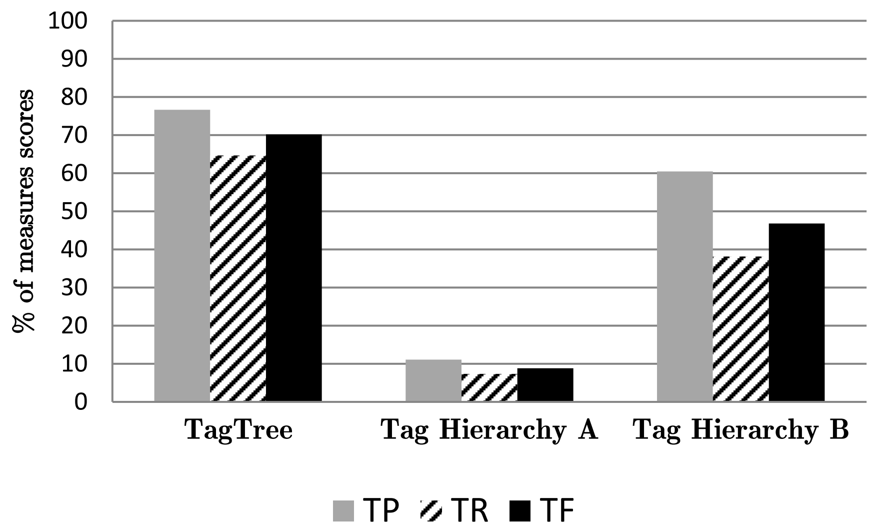


Figure 4-11: Results of semantic evaluation against reference taxonomy

4.6.1.2 Human-based Evaluation

To perform the human-based evaluation, 127 direct taxonomic pairs (t1, t2) from the tag hierarchies is extracted (all the unique direct taxonomic pairs in the generated tag hierarchies) to be manually judged as to whether they are related, and if they are, then how. The relation between each term pair can be one of the following options:

1. t1 is the same as t2.
2. t1 is a (kind of/part of) t2.
3. t1 is somehow related to t2.
4. t1 is not related to t2.
5. Unclear; because the meaning of t1 or t2 is not clear.

The insight behind this approach is that a better tag hierarchy will have a higher percentage of pairs being judged as “kind of” or “part of”, and a lower percentage of pairs being judged as “not related”. Three researchers from the WAIS group at the University of Southampton took part in a human-based evaluation, and asked to judge the relationship between 127 different tag pairs. Since they all did completely finish the study, we received 381 pair judgments. For each term pair, we computed the average (mode) of its answers over each tag hierarchy construction algorithm.

Figure 4-12 summarizes the results of semantic evaluation by human assessment of the produced tag hierarchies. The three rows correspond to the three tag hierarchies, including: TagTree, Tag Hierarchy A and B. The values on the y-axis illustrate the percentage of relation types between tags for each tag hierarchy. For example, among all judgements on direct taxonomic pairs of TagTree, the percentage of "is-a: part/kind of" answers were over than 80% (black part of the uppermost bar). Note that all direct taxonomic pairs of the three tag hierarchies were not judged as "same as" or "unclear" relations so they have not appeared on any bar.

Figure 4-12: Results of semantic evaluation by human assessment

As previously mentioned, a higher quality tag hierarchy should have a higher percentage of direct taxonomic pairs being judged as is-a: kind/part of, and a lower percentage of direct taxonomic pairs being judged as not related or unclear. Consequently, TagTree is the best tag hierarchy since it has the highest portion of "is-a" relation between pairs. In fact, as Figure 4-12 shows, there is a significant difference between the percentages of pairs being judged as "is-a" in TagTree and others. Also, all the pairs in TagTree are related, which is an interesting observation in which that the process of building TagTree has the ability to cope with the lack of consistent structure in folksonomies without doing further pre/post processing. On the other hand, Tag Hierarchy A is the worst since it has the lowest portion of "is-a" relation and the highest portion of "not related" relation between pairs. Furthermore, similar to the observation in Figure 4-11, the quality of Tag Hierarchy B is much better than the quality of Tag Hierarchy A, which confirms our expectation and validates our research methodology as well.

To sum up, the results of the semantic evaluation shows that the our proposed algorithm and tagging approach lead to tagtrees that capture a higher semantics compared to the one obtained from individual tags. The following section will discuss the results of the structural evaluation.

4.6.2 Results of Structural Evaluation

To perform this evaluation, we have used a simple but effective measure known as Area Under Tree (AUT), which considers both the width and depth of the tag hierarchy (Section 3.5.2). Figure 4-13 shows the results of AUT of the three produced tag hierarchies. Tag Hierarchy A yields the highest AUT result, which indicates that this hierarchy is bushier and deeper than other two hierarchies, whereas TagTree yields the lowest AUT result.

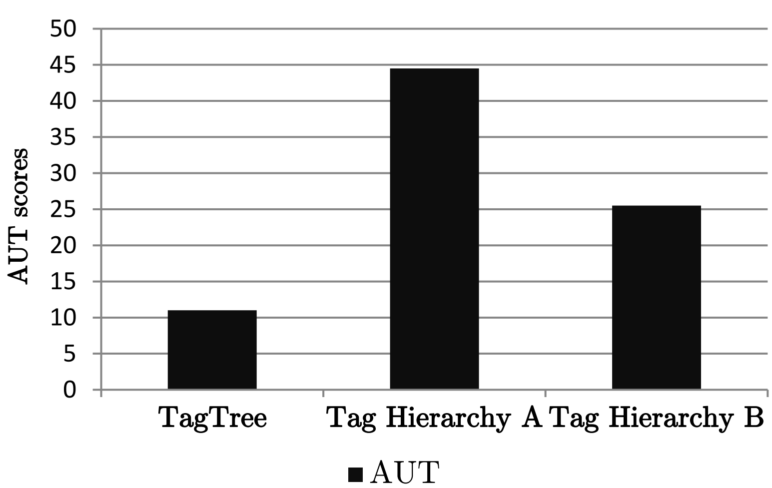


Figure 4-13: Results of structural evaluation (AUT)

Ideally, it is a better to have an approach that generates high semantic quality and expressive tag hierarchies as well. Whereas our tagging approach succeeded in tackling the lack of consistent structure in folksonomies, it generated less expressive tag hierarchy than those generated by individual tags. 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 (Chapter 5).

4.6.3 Results of Usability Evaluation

Since we propose a new tagging approach, there is a need to evaluate the usability of this approach in terms of efficiency, effectiveness and satisfaction. As mentioned in Section 4.5, the System Usability Scale (SUS), proposed by (Brooke, 1996), is adapted and used for our usability evaluation as it seen as a common standardized tool and has been used and verified in many domains. At the end of the experiment, the participants were asked to express their experience of using the new tagging approach (tag pairs) compared to the normal one (individual tags) through the adapted SUS survey. Figure 4-14 shows the average results of SUS scores for both tagging approaches.

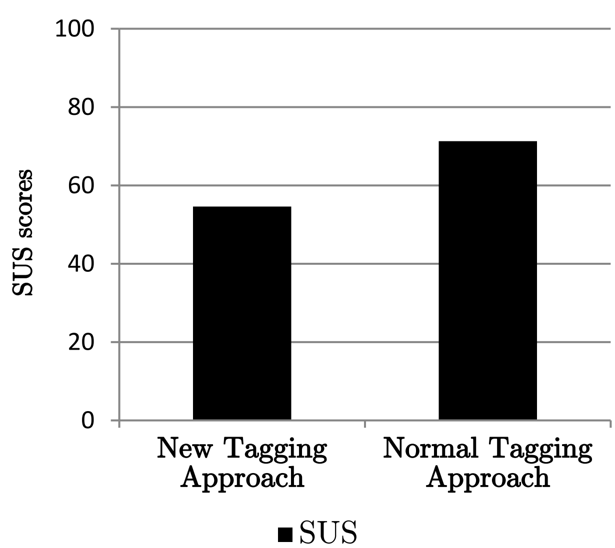


Figure 4-14: Results of usability evaluation

As mentioned earlier, Brooke did not determine when a SUS score is acceptable, and some researchers pointed out that it is more difficult to show a product is acceptable in terms of usability than if it is not. However, Bangor et al. found that a product with SUS scores below 50 will mostly have usability difficulties, whereas scores between 70 and 89, though promising, do not assure high acceptance of usability (Section 3.5.3).

The above results indicate that the average of SUS for the normal approach is 71.3%, with a standard deviation of 19.57, whereas the new approach obtains 54.6%, with a standard deviation of 16.22. These results show that the new approach is marginally acceptable since its average SUS score is over 50%, but with SUS score that is much lower than the one obtained for the normal tagging approach.

To measure the efficiency of the new tagging approach compared to the normal one, the time spent for each tagging action by users is recorded. The tagging action (*ta*) for the new tagging approach means a pair of tags typed by the user, whereas for the normal approach means one tag or more typed by the user. The average time spent for using the new tagging approach is 44.88 sec/ta, and 22.44 sec/tag. In contrast, the average time spent for using the normal tagging approach is 71.90 sec/ta, and 36.37 sec/tag.

4.7 Chapter Summary

This chapter described a pilot study to explore the impact of collecting taxonomic tag pairs, through a proposed tagging approach, rather than individual tags in tackling the "popularity-generality" problem. The research methodology is described in detail, including: the proposed tagging approach, the extended tag hierarchy construction algorithm, the experimental design and implementation of the TagTree system, and the data collection.

The results of the empirical experiment of the research are discussed. In terms of usability, the SUS results show that the new approach is marginal acceptable, with 54.6% SUS score, and the current usability level would be improved by increasing the user experience over the time. And in terms of semantics, there is a remarkable difference between the tag hierarchy built from the new knowledge structure we proposed, TagTree, (with taxonomic F-measure (TF) equal to 70.16%) and the tag hierarchy built from individual tags; i.e. Tag Hierarchy A (with taxonomic F-measure (TF) equal to 8%). Our proposed extended algorithm yields tag hierarchies from our proposed tagging approach that is more similar to the selected reference taxonomy. Also, the proposed tagging approach and algorithm succeeded in eliminating noisy tags and tackling the lack of consistent structure in folksonomies. However, the resulting tag hierarchy, TagTree, is less expressive than it should be, which requires us to perform another experiment on a much bigger data set to prove that TagTree can be a high quality and expressive hierarchy as well.

This pilot experiment has demonstrated that collecting taxonomic tag pairs increases the semantic quality of the tag hierarchy, but at the expense of expressivity, and with some degradation of user experience. The next chapter looks at how this might be overcome, by collecting individual tags from users, and looking at alternative (automatic) methods of increasing the semantic quality of the learned tag pairs from the normal tagging approach.

Chapter Five

Improving the Accuracy of Taxonomic Directions When Building Tag Hierarchies

In the previous chapter, 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.

The proposed approach extends a generality-based approach by using an existing knowledge resource to improve the accuracy of taxonomic directions when building tag hierarchies. The effectiveness of our approach is examined and evaluated in three experiments, using three different types of knowledge resources: tag relationships by users (e.g. Delicious Bundles), a closed text corpus (e.g. a download of English Wikipedia), and an open text corpus (e.g. World Wide Web via Bing).

In this chapter, we first describe the proposed approach and algorithm of tag hierarchy induction, with its specific settings, the datasets we have used, and the evaluation methodology we propose to test the performance of the proposed approach. Finally, the results of each experiment will be discussed.

5.1 Proposed Approach

As we mentioned earlier (in section 3.4) generality-based approaches to tag hierarchy construction show a superior performance compared to other approaches (Strohmaier et al., 2012). However, they suffer from the "generality-popularity" tags problem, in that popularity is assumed (sometimes inaccurately) to be a proxy for generality, i.e. high-level taxonomic terms will occur more often than low-level ones. To tackle this problem, our approach extends a generality-based algorithm, described in (Benz et al., 2010), by using knowledge resources to improve the accuracy of taxonomic directions when building tag hierarchies. Figure 5-1 shows the main steps of our proposed approach.

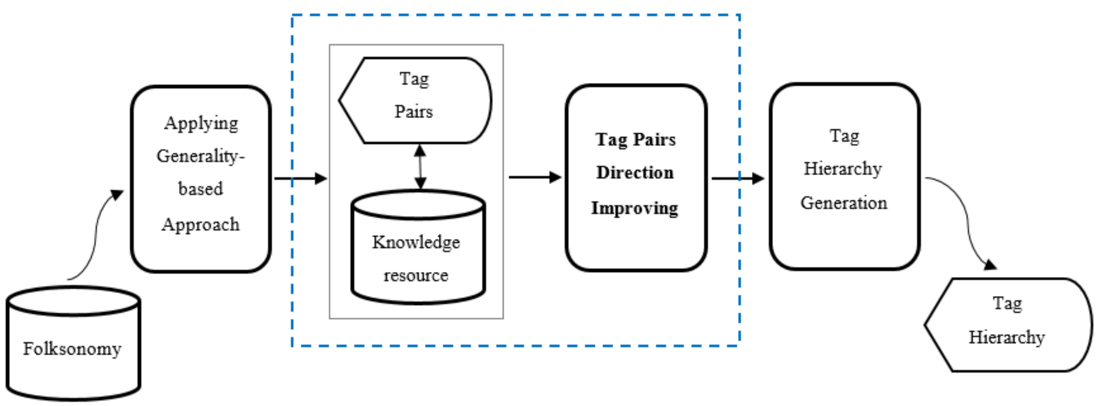


Figure 5-1: The Process diagram of our approach; the steps inside the dash box are the contribution of our approach

To maximise the testing of the effectiveness of our approach, we will separately use three different types of knowledge resources to improve the accuracy of taxonomic directions when building tag hierarchies. These types are: 1) tag relationships by users (e.g. Delicious Bundles), 2) a closed text corpus (e.g. a download of English Wikipedia), and 3) an open text corpus (e.g. the Web via Bing). While some works, e.g. (Plangprasopchok et al., 2010b), suggest to use the relationships between tags that explicitly created by users to create a common tag hierarchy, we propose to use these relationships to improve the accuracy of the taxonomic tag directions that constructed from individual tags.

For the first type we will rely on a simple match between the generated tag pairs and the ones by users in the Delicious Bundles dataset, whereas for the second and third types we will use lexico-syntactic patterns to check the direction of the generated tag pairs. The patterns that our approach uses are a combination of the well-known Hearst's lexico-syntactic patterns (Hearst, 1992), (Table 3-1), and another direct pattern:

— " *C* is a/an **P** "

From each tag pair and the proposed patterns, our proposed algorithm (Section 5.2) generates fourteen phrases (seven phrases for each taxonomic direction of that tag pair). For example, if we have "sweet - food" as a tag pair, the algorithm will generate the phrases shown in (Table 5-1).

Table 5-1: Example of using the proposed lexico-syntactic patterns with the generated the tag pairs. The "Results Count" column is the occurrences number of each phrase that found in the used knowledge resource.

sweet ← food	Results Count	*sweet → food*	Results Count
"sweet **is a** food"	10	"food **is a** sweet"	2
"food **such as** sweet"	20	"sweet **such as** food"	6
"**such** food **as** sweet"	5	"**such** sweet **as** food"	0
"sweet **or other** food"	15	"food **or other** sweet"	3
"sweet **and other** food"	17	"food **and other** sweet"	4
"food **including** sweet"	18	"sweet **including** food"	5
"food **especially** sweet"	14	"sweet **especially** food"	3
p_occ1 Total of Results Count	**99**	**p_occ2** Total of Results Count	**23**

Then all these phrases are submitted successively to a text corpus (e.g. Wikipedia or Bing) search scripts. For the closed text corpus (i.e. a download of English Wikipedia) we use full-text search in Microsoft SQL server, whereas for the open text corpus (i.e. the Web via Bing) we use Bing search API. Both of them take a phrase with double quotation mark as an input (e.g. "sweet *is a* food") and return a number of the results count (e.g. 10) for that phrase (for English Wikipedia) or web pages contains that phrase (for the Web via Bing). Then, the results count for each phrase that found in the selected text corpus is recorded. After that, the results counts for phrases 1–7 are summed up into one value, and the results counts for phrases 8–14 are summed up into another value. Finally, the right taxonomic direction (sweet ← food or sweet → food) is suggested based on these values and some parameters (for more details, see Section 5.2). Figure 5-2 shows the steps of this process.

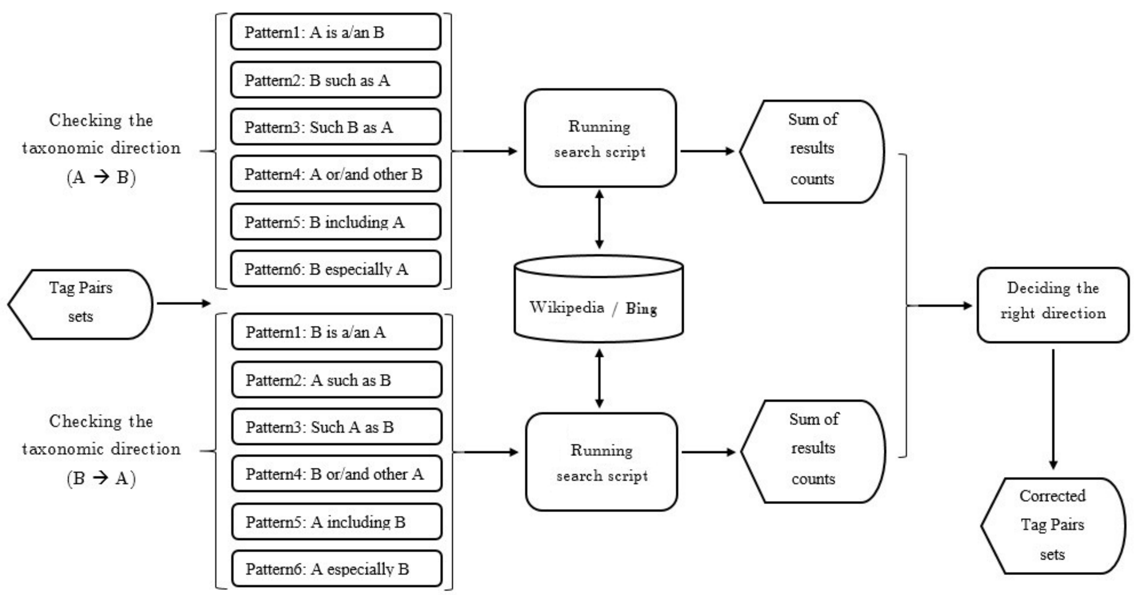


Figure 5-2: The process of checking the direction of the generated tag pairs by using the proposed lexico-syntactic patterns and the selected text corpus (English Wikipedia or the Web via Bing)

While lexico-syntactic patterns tend to achieve a very high level of precision, but low recall (Cimiano, 2006), 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.

5.2 Proposed Algorithm

The algorithm we have used in our approach is an extension of Benz's algorithm (Benz et al., 2010), which itself is an extension of Heymann's algorithm (Heymann & Garcia-Molina, 2006). Algorithm 5-1 demonstrates the pseudo-code for our proposed algorithm.

Algorithm 5-1: Pseudo-code for our algorithm for building tag hierarchies; Lines 1-17 explain the part of our algorithm that adopted from the original algorithm, whereas Lines 18-22 are the main contribution of our approach.

***Input**: user-generated terms (tags)*

***Output**: tag hierarchy*

Functions: Several functions are assumed: sim(t*i*,t*j*): Calculate the similariy (using the co-occurrence weights as similarity measure) between t*i* and t*j*. GetNodes(*TagHierarchy*): returns all nodes in the given hierarchy, *TagHierarchy*. AppendToNode(*TagHierarchy*,t*i*,t*j*): append t*j* underneath t*i* in the given hierarchy, *TagHierarchy*. CheckDirection(t*i*,t*j*): Check the taxonomic direction between t*j* and t*j* by using the suggested knowledge resource. SwapDirection(t*i*,t*j*): Swap the taxonomic direction between t*j* and t*j*.

Parameters: Several parameters are required to be set: tag occurrences threshold (*occ)*, tag-tag similarity threshold (*min˙sim)*, tag generality threshold (*min˙gen)*, taxonomic tag pair occurrences threshold (*min˙p˙occ)*, and the difference between the occurrences of two taxonomic tag pairs threshold *dif˙p˙occ*.

1. *TagList* = Filter the tags by an occurrence threshold *occ*.
2. *TagList* = Order the tags in descending order by generality (measured by degree centrality in the tag–tag co-occurrence network).
3. *TagHierarchy* = {∅, root}
4. **for** *i* = 1 … |*TagList* | - 1 **do**
5.   t*i* = *TagList* [*i*]
6.   *MostSimilarVal* = 0.
7.   **for** all t*j* ∍ GetNodes(*TagHierarchy*) **do**
8.   **if** sim(t*i*,t*j*) > *MostSimilarVal* **then**
9.   *MostSimilarVal* = sim(t*i*,t*j*)
10.   *MostSimilar* = t*j*
11.   **end if**
12.   **end for**

```
13.     if MostSimilarVal > min_sim and MostSimilarVal < min_gen then
14.         AppendToNode(TagHierarchy, ti, MostSimilar)
15.     else
16.         AppendToNode(TagHierarchy, ti,root)
17.     end if
18.     if CheckDirection(tj,ti) > min_p_occ
19.      and CheckDirection(tj,ti) - CheckDirection(ti,tj) > dif_p_occ then
20.         SwapDirection(ti,tj)
21.     end if
22. end for
```

5.2.1 Description of the Algorithm

The algorithm starts by filtering the tags (extracted from the folksonomy dataset) by an occurrence threshold *occ* (Line 1). Then, it orders the tags in descending order by generality that measured by degree centrality in the tag–tag co-occurrence network (Line 2). After that, the algorithm starts with the most general tag as the root node (Line 3). Then, it adds each tag t*i* in the tags list subsequently to an evolving tag hierarchy (Lines 4-6). It decides where to add each tag t*i* by calculating its similarity (using the co-occurrence weights as a similarity measure) to each tag currently present in the hierarchy t*j*, and appends the current tag t*i* underneath its most similar tag *MostSimilar*. If t*i* is very general (determined by a generality threshold *min_gen*) or no sufficiently similar tag exists (determined by a similarity threshold *min_sim*), the algorithm appends t*i* underneath the root node of the hierarchy (Lines 7-17).

After that, the algorithm checks the taxonomic direction by using the suggested knowledge resource. For the Delicious Bundles dataset, the checking is done by a direct match, whereas for the English Wikipedia and Bing datasets is done by using the suggested lexico-syntactic patterns. The checking step is done for both taxonomic directions (t*i* ← t*j* and t*i* → t*j*) by calculating how many occurrences found in the selected knowledge resource for both directions. Then, the algorithm corrects the taxonomic direction if: 1) the result of subtract the occurrences number of (t*i* → t*j*) from the occurrences number of (t*i* ← t*j*) more than the threshold *dif_p_occ*, and 2) the occurrences number of (t*i* → t*j*) more than the taxonomic tag pair occurrences threshold *min_p_occ* (lines 18-21). Finally, the algorithm applies 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 process is done by the steps in (Lines 4-22).

The proposed algorithm is extensible as it is possible to make several modifications to how tags are append to the growing tag hierarchy or when taxonomic tag pairs are corrected. For instance, a tag t_i can only be appended to a candidate tag t_j in the growing tag hierarchy if t_i is sufficiently similar to some parents or childs of t_j. Furthermore, the checking and correcting process of a taxonomic tag pair can be done not only between t_i and t_j but also between t_j and the parents of t_i. Moreover, the algorithm consists of several parameters that can be optimized for any tags collection (Section 5.2.3).

5.2.2 Similarity Measure

Although it has been stated in the original algorithm that the co-occurrence weight is used as a similarity measure of tags (Step 3.a in Algorithm 5-1), it was not clear how to calculate the co-occurrence weight. Thus, we have used five common tag similarity measures to compute the co-occurrence weights, and created five different versions of each algorithm (our algorithm and the original algorithm). The five common similarity measures between *Tag 1* and *Tag 2* can be mathematically defined as follows:

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

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

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

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

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

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

5.2.3 Settings of the Algorithms

The original and our algorithms are affected by several parameters (the last two are only for our algorithm), and here is a brief description of the settings used to run them:

— **Tag Occurrences Threshold** ***occ***: In general, the bigger of the number of tag occurrences, the stronger agreement between users on a proper view of the shared context. For this threshold we have chosen to include tags occurring more than 400 times (for both datasets: Delicious and Flickr); as suggested by the authors of the original algorithm (Heymann & Garcia-Molina, 2006). This gives us an adequate number of tags (17160 tags) to test our algorithm, and should also help us to make a fair comparison between the original and our algorithms.

— **Tag-Tag Similarity Threshold** ***min˙sim***: Each candidate tag from the folksonomy is appended as a child of the most similar node in the hierarchy if its similarity to that node is greater than a similarity threshold, otherwise it is appended to the root of the hierarchy. Table 5-2 shows the setting of the similarity thresholds for the five selected similarity measures (for both datasets: Delicious and Flickr). Lowering the thresholds led to relatively unrelated tags, while made it higher resulted in an unbalanced hierarchy as too many tags were appended to the root of the hierarchy.

Table 5-2: Setting of the similarity threshold parameter for each one of the used similarity measures.

Similarity measure	*Similarity threshold*
Matching	100
Dice	0.03
Jaccard	0.02
Overlap	0.099
Cosine	0.03

— **Tag Generality Threshold *min_gen***: Generality in the original and our algorithms is measured by degree centrality in the tag–tag co-occurrence network. The best value for this threshold were 0.8 for Delicious and 0.04 for Flickr. A child of a node is also a child of the root of the hierarchy, however, lowering the threshold resulted in an unbalanced hierarchy as too many tags were appended to the root of the hierarchy.

— **Taxonomic Tag Pair Occurrences Threshold *min_p_occ***: This threshold is only used for our proposed algorithm. The direction of a generated taxonomic tag pair is corrected if the occurrences number of that tag pair found in the selected knowledge resource is equal or greater than *min_p_occ* threshold. The best value for this threshold were 3, and increasing the threshold led to losing many corrections of the taxonomic directions.

— **Difference between the Occurrences of Two Taxonomic Tag Pairs Threshold *dif_p_occ***: This threshold is only used for our proposed algorithm. The direction of a generated taxonomic tag pair is corrected if the results of subtract the occurrences number of the current direction of the tag pair found in the selected knowledge resource from the occurrences number of the opposite direction of that tag pair is equal or greater than *dif_p_occ* threshold. The best value for this threshold were 1, and gradually increasing the threshold led to gradually losing many corrections of the taxonomic directions.

5.3 Datasets

In our experiments, we have used five large datasets, comprising of two tag collections and three different knowledge resources:

5.3.1 Tag Collections

To compare the performance of our proposed approach to building a tag hierarchy compared to the original approach, we have used two large-scale folksonomy datasets from the PINTS experimental datasets[38] containing a systematic crawl of Delicious and Flickr during 2006 and 2007. Table 5-3 summarized the statistics of the datasets.

Table 5-3: Statistics of the used tag collections (Delicious and Flickr).

Dataset	*Users*	*Tags*	*Resources*	*Tag assignments*
Delicious	532,924	2,481,698	17,262,480	140,126,586
Flickr	319,686	1,607,879	28,153,045	112,900,000

5.3.2 Knowledge Resources

To solve the "generality-popularity" tags problem, we have chosen three different types of knowledge resources: tag relationships by users (e.g. Delicious Bundles), a closed text corpus (e.g. a download of English Wikipedia), and an open text corpus (e.g. World Wide Web via Bing). This will allow us to suggest which a knowledge resource type is better to use with our proposed approach.

— **Delicious Bundles:** Delicious allows users to group similar tags into bundles. Although we have run a systematic crawl of Delicious bundles during February and March 2017, the bundles that we have collected have been created at different times. Table 5-4 summarized the statistics of the dataset.

Table 5-4: Statistics of the Delicious bundles dataset.

Dataset	*Users*	*Tags*	*Tag bundle assignments*
Delicious bundles	8,360	189,575	1,080,951

— **Wikipedia Dataset**: We selected to use Wikipedia since it is currently the largest knowledge repository available on the Web. Moreover, some studies show that the quality of Wikipedia is comparable to the quality of traditional encyclopaedias (Giles, 2005). The dataset that we have used is

[38] http://www.uni-koblenz-landau.de/koblenz/fb4/AGStaab/Research/DataSets/PINTSExperimentsDataSets/index˙html

the English Wikipedia articles (no talk or user pages), which we downloaded during March 2014 and contains 4,487,682 different articles[39].

— **Bing Dataset**: While the Wikipedia dataset above is a closed corpus, we want also to test our approach with an open corpus, therefore, we have chosen the World Wide Web dataset via the Bing search engine. Bing was chosen as it is one of the most popular search engines (Ritchie et al., 2016), and its API supports the highest request rate (important as our scripts need to make many thousands of calls as the tag hierarchy is constructed). Bing search API[40] provides 5000 free web search transactions per month, while other popular search engines APIs like Google search API provides 100 free transactions per day and Yahoo search API does not provide free transactions at all.

Note that for all datasets (apart from the Bing dataset), the words are passed to the normalisation process that applies two steps: 1) **Word Cleaning**, including: *Letters lower-case, symbol deleting* and *non-English letters deleting*. 2) **Plural to Singular Conversion**, using a part of the well-known Porter Stemmer (Porter, 1980).

5.4 Evaluation Methodology

To test the performance of our approach, we have performed three experiments using three different types of knowledge resources: Delicious Bundles, English Wikipedia and the Web via Bing (Section 5.3). For each experiment, we applied the original algorithm and our proposed algorithm to two large-scale folksonomy datasets collected from Delicious and Flickr (Section 5.3). And to examine the effectiveness of using similarity threshold (the minimum of sufficient similarity between tags) that is suggested by the original algorithm (Line 13 in Algorithm 5-1), we have run the experiment twice: with and without using a similarity threshold. This is because our algorithm may not need to use a similarity

39 https://en.wikipedia.org/wiki/Wikipedia:Database˙download, as collected in March 2014.

40 http://datamarket.azure.com/dataset/bing/search

threshold as it will check the similarity between tags while checking the taxonomic directions (Section 5.1). In total, we have yielded 123 different tag hierarchies from these experiments.

To evaluate our proposed approach to a building tag hierarchy against the original approach, we will use a reference taxonomy to answer the question:

1. **Which one of the two approaches produces the highest semantics quality of taxonomic tags?**

We have chosen to compare the generated taxonomic tag pairs to the WordNet (Miller, 1995) dataset as the reference taxonomy for three reasons:

1. It is considered as a gold-standard dataset for evaluating hyponym/hypernym relations (Snow et al., 2004).
2. WordNet is a reasonable reference for our purpose, i.e. solving the popularity-generality tags problem, as a significant fraction of the popular tags in the selected tag collections (Delicious and Flickr) is covered by WordNet (Figure 5-3).
3. We needed to avoid any reference dataset that was constructed automatically or based on Wikipedia as we have used it in our approach[41].

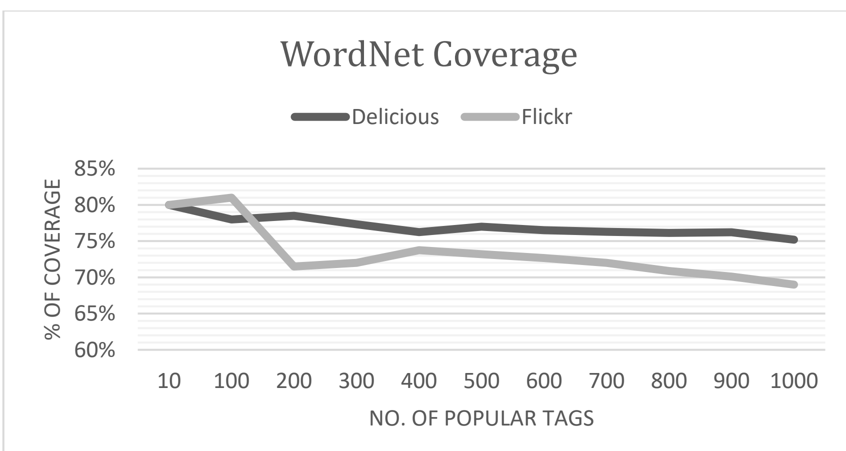


Figure 5-3: WordNet coverage of top popular tags in Delicious and Flickr

[41] Other established semantic resources, like Yago, can be used as reference taxonomies, but we avoided them as they were constructed based on Wikipedia (see Section 6.1 for more details).

Although the direction of taxonomic tags may be different based on the context, we assume that the accuracy of the taxonomic directions will be judged against WordNet, regardless of the context. This because of our proposed approach aims to construct tag hierarchies that represent shared conceptualizations that are hidden in a folksonomy, and not toward a specific domain or application. WordNet is a structured lexical database of the English language built manually by experts. It contains 206,941 terms grouped in 117,659 synsets[42]. The synsets are connected by several lexical relations. The most important and frequently used of these relations is the hyponym/hypernym relation. For our purpose we have extracted the taxonomic terms in WordNet.

5.5 Results and Analysis

To give an impression of the results, Table 5-5 shows a few examples of the produced taxonomic tag pairs from the Delicious dataset, using the five similarity measures under study.

Table 5-5: Examples of produced tag pairs from the Delicious dataset for each of the selected similarity measures.

Measure	*Rank*	*Tag A*	*Tag B*	*Rank*	*Tag A*	*Tag B*
Matching	1	Blog	Design	1000	Display	Technology
Dice			Design			LCD
Jaccard			Design			LCD
Overlap			Bloggerbeast			TFT
Cosine			Daily			LCD
Matching	100	Daily	Blog	5000	Maple	PHP
Dice			News			Willow
Jaccard			News			Willow
Overlap			Blog			Willow
Cosine			News			Willow
Matching	500	Weather	News	10000	Bridesmaid	Dress
Dice			Forecast			Bridal
Jaccard			Forecast			Bridal
Overlap			Noaa			Dress
Cosine			Forecast			Bridal

42 http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html, as visited in June 2014.

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 5-6 displays the overlap (generating common tag pairs) between the produced tag hierarchies from the Delicious dataset based on these similarity measures.

Table 5-6: Overlap between tag hierarchies generated from the Delicious dataset and using selected similarity measures.

	Matching	*Cosine*	*Overlap*	*Jaccard*
Dice	0.15	0.71	0.16	0.57
Jaccard	0.09	0.40	0.10	
Overlap	0.71	0.24		
Cosine	0.22			

In this chapter, 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 123 tag hierarchies (produced by the original and our algorithms) against a gold-standard dataset, namely: WordNet (Section 5.4). 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 5-7 shows examples of taxonomic tag pairs that the original algorithm has generated them (from the Delicious dataset and using selected similarity measures) in the form of (*Tag A* is-a *Tag B*), where they have been found in WordNet as (*Tag B* is-a *Tag A*).

For further improvement we added a *min˙p˙occ* threshold in our proposed algorithm. In other words, before we correct the direction of a generated taxonomic tag pair, we check the occurrences number of that tag pair found in the selected knowledge resource whether it is equal or bigger than *min˙p˙occ* threshold. We call this variation of our algorithm (i.e. using *min˙p˙occ* threshold) as our strict algorithm.

Table 5-7: Examples of taxonomic tag pairs generated by original algorithm that found in the form of (*Tag A* is-a *Tag B*), where they have been found in WordNet as (*Tag B* is-a *Tag A*).

Similarity Measure	*Tag A*	*Tag B*
Matching	Faith	Christian
	Footwear	Shoes
	Society	Culture
	Wealth	Money
	Poultry	Chicken

Similarity Measure	*Tag A*	*Tag B*
Dice	Meat	Beef
	Primates	Monkey
	Road	Highway
	Search	Google
	Sweet	Candy

Similarity Measure	*Tag A*	*Tag B*
Jaccard	Coffee	Espresso
	Drink	Alcohol
	Ireland	Dublin
	Pastry	Tart
	Puzzle	Sudoku

Similarity Measure	*Tag A*	*Tag B*
Overlap	Broadcast	Video
	Canine	Dog
	Footwear	Shoes
	Poultry	Chicken
	Ride	Bike

Similarity Measure	*Tag A*	*Tag B*
Cosine	Bag	Purses
	Sweet	Candy
	Meat	Beef
	Search	Google
	Broadcast	Radio

5.5.1 Delicious Bundles

Table 5-8 shows the results. The first observation that can be drawn is that the original algorithm is moderately successful (as much as 76.96% for Delicious, and 70.53% for Flickr), 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 (and found in WordNet) are in the right direction.

Table 5-8: Taxonomic tag pairs evaluation, using selected similarity measures and a similarity threshold for each measure, against WordNet; Folksonomy datasets: Delicious and Flickr; Knowledge resource: Delicious Bundles.

	% Agreement with WordNet					
	Original Algorithm		**Our Algorithm**		**Our strict Algorithm**	
	Delicious	Flickr	Delicious	Flickr	Delicious	Flickr
Matching	75.74%	69.38%	77.33%	73.87%	77.33%	70.19%
Dice	47.22%	53.52%	52.48%	61.85%	48.72%	55.90%
Jaccard	47.37%	55.22%	53.40%	62.72%	49.09%	57.72%
Overlap	**76.96%**	**70.53%**	**77.85%**	**74.07%**	**78.30%**	**71.41%**
Cosine	54.90%	59.42%	59.12%	63.03%	55.50%	58.22%

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 improving the accuracy of directions in relations constructed between tags that were generated in the wrong direction by the original algorithm. In the best case (Overlap) this leads to an accuracy of over 78%. The last column of Table 5-8 shows the impact of using the *min˙p˙occ* threshold, which was a positive with only the Overlap measure for Delicious. This was due to the low rate of the generated tag pairs occurrences found in the Delicious Bundles dataset (for more details see Section 5.5.4.1).

Table 5-9: Taxonomic tag pairs evaluation, using selected similarity measures and without using a similarity threshold, against WordNet; Folksonomy datasets: Delicious and Flickr; Knowledge resource: Delicious Bundles.

	% Agreement with WordNet					
	Original Algorithm		**Our Algorithm**		**Our strict Algorithm**	
	Delicious	Flickr	Delicious	Flickr	Delicious	Flickr
Matching	76.90%	66.67%	78.13%	71.06%	78.13%	67.68%
Dice	51.33%	52.42%	55.85%	61.26%	52.62%	55.82%
Jaccard	47.56%	52.42%	52.21%	61.26%	48.34%	55.82%
Overlap	**77.39%**	**70.37%**	**78.23%**	**72.70%**	**78.65%**	**71.15%**
Cosine	59.55%	57.14%	62.81%	60.27%	60.09%	57.92%

Table 5-9 shows the results of rerunning the experiment but with a tag similarity threshold = 0. The observation that can be drawn is that using a similarity threshold is not always help in improving the accuracy of the taxonomic directions (Delicious vs. Flickr results), as suggested by the original algorithm. Across all selected tag similarity measures, our algorithm yields taxonomic tag pairs that better match those found in WordNet.

5.5.2 English Wikipedia

Table 5-10 shows the results. The first observation that can be drawn from the results is that there is an improvement achieved by our proposed algorithm compared to the original algorithm among all folksonomy datasets and 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 5-10: Taxonomic tag pairs evaluation, using selected similarity measures and a similarity threshold for each measure, against WordNet; Folksonomy datasets: Delicious and Flickr; Knowledge resource: English Wikipedia.

	% Agreement with WordNet					
	Original Algorithm		**Our Algorithm**		**Our strict Algorithm**	
	Delicious	Flickr	Delicious	Flickr	Delicious	Flickr
Matching	75.74%	69.38%	77.38%	77.03%	79.34%	78.95%
Dice	47.22%	53.52%	55.56%	80.28%	61.11%	80.28%
Jaccard	47.37%	55.22%	64.91%	80.60%	64.04%	80.60%
Overlap	**76.96%**	**70.53%**	**81.01%**	**80.95%**	**81.11%**	**81.05%**
Cosine	54.90%	59.42%	64.71%	75.36%	64.71%	75.36%

Another observation from the results of our experiments 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 largest amount of generated tag pairs that are found in WordNet regardless of the taxonomic direction. The last column of Table 5-10 shows the improvement of using the *min˙p˙occ* threshold, which was more effective with the Matching, Dice and Jaccard similarity measures for Delicious and Matching measure for Flickr.

Table 5-11: Taxonomic tag pairs evaluation, using selected similarity measures and without using a similarity threshold, against WordNet; Folksonomy datasets: Delicious and Flickr; Knowledge resource: English Wikipedia.

	% Agreement with WordNet					
	Original Algorithm		**Our Algorithm**		**Our strict Algorithm**	
	Delicious	Flickr	Delicious	Flickr	Delicious	Flickr
Matching	76.90%	66.67%	77.81%	74.39%	80.55%	76.83%
Dice	51.33%	52.42%	66.00%	78.23%	66.67%	78.23%
Jaccard	47.56%	52.42%	66.26%	78.23%	62.60%	78.23%
Overlap	**77.39%**	**70.37%**	**81.11%**	**78.70%**	**81.30%**	**79.63%**
Cosine	59.55%	57.14%	67.42%	77.14%	67.98%	77.14%

Table 5-11 shows the results of rerunning the experiment but with a tag similarity threshold = 0. These results show that using a similarity threshold could improve the accuracy of the taxonomic directions as in Flickr results, whereas sometimes it is better to not use it as in Delicious results. Across all selected tag similarity measures, our algorithm yields taxonomic tag pairs that better match those found in WordNet.

5.5.3 The Web via Bing

Our algorithm uses seven linguistic patterns applied to a text corpus to check and correct the direction of suggested relationships. For this purpose, we have chosen the World Wide Web dataset via the Bing search engine, instead of English Wikipedia. This should increase the coverage and occurrences of the learned taxonomic tags in the selected knowledge resource (e.g. the Web), and consequently the quality of the learned tag hierarchies.

Due to the limitation of the request number that Bing API supports, we have chosen the Delicious dataset from the selected tag collections (Section 5.3), and the Overlap measure from the selected tag similarity measures (Section 5.2.2); as it produced the best performance of generating taxonomic tag pairs against WordNet in (Section 5.5.2; English Wikipedia).

Table 5-12: Taxonomic tag pairs evaluation, using (Overlap) as the selected similarity measure and without a similarity threshold, against WordNet; Folksonomy dataset: Delicious; Knowledge resource: the Web via Bing.

	% Agreement with WordNet		
	Original Algorithm	**Our Algorithm**	**Our strict Algorithm**
Overlap	76.63%	85.37%	86.09%

Table 5-12 shows the results. The First observation that can be drawn is that there is an improvement achieved by our proposed algorithm compared to the original algorithm (from 76.63% to 86.09%). The second observation is that there is also an improvement in the accuracy of the taxonomic tag pairs generated by our approach and using the Web via Bing, compared to ones that are generated by our approach and using the English Wikipedia dataset (from 81.30% to 86.09%). This is because of the increased coverage and occurrences of the tag pairs found in the Web via Bing, which consequently improves the semantic quality of learned taxonomic tag pairs.

5.5.4 Comparison of the Knowledge Resource

As a part of our proposed approach to building tag hierarchies is to use a knowledge resource, we have chosen three different types of knowledge resources to discover which one is better to use based on the performance and coverage of that knowledge resource. By "coverage" we mean how many generated tag pairs could have their direction checked, and by "performance" we mean how many checked tag pairs help in improving the accuracy of taxonomic directions when building tag hierarchies.

5.5.4.1 The Coverage of the Knowledge Resources

Table 5-13 displays the percentage of the examined taxonomic tag pairs found in Delicious Bundles, English Wikipedia and the Web via Bing. The first observation that can be drawn from the results is that the Matching and Overlap measures have much more coverage of the examined taxonomic tag pairs than other similarity measures, among all the investigated knowledge resources. The second observation that can be drawn is that the coverage of the examined taxonomic tag pairs that generated from the Flickr dataset found in Delicious Bundles is much less than the ones that generated from the Delicious dataset. This may because the datasets of Delicious and Delicious Bundles come from the same collaborative tagging system.

Table 5-13: Coverage of the generated taxonomic tag pairs in Delicious Bundles, English Wikipedia and the Web via Bing. "Found Once at Least" column shows the percentage of the examined taxonomic tag pairs found in a given knowledge resource at least one time, whereas "More than Two Times" column shows the percentage of the examined taxonomic tag pairs found in a given knowledge resource three times or more.

Knowledge Resource	*Similarity Measure*	*% of found generated tag pairs*			
		Found Once at Least (Checked)		**More than Two Times**	
		Delicious	Flickr	Delicious	Flickr
Delicious Bundles	**Matching**	**29.14%**	**4.40%**	**10.31%**	**1.52%**
	Dice	12.56%	1.73%	3.57%	0.51%
	Jaccard	9.38%	1.73%	3.25%	0.51%
	Overlap	**27.62%**	**2.26%**	**8.25%**	**0.60%**
	Cosine	15.05%	1.55%	4.20%	0.44%
English Wikipedia	**Matching**	**31.38%**	**18.08%**	**17.17%**	**9.17%**
	Dice	15.14%	7.69%	6.06%	3.80%
	Jaccard	19.85%	7.69%	9.53%	3.80%
	Overlap	**26.77%**	**12.03%**	**13.27%**	**5.32%**
	Cosine	16.90%	7.64%	7.12%	3.68%
The Web via Bing	**Overlap**	**86.57%**		**79.17%**	

Another observation from the results of our experiments is that, among all the investigated knowledge resources, the Web via Bing yields the higher coverage percentage of the examined taxonomic tag pairs (Note that because of the limitation of the request number that Bing API supports, we have investigated the Web via Bing dataset by using only the Delicious dataset and the Overlap measure; Section 5.5.3). Finally, the results above show that there is an improvement in the coverage of the examined taxonomic tag pairs from 26% in Wikipedia to 86% in the Web via Bing (using Overlap measure). This confirmed our intuition and validates our hypothesis that using an open text corpus instead of a closed text corpus in our approach should increase the coverage and occurrences of the tags in any tag collection, and consequently the quality of learned tag hierarchies (Section 5.5.3). The impact of this results on the quality of the learned taxonomic tags will be discussed next.

5.5.4.2 The Performance of the Knowledge Resources

Table 5-14 shows a comparison of using different knowledge resources (Delicious Bundles, English Wikipedia and the Web via Bing) for checking and correcting the taxonomic tag pairs that we get from our proposed algorithm, against WordNet, and compared to the ones we get from the original algorithm.

The first observation that can be drawn from the results is that, among all the selected knowledge resources, the performance of our proposed algorithm outperforms the original algorithm. This means, regardless of the knowledge resource, our approach has succeeded in correcting the direction of taxonomic tag pairs that were generated in the wrong direction by the original algorithm (correct between 1.26% to 8.70% for Delicious, and 3.54% to 10.42% for Flickr, depending on the used knowledge resource). Although there is still a space for improvement, these results show that the performance of our extended algorithm outperforms the original algorithm. This improvement will have a positive impact on the usage of the learned tag hierarchies in different tasks, like searching and browsing (Section 3.1).

Table 5-14: Comparison of the taxonomic tag pairs evaluations, using selected knowledge resources and (Overlap) as the selected similarity measure, against WordNet; Folksonomy datasets: Delicious and Flickr.

Delicious

Knowledge Resource	*% Agreement with WordNet*		
	Original Algorithm	**Our Algorithm**	**Our strict Algorithm**
Delicious Bundles	77.39%	78.23%	78.65%
English Wikipedia	77.39%	81.11%	81.30%
The Web via Bing	77.39%	85.37%	86.09%

Flickr

Knowledge Resource	*% Agreement with WordNet*		
	Original Algorithm	**Our Algorithm**	**Our strict Algorithm**
Delicious Bundles	70.53%	74.07%	71.41%
English Wikipedia	70.53%	80.95%	81.05%

Another observation from the results of our experiments is that, among all the selected knowledge resources, the Web via Bing yields the best performance of correcting taxonomic tag pairs against WordNet (improving the accuracy of taxonomic directions from 77% to 86%).

5.5.5 Comparison of the Lexico-Syntactic Patterns

While this chapter has demonstrated a successful approach to building high-quality tag hierarchies based on lexico-syntactic patterns applied to a text corpus, further investigation in this area is to explore which lexico-syntactic patterns are most successful in correcting wrong directions of the generated tag pairs, and whether any introduce errors.

Table 5-15 shows the results. The first observation that can be drawn is that all the proposed patterns succeed in improving the accuracy of taxonomic directions, across all the selected tag similarity measures and the used knowledge resources, although they introduced some errors. In other words, none of the proposed patterns introduced errors more than corrections of wrong directions of generated tag pairs.

Table 5-15: Percentage of correct and error introduced by the proposed lexico-syntactic patterns applied to English Wikipedia (using selected similarity measures) and the Web via Bing (using Overlap as the selected similarity measure) when building high-quality tag hierarchies.

English Wikipedia		is a	such as	such	or other	and other	including	especially
Matching	*Correct*	18.42%	32.89%	7.89%	23.68%	27.63%	19.74%	15.79%
	Error	13.04%	11.07%	3.95%	3.56%	9.49%	9.88%	5.53%
Dice	*Correct*	19.18%	35.62%	12.33%	19.18%	26.03%	31.51%	9.59%
	Error	14.29%	25.97%	3.90%	5.19%	11.69%	12.99%	5.19%
Jaccard	*Correct*	20.16%	35.66%	10.85%	14.73%	28.68%	31.01%	13.18%
	Error	12.82%	27.35%	3.42%	4.27%	15.38%	16.24%	5.98%
Overlap	*Correct*	13.46%	34.62%	3.85%	15.38%	28.85%	25.00%	13.46%
	Error	11.24%	8.99%	2.81%	2.81%	10.67%	8.99%	3.37%
Cosine	*Correct*	18.06%	37.50%	18.06%	23.61%	29.17%	30.56%	18.06%
	Error	12.26%	21.70%	4.72%	6.60%	14.15%	12.26%	9.43%

Bing		is a	such as	such	or other	and other	Including	especially
Overlap	*Correct*	56.98%	67.44%	19.77%	66.28%	54.65%	61.63%	58.14%
	Error	21.28%	8.87%	4.26%	10.99%	13.12%	14.18%	8.16%

Taking the Overlap measure from the selected tag similarity measures (as it yields the best performance of generating taxonomic tag pairs against WordNet), the 'such as' pattern is the most successful in correcting wrong directions of the generated tag pairs, whereas the 'such … as' is the least successful, for both English Wikipedia and the Web via Bing datasets. Another observation from the results of our experiments is that all the proposed pattern works better with the Web via Bing, due to the high coverage rate of the examined taxonomic tag pairs found in the Web via Bing compared to the ones in the English Wikipedia dataset (Section 5.5.4).

5.6 Chapter Summary

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" tags 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 experiments to measure this effect, and proposed an approach to reduce its impact. Our proposed approach extends a promising generality-based algorithm by using a knowledge resource to check the direction of hyponym/hypernym relations in order to distinguish between popular and general tags. For this purpose we have used Delicious and Flickr as tag collections, Delicious Bundles, English Wikipedia and the Web via Bing as different options of the selected knowledge resource, 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 across all the selected tag similarity measures and the used knowledge resources (correct up to 26% of the direction of taxonomic tag pairs that were wrongly generated by the original algorithm, depending on the used knowledge resource and the selected similarity measure). 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 terms of the comparison between the selected tag similarity measures, the Overlap measure yields the best performance of generating taxonomic tag pairs against WordNet. And in terms of the comparison of the selected knowledge resources, the Web via Bing yields a greater coverage and occurrences of the learned tag pairs, and consequently the best performance of correcting taxonomic tag pairs against WordNet (improving the accuracy of taxonomic directions from 77% to 86%). Although all the proposed patterns, across all the selected tag similarity measures and the used knowledge resources, succeed in improving the accuracy of taxonomic directions, the 'such as' pattern is the most successful in correcting wrong taxonomic directions, whereas the 'such … as' is the least successful, for both English Wikipedia and the Web via Bing datasets.

Based on the results we achieved in this chapter, the following chapter will focus on the best case of them (Folksonomy dataset: Delicious; Similarity measure: Overlap; Knowledge resource: the Web via Bing) to show whether this in turn has an impact on the tag hierarchy as a whole.

Chapter Six

The Impact of Correcting Taxonomic Directions on the Quality of Tag Hierarchies

In the previous chapter, we have shown that using lexico-syntactic patterns applied to an open text corpus, like the Web via Bing, could improve the *accuracy of directions* in relations constructed between tags by a generality-based approach to tag hierarchy construction. This improvement will translate to higher quality tag hierarchy structures and semantics.

Based on the results we achieved in the previous chapter, this chapter presents an extensive evaluation to assess the tag hierarchies produced using our improved approach described in the previous chapter. This should give us a measure of how the improvements in tag pair directions can be translated into improved tag hierarchies.

6.1 Datasets

In the previous chapter, we have used several large datasets to construct tag hierarchies, comprising of two tag collections (Delicious and Flickr), three knowledge resources (Delicious Bundles, English Wikipedia and the Web via Bing). For the evaluation in this chapter we will use the learned tag hierarchies that constructed from the Delicious dataset and using the Web via Bing; as they yield the best performance of correcting taxonomic tag pairs against WordNet (Section 5.5.4).

In order to perform a comparative evaluation of learned tag hierarchies by our proposed algorithm and the original algorithm we have chosen a number of reference taxonomies that are derived from established semantic resources, namely: WordNet, Yago, Freebase and Probase. While manually created knowledge resources made by experts (like WordNet) are usually smaller, but semantically more accurate, automatically or collaboratively created knowledge resources (like Yago, Freebase and Probase) are more fuzzy, but cover a greater amount of terms and domains. For our purpose, we have extracted the taxonomic terms among the concepts in these reference taxonomies.

- **WordNet Dataset**: A structured lexical database in English built manually by experts. It contains 206,941 terms grouped in 117,659 synsets[43], and is considered as a gold-standard dataset for testing hyponym/hypernym relation building algorithms (for more details see Section 5.4).

- **Yago Dataset:** A large ontology that is automatically extracted from Wikipedia, WordNet and GeoNames[44]. At present, it contains more than 10 million entities and more than 120 million facts about these entities[45]. Manual evaluation has shown that the accuracy of YAGO lies at around 95% (Suchanek et al., 2007).

- **Freebase Dataset**: An open collaborative knowledge base that launched in 2007 by Metaweb (Bollacker et al., 2008) and acquired in 2010 by Google. At the time of writing it contains more than 57 million topics and over 3

[43] http://wordnet.princeton.edu/wordnet/man/wnstats.7WN.html, as visited in June 2014.

[44] http://www.geonames.org

[45] http://www.mpi-inf.mpg.de/departments/databases-and-information-systems/research/yago-naga/yago, as visited in November 2015.

billion facts[46], and is considered the most comprehensive publicly available source of general knowledge facts (Bast et al., 2014).

— **Probase Dataset**: A probabilistic taxonomy that contains 2.7 million concepts harnessed automatically from a text corpus of 1.68 billion web pages[47]. It is reported that Probase is the largest and most comprehensive taxonomy in terms of the number of concepts included (Wu et al., 2012).

6.2 Evaluation Methodology

We have run the experiment in two phases. First, we applied the Heymann–Benz algorithm (using Overlap as the similarity measure) and our proposed algorithm (using Overlap as the selected similarity measure, and the Web via Bing as the used knowledge resource) to the Delicious dataset, to build two tag hierarchies. Second, we evaluated the performance of our proposed approach to building tag hierarchy against the original approach using the evaluation metrics mentioned in the previous section.

To evaluate the performance of our proposed approach to building tag hierarchy against the original approach, we will use the proposed evaluation metrics as explained in section 3.5. This evaluation is needed to answer two questions:

1. **Which one of the two approaches produces the highest semantic quality of tag hierarchies?** (evaluation metrics: 3.5.1.1 and 3.5.1.2)
2. **Which one of the two approaches produces the most expressive tag hierarchies (hierarchy width and depth)?** (evaluation metrics: 3.5.2)

[46] http://www.freebase.com, as visited in December 2015.

[47] http://research.microsoft.com/en-us/projects/probase/default.aspx, as visited in December 2015.

To perform the human-based evaluation, a subset of direct taxonomic pairs (t1, t2) from the learned tag hierarchies is extracted to be manually judged as to whether they are related, and if they are, then what is that relation (Section 3.5.1.2). For choosing the subset of tag pairs, we have followed this a priori filtering:

1. The tag pairs presented in both tag hierarchies (from our and original algorithms) are selected.
2. The tag pairs presented in the previous step and in at least one of the selected reference taxonomies are selected.
3. Then we only kept those tag pairs as candidates for the study where both terms t1 and t2 were present in a popular list of common words that used in the Brown corpus[48]. This step has been also used by other related works, like (Strohmaier et al., 2012).

The first step of the above filtering is needed as we are evaluating the performance of both algorithms. The rest of the filtering process (Steps 2-3) is performed to allow as many meaningful answers as possible from a broad audience. This a priori filtering leads to a list of 247 tag pairs. As a control condition, we also added 50 random term pairs and another 50 term pairs randomly sampled from a gold-standard dataset (WordNet), leading to a final list of 347 term pairs to be judged by human subjects. To motivate more people to be involved in this study, we decided to ask each person to judge only 20 tag pairs of the final list.

A link[49] pointing to the human-based study was disseminated via emails and social media, and was live for a period of four weeks. A total of 450 participants took part in the study. The Ethics form of the experiment has been approved by FPAS Ethics Committee at the University of Southampton (Reference No. 16781, on 9/7/2015; for more details see the Appendix). The following sections will show the results and discussions of this broad evaluation.

---

[48] http://clu.uni.no/icame/manuals/BROWN/INDEX.HTM

[49] http://tagtrees.ecs.soton.ac.uk

6.3 Results and Analysis

This section presents an evaluation of the quality of the tag hierarchy semantics and structure that produced by our proposed approach, compared to the one that produced by the original approach. The following sections will show the results and discussions of this evaluation.

6.3.1 Results of Semantic Evaluation

The semantic evaluation was undertaken in two steps. Firstly by comparing how similar a produced taxonomy is to a related reference taxonomy, and secondly by comparing a sample of relationships in the hierarchy to human judgment.

6.3.1.1 Reference-based Evaluation

Figure 6-1 shows the results of the first semantic evaluation against four reference taxonomies, namely: WordNet, Yago, Freebase and Probase. The y-axis of each figure illustrates the similarity between each tag hierarchy and a reference taxonomy. The similarity is measured by using several measures that were explained in Section 3.5.1.1, including: taxonomic precision (TP), taxonomic recall (TR) and taxonomic F-measure (TF). The insight behind this evaluation is that the more similarity between a tag hierarchy and a reference taxonomy, the higher the semantic quality of that tag hierarchy.

The results from our experiments provide a consistent picture as across all similarity measures (TP, TR and TF), and against all reference taxonomies, the tag hierarchy induced by our algorithm outperforms the original one (Heymann-Benz). In other words, our proposed extended algorithm yields a tag hierarchy that is more similar to all the selected reference taxonomies. It should be noted that (TR) for both algorithms (the original and our algorithms) against WordNet is relatively low, as the size of WordNet is much smaller than other selected reference taxonomies.

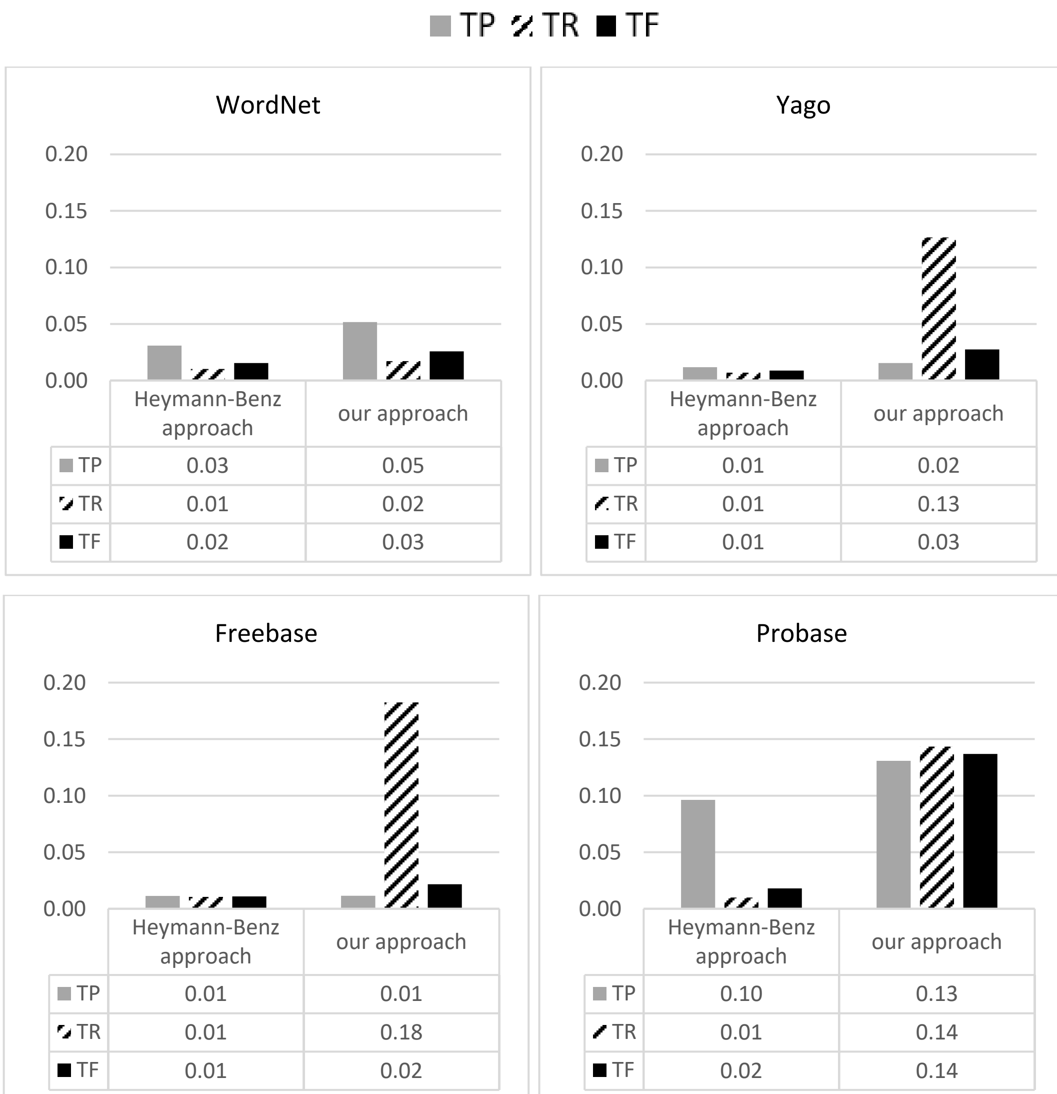


Figure 6-1: Results of semantic evaluation against reference taxonomies

To test the statistical significance of all the results, we performed a two-sample t-Test with a significance level of $\alpha = 0.05$ on each reference taxonomy dataset. The null hypothesis is that there is no difference between the learned tag hierarchies by our algorithm and the original algorithm against the selected reference taxonomy. This hypothesis is rejected, across all the selected taxonomy references, for the similarity measures TP (WordNet: $\alpha = 0.032$, Yago: $\alpha = 0.018$, Freebase: $\alpha = 0.001$, Probase: $\alpha = 0.029$) and TF (WordNet: $\alpha = 0.004$, Yago: $\alpha = 0.018$, Freebase: $\alpha = 0.001$, Probase: $\alpha = 0.031$). This means our algorithm outperforms the original algorithm significantly with respect to TP and TF metrics. For the similarity measure TR, the values of α indicate that there is no statistical evidence for rejecting the null hypothesis (WordNet: α= 0.144, Yago: α= 0.761, Freebase: α= 0.574, Probase: α= 0.576).

As the process of automatically or collaboratively creating and maintaining knowledge resources (like Yago, Freebase and Probase) are facing difficulties and potential errors, a human assessment of the learned tag hierarchies was performed (whose results will be discussed next) as a further check the validity of the results in this section.

6.3.1.2 Human-based Evaluation

To perform the human-based evaluation, a subset of 347 direct taxonomic pairs (t1, t2) from the learned tag hierarchies is extracted (following a priori filtering; Section 6.2) to be manually judged as to whether they are related, and if they are, then how. The relation between each term pair can be one of the following options:

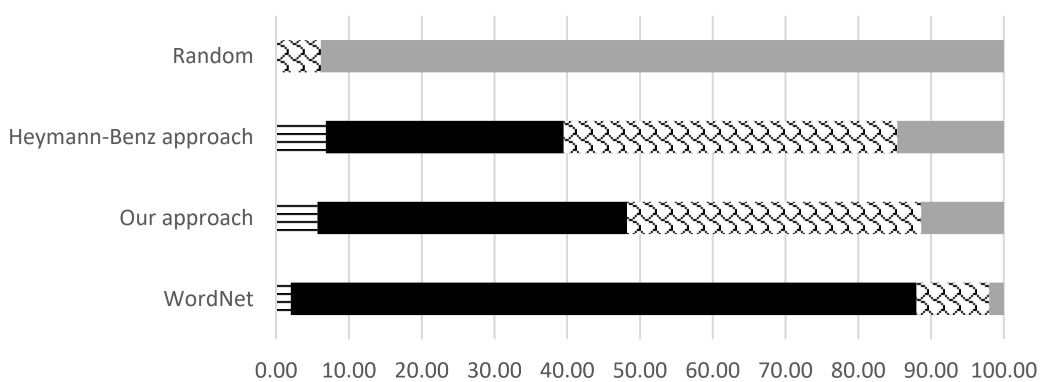
1. t1 is the same as t2.
2. t1 is a (kind of/part of) t2.
3. t1 is somehow related to t2.
4. t1 is not related to t2.
5. Unclear; because the meaning of t1 or t2 is not clear.

The insight behind this approach is that a better tag hierarchy will have a higher percentage of pairs being judged as “kind of” or “part of”, and a lower percentage of pairs being judged as “not related”.

450 participants took part in a human-based evaluation. Participants were recruited through emails and social media, and asked to judge the relationship between 20 different tag pairs. Since some of them did not completely finish the study, we received 5265 pair judgments, including 233 that the participants classified as “unclear”, leading to a total of 5032 usable judgments for our study. All the selected term pairs in the study were judged by at least 15 participants, and for each term pair, we computed the average (mode) of its answers over each

tag hierarchy construction algorithm. In order to have a reliable assessment, we removed 17 tag pairs that have very sparse voting of judgments, leading to a final list of 330 tag pairs.

Figure 4-12 summarizes the results of semantic evaluation by human assessment of the produced tag hierarchies. The middle two rows correspond to the two tag hierarchies (generated by our and the original algorithms), while the topmost and the lowermost rows depicts a control condition based on random dataset (expected to be poor quality) and the WordNet taxonomy (expected to be high quality) respectively; these are included to give the results for the two algorithms some context. The values on the y-axis illustrate the percentage of relation types between tags for each tag hierarchy. As previously mentioned, a higher quality tag hierarchy should have a higher percentage of direct taxonomic pairs being judged as is-a: kind/part of, and a lower percentage of direct taxonomic pairs being judged as not related.



	WordNet	Our approach	Heymann-Benz approach	Random
same as	2.00	5.67	6.83	0.00
kind of/part of	86.00	42.55	32.68	0.00
somehow related	10.00	40.43	45.85	6.12
not related	2.00	11.35	14.63	93.88

% of the judgment

same as   kind of/part of   somehow related   not related

Figure 6-2: Results of semantic evaluation by human assessment

As expected, the term pairs produced by the random algorithm are judged as the worst, with only 6.12% of somehow related tag pairs. Also as expected the term pairs selected from the WordNet taxonomy are judged as the best, with all positive relations (same as/kind of/part of/somehow related) adding up to 98%, including a large portion (86%) of the most desired "kind of/part of" relations.

In between these extremes, the results show that our approach yields a higher percentage of "kind of/part of" answers (42.55%) and a lower percentage of "not related" answers (11.35%) compared to the original algorithm (32.68% of "kind of/part of" and 14.63% of "not related" answers). These results confirmed the results we have achieved from the reference-based evaluation.

In conclusion the results of the semantic evaluation shows that our proposed algorithm leads to tag hierarchies that capture a higher semantics compared to the one obtained from the original algorithm. The following section will discuss the results of the structural evaluation.

6.3.2 Results of Structural Evaluation

Figure 6-3 shows the results of a structural evaluation of the two produced tag hierarchies that produced by our algorithm and the original algorithm. The y-axis shows the AUT results of the two tag hierarchies. The tag hierarchy produced by our algorithm yields the highest AUT result, with 10967 score, which indicates that this hierarchy is bushier and deeper than the one by the original algorithm, with 9197.5 AUT score.

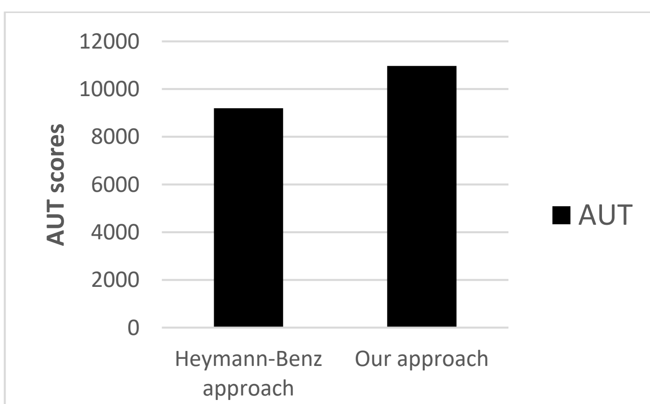


Figure 6-3: Results of structural evaluation (AUT)

6.4 Chapter Summary

This chapter presents an extensive evaluation to assess the tag hierarchies produced by our proposed approach, using lexico-syntactic patterns applied to an open text corpus, such as the Web via Bing. The evaluation methodology, the used datasets and the results are described in detail.

The results of this extensive evaluation shows that our proposed algorithm leads to tag hierarchies that capture a higher semantics compared to the one obtained from the original algorithm. We have shown that across all similarity metrics (TP, TR and TF) our approach generated tag hierarchies that were more similar to the reference taxonomies than the original algorithm. Moreover, our algorithm outperforms the original algorithm significantly with respect to TP and TF metrics. In comparison to human judgment our algorithm increases the number of relations judged as kind of/part of from 32.68% (the original algorithm) to 42.55%, and reduces the number of relations judged as non-related from 14.63% to 11.35%. In terms of the tag hierarchy structure, the tag hierarchy produced by our algorithm yields a higher AUT score, which indicates that this hierarchy is bushier and deeper than the one produced by the original algorithm.

The next chapter will summarise the research, its key findings, and the potential future work that can be done to take the research forward.

Chapter Seven

Conclusions and Future Work

In this thesis I have detailed my work exploring the issues around automatically building effective knowledge structures based on collective intelligence, and documented a set of experiments that investigate the validity of our proposed approaches to building tag hierarchies. This final chapter covers the main conclusions and findings of this research. It also identifies key areas for future work that could take the research forward.

7.1 Research Summary

The main objective of this thesis has been to propose an automatic approach to building high-quality tag hierarchies. To achieve this, a comprehensive review of the current approaches to constructing tag hierarchies from folksonomies, and their limitations, were discussed. In addition, a broad evaluation process for automated hierarchical structures construction was presented.

Due to the lack of consistent structure in folksonomies, which causes many problems, such as homonym, synonym and basic level variation, many researchers have been working on approaches 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 tag hierarchy algorithms suffer from the "generality-popularity" tags problem; as they (sometimes inaccurately) assume that because a tag occurs more frequently it must be more general and thus appear higher in the hierarchy. Consequently, this thesis has presented solutions to reduce the impact of this problem and build better tag hierarchy structure and semantics. These proposed solutions have been developed through the stages and activities listed in the following sections.

7.1.1 An Approach to Building High-Quality Tag Hierarchies from Taxonomic Tag Pairs

A first investigation to tackle the "generality-popularity" tags problem was by introducing a new tagging approach for moving from collective folksonomies to collective taxonomies. In this new tagging approach, we propose making a small change to the current tagging approach, by asking participants to tag in the form of "is-a" relationship, in order to make a big change to the type of knowledge structure that can be built. Although this tag pairs approach shares some of the issues of individual tags, such as spelling errors, it also provides additional semantics between tags. The algorithm we have developed to building tag hierarchy from taxonomic tag pairs (Algorithm 4-1) is an extension of Benz's, which itself is an extension of Heymann's algorithm (Chapter 4).

7.1.2 Experiment 1: The Impact of the New Tagging Approach

A pilot study was conducted to ascertain whether the new tagging approach has a genuine impact on the semantic of the learned taxonomic tags, and whether this in turn has an impact on the tag hierarchy as a whole. The results of the empirical experiment showed that there was a remarkable difference in the semantic level between tag hierarchies constructed from the new tagging approach and the normal one (individual tags). The tag hierarchy constructed from our proposed tagging approach is much more similar to an expert-crafted taxonomy (as a reference taxonomy), with taxonomic F-measure equal to 70%, than ones constructed from individual tags, with taxonomic F-measure equal to

8%. The proposed algorithm succeeded in eliminating noisy tags and tackling the lack of consistent structure in folksonomies. However, the resulting tag hierarchy from the new tagging approach is less expressive, with 11 AUT score, than those generated by individual tags, with 44.5 AUT score. This leads us to the insight of our second approach that if we could improve the accuracy of directions in relations constructed between tags by a generality-based automatic approach, we would be able to improve the quality of the resulting tag hierarchy structure and semantics without sacrificing richness.

7.1.3 An Improved Approach to Building High-Quality Tag Hierarchies

Based on the results achieved by Experiment (1), the thesis introduced a new approach to building tag hierarchy by using an existing knowledge resource to improve the accuracy of taxonomic directions when building tag hierarchies. The effectiveness of our approach is examined and evaluated in three experiments, using three different types of knowledge resources: tag relationships by users (e.g. Delicious Bundles; Experiment 2), a closed text corpus (e.g. a download of English Wikipedia; Experiment 3), and an open text corpus (e.g. World Wide Web via Bing; Experiment 4). For the first type we will rely on a simple match between the generated tag pairs and the ones by users in the Delicious Bundles dataset, whereas for the second and third types we will use lexico-syntactic patterns (Section 5.1). While lexico-syntactic patterns tend to achieve a very high level of precision, but low recall, 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.

7.1.4 Experiment 2: Using Personal Tag Relationships with our Improved Approach to Building High-Quality Tag Hierarchies

The key aim of this experiment was to explore whether personal tag relationships created by users (e.g. Delicious Bundles) would improve the accuracy of taxonomic tag directions that built from individual tags by using a promising generality-based approach. To test the performance of our approach, we applied the original approach and our proposed approach, using five common tag similarity measures and with different similarity thresholds, to two large-scale folksonomies collected from Delicious and Flickr, yielding 60 different tag hierarchies. Since in this experiment we were focusing on checking and correcting the taxonomic tag pairs that we get from our proposed algorithm, we evaluated all the taxonomic tag pairs from all the resulting tag hierarchies against a gold-standard dataset (WordNet). The results of the experiment have shown that generality acts as a successful proxy for popularity in 47% to 76% of cases (depending on the similarity measure used), and that there is a modest improvement achieved by our proposed algorithm compared to the original algorithm among all the selected tag similarity measures (correct between 0.90% to 5.28% for Delicious, and 3.57% to 8.37% for Flickr, depending on the similarity measure used). This improvement will result in building higher quality tag hierarchy structure and semantics.

The following experiment explored the use of lexico-syntactic patterns applied to a closed large text corpus (e.g. English Wikipedia) for improving the accuracy of taxonomic tag directions that built from individual tags.

7.1.5 Experiment 3: Using a Closed Text Corpus with our Improved Approach to Building High-Quality Tag Hierarchies

In this approach we extended a promising generality-based approach by using lexico-syntactic patterns applied to a large text corpus specifically the text of English Wikipedia. The patterns that our approach uses are a combination of the well-known Hearst's lexico-syntactic patterns (Table 3-1), and another direct pattern: "*Tag t1* is a/an *Tag g1*".

To test the performance of our approach, we applied the original approach and our proposed approach, using five common tag similarity measures and with different similarity thresholds, to two large-scale folksonomies collected from Delicious and Flickr, yielding 60 different tag hierarchies. In this experiment we were focusing on checking and correcting the taxonomic tag pairs that we get from our proposed algorithm, therefore, we evaluated all the taxonomic tag pairs from all the resulting tag hierarchies against a gold-standard dataset (WordNet).

The results of the experiment have shown that the performance of our proposed algorithm outperforms the original algorithm among all the selected tag similarity measures (correct between 3.63% to 16.75% for Delicious, and 9.64% to 26.90% for Flickr, depending on the similarity measure used). This improvement will result in building higher quality tag hierarchy structure and semantics.

The following experiment explored the use of an open knowledge repository (the Web via Bing) instead of a closed knowledge resource (a download of English Wikipedia).

7.1.6 Experiment 4: Using an Open Text Corpus with our Improved Approach to Building High-Quality Tag Hierarchies

The key aim of this experiment was to increase the coverage and occurrences of the tags in any tag collection by using an open knowledge repository, i.e. the Web via Bing, instead of a closed knowledge resource, i.e. a download of English Wikipedia. The results of this empirical experiment have shown that there is an improvement in the coverage of the learned taxonomic tag pairs from 26% in Wikipedia to 86% in the Web via Bing. This improvement has a positive impact on using our proposed approach to improve the semantic quality of the learned tag hierarchies as the results of the semantic evaluation shows that there is an improvement achieved by our proposed algorithm compared to the original algorithm (from 76.63% to 86.09%). Also, there is an improvement in the accuracy of the taxonomic tag pairs generated by our approach and using the Web via Bing, compared to ones that are generated by our approach and using the English Wikipedia dataset (from 81.30% to 86.09%). This is because of the increased coverage and occurrences of the tag pairs found in the Web via Bing, which consequently improves the semantic quality of learned taxonomic tag pairs.

7.1.7 The Impact of Correcting Taxonomic Directions on the Quality of Tag Hierarchies

Based on the results we achieved in the previous experiments, an extensive evaluation to assess the tag hierarchies produced using our improved approach was conducted. This would give us a measure of how the improvements in tag pair directions can be translated into improved tag hierarchies.

The results of this extensive evaluation shows that our proposed algorithm leads to tag hierarchies that capture a higher semantics compared to the one obtained from the original algorithm. We have shown that across all similarity metrics (TP, TR and TF) our approach generated tag hierarchies that were more similar to the reference taxonomies than the original algorithm. In comparison to human judgment our algorithm increases the number of relations judged as kind of/part of from 32.68% (the original algorithm) to 42.55%, and reduces the number of relations judged as non-related from 14.63% to 11.35%. In terms of the tag hierarchy structure, the tag hierarchy produced by our algorithm yields a higher AUT score, with a 10967 score, which indicates that this hierarchy is bushier and deeper than the one by the original algorithm, with a 9197.5 AUT score.

7.2 Research findings

The research contributions have been created during our work to investigate the original hypothesis which was:

Lexico-syntactic patterns applied to a large text corpus can be used to improve the accuracy of directions in relations constructed between tags by an approach to tag hierarchy construction, and to improve the quality of the resulting tag hierarchy structure and semantics.

This hypothesis was broken down into three research questions:

1. To what extent do high quality tag pairs captured directly from users change the quality of constructed tag hierarchies?
2. Can lexico-syntactic patterns applied to a closed text corpus improve the direction of automatically derived tag pairs, and how is this affected when the lexico-syntactic patterns are applied to an open text corpus, such as the open web?
3. Will the improvement of the accuracy of taxonomic tag directions translate to higher quality tag hierarchy structure and semantics?

Evidence can be found in this thesis to answer these questions. **Question one** aimed to explore the impact of two things: First, the impact of gathering taxonomic tag pairs from users rather than individual tags in the quality of the learned tag hierarchies. Second, the impact of using personal tag relationships created by users on improving the accuracy of taxonomic tag directions constructed from individual tags by a generality-based approach. For the first exploration, we proposed a new tagging approach that takes the form of "is-a" relationship, where users should type two related tags. And for the second exploration, we performed an experiment to use personal tag relationships for improving the accuracy of the taxonomic tag directions built from individual tags. In Chapter 4 and 5 this question was answered with five main results found:

1. Our extended algorithm succeeded in eliminating noisy tags and tackling the lack of consistent structure in folksonomies.
2. The tag hierarchy constructed from our new tagging approach (tag pairs) is much more similar to an expert-crafted taxonomy (as a reference taxonomy), with taxonomic F-measure equal to 70%, than ones constructed from the regular approach (individual tags), with taxonomic F-measure equal to 8%. This means that our new tagging approach succeeded in improving the semantic quality of the learned tag hierarchy.
3. On the other hand, the new tagging approach generated less expressive tag hierarchy, with 11 AUT score, than those generated by individual tags, with 44.5 AUT score.
4. The usability of the new tagging approach (in terms of efficiency, effectiveness and satisfaction) is marginal acceptable, with 54.6% SUS score.
5. Using personal tag relationships with our extended algorithm resulted in a modest improvement in the semantic quality of the learned tag hierarchies compared to the ones constructed by the original algorithm, among all the selected tag similarity measures.

The above results have shown that collecting taxonomic tag pairs increases the semantic quality of the tag hierarchy, but at the expense of expressivity, and with some degradation of user experience. Moreover, using personal tag relationships would improve the accuracy of the taxonomic tag pairs constructed from individual tags.

Question Two aimed to explore two things: First, the impact of using lexico-syntactic patterns applied to a closed text corpus in the semantic quality of the learned taxonomic tags. Our proposed approach extended a promising generality-based algorithm by using lexico-syntactic patterns to check the direction of hyponym/hypernym relations in order to distinguish between popular and general tags. Second, the impact of using an open knowledge repository (e.g. the Web) instead of a closed text corpus. In Chapter 5 this question was answered with three main results found:

1. Our proposed approach outperformed the original algorithm, among all the selected knowledge resources and tag similarity measures (correct between 3.63% to 16.75% for Delicious, and 9.64% to 26.90% for Flickr).
2. The coverage of the examined taxonomic tag pairs has increased from 26% in Wikipedia to 86% in the Web via Bing.
3. The accuracy of the taxonomic tag pairs generated by our approach and using the Web via Bing is improved, with an accuracy of 85.37%, compared to ones that generated by our approach and using English Wikipedia, with an accuracy of 81.11%.

Question Three aimed to explore whether the improvement in the *accuracy of directions* in relations constructed between tags would subsequently improve the *quality of the resulting tag hierarchy* in terms of structure and semantics. In Chapter 6 this question was answered with two main results found:

1. Across all similarity measures (TP, TR and TF) and the selected reference taxonomies, our approach generated higher semantic quality tag hierarchies by improving the accuracy of the constructed tag pair directions. Taking TF measure as an example, across all experimental settings the tag hierarchy constructed by our approach is more similar to the four selected reference taxonomies, with TF= 0.03, 0.03, 0.02 and 0.14, compared to the one constructed by the original algorithm, with TF=

0.02, 0.01, 0.01 and 0.02 respectively. Similar evidences of the improvement can be found by the other measures (TP and TR).

2. By improving the accuracy of the learned tag pair directions, the resulting tag hierarchy has yielded a higher AUT score, with 10967 score. This indicates that the resulting hierarchy is bushier and deeper than the one constructed by the original approach, with 9197.5 AUT score.

The work in this thesis has shown a positive outcome for the hypothesis we started from. We have demonstrated a successful approach to building tag hierarchy that we developed based on existing work in tag hierarchy building and lexico-syntactic patterns. We have also shown that this approach improved further by using an open text corpus, such as the Web, instead of a closed text corpus.

7.3 Future Work

Although significant conclusions have been achieved in this thesis, the work has raised further questions that deserve further research. In this section we identify three key areas for future work.

7.3.1 Building Tag Hierarchies from Different Sources

While the research has provided a dynamic approach to building tag hierarchies from any tag collection, it has only been used in datasets extracted from two collaborative tagging system; i.e. Delicious and Flickr. The success and spread of Web 2.0 systems has led to the emergence of new forms and applications of social tagging systems. As the characteristics of these systems are varied, the way to treat them might be different. Thus, the question arises to what extent our approach helps in building high-quality tag hierarchies from a particular kind of social tagging systems. For instance, microblogging platforms, such as Twitter, provide a different tagging approach, where users type hashtags within the content they post. A comparative study of using different kinds of social tagging systems should help in providing a framework by which latent structures and semantics from those different systems can be categorised.

7.3.2 Expanding the Use of Lexico-Syntactic Patterns

This thesis has demonstrated a successful approach to building high-quality tag hierarchies based on lexico-syntactic patterns applied to an open text corpus, such as the Web. Also, the thesis has shown an exploration of which lexico-syntactic patterns are most successful in correcting wrong directions of the generated tag pairs, and whether any introduce significant errors, based on the selected similarity measure. Further investigation in this area might perform more empirical experiments and deep analysis to provide suggestions on which lexico-syntactic pattern that we proposed should be excluded from our approach.

As the proposed approach presented in this thesis does not always produce sharp taxonomic (is-a) relations, the research question arises can lexico-syntactic patterns be used to eliminate non-taxonomic tag pairs. Another relevant question is to what extend lexico-syntactic patterns can help in detecting the kind of semantic relation between constructed tag pairs. This detection would provide further improvement to content retrieval and enriching knowledge bases tasks.

7.3.3 Case Study based Evaluation of Tag Hierarchy Construction

This thesis presented a broad evaluation process that involves a mix of objective and subjective metrics to evaluate the quality of tag hierarchies in terms of structure and semantics. Future work in this area would involve a task-based evaluation where users will be involved in assessing the usefulness of a tag hierarchy for searching or browsing a social tagging system. This evaluation could also aid in discovering the weaknesses of an approach of tag hierarchy construction in order to improve its performance.

7.4 Final Conclusions

In this thesis I have explored the issues around automatically building effective knowledge structures based on collective intelligence, and proposed changes to the state-of-the-art approaches that improve their performance. This improvement was achieved by providing three approaches to tackle the "generality-popularity" tags problem, in that popularity is assumed (sometimes inaccurately) to be a proxy for generality, i.e. high-level taxonomic terms will occur more often than low-level ones.

In the first we propose a change to the current tagging approach with the aim of leading to a big change to the type of knowledge structure that can be built. The new tagging approach takes the form of "is-a" relationship, where users should type two related tags (e.g. Tag t1 is-a Tag g1). The results of our experiments have demonstrated that collecting taxonomic tag pairs increases the semantic quality of the tag hierarchy, but at the expense of expressivity, and with some degradation of user experience. This leads us to the insight of our second approach to tackle the "generality-popularity" tags problem 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 or changing the user experience.

Our second approach proposes to use personal tag relationships created by users for improving the accuracy of the taxonomic tag directions built from individual tags. We tested this approach by using Delicious Bundles and tag hierarchies constructed from a Delicious dataset, as well as tag hierarchies constructed from another folksonomy dataset (e.g. a Flickr dataset). Our results have shown that using personal tag relationships with our extended algorithm would improve the semantic quality of the learned tag hierarchies. However, the coverage of the examined taxonomic tag pairs generated from a Flickr dataset found in Delicious Bundles is much less than the ones generated from the Delicious dataset. This indicates that is better to use personal tag relationships from the same collaborative tagging system that used to construct tag hierarchies. Yet, not every tagging system allows users to organize content hierarchically.

Our third approach extends a promising generality-based algorithm by using lexico-syntactic patterns to check the direction of hyponym/hypernym relations in order to distinguish between popular and general tags. Our experiments have shown that our proposed algorithm succeeds in generating more high-quality taxonomic tag pairs compared to the original algorithm. This improvement results in building higher quality tag hierarchy structure and semantics. Also, our experiments have confirmed that using an open text corpus instead of a closed text corpus increases the coverage and occurrences of the tags in the investigated tag collection, which consequently improves the quality of the learned tag hierarchies in terms of structure and semantics.

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 help us get new value from the Social Web.

References

Abel, F., Henze, N. & Krause, D., 2008. A Novel Approach to Social Tagging: GroupMe! In *the 4th International Conference on Web Information Systems and Technologies*. Madeira, Portugal. Springer.

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2013. An Approach to Building High-Quality Tag Hierarchies from Crowdsourced Taxonomic Tag Pairs. In *the 5th International Conference on Social Informatics (SocInfo 2013)*. Kyoto, Japan. Springer.

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2014. Improving on Popularity as a Proxy for Generality When Building Tag Hierarchies from Folksonomies. In *the 5th International Conference on Social Informatics (SocInfo 2014)*. Barcelona, Spain. Springer.

Almoqhim, F., Millard, D.E. & Shadbolt, N., 2015. The Horse before the Cart: improving the accuracy of taxonomic directions when building tag hierarchies. In *the 8th Saudi Students Conference in the UK*. London. Imperial College Press.

Ames, M. & Naaman, M., 2007. Why We Tag: motivations for annotation in mobile and online. In *the SIGCHI Conference on Human factors in Computing Systems*. San Jose, California, USA. ACM.

Andrews, P. & Pane, J., 2013. Sense Induction in Folksonomies: a review. *Artificial Intelligence Review*, 40(2), pp.147-74.

Angeletou, S., Sabou, M. & Motta, E., 2008. Semantically Enriching Folksonomies with FLOR. In *the 1st International Workshop on Collective Semantics:*

Collective Intelligence & the Semantic Web (CISWeb 2008) at the 5th Annual European Semantic Web Conference (ESWC 2008). Tenerife, Spain.

Angeletou, S., Sabou, M., Specia, L. & Motta, E., 2007. Bridging the Gap between Folksonomies and the Semantic Web: an experience report. In *the Workshop on Bridging the Gap between Semantic Web and Web 2.0 at the 4th European Semantic Web Conference (ESWC 2007)*. Innsbruck, Austria.

ANSI/NISO, 2005. Guidelines for the Construction, Format, and Management of Monolingual Thesauri. *National Information Standards Organization (Z3919-1993)*.

Bangor, A., Kortum, P.T. & Miller, J.T., 2008. An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction*, 24(6), pp.574-94.

Bast, H., Bäurle, F., Buchhold, B. & Haußmann, E., 2014. Easy Access to the Freebase Dataset. In *the 23rd International Conference on World Wide Web*. Seoul, Korea. ACM.

Begelman, G., Keller, P. & Smadja, F., 2006. Automated Tag Clustering: Improving search and exploration in the tag space. In *the Collaborative Web Tagging Workshop at the 15th International World Wide Web Conference (WWW2006)*. Edinburgh, Scotland.

Benz, D., Hotho, A., Jäschke, R., Krause, B., Mitzlaff, F., Schmitz, C. & Stumme, G., 2010. The Social Bookmark and Publication Management System BibSonomy. *The VLDB Journal*, 19(6), pp.849-75.

Benz, D., Hotho, A., Stützer, S. & Stumme, G., 2010. Semantics Made by You and Me: Self-emerging ontologies cancapture the diversity of shared knowledge. In *2nd Web Science Conference (WebSci10)*. Raleigh, NC, USA.

Benz, D., Körner, C., Hotho, A., Stumme, G. & Strohmaier, M., 2011. One Tag to Bind Them All: Measuring term abstractness in social metadata. In *the 8th Extended Semantic Web Conference on The Semanic Web: Research and Applications*. Crete, Greece. Springer.

Berland, M. & Charniak, E., 1999. Finding Parts in Very Large Corpora. In *the 37th Annual Meeting of the Association for Computational Linguistics (ACL)*. Stroudsburg, PA, USA.

Berners-Lee, T., Hendler, J. & Lassila, O., 2001. The Semantic Web. *Scientific American*, pp.34-43.

Bloehdorn, S., 2008. Kernel Methods for Knowledge Structures. *Ph.D. thesis, Karlsruhe Institute of Technology*.

Bollacker, K., Evans, C., Paritosh, P., Sturge, T. & Taylor, J., 2008. Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge. In *the 2008 ACM SIGMOD international conference on Management of data*. Vancouver, Canada. ACM.

Borst., W., 1997. Construction of Engineering Ontologies. *Ph.D. thesis, University of Twente, TheNetherlands*.

Brooke, J., 1996. SUS: A quick and dirty usability scale. In P.W.Jordan, B. Thomas, B.A. Weerdmeester & a.I.L. McClelland, eds. *Usability Evaluation in Industry*. London: Taylor and Francis. pp.189-94.

Buchanan, B.G., 2005. A (Very) Brief History of Artificial Intelligence. *AI Magazine*, 26(4), pp.53-60.

Cai, Y., Chen, W.-H., Leung, H.-F., Li, Q., Xie, H., Lau, R.Y.K., Min, H. & Wang, F.L., 2016. Context-aware Ontologies Generation with Basic Level Concepts from Collaborative Tags. *Neurocomputing*, 208, pp.25 - 38.

Cantador, I., Szomszor, M., Alani, H., Fernández, M. & Castells, P., 2008. Enriching Ontological User Profiles with Tagging History for Multi-domain Recommendations. In *the 1st International Workshop on Collective Semantics: Collective Intelligence & the Semantic Web (CISWeb 2008)*. Tenerife, Spain.

Cardoso, J., 2006. Approaches to Developing Semantic Web Services. *International Journal of Computer Science*, 1(1), pp.8-21.

Cardoso, J., 2007. The Syntactic and the Semantic Web. In *Semantic Web Services: Theory, tools and applications*. Hershey, PA: IGI Global. pp.1-23.

Cattuto, C., Benz, D., Hotho, A. & Stumme, G., 2008. Semantic Grounding of Tag Relatedness in Social Bookmarking Systems. In *the 7th International Semantic Web Conference (ISWC 2008)*. Karlsruhe, Germany. Springer.

Chandrasekaran, B., Josephson, J.R. & Benjamins, V.R., 1999. What are Ontologies, and Why Do We Need Them? *IEEE Intelligent Systems*, 14(1), pp.20–26.

Chen, J., Feng, S. & Liu, J., 2014. Topic Sense Induction from Social Tags based on Non-Negative Matrix Factorization. *Information Sciences*, 280, pp.16 - 25.

Cimiano, P., 2006. *Ontology Learning and Population from Text: Algorithms, Evaluation and Applications*. Springer.

Cimiano, P., Hotho, A. & Staab, S., 2005. Learning Concept Hierarchies from Text Corpora using Formal Concept Analysis. *Journal of Artificial Intelligence Research*, 24(1), pp.305-39.

Corbett, D., 2004. Interoperability of Ontologies Using Conceptual Graph Theory. In K.E. Wolff, H.D. Pfeiffer & H.S. Delugach, eds. *Conceptual Structures at Work, (LNCS), volume 3127*. Springer. pp.375-87.

Crampes, M. & Ranwez, S., 2000. Ontology-Supported and Ontology-Driven Conceptual Navigation on the World Wide Web. In *the 11th ACM conference on Hypertext and hypermedia (HYPERTEXT '00)*. San Antonio, Texas, USA. ACM.

d'Aquin, M. & Noy, N.F., 2012. Where to Publish and Find Ontologies? A survey of ontology libraries. *Web Semantics: Science, Services and Agents on the World Wide Web*, 11, pp.96 - 111.

Dellschaft, K. & Staab, S., 2006. On How to Perform a Gold Standard based Evaluation of Ontology Learning. In *the 5th International Semantic Web Conference (ISWC 2006)*. Athens, GA, USA. Springer.

Dix, A., Levialdi, S. & Malizia, A., 2006. Semantic Halo for Collaboration Tagging Systems. In *the Workshop on Social Navigation and Community-based Adaptation Technologies at the Int. Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*. Dublin, Ireland.

Dong, H., Wang, W. & Liang, H.-N., 2015. Learning Structured Knowledge from Social Tagging Data: A critical review of methods and techniques. In *IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*. Chengdu, China. IEEE.

Fang, Q., Xu, C., Sang, J., Hossain, M.S. & Ghoneim, A., 2016. Folksonomy-Based Visual Ontology Construction and Its Applications. *IEEE Transactions on Multimedia*, 18(4), pp.702-13.

Farooq, U., Kannampallil, T.G., Song, Y., Ganoe, C.H., Carroll, J.M. & Giles, C.L., 2007. Evaluating Tagging Behavior in Social Bookmarking Systems: metrics and design heuristics. In *the International ACM Conference on Supporting Group Work (GROUP '07)*. Sanibel Island, Florida, USA. ACM.

Faure, D. & Nedellec, C., 1998. 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 at the 1st International Conference on Language Resources and Evaluation*. Granada, Spain.

Frey, B.J. & Dueck, D., 2007. Clustering by Passing Messages between Data Points. *science*, 315(5814), pp.972-76.

Gamper, J., Nejdl, W. & Wolpers, M., 1999. Combining Ontologies and Terminologies in Information Systems. In *the 5th International Congress on Terminology and Knowledge Engineering*. Innsbruck, Austria.

Garcia-Silva, A., Corcho, O., Alani, H. & Gomez-Perez, A., 2012. Review of the state of the art: Discovering and Associating Semantics to Tags in Folksonomies. *The Knowledge Engineering Review*, 27(1), pp.57-85.

García-Silva, A., García-Castro, L.J., García, A. & Corcho4, O., 2015. Building Domain Ontologies Out of Folksonomies and Linked Data. *International Journal on Artificial Intelligence Tools*, 24(2).

Garcia, A., Szomszor, M., Alani, H. & Corcho, O., 2009. Preliminary Results in Tag Disambiguation Using DBpedia. In *the 1st International Workshop on Collective Knowledge Capturing and Representation (CKCaR'09)*. California, USA.

Garshol, L.M., 2004. Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of It All. *Journal of InformationScience*, 30(4), pp.378–91.

Giannakidou, E., Koutsonikola, V., Vakali, A. & Kompatsiaris, Y., 2008. Co-clustering Tags and Social Data Sources. In *the 9th International Conference On Web-Age Information Management*. Zhangjiajie, China.

Gilchrist, A., 2003. Thesauri, Taxonomies and Ontologies – An etymological note. *Journal of Documentation*, 59(1), pp.7-18.

Giles, J., 2005. Internet Encyclopaedias Go Head to Head. *Nature*, 438(7070), pp.900-01.

Giunchiglia, F., Marchese, M. & Zaihrayeu, I., 2006. Encoding Classifications into Lightweight Ontologies. *Journal on Data Semantics*, VIII(4380), pp.57–81.

Giunchiglia, F. & Zaihrayeu, I., 2009. Lightweight ontologies. In L. Liu & M.T. Özsu, eds. *Encyclopedia of Database Systems*. Springer. pp.1613-19.

Godoy, D. & Corbellini, A., 2016. Folksonomy-based Recommender Systems: A state-of-the-art review. *International Journal of Intelligent Systems*, 31(4), pp.314–46.

Golbeck, J., Koepfler, J. & Emmerling, B., 2011. An Experimental Study of Social Tagging Behavior and Image Content. *Journal of the American Society for Information Science & Technology*, 62(9), pp.1750–60.

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

Greene, K.K., Byrne, M.D. & Everett, S.P., 2006. A Comparison of Usability between Voting Methods. In *the USENIX/ACCURATE Electronic Voting Technology Workshop*. Vancouver, Canada.

Gruber, T.R., 1993. A Translation Approach to Portable Ontologies. *Knowledge Acquisition*, 5(2), pp.199–220.

Gruber, T., 2008. Collective Knowledge Systems: Where the Social Web meets the Semantic Web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6(1), pp.4-13.

Gruber, T., 2009. Ontology. In L. Liu & M.T. Özsu, eds. *Encyclopedia of Database Systems*. Springer. pp.1963-65.

Gruninger, M., Bodenreider, O., Olken, F., Obrst, L. & Yim, P., 2008. Ontology Summit 2007 – Ontology, Taxonomy, Folksonomy: Understanding the distinctions. *Applied Ontology*, 3(3), pp.191–200.

Guarino, N., 1998. Some Ontological Principles for Designing Upper Level Lexical Resources. In *the 1st International Conference on Language Resources and Evaluation*. Granada, Spain.

Gupta, M., Li, R., Yin, Z. & Han, J., 2011. An Overview of Social Tagging and Applications. In C.C. Aggarwal, ed. *Social Network Data Analytics*. Springer. pp.447–97.

Gu, C., Yin, G., Wang, T., Yang, C. & Wang, H., 2015. A Supervised Approach for Tag Hierarchy Construction in Open Source Communities. In *the 7th Asia-Pacific Symposium on Internetware*. Wuhan, China. ACM.

Guy, M. & Tonkin, E., 2006. Tidying up Tags. *D-Lib Magazine*, 12(1).

Hamasaki, M., Matsuo, Y., Nishimura, T. & Takeda, H., 2007. Ontology Extraction using Social Network. In *International Workshop on the Semantic Web for Collaborative Knowledge Acquisition*. Hyderabad, India.

Hammond, T., Hannay, T., Lund, B. & Scott, J., 2005. Social Bookmarking Tools (I) A general review. *D-Lib Magazine*, 2(4).

Harmelen, F.v., Lifschitz, V. & Porter, B., 2008. *Handbook of Knowledge Representation*. Oxford, UK: Elsevier B.V.

Harris, Z.S., 1968. *Mathematical Structures of Language*. John Wiley and Son.

Hearst, M.A., 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *the 14th conference on Computational linguistics (COLING '92)*. Nantes, France. ACM.

Hearst, M.A., 1998. Automated Discovery of WordNet Relations. In C. Fellbaum, ed. *WordNet: An Electronic Lexical Database*. Cambridge, MA.: MIT Press. pp.131-53.

Heymann, P. & Garcia-Molina, H., 2006. *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*. InfoLab Technical Report. Stanford.

Hong, L., Chi, E.H., Budiu, R., Pirolli, P. & Nelson, L., 2008. SparTag.us: A low cost tagging system for foraging of web content. In *the Working Conference on Advanced Visual Interfaces (AVI 2008)*. Napoli, Italy. ACM.

Horrocks, I., 2013. What Are Ontologies Good For? In B.-O. Küppers, U. Hahn & S. Artmann, eds. *Evolution of Semantic Systems*. Springer. pp.175-88.

Hotho, A., JΥaschke, R., Schmitz, C. & Stumme, G., 2006. Information Retrieval in Folksonomies: Search and ranking. In Sure, Y. & Domingue, J., eds. *The Semantic Web: Research and Applications, volume 4011 of Lecture Notes in Computer Science*. Budva, Montenegro. Springer.

ISO, 1998. Ergonomic Requirements for Office Work with Visual Display Terminal (VDT's)–Part 11: Guidance on Usability (ISO 9241–11(E)). Geneva, Switzerland.

Jabeen, F., Khusro, S., Majid, A. & Rauf, A., 2016. Semantics Discovery in Social Tagging Systems: A review. *Multimedia Tools and Applications*, 75(1), pp.573 – 605.

Joorabchi, A., English, M. & Mahdi, A.E., 2015. Automatic Mapping of User Tags to Wikipedia Concepts: The case of a Q&A website – StackOverflow. *Journal of Information Science*, 41(5).

Kalboussi, A., Omheni, N., Mazhoud, O. & Kacem, A.H., 2015. How to Organize the Annotation Systems in Human-Computer Environment: Study, Classification and Observations. In Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P. & Winckler, M., eds. *the 15th IFIP TC.13 International Conference on Human-Computer Interaction (INTERACT 2015)*. Bamberg, Germany. Springer.

Kiu, C.-C. & Tsui, E., 2010. TaxoFolk: A hybrid taxonomy–folksonomy classification for enhanced knowledge navigation. *Knowledge Management Research & Practice*, 8(1), pp.24-32.

Körner, C., Benz, D., Hotho, A., Strohmaier, M. & Stumme, G., 2010. Stop Thinking, Start Tagging: Tag semantics emerge from collaborative verbosity. In *the 19th International Conference on World Wide Web (WWW '10)*. Raleigh, North Carolina. ACM.

Laniado, D., Eynard, D. & Colombetti, M., 2007. Using WordNet to Turn a Folksonomy into a Hierarchy of Concepts. In *th 4th italian semantic web workshop: Semantic Web Application and Perspectives*. Bari, Italy.

Lee, C.S., Goh, D.H.-L., Razikin, K. & Chua, A.Y.K., 2009. Tagging, Sharing and the Influence of Personal Experience. *Journal of Digital Information*, 10(1).

Lee–Smeltzer, K.-H.(., 2000. Finding the Needle: Controlled vocabularies, resource discovery, and Dublin Core. *Library Collections, Acquisitions, & Technical Services*, 24(2), pp.205–15.

Leimeister, J.M., 2010. Collective Intelligence. *Business & Information Systems Engineering*, 2(4), pp.245-48.

Lenzerini, M., 2002. Data Integration: A theoretical perspective. In *the 21st ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS '02)*. Madison, Wisconsin, USA. ACM.

Limpens, F., Gandon, F. & Buffa, M., 2008. Bridging Ontologies and Folksonomies to Leverage Knowledge Sharing on the Social Web: a brief survey. In *the 1st Workshop on Social Software Engineering and Applications (SoSEA) at the 23rd IEEE/ACM International Conference on Automated Software Engineering*. L'Aquila, Italy. IEEE.

Lin, H. & Davis, J., 2010. Computational and Crowdsourcing Methods for Extracting Ontological Structure from Folksonomy. In *the 7th Extended Semantic Web Conference (ESWC 2010)*. Heraklion, Crete, Greece. Springer.

Lin, H., Davis, J. & Zhou, Y., 2009. An Integrated Approach to Extracting Ontological Structures from Folksonomies. In *the 6th European Semantic Web Conference (ESWC 2009)*. Heraklion, Crete, Greece. Springer.

Liu, K., Fang, B. & Zhang, W., 2010. Ontology Emergence from Folksonomies. In *the 19th ACM International Conference on Information and Knowledge Management (CIKM'10)*. Toronto, Canada. ACM.

Lykourentzou, I., Vergados, D.J., Kapetanios, E. & Loumos, V., 2011. Collective Intelligence Systems: Classification and modeling. *Journal of Emerging Technologies in Web Intelligence*, 3(3), pp.217-26.

Macgregor, G. & McCulloch, E., 2006. Collaborative Tagging as a Knowledge Organisation and Resource Discovery Tool. *Library Review*, 55(5), pp.291–300.

Maleewong, K., Anutariya, C. & Wuwongse, V., 2008. A Collective Intelligence Approach to Collaborative Knowledge Creation. In *the 4th International Conference on Semantics, Knowledge and Grid (SKG '08)*. Beijing, China. IEEE.

Malone, T.W., 2008. What is Collective Intelligence and What will we Do About it? In M. Tovey, ed. *Collective Intelligence: Creating a Prosperous World at Peace*. Oakton, Virginia: Earth Intelligence Network.

Malone, T.W., Laubacher, R. & Dellarocas, a.C., 2009. Harnessing Crowds: Mapping the Genome of Collective Intelligence. *MIT Sloan Research Paper No. 4732-09*.

Marlow, C., Naaman, M., boyd, d. & Davis, M., 2006. HT06, tagging paper, taxonomy, Flickr, academic article, to read. In *the 17th conference on Hypertext and hypermedia*. Odense, Denmark. ACM.

Mathes, A., 2004. Folksonomies - Cooperative classification and communication through shared metadata. *Computer Mediated Communication*, 47(10), pp.1-13.

McGuinness, D.L., 2003. Ontologies Come of Age. In D. Fensel, J. Hendler, H. Lieberman & W. Wahlster, eds. *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*. Cambridge, MA: MIT Press. pp.171-94.

McLellan, S., Muddimer, A. & Peres, S.C., 2012. The Effect of Experience on System Usability Scale Ratings. *Journal of Usability Studies*, 7(2), pp.56-67.

Mika, P., 2007. Ontologies are us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(1), pp.5-15.

Millard, D.E., Gibbins, N.M., Michaelides, D.T. & Weal, M.J., 2005. Mind the Semantic Gap. In *the 16th ACM Conference on Hypertext and Hypermedia*. Salzburg, Austria. ACM.

Miller, G.A., 1995. WordNet: A lexical database for English. *Communications of the ACM*, 38(11), pp.39-41.

Morrison, P.J., 2008. Tagging and Searching: Search retrieval effectiveness of folksonomies on the World Wide Web. *Information Processing and Management*, 44(4), pp.1562–79.

Mugnier, M.C.e.M.-L., 1992. Conceptual Graphs: Fundamental notions. *Revued'Intelligence Artificielle*, 6(4), pp.365–406.

Munk, T.B. & Mork, K., 2007. Folksonomies, Tagging Communities, and Tagging Strategies: An empirical study. *Knowledge Organization*, 34(3), pp.115 - 127.

Nicola, A.D. & Missikoff, M., 2016. A Lightweight Methodology for Rapid Ontology Engineering. *Communications of the ACM*, 59(3), pp.79 - 86.

Nov, O. & Ye, C., 2010. Why Do People Tag? Motivations for Photo Tagging. *Communications of the ACM*, 53(7), pp.128 - 131.

Obrst, L., 2003. Ontologies for Semantically Interoperable Systems. In *the 12th International Conference on Information and Knowledge Management (CIKM '03)*. New Orleans, LA, USA. ACM.

Oramas, S., 2014. Harvesting and Structuring Social Data in Music Information Retrieval. In *the 11th European Semantic Web Conference (ESWC 2014)*. Crete, Greece. Springer.

Parunak, H.V.D., 1993. Hypercubes Grow on Trees (and Other Observations from the Land of Hypersets). In *the 5th ACM Conference on Hypertext*. Seattle, Washington, USA. ACM.

Pepper, S., 2000. The TAO of Topic Maps - Finding the way in the age of infoglut. In *the XML Europe 2000 Conference*. Paris, France.

Petrucci, G., 2015. Information Extraction for Learning Expressive Ontologies. In Gandon, F., Sabou, M., Sack, H., d'Amato, C., Cudré-Mauroux, P. & Zimmermann, A., eds. *the 12th European Semantic Web Conference (ESWC 2015)*. Portoroz, Slovenia. Springer.

Plangprasopchok, A., 2010. *Statistical Approaches for Inferring Category Knowledge from Social Annotation. PhD Thesis*. University of Southern California.

Plangprasopchok, A. & Lerman, K., 2009. Constructing Folksonomies from User-Specified Relations on Flickr. In *the 18th International World Wide Web conference*. Madrid, Spain. ACM.

Plangprasopchok, A., Lerman, K. & Getoor, L., 2010a. From Saplings to a Tree: Integrating structured metadata via relational affinity propagation. In *the AAAI workshop on Statistical Relational AI*. Menlo Park, CA, USA.

Plangprasopchok, A., Lerman, K. & Getoor, L., 2010b. Growing a Tree in the Forest: Constructing Folksonomies by Integrating Structured Metadata. In *the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '10)*. Washington DC, DC, USA. ACM.

Polovina, S., 2007. An Introduction to Conceptual Graphs. In *the15th International Conference on Conceptual Structures (ICCS '07)*. Sheffield, UK. Springer.

Porter, M., 1980. An Algorithm for Suffix Stripping. *Program*, 14(3), pp.130-37.

Quintarelli, E., 2005. Folksonomies: Power to the people. In *the ISKO-Italy Uni-MIB Meeting*. Milan, Italy.

Ravendran, R., MacColl, I. & Docherty, M., 2012. Usability Evaluation of a Tag-Based Interface. *Journal of Usability Studies*, 7(4), pp.143-60.

Rêgo, A.S.C., Marinho, L.B. & Pires, C.E.S., 2015. A Supervised Learning Approach to Detect Subsumption Relations between Tags in Folksonomies. In *the 30th Annual ACM Symposium on Applied Computing*. Salamanca, Spain. ACM.

Ritchie, L., Tornari, C., Patel, P. & Lakhani, R., 2016. Glue Ear: How good is the information on the World Wide Web? *The Journal of Laryngology & Otology*, 130(2), pp.157-61.

Sanderson, M. & Croft, B., 1999. Deriving Concept Hierarchies from Text. In *the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. Berkeley, CA, USA. ACM.

Schmitz, P., 2006. Inducing Ontology from Flickr Tags. In *Collaborative Web Tagging Workshop at the 15th WWW Conference (WWW2006)*. Edinburgh, Scotland.

Schmitz, C., Hotho, A., Jäschke, R. & Stumme, G., 2006. Mining Association Rules in Folksonomies. In *the 10th IFCS Conference on Studies in Classification, Data Analysis and Knowledge Organization*. Ljubljana, Slovenia. Springer.

Schwarzkopf, E., Heckmann, D., Dengler, D. & Kröner, A., 2007. Mining the Structure of Tag Spaces for User Modeling. In *the Workshop on Data Mining for User Modeling at the 11th International Conference on User Modeling*. Corfu, Greece.

Sen, S., Lam, S.K.(., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M. & Riedl, J., 2006. Tagging, Communities, Vocabulary, Evolution. In *the 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW '06)*. Banff, Alberta, Canada. ACM.

Shadbolt, N., Berners-Lee, T. & and Hall, W., 2006. The Semantic Web Revisited. *IEEE Intelligent Systems*, 21(3), pp.96-101.

Shirky, C., 2005. *Ontology is Overrated: Categories, links and tags*. [Online] Available at: http://www.shirky.com/writings/ontology-overrated.html [Accessed May 2013].

Snow, R., Jurafsky, D. & Ng., A.Y., 2004. Learning Syntactic Patterns for Automatic Hypernym Discovery. In *the 18th Annual Conference on Neural Information Processing Systems (NIPS 2004)*. Vancouver, Canada.

Solskinnsbakk, G. & Gulla, J.A., 2010. A Hybrid Approach to Constructing Tag Hierarchies. In *the Confederated International Conferences: CoopIS, IS, DOA and ODBASE*. Hersonissos, Crete, Greece. Springer.

Solskinnsbakk, G. & Gulla, J.A., 2011. Mining Tag Similarity in Folksonomies. In *the 3rd International Workshop on Search and Mining User-generated Contents (SMUC '11)*. Glasgow, Scotland, UK. ACM.

Sowa, J.F., 1976. Conceptual Graphs for a Data Base Interface. *IBM Journal of Research and Development*, 20(4), pp.336-57.

Sowa, J.F., 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Reading, MA, USA: Addison-Wesley.

Sowa, J., 2000. *Knowledge Representation: Logical, Philosophical, and Computational Foundations*. Cambridge, USA: MIT Press.

Sowa, J.F., 2008. Conceptual Graphs. In F.v. Harmelen, V. Lifschitz & B. Porter, eds. *Handbook of Knowledge Representation*. Elsevier Science. pp.213 - 237.

Specia, L. & Motta, E., 2007. Integrating Folksonomies with the Semantic Web. In *the 4th European conference on The Semantic Web: Research and Applications*. Innsbruck, Austria. Springer.

Spiteri, L.F., 2007. The Structure and Form of Folksonomy Tags: The road to the public library catalogue. *Webology*, 4(2).

Staab, S. & Studer, R., 2009. *Handbook on Ontologies*. 2nd ed. Berlin–Heidelberg, Germany: Springer.

Strohmaier, M., Helic, D., Benz, D., Körner, C. & Kern, R., 2012. Evaluation of Folksonomy Induction Algorithms. *ACM Transactions on Intelligent Systems and Technology*, 3(4).

Strohmaier, M., Körner, C. & Kern, R., 2012. Understanding Why Users Tag: A survey of tagging motivation literature and results from an empirical study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 17, pp.1 - 11.

Suchanek, F.M., Kasneci, G. & Weikum, G., 2007. YAGO: A Core of Semantic Knowledge. In *the 16th international conference on World Wide Web*. Banff, Alberta, Canada. ACM.

Tapscott, D. & Williams, A., 2008. *Wikinomics: How mass collaboration changes everything*. New York: Portfolio.

Tesconi, M., Ronzano, F., Marchetti, A. & Minutoli, S., 2008. Semantify del.icio.us: Automatically Turn your Tags into Senses. In *the Social Data on the Web Workshop at the 7th International Semantic Web Conference*. Karlsruhe, Germany.

Tommasel, A. & Godoy, D., 2015. Semantic Grounding of Social Annotations for Enhancing Resource Classification in Folksonomies. *Journal of Intelligent Information Systems*, 44(3), pp.415 – 446.

Torres, D., Diaz, A., Skaf-Molli, H. & Molli, P., 2011. Semdrops: A Social Semantic Tagging Approach for Emerging Semantic Data. In *the IEEE/WIC/ACM International Conference on Web*. Lyon, France. IEEE.

Tullis, T.S. & Stetson, J.N., 2004. A Comparison of Questionnaires for Assessing Website Usability. In *the Usability Professional Association (UPA) Conference*. Minneapolis, USA.

Uschold, M. & Gruninger, M., 2004. Ontologies and Semantics for Seamless Connectivity. *ACM SIGMod Record*, 33(4), pp.58-64.

Van Damme, C., Hepp, M. & Siorpaes, K., 2007. Folksontology: An integrated approach for turning folksonomies into ontologies. In *the Workshop on Bridging the Gap Between Semantic Web and Web 2.0 at the 4th European Semantic Web Conference*. Innsbruck, Austria.

Vander Wal, T., 2007. *Folksonomy Coinage and Definition*. [Online] Available at: http://www.vanderwal.net/folksonomy.html [Accessed March 2012].

Velardi, P., Navigli, R. & D'Amadio, P., 2008. Mining the Web to Create Specialized Glossaries. *IEEE Intelligent Systems*, 23(5), pp.18-25.

Wang, S., Wang, W., Zhuang, Y. & Fei, X., 2015. An Ontology Evolution Method based on Folksonomy. *Journal of Applied Research and Technology*, 13(2), pp.177 - 187.

Wasserman, S. & Faust, K., 1994. *Social Network Analysis: Methods and applications.* Cambridge University Press.

Wikipedia, 2016. *Wikipedia: the free encyclopedia*. [Online] Available at: http://en.wikipedia.org/wiki/Wikipedia [Accessed 10 April 2016].

Wu, W., Li, H., Wang, H. & Zhu, K.Q., 2012. Probase: A Probabilistic Taxonomy for Text Understanding. In *the 2012 ACM SIGMOD International Conference on Management of Data (SIGMOD '12)*. Scottsdale, Arizona, USA. ACM.

Wu, D., Xu, X. & Yu, W., 2016. Comparing Collaborative Annotations on Books between Libraries and Social Community Sites: A case study. *The Electronic Library*, 34(2), pp.178 - 195.

Wu, H., Zubair, M. & Maly, K., 2006. Harvesting Social Knowledge from Folksonomies. In *the 17th conference on Hypertext and hypermedia*. Odense, Denmark. ACM.

Yang, H. & Callan, J., 2009. A Metric-based Framework for Automatic Taxonomy Induction. In *the Joint conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing*. Suntec, Singapore. ACM.

Yao, J., Cui, B., Cong, G. & Huang, Y., 2012. Evolutionary Taxonomy Construction from Dynamic Tag Space. *World Wide Web*, 15(5-6), pp.581-602.

Zahia, M. & Mohamed, B.S., 2013. Fuzzy Clustering-based Approach to Derive Hierarchical Structures from Folksonomies. In *2013 ACS International Conference on Computer Systems and Applications (AICCSA)*. Ifrane, Morocco. IEEE.

Zhang, Y., Gao, K., Zhang, B. & Gao, L., 2014. A Semantic Case Based Web 2.0 Tag Hierarchy Construction Framework. In *11th International Conference on Fuzzy Systems and Knowledge Discovery*. Xiamen, China. IEEE.

Zhang, L. & Yu, Y., 2001. Learning to Generate CGs from Domain Specific Sentences. In Delugach, H.S. & Stumme, G., eds. *Conceptual Structures: Broadening the Base, LNCS (LNAI), vol. 2120*. Springer.

Zheng, H., Wu, X. & Yu, Y., 2008. Enriching WordNet with Folksonomies. In *the 12th Pacific-Asia Conference (PAKDD 2008)*. Osaka, Japan. Springer.

Zhuhadar, L., Kruk, S.R. & Daday, J., 2015. Semantically Enriched Massive Open Online Courses (MOOCs) Platform. *Computers in Human Behavior*, 51(Part B), pp.578 - 593.

Zwol, R.v., Murdock, V., Pueyo, L.G. & Ramirez, G., 2008. Diversifying Image Search with User Generated Content. In *the 1st ACM international conference on Multimedia information retrieval*. Vancouver, Canada. ACM.

Appendix

Ethics Approval Granted For Experimentation

Two of our experiments that included human participants were required to seek approval from an internal ethics committee at the University of Southampton before being carried out. This appendix shows the Participant Information Sheets that submitted to the ethics committee and for reference the code under which each was approved.

A.1 Building Tag Hierarchies from Crowdsourced Taxonomic Tag Pairs

Participant Information Sheet

Researcher: Fahad Ibrahim Bin Moqhim **Ethics Reference Number**: 4864

What is the research about?

This research is conducted by Fahad Ibrahim Bin Moqhim, a PhD student in WAIS Group, ECS, University of Southampton, and under supervision of Dr David Millard and Prof. Nigel Shadbolt.

The aim of this research is to test our new tagging approach (pair tags in the form is-a relationship) and compare it to the normal (flat) tagging approach in terms of building high level knowledge structures.

Why have I been chosen?

You are included in this study because you express your interest in participating and want to help us.

What will happen to me if I take part?

- A small website is designed to allow you to tag 10 pictures (Top 10 London Attractions by visitlondon.com) by using the two tagging approaches: tagging 5 pictures by the normal (flat) tagging and another 5 pictures by the new (is-a form) tagging (Figure A-1).

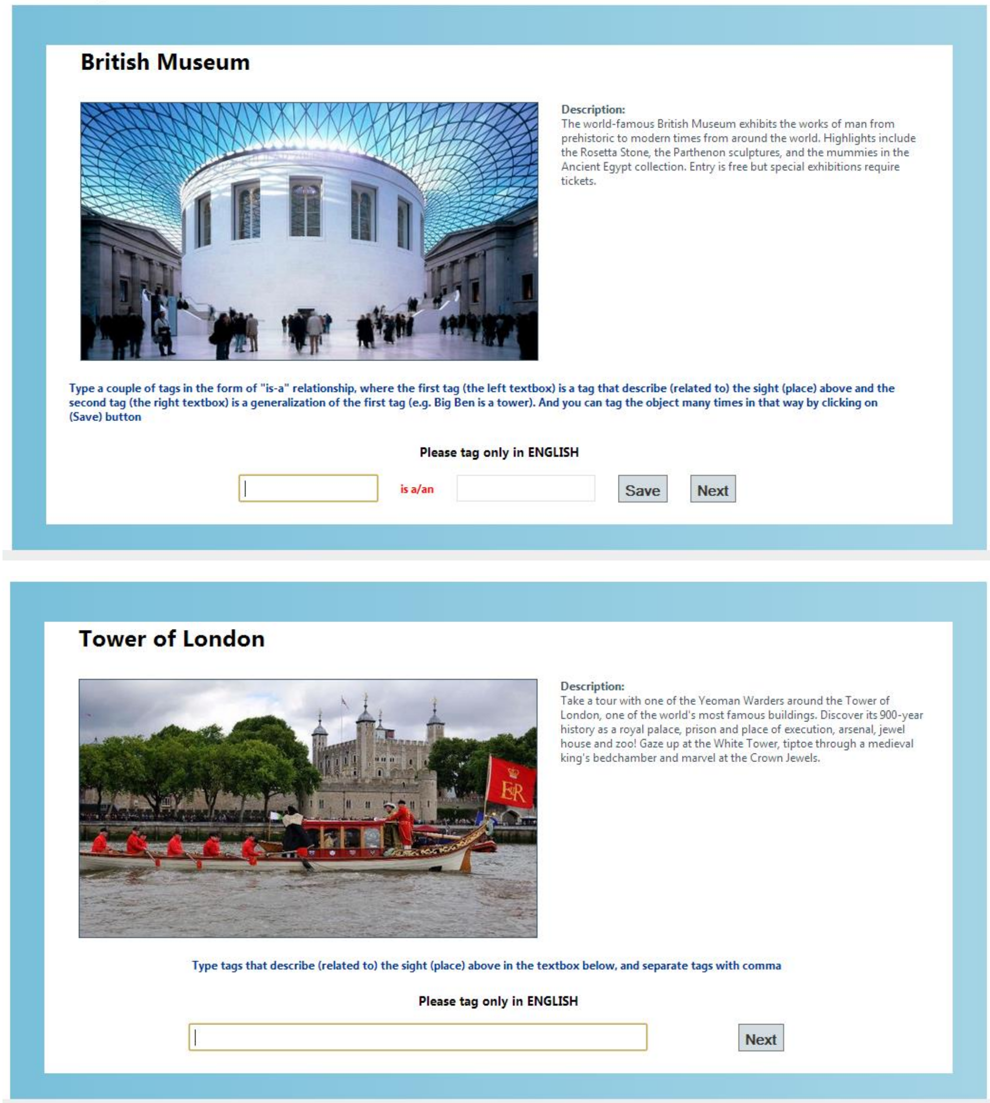


Figure A-1: The New (top) and The Normal (bottom) Tagging Approaches

- After you complete the previous step, the website asks you to complete an online questionnaires (10 scale questions and an open-ended question for genral comments) in order to evaluate each tagging approach in terms of usability (Table A-1).

Table A-1: Usability Evaluation of using the new and regular tagging approaches

Statements From SUS	▪ I think that I would like to use this approach frequently. ▪ I found this approach unnecessarily complex. ▪ I thought this approach was easy to use. ▪ I think that I would need the support of a technical person to be able to use this approach. ▪ I would imagine that most people would learn to use this approach very quickly. ▪ I found this approach very cumbersome to use. ▪ I felt very confident using this approach. ▪ I needed to learn a lot of things before I could get going with this approach.
Additional Statements	▪ I could express the ideas that I want to by using this approach. ▪ I was satisfied with the quality of what I wrote.

Each statement has to be rated on a five-point scale of “Strongly Disagree” to “Strongly Agree”.

Are there any benefits in my taking part?

Although there may be no direct benefit, your participation will contribute to knowledge and improve the use of the Web. Your participation will be highly appreciated.

Are there any risks involved?

Nothing since the study will be performed through online web pages and questionnaires

Will my participation be confidential?

The study is completely anonymous. No personal information will be collected or recorded and any information you give will be kept on a password-protected computer.

What happens if I change my mind?

Your participation in this study is voluntary. You may decide not to participate or you may leave the study at any time. And all your data will be destroyed immediately after your withdrawal.

What happens if something goes wrong?

In the unlikely case of concern or complaint, you should contact:

- Fahad Ibrahim Bin Moqhim, the researcher (fibm1e09@ecs.soton.ac.uk)
- Dr David Millard, the researcher's supervisor (02380 595567, dem@ecs.soton.ac.uk)
- Dr Martina Prude, Head of Research Governance (02380 595058, mad4@soton.ac.uk)

Where can I get more information?

If you have any questions about this study or your participation, please contact:

Fahad Ibrahim Bin Moqhim

WAIS Group,

Electronic and Computer Science,

University of Southampton, SO17 1BJ, UK

Email: fibm1e09@ecs.soton.ac.uk

A.2 Semantic Evaluation by Human Assessment

Participant Information Sheet

Ethics reference number: **ERGO/FPSE/16781**	Version: 1.0	Date: 2015-07-09
Investigator: Fahad Ibrahim Bin Moqhim		

What is the research about?
This is a student project which aims to evaluate the quality of the knowledge structures that the researcher has created based on his proposed approach for acquiring latent hierarchical structures from folksonomies.

Why have I been chosen?
You have been approached because you express your interest in participating and want to help us. You are part of a randomly selected sample.

What will happen to me if I take part?
You will first do agree to take part in this study. Then will be asked to evaluate the relation quality of term pairs through this web page (Figure A-2). One term pair (A,B) at a time will be presented, asking "What's the relation between the two terms A and B?". As an answer, the participant could choose between selecting one of the following options:

(1) A is the same as B.
(2) A is a kind of/part of B.
(3) A is somehow related to B.
(4) A is not related to B.
(5) I don't know the meaning of A or B.

It should take about 5 mins in total.

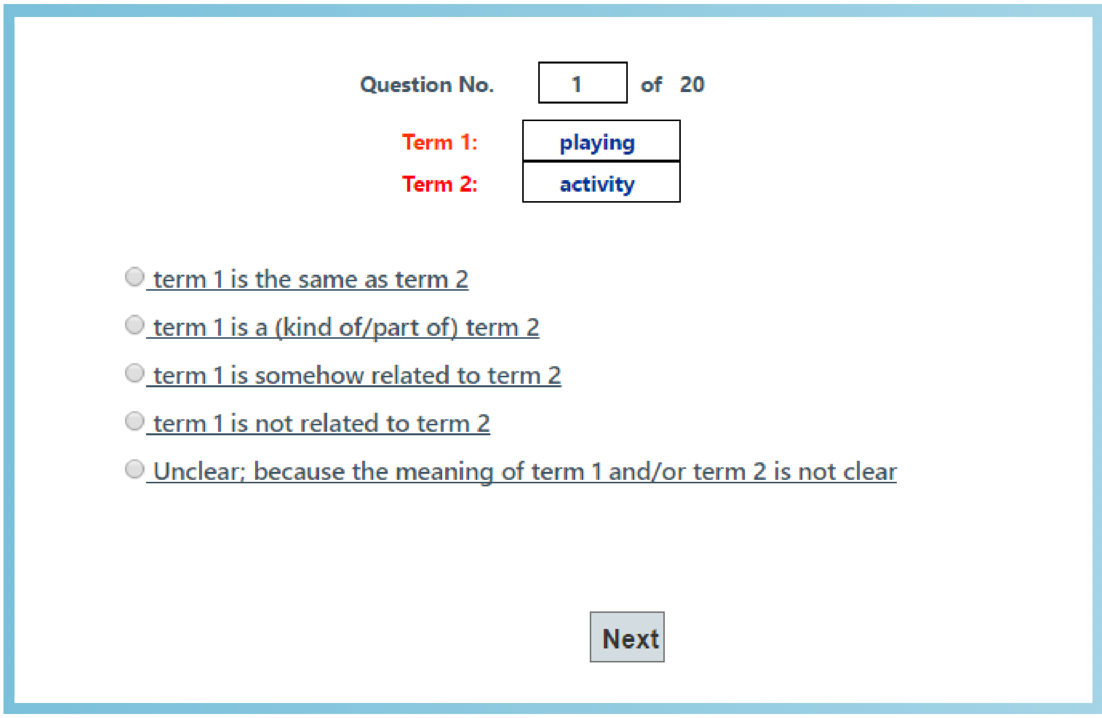


Figure A-2: Example of the tag pairs needed to be judged by the participants

Are there any benefits in my taking part?

Although there may be no direct benefit, your participation will contribute to knowledge and improve the use of the Web. Your participation will be highly appreciated.

Are there any risks involved?

There are no particular risks associated with your participation.

Will my data be confidential?

The study is completely anonymous. No personal information will be collected or recorded and any information you give will be kept on a password-protected computer.

What happens if I change my mind?

Your participation in this study is voluntary. You may decide not to participate or you may leave the study at any time. Once you finished the study, your answers cannot be withdrawn as the study is completely anonymous.

What happens if something goes wrong?

Should you have any concern or complaint, contact me if possible (fibm1e09@ecs.soton.ac.uk) or the Head of Research Governance (rginfo@soton.ac.uk).