SEMANTICALLY-ENHANCED IMAGE TAGGING SYSTEM

Ph.D. Thesis

Awatef Ruhuma

Software Technology Research Laboratory

De Montfort University

March 2013

ABSTRACT

In multimedia databases, data are images, audio, video, texts, etc. Research interests in these types of databases have increased in the last decade or so, especially with the advent of the Internet and Semantic Web. Fundamental research issues vary from unified data modelling, retrieval of data items and dynamic nature of updates.

The thesis builds on findings in Semantic Web and retrieval techniques and explores novel tagging methods for identifying data items. Tagging systems have become popular which enable the users to add tags to Internet resources such as images, video and audio to make them more manageable. Collaborative tagging is concerned with the relationship between people and resources. Most of these resources have metadata in machine processable format and enable users to use free- text keywords (so-called tags) as search techniques. This research references some tagging systems, e.g. Flicker, delicious and myweb2.0. The limitation with such techniques includes polysemy (one word

and different meaning), synonymy (different words and one meaning), different lexical forms (singular, plural, and conjugated words) and misspelling errors or alternate spellings. The work presented in this thesis introduces semantic characterization of web resources that describes the structure and organization of tagging, aiming to extend the existing Multimedia Query using similarity measures to cater for collaborative tagging. In addition, we discuss the semantic difficulties of tagging systems, suggesting improvements in their accuracies.

The scope of our work is classified as follows:

- Increase the accuracy and confidence of multimedia tagging systems.
- Increase the similarity measures of images by integrating varieties of measures.

To address the first shortcoming, we use the WordNet based on a tagging system for social sharing and retrieval of images as a semantic lingual ontology resource. For the second shortcoming we use the similarity measures in different ways to recognise the multimedia tagging system.

Fundamental to our work is the novel information model that we have constructed for our computation. This is based on the fact that an image is a rich object that can be characterised and formulated in n-dimensions, each dimension contains valuable information that will help in increasing the accuracy of the search. For example an image of a tree in a forest contains more information than an image of the same tree but in a different environment.

In this thesis we characterise a data item (an image) by a primary description, followed by n-secondary descriptions. As n increases, the accuracy of the search improves. We give various techniques to analyse data and its associated query.

To increase the accuracy of the tagging system we have performed different experiments on many images using similarity measures and various techniques from VoI (Value of Information).

The findings have shown the linkage/integration between similarity measures and that VoI improves searches and helps/guides a tagger in choosing the most adequate of tags.

Dedication

To my mother

Aisha AbdAllah

For her endless love, continual prayers, unconditional encouragement, she has done more than words can express to take care of me, may Allah reward her, and thank you for everything you have done to me in my life.

To my father

Mohamed Rahuma

To the soul of my father, who has departed from this world, was strong and faithful, but never forgotten you, wish you were here.

To my brother

Dr. Abdurahman Rahuma

For dedicating his life to be fatherly figure, without his unconditional support, loving, the dream of my Ph.D. would not be reality.

To my husband

Gamal Bayouk

For his love, support and encouragement during my study Thanks for everything.

To my sons

Abdulmaalk, Abdul-Aziz, Abdurhman, Abdulmohiemen For their endless love and when looking to them gave me support and happiness to have my Ph.D., thanks my dearest sons.

DECLARATION

I hereby declare that the work described in this thesis is my own original work undertaken for the degree of Doctor of Philosophy, at the Software Technology Research Laboratory (STRL), at De Montfort University, United Kingdom.

No part of the material described in this thesis has been submitted for any award of any other degree or qualification in this or any other university or college of advanced education.

This thesis is written by me and produced using LaTeX.

ACKNOWLEDGEMENT

First and most importantly, I truly and sincerely thank **Allah**, for answering all my prayers for giving me the strength to complete my research.

I would like to thank my team of supervisors, Dr. Martin Ward and Dr. Amelia Platt, who were more than generous with their expertise and precious time.

I want to express my deepest thanks to the one who all the good words would not be enough to describe, for her love and caring for me, whom without, I would not have my Ph.D., to my beloved mother Aisha, she was the one to push me forward, to plant every goodness in me, I owe it all to you mother.

I would also to express my thanks to my lovely, beautiful and amazing family, dearest brothers Dr.Abdurhman, Ibrahim and my dearest sisters Salma, Wedad, Najmia, for their love, care, advice, support, prayer and encouragement have always enhanced my heart, thank you for everything.

Moreover, my innermost thanks and my appreciation goes to my husband Gamal, and my sons Abdulmaalk, Abdulaziz, Abdurhman , Abdulmohiemen , than you for sharing the happy and sad times, good and bad times with me, thank you.

Thank you also to Najmias husband, Electrical Engineer Rashed Marwan, for helping to expedite and sort out my documents from Libya during my study.

I would like to thank to my dearest friend Raja Elmahmmody for her tender loving care both to me and my children, and to also thank Dr. Waheeda Elmayyan, Dr. Somiah Abdulaziz for supporting me during my study, so our friendship started here in Leicester and will continue forever.

A special thanks goes to my friend, colleague and the team of staff at STRL for being accommodating to my queries and for their assistance in all my research needs. A special mention goes to Mrs Lynn Ryan and Mrs Lindsey Trent for their care and collaboration.

Thank you to all of you.

Contents

List of Figures

List of Tables

Chapter 1

INTRODUCTION

Objectives

- Motivating the research work presented in the thesis.
- Articulating the research question and contribution to knowledge.
- Outlining the research methodology.
- Giving a detail organisation of thesis

1.1 Introduction

The ubiquity of the internet and the overwhelming success of mobile communications has affected the environment of individuals as well as organisations [18]. The abundant availability of internet enabled mobile phones, desktops, laptops, tablets, mini tablets, PDAs, set-top boxes, game consoles and smart TVs has facilitated the ease of access for any user to upload and modify content, images, video and audio media. For example, 7 Petabytes (one Petabyte is equal to 1 million gigabytes) - is the amount of photo content added to Facebook every day; 100 billion is the estimated number of photos on Facebook by the middle of 2011 ([104]); 300 million is the number of new photos added every day to Facebook during 2012; 4.5 million is the number of photos uploaded to Flickr each day (2012); 6 billion is the number of photos hosted on Flickr (as of August 2011); 5 billion is the total number of photos uploaded to Instagram since its start, reached in September 2012 and 58 is the number of photos uploaded every second to Instagram. These numbers illustrate the importance of accurate tagging and hence searching.

The user's interaction with the internet has also been augmented by the high availability and low cost of wireless communications (Wi-Fi) everywhere, from Cafetarias and restaurants to universities, hotels, trains and airplanes.

However, such an advancement in user generated content has caused performance and accuracy issues for search engines due to the surge in volume of the content and the poor annotation of text and media. To solve this issue, Social (also known as user generated) tagging was introduced and gained popularity with the launch of social sites such as Delicious and Flickr.

User generated tags were easy to generate and required to special skill or training. Users could add one or more words to a content or to an image, which would be grouped under a specific category. However, with the absence of standards and guidelines, the user generated tags rapidly became inconsistent and ambiguous.

The inconsistency of the tags was also compounded by the ambiguity of uploaded images. Social photo sharing web sites such as Flickr and Delicious contained a large number of inaccurate, ambiguous and poorly tagged images which warrants a solution that would improve and increase the accuracy of images' Information Retrieval (IR) [42].

1.2 Motivation

The ambiguity of images may be attributed, in part, to polysemy, which can be defined as "the capacity for a word or phrase to have multiple related meanings". As an example, consider a user searching for images that are tagged as 'scales' in Flickr. The search results returned contain images for bathroom scales, reptile scales and fish scales. The search results in this case may be deemed un-necessary and wasteful by the user, and thus, a need exists that requires images to be tagged accurately, efficiently and relevantly.

One reason why such inaccurate and ambiguous tags were introduced into Flickr, and other image sharing sites, can be gleaned from the users' motivation to tag. These numerous motivations can be summarised as follows [114]:

- Contribution to the community: Social tags can be used to share and promote common aspects that would be of use to an on-line virtual community.

- Self attention: Some users may attempt to exploit tags to promote their own products and services, and hence, may add irrelevant tags, which can be deemed as a form of 'spamming'.

- Sharing with other users: Tags can be used to share resources with other users or groups of the same interests. Such tags could have a higher level of accuracy than tags for personal reasons.

- Future Retrieval: Images can be tagged to act as a reminder or to ease future retrieval by oneself or by others. Such tags would have a high level of descriptive text that act as metadata about the images that have no other associated tags.

- Expressing one's opinion: Tags can be used to reflect the user's own opinions and beliefs, which in turn will promote the user's own standing and reputation in the community. This can be considered as a form of expert tagging which would be of high value to the images (or any other objects) being uploaded.

- Organising one's tasks: Tags can be used to organise a user's daily tasks. E.g. "to-do", "to-reply", "to-schedule" and "to-use".

It is obvious from the above reasons that the motivation behind this research, is not only to solve the issues arising from ambiguity, but also to solve similar issues caused by misspelling, shorthand writing and slang or abbreviated words, to name but a few. These issues are further complicated by the users ability to annotate images freely with any chosen words (tags), thus implicitly causing tags to appear random and unstructured [52].

1.3 Research questions

As stated in the previous section, the main question concerning this research is how to improve the accuracy of Information Retrieval for ambiguous images that are uploaded to photo sharing sites. One possible solution is through the use of similarity measures[1], such that the tags from similar images will be "recommended" to users who intend to tag similar images to the ones obeying the used similarity measures.

Tags that show no improved accuracy as a result of using one or more similarity measures, will have a measured weight applied to them followed by the application of the semantic ontology from WordNet.

The purpose of tag recommendation is to solve the problems arising from the practice of tagging that were highlighted in the previous section.

The main questions and sub-questions that will be answered by this research are discussed below:

- **Question 1:**

  Why are the current images on photo sharing web sites inaccurate?

- **Question 2:**

[1] It is important to note the distinction between "similarity" between objects and their "semantic equivalence". The later requires a complete characterisation of contexts and intention of objects.

How to improve the accuracy of Information Retrieval from existing images?

- Can colours be extracted from existing images to help improve tag-based search accuracy?
- Can the action depicted in the image be used to improve tag-based search accuracy?
- Does a single similarity measure work?
- Is it possible to integrate two or more similarity measures in the form of "weighting" to improve the accuracy of tagging-based search?

The above research questions will help us pinpoint the type and volume of data required to complete the investigation.

1.4 Contribution to knowledge

The ultimate aim of this research is to contribute a deeper understanding of the methods and approaches of image tagging and image information retrieval to the community at large.

The current state of content related to image storage and retrieval is random by nature, irrelevant in many cases, unstructured and haphazard.

There have been many attempts by other researchers to improve the existing knowledge base of multimedia tagging, however the research community is still a long way away from achieving a thorough understanding of the underlying knowledge infrastructure required to solve these issues, by offering a multi-dimensional approach.

To achieve these stated objectives, this research will examine many existing approaches with the ultimate aim being the improvement of image search accuracy within the existing Internet search engines such as Google Images, Yahoo and Bing, as well as social image and content sharing sites such as deli.cio.us, Flickr and Facebook.

1.5 Measures of success

To gauge the success of this research, the questions raised in section 1.3 must be answered in an acceptable manner. The proposed approach of returning search results based on recommended tags must yield images that have a high degree of similarity to the images being searched for and compared against.

The proposed multi-dimensional approach must also return a much improved search results when compared to existing image tagging and search approaches.

1.6 Research methodology

To ensure the success of this research, the following overview summarises the methodology followed in this thesis:

- **Research literature background:**

  The research starting point was the review of the existing literature in the area of the early web, the current Social Web and upcoming Semantic Web. Once the required background was gathered and analysed, a new and original approach was planned and designed for applying several similarity measures to existing image tags such that the existing inaccuracies and ambiguities in social sites tagging systems can be addressed. Following that, the research identified the set of criteria for an efficient approach in developing a tagging system.

- **Classification of Research Methodology:**

  To bolster the knowledge acquired from examining the existing literature pertaining to this research, a detailed analysis of the current methods that are utilised to improve the accuracy of image tagging and information retrieval was carried out.

- **Statistical and Experimental Analysis:**

  Having completed the classification of the research methodology and the review of the literature in the scope of this research, a detailed analysis of the current statistics of uploaded images were carried out. The role of aggregated usage statistics, the aggregation of usage voting (i.e. selecting the same set of tags for tagging similar images) and the influence of the crowd on tagging statistics were also examined.

A generic architecture of the experimental work flow was also carried out to design the experiment by which tags will be selected and recommended to the users. An outline of the actual experiments to be carried out during the research is also proposed.

The methodology and research architectures affords guidance and directions for addressing the outlined image tagging issues and outlines the challenges caused by the motivation and the diversity of cultures of the interest group.

- **Architecture:**

  The main methodology of the architecture was designed next, where a model consisting of the main components was outlined. The main components used in this model were a tag and content component, a visual correlation component, a tag co-occurrence component, a visual language component and a tag selection component.

To cement the choice of this architecture, an example was also identified, where a defined list of tags were classified by the types intended to be used in this research, namely Primary object, Secondary object, Action and Colour. User voting strategy acceptance ratios based on WordNet were then applied to arrive at the final set of recommended tags.

- **Gather Corpuses of Images:**

  As this research is experimental in nature, image corpuses were gathered from the Internet and used in the main body of the experiments. Before starting the experiments, a benchmark for the proficient development of an image tagging system was established.

- **Experimentation:**

  In the first experiment, five similarity measures were combined to arrive at a single value that may give an indication of how similar one image is to another. Additionally, the method of using Domains, Values and Thresholds were also examined in which sets were used to compare the proximity (or distance) of the returned images.

  In the second experiment, a weight measure was applied to only the inaccurate results from the first experiment to further improve the accuracy of the tags.

Finally, in the third experiment, WordNet cognitive synonym sets (synsets) were applied in addition to the weight measure that was applied in the second experiment.

- **Evaluation of Results:**

  The large sample of images collected from the Internet were all subjected to the three aforementioned experiments to evaluate whether there is a significant improvement in the accuracy of tags between the results of the experiments and whether there is an overall improvement over existing, previously published methods.

1.7 Thesis structure

The total number of chapters in this thesis, including this one, is seven. Below is a summarised brief description of each chapter, starting from chapter 2:

- **Chapter 2:**

  This chapter discusses and presents a time-line of the evolution of the World Wide Web, starting with Web 1.0 to Web 3.0. There is also a presentation of the methods of image metadata generation approaches, an overview of the various semantic similarity measures and the measures of relatedness. Finally, there will be an overview of the limitation of the various versions of the web.

- **Chapter 3:**

  This chapter provides a presentation of the equations that are used in research methodology, the methods used to generate tag clouds, Clustering, classification and n-dimension similarity measure. The chapter also presents an overview of the main challenges that need to be addressed for social tagging and the extraction of relevance information by the aggregation of user defined tags.

- **Chapter 4:**

  This chapter presents an overview of the differences between image and text tagging, the added value of information to tag values, symmetric and asymmetric measures, tag ambiguity, the use of tag recommendation based on tag co-occurrence and the concept of tag clouds and their use by photo sharing websites.

- **Chapter 5:**

  This chapter presents a discussion of the issues associated with user generated image tags, the benchmarks used for the development of image tagging system, the richness inherent of the image tag values, the problems with the current image tags, an initial experiment combining several similarity measures, the usability of values within image tags and the domains and threshold associated with image tags.

- **Chapter 6:**

  This chapter provides an overview of an integrated approach to accuracy improvement based on WordNet, comparative analyses of search results and a follow up second experiment with improved results. There is also a discussion on whether the visual objects and features that constitutes the richness of images can be categorised as Bag of Words representations for the purpose of image object recognition.

  A third experiment was also carried out in a final attempt to meet the research criteria and answer the research questions by further improving the accuracy of the tags.

- **Chapter 7:**

  This chapter provides a conclusion to this research, a summary of this research's achievement, the areas of limitation and the possible future research areas and direction resulting from this research.

Chapter 2

TAGGING SYSTEMS: STATE-OF-THE ART

Objectives

- Discussing the Evolution of the World Wide Web.
- Discussing the methods of Metadata generation.
- Discussing the various similarity measures.
- Discussing the Measures of Relatedness.
- Discussing the limitation of the various versions of the web.

2.1 Introduction

The work in this chapter will entail a detailed presentation of how the web has evolved since its inception more than two decades ago. We will also be looking at Social Bookmarking, the methods used to generate image metadata, similarity measures and the use of WordNet for tagging and Semantic Similarity Measures. The limitations of the different versions of the web will also be detailed.

With reference to the measures for the relatedness of tags, we will present three methods, namely the co-occurrence count, the cosine similarity of co-occurrence distributions, and FolkRank. The benefits resulting from these discussions will also be presented.

2.2 The Evolution of the Web

The ancient Greeks thought that all the knowledge were held by their Gods and their only method of grasping this information was through the Oracles who were thought to be portals through which the gods spoke directly to the people. However, in today's society, a lot of the knowledge is held by computers, such as Web servers, Relational Databases and file systems. The interaction of computers with this information has evolved since the inception of the Internet in the early nineties.

The bulk of this information is held on these computers in the form of Natural Language, of which computers are neither able to understand, nor able to access in a useful way to the users. The early form of the web (also known as web 1.0) allowed users to search this information, but without the ability to influence the results of the search. Hence, when users used the internet to search for their required resources, most of the retrieved results were irrelevant. This issue was further compounded by the fact that not all the relevant results were retrieved.

Additionally, the Internet is now widely used as a social medium, where users with similar interest are encouraged to participate in many of the wide plethora of social networks that enables them to socialise and to exchange and share ideas via multimedia, such as video and photos. This concept is referred to as the Social Web or web 2.0.

The Semantic Web, on the other hand, attempts to address the problem of accessing the considerable amount of un-structured data stored on the Internet by expressing Web content in machine processable forms which software applications can maintain more efficiently. This would enable search engines to enhance search precision and enable logical reasoning.

2.3 Web 1.0

The early version of the web was characterised as a static read-only technology that allowed users to mainly search for textual information with very little control over what other type of data, such as images, were returned.

This version of the web did not facilitate user or site interaction and had little or no linking structure. It did not offer bookmarks or tagging solutions and the returned results were in the main impersonal, descriptive and statements of fact. Additionally, the search engine technologies were characterised by large indexes but crude retrieval techniques and the search results were focused purely on size of index with the relevance mostly ignored.

The web 1.0 issues highlighted the need for proactive research to further evolve the web into an interactive and dynamic tool, thus web 2.0 was born.

2.4 Web 2.0 - The Social Web

The term, Web 2.0, began to rise in popularity in 2004 as a successor to the early Web 1.0 version. This new version of the web, initially, enabled users to interact with the visited sites, then evolved into a network of social sites. Such sites offered users the opportunity to blog, chat, share photos, make new friends, buy and sell goods and services, contribute to wikis and even plan and organise the recent ”Arab spring” in the middle east.

2.4.1 Social Bookmarking

Social bookmarking sites such as deli.cio.us allow their users to submit, share and tags of web pages and images. The combination of links and tags become part of the community pool and are made available for other users to browse.

The use of tag metadata can leverage the identification of keywords which in turn will improve search engine rankings and web site navigation. Additionally, a significant fraction of the users provide tag metadata for their content such as images and photographs. Semantic similarity measures of tags and metadata among users based solely on their annotation patterns can be employed to improve the accuracy of image searches.

Tag metadata annotations are provided mostly by the content creator, i.e., the tags associated with an image are typically provided by the user who posted that image. Alternatively, image metadata can be automatically generated, as will be seen in the next sections.

2.4.2 Image metadata generation approaches

Automatic metadata generation approach:

With this approach, the metadata is automatically generated by analysing the content and text on the web pages [103]. This approach is efficient in that it costs very little in terms of human effort. However, this statistical model based approach is generally unsatisfying, as the metadata generated may be of poorer quality than professionally generated metadata [46], due to inaccuracy and general noise. This method is also are dependent on having a large enough tagged corpus for training and, in some instances, does not fare well with tagging at the sentence level.

Additionally, Esner [89] cites two examples that explain the inferiority of automated annotation (and more so for automated image annotation). Firstly, the so-called visibility limitation, attempts to describe how automated image tagging algorithms typically depend on successfully linking visible image features to words. It is very difficult for automated algorithms to capture content and contextual information from images that do not have any associated image features. Enser [89] provides the CBIR query, "find a picture of the first public engagement of Prince Charles" as a prime example of content that would be hard to automatically extract from images. In addition, the author goes on to mention another significant limitation in the form of generic object limitation, which questions the use of very generic tags for the images such as "sun", "grass" and "tiger" as "they have the common property of visual stimuli which require a minimally-interpretive response from the viewer." [89].

Manual metadata generation approach :

The manual metadata generation approach is, potentially, more accurate and practical than automatic annotation. A collaborative tagging systems was described by Golder and Huberman (2006) [28]. Users tag primarily for their own benefit, but the software makes it possible to see all the tags used for a resource so that all users can utilise tags from each other.

With this approach, the folksonomy becomes a common vocabulary grown from the ground up. As the number of uses increases, each resource develops a "tag cloud" or a cluster of tags denoting popularity. Furthermore, the most popular resources are tagged the most frequently which, in turn, influences other users in their choice of tags.

Users, tagging for themselves, collectively create useful sets of subject descriptors in the form of tags for the resources they are tagging and this user-added metadata can then be leveraged for information retrieval on a general as well as a personal level.

2.4.3 Similarity measures and the Semantics of Social Tagging

The increased popularity of social bookmarking systems such as GiveALink.org, BibSonomy.org, CiteULike.org and Delicious.com can be attributed to the way that users share resources by adding keywords in the form of tags, thus leading to the creation of an aggregated tag-index (folksonomy).

This system of social bookmarking has been built on 3 dimensions: R(Resource), U(User) and T(Tags). Hotho et.al. [91] formally defines folksonomy as a Tuple F := (U, T, R, Y) where U, T, and R are finite sets, whose elements are users, tags, resources and Y is a ternary relation between them.

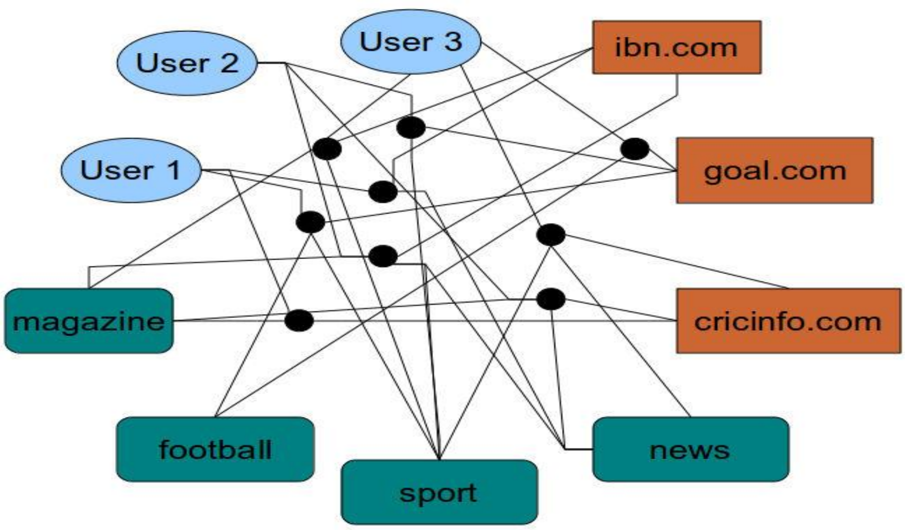


Figure 2.1: Example of Folksonomy represented as a Network

A folksonomy can be represented as a network structure, as shown in Figure 2.1. In this Figure there are 3 users, 3 resources and 4 tags. Each dot in the Figure represents an annotation (tag posting). The similarity (or relatedness) of the tags can be measured by Context (Distribution) which is based on 3 different vector space representation for the tag: Tag Context, Resource Context and User Context.

2.4.4 Measures of Relatedness

As there are multiple notions of explicitly representing folksonomies, they can all be thought of as special cases of three-mode data. Since measures for similarity and relatedness are not well developed for three-mode data yet, only two and one-mode views on the data will be considered. These two views will be complemented by a graph-based approach for discovering related tags (FolkRank) which makes direct use of the three-mode structure.

Co-Occurrence

Given a folksonomy (U,T,R,Y), we define the tag-tag co-occurrence graph as a weighted, undirected graph, whose set of vertices is the set T of tags, and where two tags t1 and t2 are connected by an edge, if there is at least one post (u,T_{ur} , r). The weight of this edge is given by the number of posts that contain both t1 and t2. Co-occurrence relatedness between tags is given directly by the edge weights.

Cosine Similarity

The method is a distributional measure of tag relatedness by computing the cosine similarity of tag-tag co-occurrence distributions. Specifically, the cosine similarity is computed in a vector space where each tag is represented by a vector. The weight between a node and itself is given as zero as any two tags are considered as related when they occur in a similar context, and not when they occur together.

FolkRank

The PageRank algorithm reflects the idea that a web page is only important if there are many pages linking to it, and if those pages are important themselves. The same principle was employed for folksonomies i.e. a resource which is tagged with important tags by important users becomes important itself. The same holds, symmetrically, for tags and users. By modifying the weights for a given tag in the random surfer vector, FolkRank can compute a ranked list of relevant tags.

One other similarity measure is the Jaccard coefficient which employs the use of WordNet to unify the tags described by the users. The Jaccard measure can be used to work out the similarity between images based on a set of words that represent tags containing certain criteria that would have been specified by the search.

The Jaccard coefficient is a measure of the similarity between sample sets. WordNet is a semantic network and will be described in more details in the next section.

To illustrate the use of the Jaccard measure, we use an example of an image search from Flickr, combined with add-on or core tagging/bookmarking features that form an integral part of most of the currently popular web browsers, such as Firefox, Internet Explorer or Chrome. If we search for a photo of a Barking Dog using Flickr, then the search results will return images that may or may not be relevant to the search. Examples include, a dog barking up a tree, barking Sea Lion and a dog barking at a postman (see examples 2.2).

The bookmarking feature would allow users to tag the images of a barking dog in a more precise manner. This is achieved by displaying a form to the user that exposes three fields, namely 'object, 'action' and 'background'. The object represents the name of the object within the image, i.e. dog in this case. Similarly, the action would be barking and the background would be

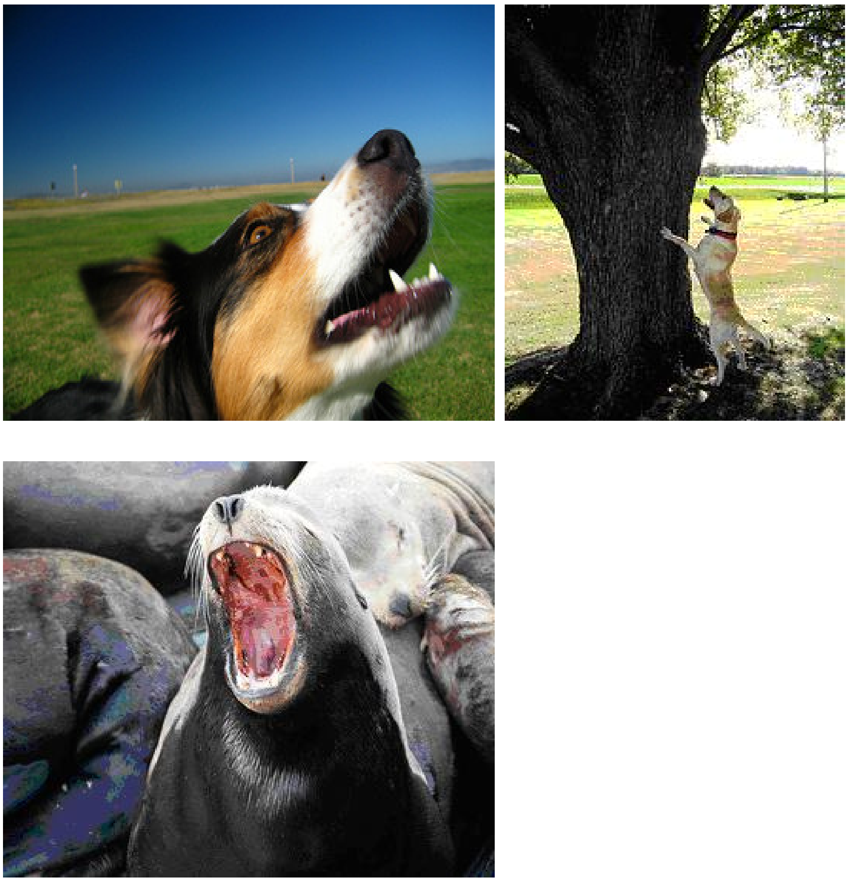


Figure 2.2: Examples of Flickr search for a barking dog

tree. The accuracy of the tagging would then be improved by developing an application that compares the saved tags that are used to identify the content of the image with the matching words that are stored on the WordNet database.

When a user submits a search for an image, the tag information for the potential returned images, which are normally stored in the *alt* property and the surrounding innermost element content, would be compared with similar words stored in the WordNet database.

The contents of WordNet are stored on a relational database. One useful function of the WordNet database is the grouping of stored words into categories, which are defined as noun (object), adjective, adverb and verb (action). Words are stored in a table with their own category identifiers that references a synset table. When a word that identifies the tag of an image is submitted, the WordNet database is queried to check if this specified word exists. If it does exists, its category type is obtained from the database and all similar words of the same category type are also returned. This process is repeated for the all three tag fields, i.e. object, action and background. If the majority of the words that have been returned are of a certain type (more nouns than verbs, for example), then the tag word is accepted as being of that majority type.

The returned list of words are then used to employ the Jaccard similarity measure, which are applied two images at a time. Given a set of words as tag information, we find the Jaccard similarity between two images in terms of set of words as,

Jaccard similarity will give back a similarity score between 0 and 1. For our example, the Jaccard similarity is modified. It uses WordNet to find the

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Where $A=\{w_{A1},\ w_{A2},\ w_{A3},\ \ldots w_{An}\ \}$ and $B=\{w_{B1},\ w_{B2},\ w_{B3},\ \ldots w_{Bm}\ \}$

number of synonym words. As the number of synonym words increase, the contribution of that word to similarity result decreases. This method is the combination of WordNet and Jaccard similarity. In our example, given the tags of image A and image B,

A = (dog, barking)

B = (dog, barking, tree)

The Jaccard similarity scored returned will be a number fraction between 0 and 1. In extreme cases, the score may be returned as 0, which indicates no similarity is present between the images. However, the closer the score is to 1, the more likely that the compared images are similar.

This process outlined in the above example illustrates how it would significantly enhance the search results of images. Figure (2.3) displays the search results that would be returned if such a method was applied.

Figure 2.3: Examples of Accurate search for a barking dog

2.4.5 Limitations of Web 2.0 tagging

Ambiguity:

As an uncontrolled vocabulary that is shared across an entire system, the terms in a folksonomy have inherent ambiguity as different users apply terms to documents in different ways. There are no explicit systematic guidelines and no scope notes. Additionally, when users come together to collaborate on the same interest, biases might emerge. This is because people with the same tendencies and preferences when using classification methods might encourage one another to propagate these biases. As a result, the objective view of content might suffer. Therefore, ambiguity might occur and become prevalent.

Synonyms :

There is no synonym control in the tagging system. There is a possibility that allowing freely annotated and distributed content can produce inconsistency and unreliability. Over a period of time, tags for a single concept (synonyms), a single tag with multiple meanings (homonymy) and a single tag for several different but related meanings (polysemy) might emerge. This might lead to the problem of inefficiency in terms of content search and indexing. Meta noise, which refers to tags that are irrelevant and imprecise, might also increase.

These sorts of problems are the reasons why controlled vocabularies are used in many settings. Generally, any of the classic problems that controlled vocabularies help deal with will be present in these systems to varying degrees. However, it is likely that a controlled vocabulary would be impossible in the context of systems like Delicious and Flickr.

2.5 Web 3.0 - The Semantic Web

Tim Berners-Lee further defined the Semantic Web as "the development of machines to become much better able to process and understand the data that they merely display at present" [5].

The architecture of the semantic web is based on ontologies and machine-processable metadata. Additionally, it contains layers that refers to logical reasoning, proof, and trust. Such layers are crucial to enable the exploitation of information offered by ontologies and metadata for the delivery of knowledge and to enable automated or semi-automated decision making.

The semantic web (also referred to as Web 3.0) can be thought of as the existing social web (also referred to as web 2.0) with an offering to define extensible and flexible standards for information exchange and interoperability.

2.5.1 From Tags to Folksonomy

The eruption of the tagging phenomena over the last few years was caused by many thousands of users adding tags to organise web resources. The tags (natural language terms) can range from tagging bookmarks at deli.cio.us to tagging photographs and images on Flickr. This has lead to the availability of an enormous amount of tagged data on the web. Tagged data is usually for that data to be retrieved later and found by others, and so the scheme used to classify the data is essentially a convention that is given a social meaning, as "People will in general use the minimum amount of convention to solve their co-ordination problem. This rule-of-thumb might explain the slowness of the Web community to embrace model-theoretic semantics." [34]. Tagging is popular precisely because it uses a minimal amount of convention: "Groups of users do not have to agree on a hierarchy of tags or detailed taxonomy, they only need to agree, in a general sense, on the "meaning" of a tag enough to label similar material with terms for there to be cooperation and shared value." [69]. The low cognitive load of using tagging in comparison with ontologies is one reason for its success, since "picking topics from a pull-down menu is arduous, the topics we currently employ are not sufficient, and updating the tool with new topics is too time consuming" [71]. This lead tagging to have a high cost-benefit analysis in terms of being able to retrieve the data and share it in comparison with the time consumed classifying it: "Free typing loose associations is just a lot easier than making a decision about the degree of match to a pre-defined category (especially hierarchical ones)[11].

One issue with tagging is due to the natural language nature of tags themselves. Tags are not normalized for synonymity, morphology, or even just different manners of specifying the exact same meaning, such that "if you want to find all references to New York City on Del.icio.us, you'll have to look through nyc, newyork, and newyorkcity [71]. Second, heteronymity runs rampant on tagging systems, with users employing "the same term for disparate concepts" such that words like "flow" can mean either "optimal experience" or the movement of liquids like rivers [71]. The lack of an explicit hierarchy makes many of the tags redundant, such that a web-page about pianos must be labelled both as about pianos and about music, despite the fact that every piano is about music. Any sort of structured data becomes impossible, yet certain types of data lead inevitably towards structure. For example, the concept of a "date" or "time-stamp" of a URI makes no sense without the actual date, such as "February 18th 2006." The fact that often data comes with a natural structure, as given by frame or facet-based systems, is handled easily by the Semantic Web. It is also impossible to express complex relationships using only tags. For example, a web-page may inform the results of an election, but it can not distinguish by its tags alone who won the election and by what margin. Unlike Semantic Web ontologies, collaborative folksonomies that use tags alone cannot in general be shared across collaborative tagging systems to another without the use of at least an ontological layer to resolve the "tags" to URIs and even then the problems cited above still make it impossible [32]. In that regard, each tagging system is stranded from interaction with the greater Web, and the data itself is usually held hostage behind firewalls.

The term "folksonomy" was coined by Thomas Vander Wal and is a combination of "folk" and "taxonomy." hello world Collaborative folksonomies utilises large scale human annotations of human of web resources and thus plays a crucial role in the notion of similarity. The definition and analysis of semantic similarity relationships and measures form the bulk of our work and will be discussed in a later section.

An important aspect of a folksonomy is that is comprised of terms in a flat namespace: that is, there is no hierarchy, and no directly specified parent-child or sibling relationships between these terms. There are, however, automatically generated "related" tags, which cluster tags based on common URLs. This is unlike formal taxonomies and classification schemes where there are multiple kind of explicit relationships between terms. These relationships include things like broader, narrower, as well as related terms. These folksonomies are simply the set of terms that a group of users tagged content with, they are not a predetermined set of classification terms or labels.

2.5.2 Wordnet and Wordnet-Based Semantic Similarity Measures

WordNet is a semantic network, which is organised in such a way that synsets and wordsenses are the nodes of the network, and relations among the synsets and wordsenses are the edges of the network. In WordNet, each meaning of a word is represented by a unique wordsense of the word, and a synset (stands for "synonym set") consisting of a group of wordsenses sharing the same meaning. More than two thirds of the nodes in WordNet are synsets. Hyponym is the key relationship for noun synsets in WordNet, which has been widely used to estimate the semantic relatedness among nouns.

WordNet has been commonly used to measure semantic similarity among words since it has the inherent advantages of being structured in the way of simulating human recognition behaviours. There are currently three categories of WordNet-based semantic similarity measures.

A. Node-based methods

Node-based methods use the amount of information contained by related nodes (i.e., related concepts) in WordNet to estimate semantic similarity between the concepts of interest, i.e., c1 and c2. These kinds of methods are also called as information-based methods.

Most of node-based methods employ the information content to quantify the amount of information that a concept contained. According to the definition in the information theory [95], the Information Content (IC) of a concept c can be quantified by $\mathrm{IC(c)} = \ \log(\mathrm{P(c)})$, where P(c) is the probability of c appearing in a corpus.

Resnik [86] believed that the similarity of c1 and c2 is determined by the closest common superordinate concepts (i.e., hypernyms) of c1 and c2 in WordNet. Thus, Resnik proposed to use the IC of the lowest hypernyms of c1 and c2 to calculate the semantic relatedness between c1 and c2.

The drawbacks of node-based methods include: (i) it is a time-consuming work to analysis the corpora for estimating the IC values; (ii) unbalanced contents of the employed corpora may significantly decrease the accuracy of the IC values.

B. Edge-based methods

Edge-based methods utilise the shortest path between concepts (i.e., c1 and c2) in WordNet to estimate the semantic relatedness between c1 and c2. Lengths of all edges on the shortest path are accumulated to quantify the semantic similarity. It is the way of calculating the length of edges that differentiates methods in this category.

C. Hybrid methods

Hybird methods combine the information from different resources to estimate the semantic similarity between concepts, e.g., combining the IC of concepts with the structure information retrieved from WordNet to conduct the estimation.

2.5.3 WordNet Based Tagging

The words and synsets of WordNet have been used for tagging. Tag content is provided for end users and tag groups through WordNet's definition in it's relational database. The synonym set of a word can be used by doing inquiries on the database.

In the case of multimedia, such as images stored on a website, the contents of the image can be identified by using the tag information of the said image. This information is contained in the 'alt' property and the surrounding innermost content of the image.

Whenever potential tag information has been received, each word that has been obtained will be matched with similar words existing in the wordNet database and potential tagging content is intended to be increased. Here, type info keeps the type of the word(noun, verb, adverb, etc.) and superid field keeps the broader meaning of word(apple is a fruit).

2.6 Summary

With the deep academic challenges associated with recognising real world objects within images, it is not surprising to find that there has been great interest amongst the computer vision and information retrieval communities in the development of robust, accurate and efficient image tagging systems. The main purpose of tagging images is to allow for the retrieval of images based on the similarity measures of similarity among words. In this chapter, we detailed the plethora of current literature about the methods used to enhance image tagging systems along with the advantages and limitations of employing such methods.

A background of the evolution of the Web and a comparison of its three main versions, along with their limitations was also presented. The adoption of similarity measures to enhance the accuracy of tagging images in current and future versions of the web will, significantly, enhance user experiences.

The current trend for instant image and video creation by means of mobile devices, such as smart phones and tablets, and the ability to store such media, instantly, on image sharing sites such as Facebook and MySpace is set to increase, disproportionately, over the next few years due to the ease of use facilitated by the advancement of technology.

The social study conducted by Ames et al. [2] provided some insights into the motivations that drive private individuals to annotate their images. This study revealed a changing opinion of the usefulness of tagging, from it being nearly completely avoided for personal off-line collections through to it being heartily embraced for on-line collections such as those on Flickr.com.

Additionally, for commercial organizations, the correct tagging of images has a direct effect on their revenues and efficiency in satisfying the needs of their consumers, as an incorrectly or insufficiently labelled or tagged image is unlikely to be found, particularly within the stringent deadlines commonly experienced within the commercial world, thereby leading to a loss in operational efficiency.

The above reasons highlight the need for further research into the improvement of the accuracy of image tagging, which will go some way towards answering the criticism directed at the visual image retrieval research community by many other researchers, such as Jrgensen, who has expressed concern that "the emphasis in the computer science literature has been largely on what is computationally possible, and not on discovering whether essential generic visual primitives can in fact facilitate image retrieval in 'real-world' applications." [49].

Chapter 3

THE FRAMEWORK

Objectives

- Discussing the equations used in our work.
- Discussing the methods of Tag Clouds, Clustering, classification and n-dimension similarity measure.
- Discussing the main challenges that need to be addressed for social tagging.
- Discussing the extraction of relevance information by the aggregation of user defined tags.

3.1 Introduction

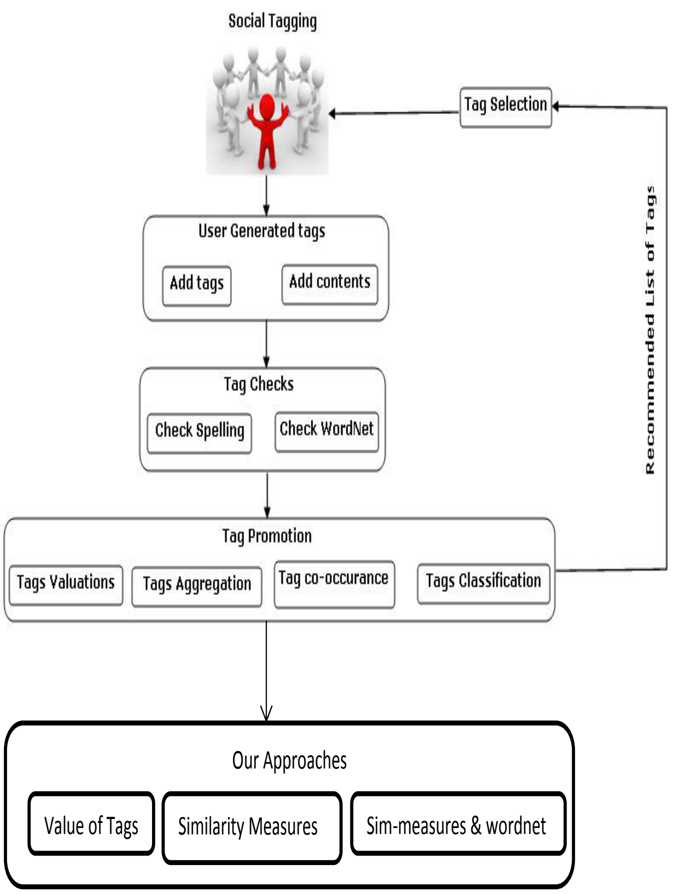
Our recent Information Revolution has contributed to our accelerated scientific progress which was, subsequently, driven by our ability as humans to make sense of the enormous data collections, and harness the resulting findings in a continued sense-making loop. Appropriately, humans were termed informavores: *species that consume information to accelerate their technical evolution* [27]. However, the rate of expansion of the information consumption is limited by the organisation of underlying data, hence, it is important to design and develop systematic and meaningful methods for data *storage* and *retrieval*. To-date, significant progress has been made in the area of textual information retrieval, where numerous models, algorithms and systems governing large text collections have been developed and published. Multimedia, on the other hand, remain largely under-developed due to poorly-understood theories of perception and cognition. In this thesis, our focus will be on images to enable us to study the development of semantic understanding of how the storage and retrieval of a large number of image collections can be improved. Semantics, with respect to images, represent the association between low-level visual features and high-level concepts that can be described in words. Such knowledge possibly arises from the awareness of the context in which photographs are shot. Thus, our objective of image understanding encompasses capturing abstract notions of events, locations, and personalized references that situate images beyond the realm of visual features.

In this chapter, we will briefly outline our approach which is *experimentally-based* and give an account of similarity measures and statistical rationale that are used.

3.2 Our General Approach and Experimental Analyses

Our general approach is depicted in Figure 7.7 and in the next few chapters, we will be carrying out 3 experiments for the purpose of evaluation and validation.

Figure 3.1: General Approach



- **Experiment 1**: The richness and variety of information embedded within an image (also known as Value of Information), will be used to enrich the value of the image tags and, thus increase their accuracy, as explained in Figure 3.2.

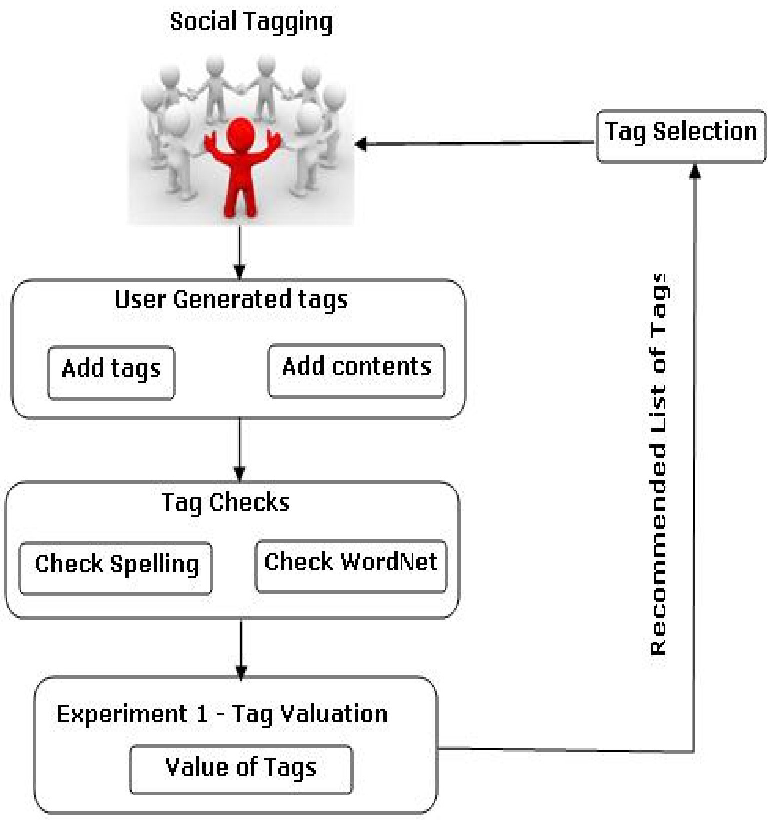
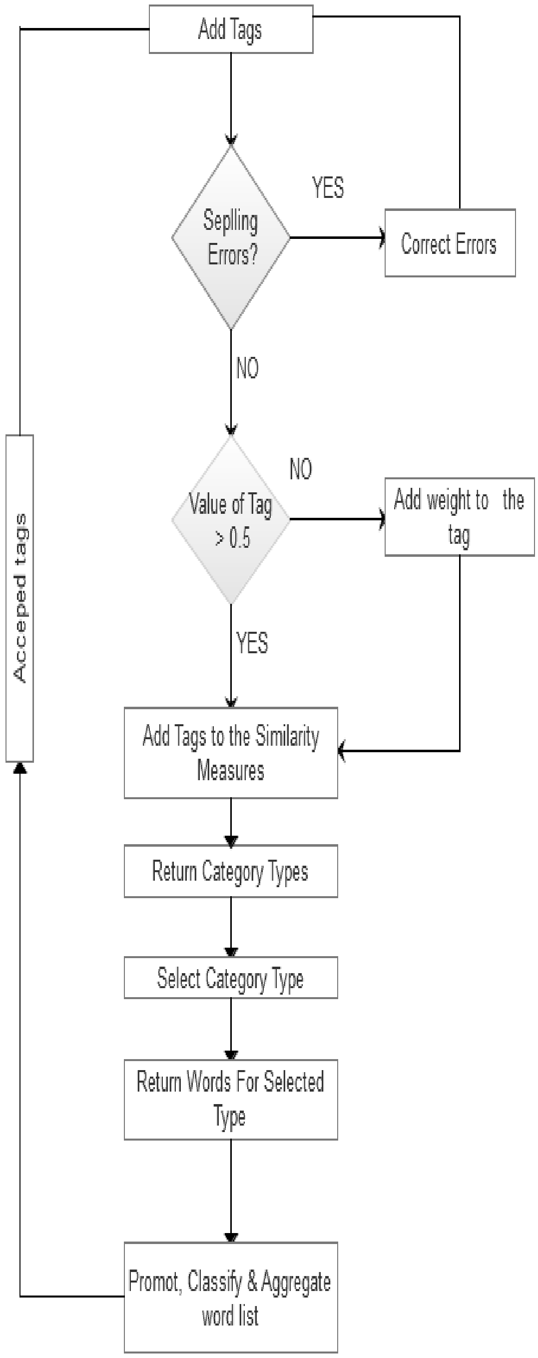


Figure 3.2: Value of Tags (Experiment 1)


Add Tags
Seplling Errors?
YES
Correct Errors
NO
Value of Tag > 0.5
NO
Add weight to the tag
YES
Add Tags to the Similarity Measures
Return Category Types
Select Category Type
Return Words For Selected Type
Promot, Classify & Aggregate word list
Acceped tags

- **Experiment 2**: The various similarity measures and weighting techniques will be employed to increase the accuracy of image tags, as explained in Figure 3.4.

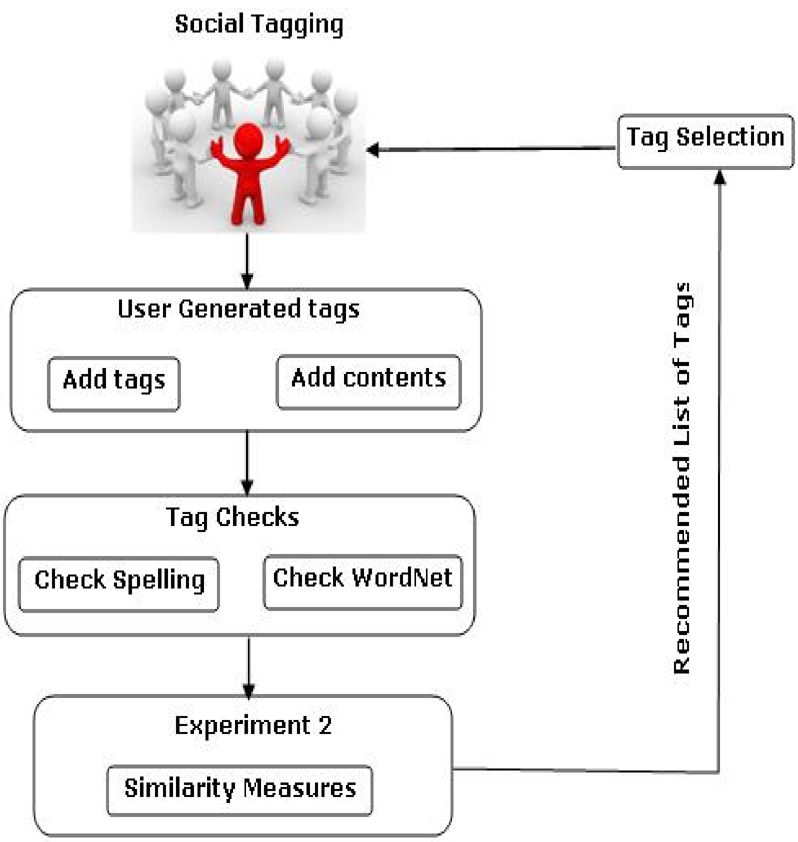
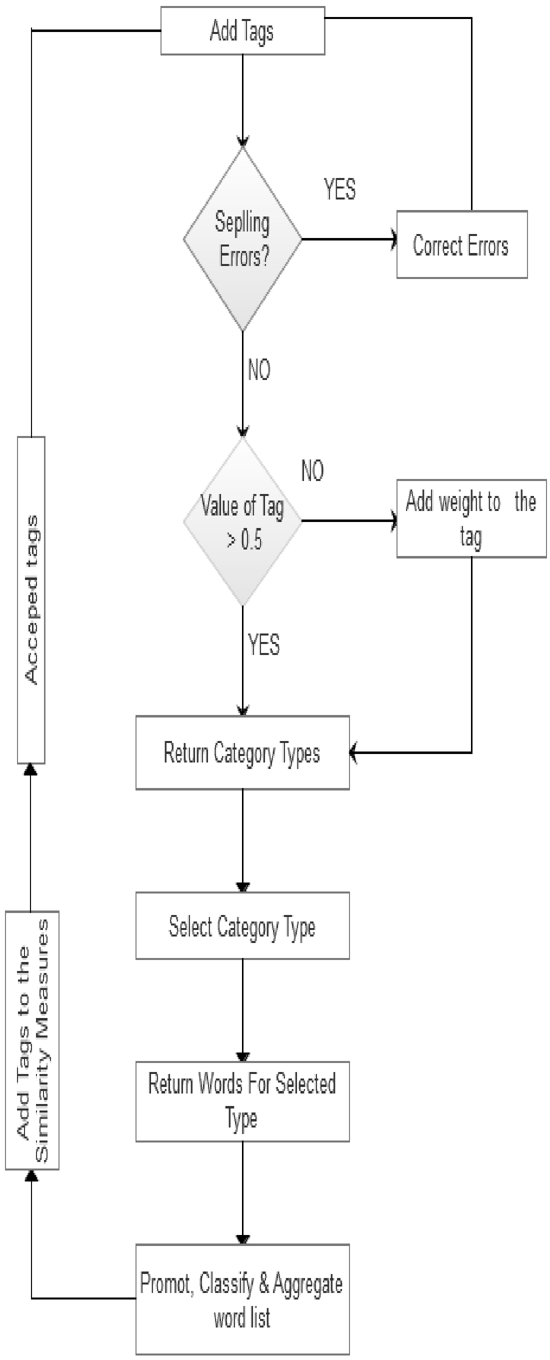


Figure 3.4: Similarity Measures (Experiment 2)



Add Tags
Seplling Errors?
YES
Correct Errors
NO
Value of Tag > 0.5
NO
Add weight to the tag
YES
Return Category Types
Select Category Type
Return Words For Selected Type
Promot, Classify & Aggregate word list
Add Tags to the Similarity Measures
Acceped tags

- **Experiment 3**: Finally, all the methods and resources listed above (or n-dimensions) will be used as a single unified process to improve the accuracy of the tags, as explained in Figure 3.6.

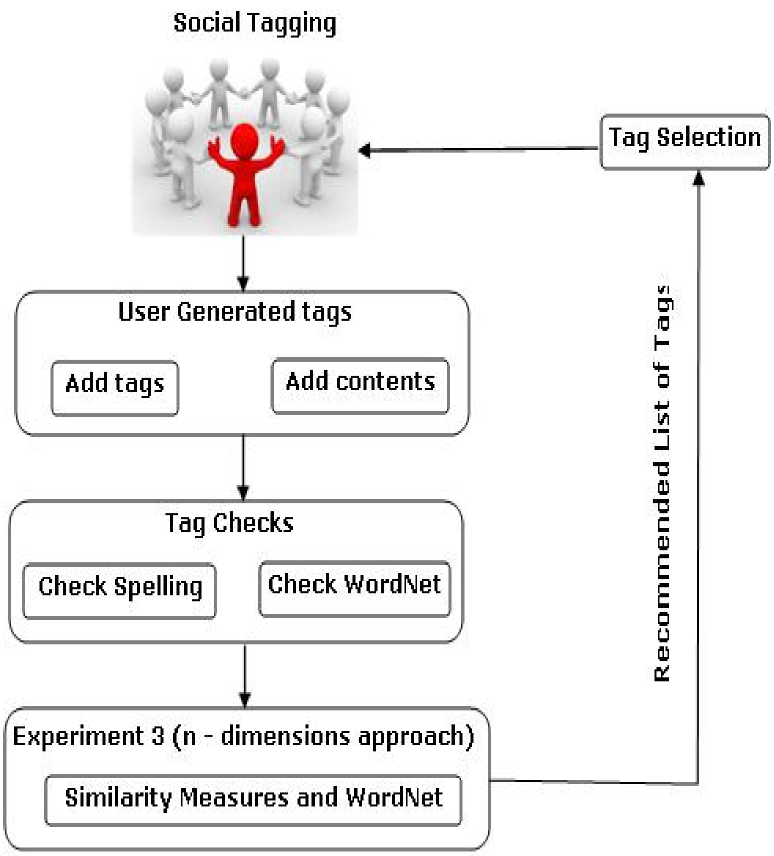
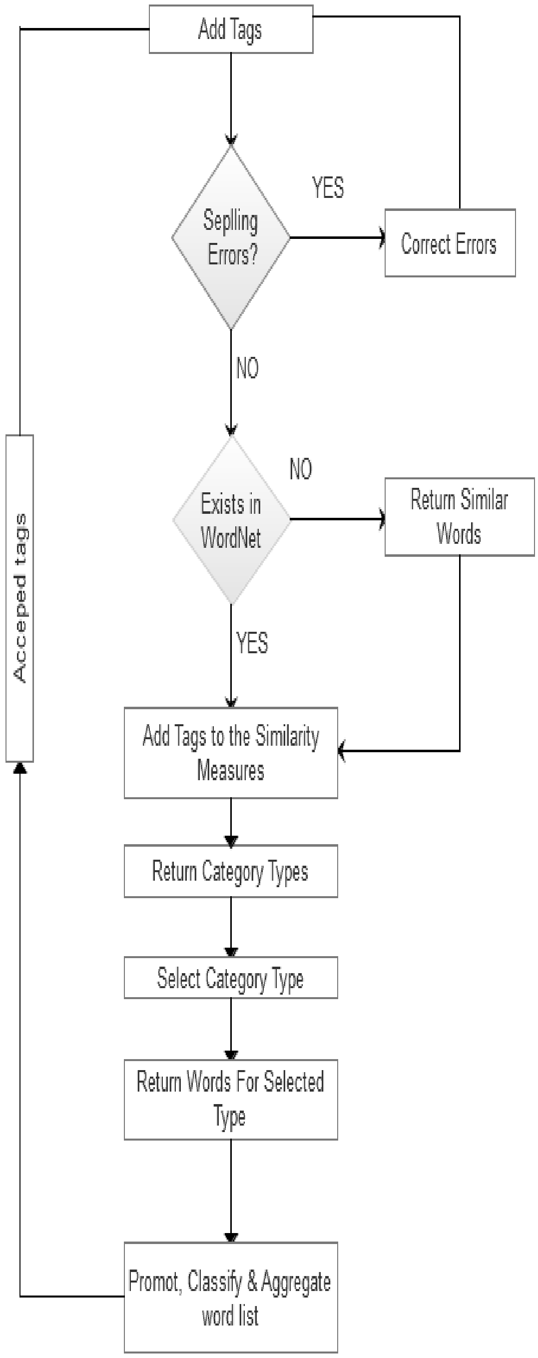


Figure 3.6: n dimensions (Experiment 3)


Add Tags
Seplling Errors?
YES
Correct Errors
NO
Exists in WordNet
NO
Return Similar Words
YES
Add Tags to the Similarity Measures
Return Category Types
Select Category Type
Return Words For Selected Type
Promot, Classify & Aggregate word list
Acceped tags

3.3 Similarity Measures

In chapter 4, we will be discussing the employment of both a single method and a combination of several methods to increase the accuracy of image tagging and retrieval. Our main work will concentrate on evaluating and using several forms similarity measures, which can be defined as the 'Measure of distances between data or sets of data'. In the context of this work, the similarity between two images can further be defined by the three assumptions below:

- The similarity between two images A and B is related to their commonality. The more commonality of attributes they share, the more similar the two images are.

- The similarity between two images A and B is related to the differences between them. The more differences their attributes have, the less similar the two images are.

- The maximum similarity between two images A and B can only be reached when A and B are identical, no matter how much commonality they share.

3.3.1 Summary of Measures

In our case, we will use the various types of Similarity Measures to measure the distance (or similarity) between the sets of attributes of two objects (or photos). In this section, we start by listing and defining the measures that will be used in the following chapters.

- ***Jaccard coefficient***:

As discussed in the previous chapter, we use equations belonging to the Jaccard coefficient to normalise the co-occurrence between two tags. The Jaccard coefficient, sometimes referred to as the "Jaccard similarity coefficient", can be defined as a *statistic* used for comparing the similarity and diversity of sample sets. That is, given two objects, X_1 and X_2, each with n binary attributes, the Jaccard coefficient is a useful measure of the overlap that X_1 and X_2 share with their attributes. Each attribute of X_1 and X_2 can either be 0 or 1.

$$\boldsymbol{\sigma(X_1, X_2) = \frac{|X_1 \cap X_2|}{|X_1 \cup X_2|}}$$

For example if we consider the following attributes for a fruit: *Sphere, sweet, sour* and *crunchy.* Then, an *Apple* ($\boldsymbol{X_1}$) and a *Banana* is represented as

$$\boldsymbol{Apple\ = \{1, 1, 1, 1\}\ and\ \mid Apple \mid =\ 4}$$

$$\boldsymbol{Banana\ = \{0, 1, 0, 0\}\ and\ \mid Banana \mid =\ 4}$$

Here we have

$\boldsymbol{Apple} \cup \boldsymbol{Banana} = \{\mathbf{1, 0}\}$ with $| \boldsymbol{Apple} \cup \boldsymbol{Banana} | = \mathbf{2}$, and

$\boldsymbol{Apple} \cap \boldsymbol{Banana} = \{\mathbf{1}\}$ with $| \boldsymbol{Apple} \cap \boldsymbol{Banana} | = \mathbf{1}$.

$$\boldsymbol{\sigma}(\boldsymbol{Apple}, \ \boldsymbol{Banana}) = \frac{|\boldsymbol{Apple} \cap \boldsymbol{Banana}|}{|\boldsymbol{Apple} \cup \boldsymbol{Banana}|} = \mathbf{0.5}$$

- ***Dice***:

For two sets, X and Y, we can define Dice similarity as:

$\boldsymbol{sim}$ *(X, Y)* $= \frac{\mathbf{2|X \cap Y|}}{\mathbf{|X|+|Y|}}$ In our example above, $\boldsymbol{sim}$ *(Apple, Banana)* $=$ $\mathbf{0.25}$.

- ***Matching***:

$$\boldsymbol{\sigma}(\boldsymbol{x_1}, \boldsymbol{x_2}) = \boldsymbol{\Sigma_y w_{x1y} w_{x2y}} = | \boldsymbol{x_1} \cap \boldsymbol{x_2} |$$

- ***Overlap coefficient***:

Projection-aggregated overlap similarity can be defined as:

$$\boldsymbol{\sigma}(\boldsymbol{x_1}, \boldsymbol{x_2}) = \frac{|\boldsymbol{X_1} \cap \boldsymbol{X_2}|}{\boldsymbol{min}(|\boldsymbol{X_1}|, |\boldsymbol{X_2}|)}$$

The overlap coefficient is a similarity measure that computes the overlap between two sets, or the attributes of two images.

- ***Cosine coefficient***:

The Cosine similarity for two tags t1, t2 can be defined as:

$$\boldsymbol{\sigma(X_1, X_2) = \frac{X_1}{\sqrt{|X_1|}} . \frac{X_2}{\sqrt{|X_2|}} = \frac{X_1 \cap X_2}{\sqrt{|X_1| . |X_2|}}}$$

The resulting similarity between two images ranges from 1 meaning the images are exactly opposite, to 1 meaning the two images are exactly the same, with 0 usually indicating independence, and in-between values indicating intermediate similarity or dissimilarity.

- ***Mutual Information***:

With projection and distributional aggregation we define the *Mutual Information* measure as:

$$\boldsymbol{\sigma(X_1, X_2) = \sum_{y1 \in X1} \sum_{y2 \in X2} p(y_1, y_2) log \frac{p(y_1, y_2)}{p(y_1) p(y_2)}}$$

where $\boldsymbol{0 \ < \ p(y_i) \ \leq \ 1}$.

where for the projection case the probabilities p(y) are defined to perform resource/tag normalization to prevent very popular items from dominating the similarity, and the joint probabilities p(y1,y2) are also based on resource/tag normalization.

3.3.2 Discussion

In our analysis of tag similarity, we can use this coefficient to work out the similarity between the attributes of the tags of two images. Each of the attributes, such as the image's object, colour, action and background, can be compared separately to arrive at a final list of recommended image tags.

The Jaccard equation above can be used in two different measures:

- *Symmetric*, which calculates the most co-occurring (or voted) tags, and
- *Asymmetric*, which calculates the probability of an image being annotated with a tag that is similar to the tag it already annotated by.

Another "voting" method of refining the list of recommended tags, is the concept of 'tag clouds', where the size, colour and font of every tag is determined by its frequency of occurrence (or votes).

The method of valuing tags is also employed, where each user defined tag is 'valued' against a set of criteria, such as Popularity, Topicality, Uniqueness and spelling errors.

The next method to use is the classification of tags, where photos are classified by their number of tags, which are grouped (or summed).

Finally, all of the methods defined above will be used in an n-dimensional analysis, where the value of tags, the similarity measures, tag weights and WordNet will be combined to produce the most accurate list of recommended tags. This approach is further illustrated by Figure 3.8.

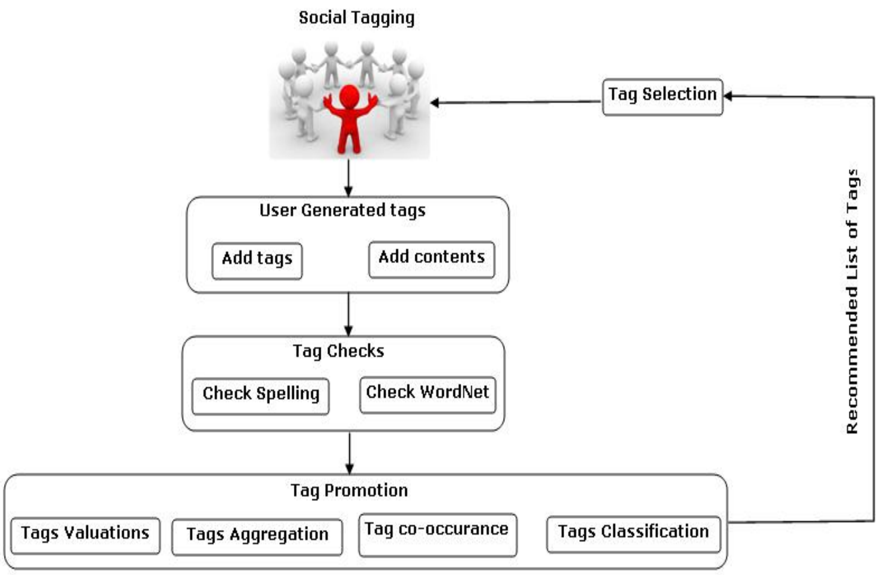


Figure 3.8: Social Tagging Methods

3.4 Statistical Analysis

For social tagging to succeed, the methods employed must work across the whole image tagging spectrum, including the enormous number of Flickr images, which has surpassed 6 billion photos in August 2011 [25]. In April 2012, Flickr claimed that its users have uploaded more than 7 billion photos [102].

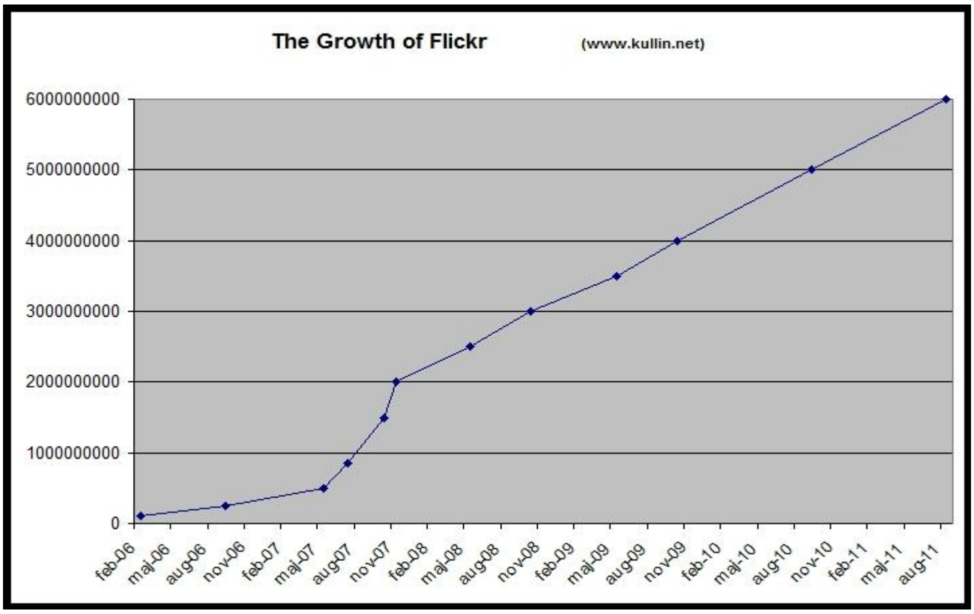


Figure 3.9: Flicker's 6th billion milestone

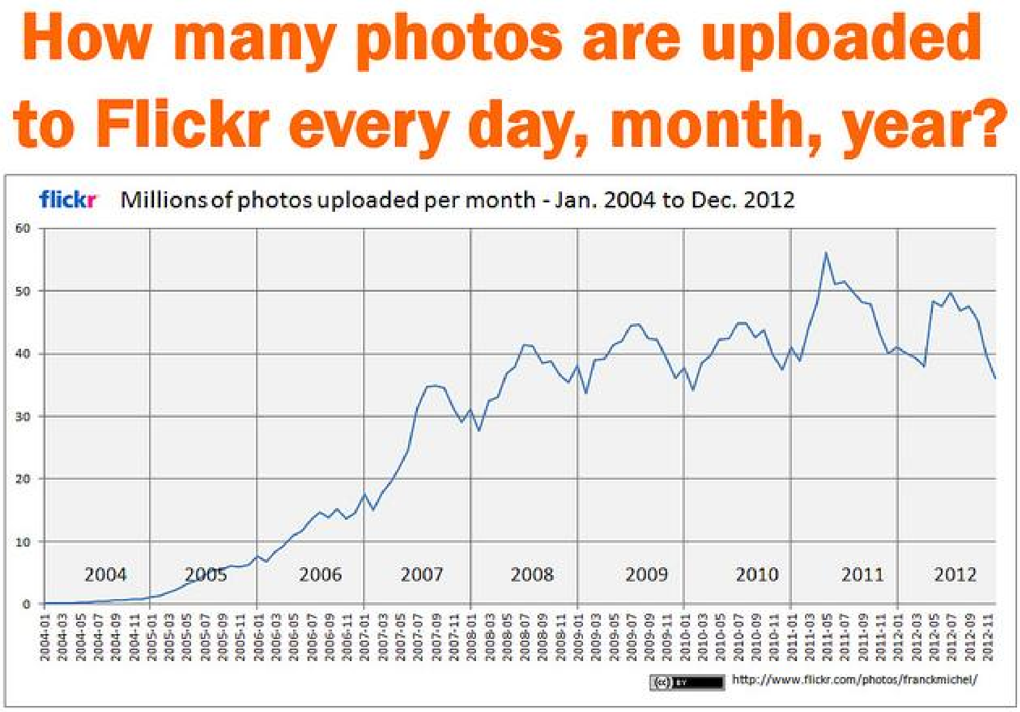


Figure 3.10: Flickr's upload per month and year

Over the last few years, flickr's upload have been increasing 20% year-on-year [72] (Figure 3.9 and 3.10). Such a milestone [110] can be attributed to how the uploads were organised in terms of tags, bookmarks and annotations. Such an enormous amount of tags must be aggregated, clustered, classified, promoted and recommended to the users to improve the efficiency of searching through them [21].

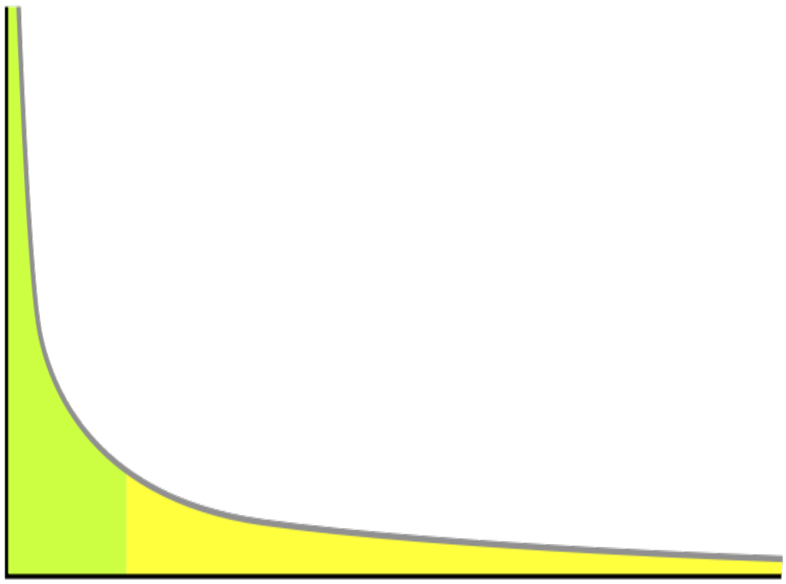


Figure 3.11: The long tail graph

The distribution of the tags within Flickr, Del.icio.us and many other image hosting social sites will, generally, follow power law distribution (Figure 3.11), where a high number of tags are used in low frequency and a low number of tags are used in high frequency. This was achieved in a study by [88]. This study showed that, for a sufficient number of active users, over a period of time, a stable distribution with a limited number of stable tags and a much larger "long-tail" of more idiosyncratic tags develops. Such a development of the tags would be of great use to our intended objectives of classifying and categorising image tags, such that further tagging will only reinforce the pre-existing categorisation scheme given by the current number of stable tags.

The process of tag selection by users can further be explained by the flow diagram in Figure 3.12.

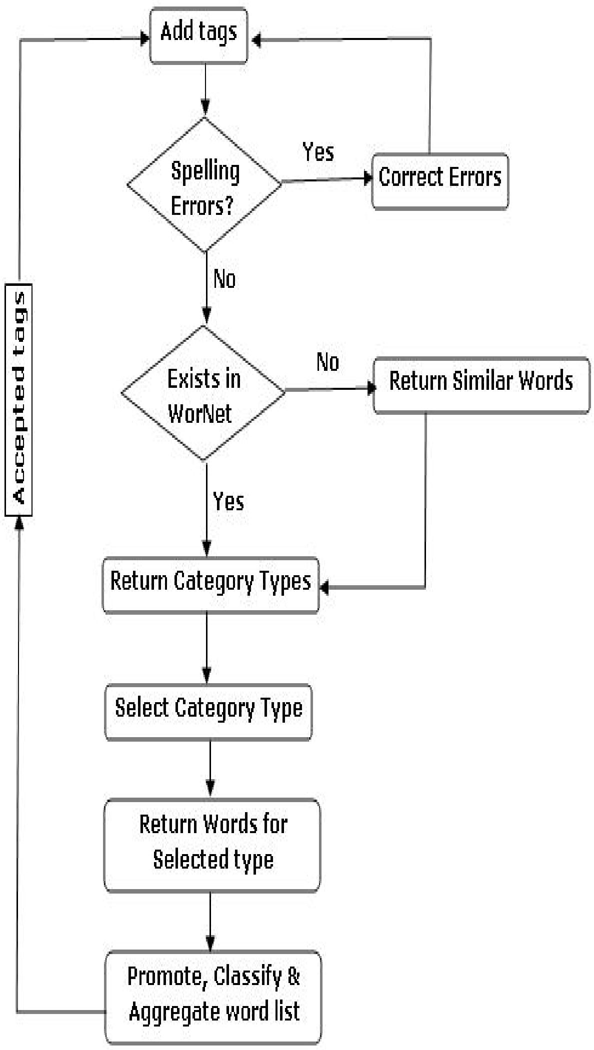


Figure 3.12: The long tail graph

Statistical Semantics

Large volumes of data containing tag vocabulary can be used to form patterns once statistical regularity is reached. This can only be achieved once the social tagging system reaches stability. Emerging patterns and trends can be separated from noise by the volume of usage statistics and can be used to understand the meaning of user defined tags, at a minimum to a level sufficient for information access. The process of the studying of statistical patterns of human word usage for semantic interpretation is referred to as "statistical semantics".

Usage statistics aggregated over a large number of independent users play a pivotal role in a number of information retrieval applications. Using the relevance feedback mechanism, user clicks on Web search results are used to tune future result ranking. In our case, multiple users clicking on the recommended tags that have been promoted by our solution is computed and aggregated, such that it can be used to further improve the results of tag recommendation, by acting as the input for further user tagging.

The extraction of relevance information by the aggregation of clicks per user sessions has been utilised to cluster results with similar semantics, particularly in the reduction of the ambiguity of polysemous queries. The phenomena of social actions and annotations share many similarities with the proposition of wisdom of crowds, that the aggregated verdict of a group of independent people is closer to the truth than that of any individual in the group. The origin of this theory goes back several decades. At a country fair in 1906, Sir Francis Galton observed that when hundreds of people were asked to guess the weight of an Ox, none of the individuals - even the cattle experts, could correctly guess the weight. On the other hand, the average of all estimates was closer to the real weight of the Ox. This occurrence has since then become a famous anecdote for wisdom of crowds. In case of social media annotations, similar analogy exists. When a tag is applied by a large number of users to similar visual content, such relationship is significant from the point of view of tag visual semantics. Drawing a parallel with wisdom of crowds, four main characteristics must be discussed:

- **Diversity of opinion** - Every person is entitled to a personal opinion. In case of social image tagging, each person is entitled to their own subjective interpretation of image content and corresponding use of annotations.

- **Independence** - A person's opinion is not influenced by that of the others. In case of social image tagging, each person can independently provide zero or more tags to zero or more images belonging to self and others. However, complete independence cannot be guaranteed when social influences are strong.

- **Decentralization** - Each person can operate in a local setting and have a different view of the system. In case of social image tagging, decentralization is ensured as users have control of their own tagging activity, without being exposed to tags given by other users to content-similar images.

- **Aggregation** - A mechanism to convert the opinions into an aggregated verdict must exist. As the population size increases, the confidence in the verdict increases as well. Consider for example, the task to compute similarity between two tags.

One simple mechanism is to count the number of images tagged with both tags. Other mechanisms can be devised by considering complex relationships of tags with other tags and users in folksonomy. The assumption of statistical semantics is that a typical user makes rational choices. In such a case, the actions and annotations of a few idiosyncratic users are reduced to noise when a large number of users are considered.

Three statistical techniques will be used in the next chapter, namely co-occurrence, clustering and classification.

The quality of tags that are assigned by various users are affected by their personal choices and the context of the their social network. For example, the type of tagging motivation is correlated with the number and types of tags by the user. Also, the number of tags is proportional to the size of the user's network and the number of social groups to which he belongs [138, 139]. An estimation of the idiosyncrasies helps assess the quality level of a user's annotations. In this section, we summarize a number of descriptive features such as expertise, reputation and reliability.

- **Expertise**: An expert is a provider of high-quality annotated resources. Topic experts can be identified by a substantive contribution of relevantly tagged resources or by a membership to special interest groups related to that topic. Noll et al. defined experts using the Hyperlink-Induced Topic Search (HITS) algorithm and distinguished tag spammers from experts. Members of special interest groups are expected to possess specialized knowledge as compared to non-members. It is possible to identify the topic of expertise and vocabulary by jointly analysing visual content and tagging behaviour of group members using techniques like probabilistic latent semantic analysis.

- **Reputation**: Expertise is a topic-specific feature. Reputation, on the other hand, is a more general property that assimilates overall activities of networked users into a social order. The degree to which a member's work is recognized in the network and a user's social influence can be used as an indicator of reputation. For example, in the computation of Flickr Interestingness, a user's tagging and social activity plays a major role, such that professional and active members are qualitatively ranked higher. Tags, comments and views by high ranked users are considered more useful and can be employed in determining image interestingness. The prestige of special interest groups in which the photo appears is also a contributing factor.

- **Reliability**: A tag assignment is considered reliable if similar associations are consistently observed over a large user collection. Unreliable tag assignments should be treated carefully in relevance ranking applications. The reliability of a specific user's annotations can be modelled using game-theoretic techniques as well.

3.5 Corpus

The API below was used to return tags from Flickr about any searched word.

```
import java.io.IOException;
import java.util.ArrayList;
import org.xml.sax.SAXException;
import com.aetrion.flickr.*;
Import com.aetrion.flickr.tags.*;
public class FlickrJ_Test {
  public static void main(String args[]) throws IOException,
```

```
  SAXException, FlickrException{
    String word = "baby";
    String apiKey = "d2dad4c83ffa2423d88ba197453341a9";
    Flickr flickr = new Flickr(apiKey);
    TagsInterface tag_interface =
      flickr.getTagsInterface();
    ClusterList cluster_list =
      tag_interface.getClusters(word);
    ArrayList<Cluster> cluster =
      cluster_list.getClusters();
    int tag_counter = 0;
    for (Cluster each_cluster : cluster){
        ArrayList<Tag> tag = each_cluster.getTags();
        System.out.println("");
    for (Tag each_tag : tag){
            tag_counter++;
            System.out.print(each_tag.getValue() +
                                              " , ");
            if (tag_counter == 20) break; // to
               retrieve the first 10 tags only in
               each cluster
```

```
              }
         tag_counter = 0;
         System.out.println("");
       }// End of the outer loop
     }// End of main
}// End of the class
```

Tables (3.13) and (3.14) represent a search for a tag that yields a corresponding set of returned words when the API is used.

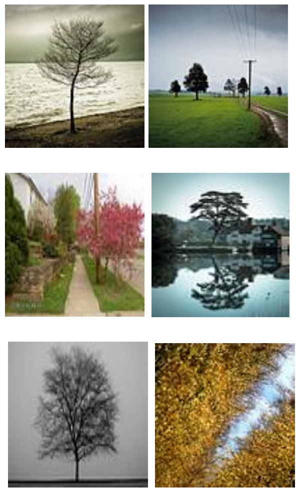
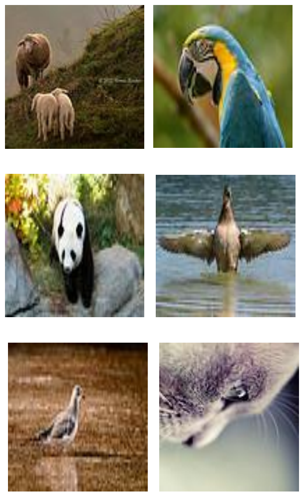
Tag	Tag From Flicker	Examples of Images
Tree	sky , white , orange , blue , grass , sunset , snow , black , bw ,nature , trees , light , sun , autumn , forest , park , macro , shadow , leaves , green , fall , leaf , water , landscape , clouds , red , yellow , canon ,spring , flower , blossom ,	
Animal	cats , pets , kitten , kitty , kittens ,dog , dogs , puppy ,nature , zoo , bird , canon , macro , wild , nikon , blue , closeup , tiger , monkey , lion , giraffe , flowers , bear , park , landscape , sky , trees , deer ,animal , cat , birds , pet , cute , water , wildlife , white , black , eyes , horse , portrait ,fish feline , green , horses , gato , ducks , gatto , baby ,	

Figure 3.13: Tree and Animal

Tag	Tags From Flickr	Examples of images
Baby	girl , children , kids , pink , toddler , beautiful , daughter , portrait , boy , family , newborn , mother , infant , love , kid , blackandwhite , canon , people , nikon , mom , son , cute , child , eyes , smile , bw , blue , white , happy , face , black , dog , adorable , sweet , animal , puppy , little , babies , funny ,	   
Beautiful	flowers , green , macro , flower , water , pink , yellow , ocean , rose , garden , sea , closeup , portrait , eyes , pretty , beauty , cute , face , female , people , canon , nature , blue , sky , landscape , sun , orange , clouds , color , beach , trees , tree , hot, summer , girl , woman , model , black , red , fashion , hair , white , bw , lady , light , love ,	   

Figure 3.14: Baby and Beautiful

3.6 Summary

Since the start of the exponential increase of the image tagging phenomena, many studies have been developed to explore the method of increasing the accuracy of image tagging. The majority of the studies centre on using various methods of promoting tags to users, who will accept and select one or more recommended tags to add to their images.

In contrast, the bulk of our research will centre around employing any number (or n-dimensions) of methods, as no other study has addressed the development of such an approach. This approach will be developed, initially, using similarity measures and weighting techniques, followed by an approach that takes advantage of the rich value of information within an image and, finally, using the n-dimensions approach.

Chapter 4

THE INFORMATION MODEL

Objectives

- Discussing the ways in which image tagging and text tagging differ.
- Discussing the added value of information to tag values.
- Discussing the use of tag recommendation based on tag co-occurrence.
- Discussing the concept of tag clouds and their use by photo sharing websites.

4.1 Introduction

In Chapter 2, we looked at several attempts to solve the issues of the lack of accuracy of image tagging. In this chapter, we will describe the adoption of the similarity measures to improve the accuracy of image tagging. We will also describe studies that the methods that have been used to promote tag recommendation for user-defined tags, based on tag co-occurrence, tag visual correlation and tag aggregation and promotion.

The information theory will detail the tag recommendation process by employing either Symmetric or Asymmetric measures. An expansion of the chapter 2 example that used the Jaccard coefficient will also be discussed, as will as the concept of the Value of Information (or tags).

The Rationale section will explain why these ranking methods were selected, and finally the Examples section will outline the method used to combine many of the methods detailed in the previous sections.

Web browsers read HTML documents from top to bottom, left to right. Whenever a browser finds a tag, the tag is rendered accordingly. Paragraph tags render paragraph text, image tags render images, etc. A comparison between these various tag types, namely the text and image tags, will be carried out.

The use of the tag cloud concept by photo sharing websites, such as Flickr is emphasised with visual examples.

4.2 Tagging

To recall, tagging is the act of assigning a keyword (a tag) to a piece of information, such as a web page content or a digital image. Tags are a type of meta-data that allows an item to be described and thus, be found again by search engines. This makes tags to be very useful, both to its creator and to the larger community of web users. Tags are generally chosen informally and personally by the item's creator or by its viewer, depending on the system. Websites that include tags often display collections of tags as *tag clouds*.

A tag cloud can be defined as a visual representation of text data. Such a concept is typically used to depict keyword meta-data (tags) on websites, or to visualize free form text. Tags are represented by single words, with the size and colour of the font representing the importance of the tag. They can be displayed on a website as navigation aids, where the terms are hyper-linked to items associated with the tag.

There are many types of tag cloud applications. The main and most used type, which is used by Flickr, is "Frequency" type where size represents the number of items to which a tag has been applied, as a presentation of each tag's popularity. This is useful as a means of displaying meta-data about an item that has been democratically "voted" [112] [85].

The tag cloud image below represents Flickr's all time most popular tags. The tags are entered by the users who can assign up to 75 tags to each photo.

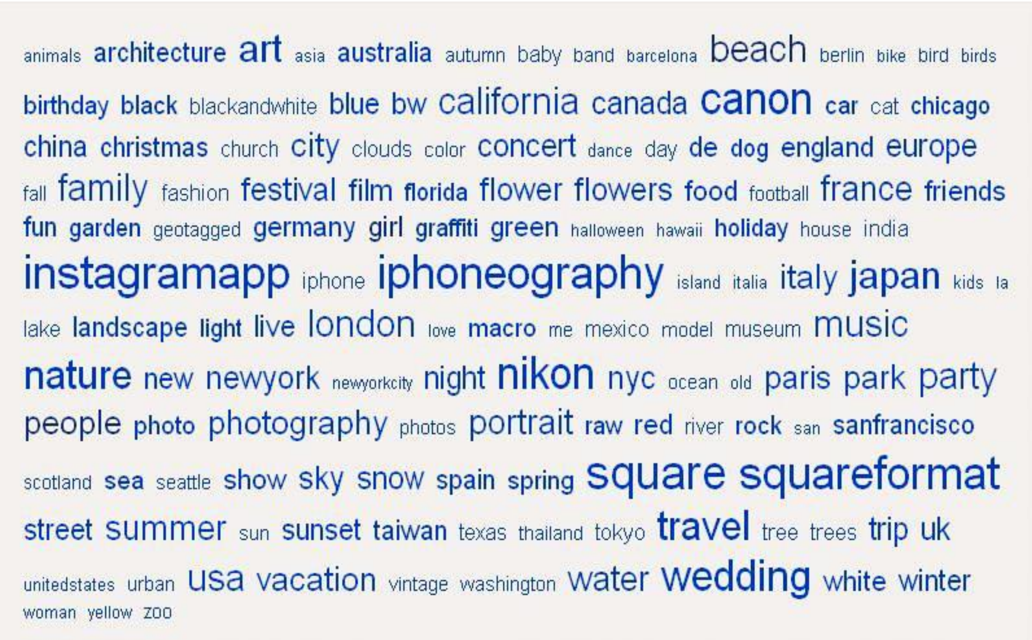


Figure 4.1: Flickr's [24] all time most popular tags

Tag types on websites vary markedly between normal *text* tagging and *image* tagging. In what follows we discuss both in detail.

4.2.1 Text and Image Tagging

Text Tagging

A website's contents are made up of words that are usually trawled and indexed by most search engines for categorisation. For example, a food recipes site that utilises words like "tips", "baking" and "cakes" several times within its content, will feature high on the list of search results for anyone looking for tips on bread baking. This type of tagging requires the user to use words that are relevant to the contents and to the website audience, and not use too many abbreviations or slang phrases. This is particularly true for article titles as vague or conceptual headlines will not be the exact phrases that people are searching for. An example of a site that uses relevant keywords as tags to climb up the search engines' ranking is the about.com site for 'German Baking', http://germanfood.about.com/od/breadbaking101/a/bread-baking-101.htm. The keywords 'bread', 'bake' and 'tips' are repeated 14 times, 8 times and 5 times respectively. Adding a good variation of the keywords to the website may make a significant benefit to search engine optimization.

Image Tagging

Images are not inserted into an HTML page. Instead, they are linked to the HTML page by the source ("src") attribute of the image tag. The value of the src attribute is the URL of the image to be displayed. The URL points to the location where the image is stored. The required "alt" attribute of the image tag, specifies an alternate text for an image, if the image cannot be displayed. The value of the alt attribute is a user-defined text, and usually contains the keywords pertaining to the image contents. The inclusion of relevant keywords within the *alt* attribute and its surrounding innermost element content will help search engines to find the image and the page that contains it.

Digital images that are loaded to a web page can be more searchable if they are tagged appropriately. Images are as searchable as the articles they support, and can significantly improve a website's ranking by increasing the amount of traffic through search engines such as Google images. Images are usually tagged when they are uploaded, so relevant captions (tags) will help users to easily find them. Images that do not have titles or a caption with the relevant tag, will be very difficult to find through a search engine.

4.2.2 Tags Challenges

Social tagging classifications differ significantly from expert tagging as they are often performed as a result of personal motivation or agenda. Additionally, community trends influence and affect the quality of tags. The sub-sections below, will detail the main challenges that need to be addressed before social tagging can be suitably utilised. We also look at the effect of these challenges on the usage statistics.

- **The influence of the users' culture**: Ethnicity and cultural differences guide perception and cognition differently. For example, an analysis of image tags created by European, American and Chinese participants concluded that whereas Westerners focus more on foreground objects, the Easterners have a more holistic way of viewing images early on. This was discovered through the analysis of tag assignment order. For Easterners, the specificity of tags increased from holistic scene description to individual objects. On the other hand, the tags given by the Westerners focused on individual objects first and then on overall scene content.

- **The influence of Motivation**: Motivation of, probably, forms a major influence on the usability of tags for all purposes. Tags that arise from the need for future retrieval and contribution, particularly for the benefit of an external audience, are likely to be visually more relevant compared to tags used for personal references. Images that are annotated and shared within special interest groups are very likely to be specifically annotated and heavily monitored. They would also be heavily influenced by the motivation of the interest group.

- **The Users' Domain knowledge**: Some users who tag their images with non-understandable words, characters, personal references or numeric symbols, can only be thought of as doing so because they have a praticular knowledge about the domain that caused them to save and annotate the images in the first place. Such tags have no use or meaning to the wider audience, and should be filtered out, so as not to affect usage statistics.

- **The issue of Semantic loss**: An annotator in folksonomies is not obliged to associate all relevant tags with an image, leading to semantic loss in the textual descriptions. The batch-tag option provided by most photo sharing sites adds to this problem by allowing users to annotate an entire collection of photos with a set of common tags. Even if such tags are potentially useful to provide a broad personal context, they cannot be used to identify image-level differences, thus leading to semantic loss. One consequence of this fact is that the absence of a tag from an image description cannot be used to confirm the absence of the concept in that image. Hence, such images cannot be directly used as negative examples for training.

- **The issue of Vocabulary**: The spontaneous choice of words to describe the same content varies among different people, and the probability of two users using the same term is very little. Known as the vocabulary problem, this issue is often cited as a common characteristic of folksonomic annotations. The different word choices introduce problems of polysemy (one word with multiple meanings), synonymy (different words with similar meanings) and basic level variation (use of general versus specialized terms to refer to the same concept).

4.2.3 Discussion

For both of the tag types above, repetition must be restricted to relevant keywords only, as spamming with non-relevant keywords to force traffic to the sites may force the search engine to penalise the site and lower its ranking. For the most optimised search results, keywords should identify elements that users are likely to use as search items. Keywords must be used strategically and sparingly.

To illustrate the approach of manual metadata generation, we look at social tagging, which requires all the users in the social network to label web resources with their own keywords and share with others. This approach is best demonstrated on the social photo sharing site, Flickr, which has grown into one of the premier photo hosting and sharing sites on the internet, boasting an upload rate of up to 2,504 uploads per minute.

Flickr has simplified social tagging, such that users can enter any tag for their photos. However, such a simplified approach has introduced the issue of ambiguity, where different users may tag similar images with different words and may also use a single general tag to represent different images. Therefore, many images may not be effectively retrieved. A good example is the search for "jelly bean", which retrieved the two images below:

Figure 4.2: Example of an Ambiguous search for "jelly bean"

In general, it is quite difficult for the web users to realize the existence of ambiguity, hence users continue to generate and retrieve many irrelevant tags.

Another issue associated with manually entered image tags is the problem of misspelling. Users who enter incorrectly spelt tags will make the act of finding these images very difficult for other users. It is estimated that, in the Flickr tag distribution, around 60 % of tags in the tag corpus are misspelling or meaningless words. Nardini et all [106] proposed a spell checking system on tags to manage sets of terms (with their relative co-occurrence patterns). The method exploits correlation between tags associated with the same resource. This method is then able to detect and correct common variations of tags by proposing the "right", i.e., the most commonly used, versions. However, although such a method may increase photo tagging accuracy, it does not completely eliminate the synonymy and tag ambiguity problems.

One good method to avoid noise and compensate for the semantic loss, is the proposal of tag recommendation by combining both visual correlation in concept level and tag co-occurrence information. The semantically or visually related tags are recommended to the users to improve the tagging quality. The recommendation system will remind the users of the alternative tags and it can also help clarify the true semantic of the images. For example, when the user tags an image with word "la Sagrada Famlia", the recommendations system will list more rich and precise tags based on the input tags, such as

"Gaudi", "Barcelona" and "church". These recommendations will help users clarify the image content as well as reminding them of related semantics which may otherwise be ignored. They will also help with tag misspelling, where users can tag an image by choosing rather than typing, which effectively avoids spelling errors.

The quality of tag recommendation is quite important to social tagging and the consequent performance of image search. Firstly, high quality tag recommendation will motivate users to contribute more useful tags to an image. The average number of tags for each image on Flickr is relatively small. One of the reasons for that is due to users not entering a large amount of tags as they generally cannot think of too many words, and only a few people would spend much time thinking about alternative tags very precise tags. With the help of high quality tag recommendation, users can provide many more of useful tags. Thus the average number of correct tags for each image is expected to increase. Additionally, tag recommendation will remind the users of more rich and specific tags. The distribution of tags on Flickr follows a power law distribution. Most of the users only use the popular keywords, which are only around 5.82% of the whole tag collection. These tags are popular because they are common vocabulary and easily come to mind. Another 33.21% of the tags which appear 50 to 5,000 times are also informative while generally ignored by most users, because these words are more professional terms or only used for specific object or situations. The tag recommendation will help remind the user to use both popular and specific tags for social tagging. This reminder will also help to create more precise tags.

4.3 Information theory and modelling

Our first approach for this thesis is to use a system of tag recommendation strategies by utilising a combination of different kinds of correlations to rank image tags, namely tag co-occurrence, tag visual correlation and Tag Aggregation and Promotion.

Our model of image ranking can further be detailed by the flowchart above, where web users tag images with semantically related words, such as "Jelly Bean" together with "Android". Within a large photo sharing social website containing numerous independent users, such as Flickr, the semantic relationship can be captured and utilised. However, this method alone is not sufficient to all the relationships between the tags such as "window" in the photo of a "house". The photos containing both "house" and particular style of "window" may be tagged as "house" only. Such an issue can be solved by

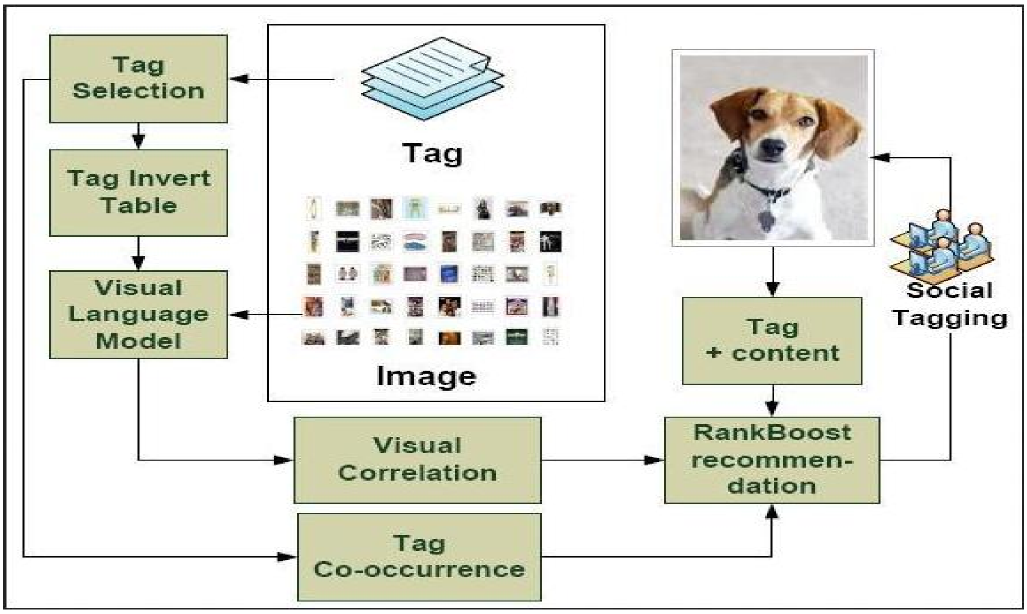


Figure 4.3: Social tagging recommendation system

applying tag visual correlation to measure the tags visual similarity. These two methods of correlations only use the relation between tags, which can be combined in the Rankboost framework [116] [26], which in turn uses the order of instances rather than the absolute distance.

The tag recommendation process can be explained by an example, where a selected photo with user-defined tags and an ordered list of candidate tags is derived for each of the user-defined tags, based on tag co-occurrence. The lists of candidate tags are then used as input for tag aggregation and ranking, which ultimately produces the ranked list of recommended tags. For example, the photo of Sacr-Coeur Figure(4.4) may have two user-defined tags, namely Sacr-Coeur and Paris. Using Tag Co-occurrence, a list of co-occurring tags (church, architecture, montmarte, seine, Europe, travel and night) is derived 4.4. They have some tags in common, such as France and Paris. After aggregation and ranking four tags are recommended: *Paris*, *Church*, *Architecture* and *France*. The actual number of tags being recommended should, of course, depend on the relevancy of the tags, as we will see in the example case of using the 'value of tags' (section 5.3).

Tag co-occurrence is the pillar that the tag recommendation approach is built upon, and as a consequence, only works reliably when a large quantity of supporting data can be captured and accessed [96]. Fortunately, the amount of user-generated content that is created by Flickr users, satisfies this demand and provides the collective knowledge base that is needed to make tag recommendation systems work in practice. There exists various methods to calculate co-occurrence coefficients between two tags. The co-occurrence between two tags is defined as the number of photos, in our collection, where both tags are used in the same annotation.

Using the raw tag co-occurrence for computing the quality of the relationship between two tags is not very meaningful, as these values do not take the frequency of the individual tags into account. Therefore it is common to normalise the co-occurrence count with the overall frequency of the tags. There are essentially two different normalisation methods: symmetric and asymmetric.

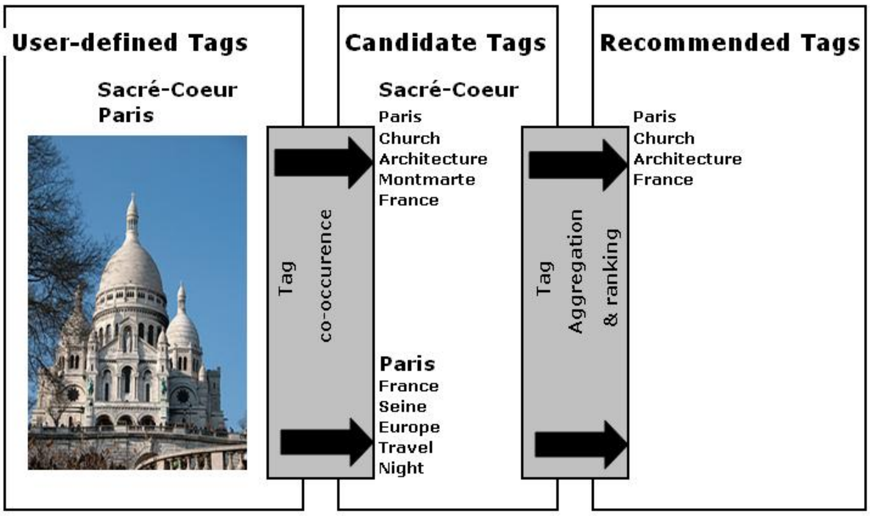


Figure 4.4: The tag recommendation process

Symmetric measures:

We use the Jaccard coefficient, introduced in chapter 2, to normalise the co-occurrence of two tags $\boldsymbol{t_i}$ and $\boldsymbol{t_j}$ by calculating:

$$J(t_i, t_j) := \frac{|t_i \bigcap t_j|}{|t_i \bigcup t_j|}$$

The coefficient takes the number of intersections between the two tags, divided by the union of the two tags. The Jaccard coefficient is known to be useful to measure the similarity between two objects or sets. In general, we can use symmetric measures, like Jaccard, to deduce whether two tags have a similar meaning.

Asymmetric measures:

Alternatively, tag co-occurrence can be normalised using the frequency of one of the tags. We can use the equation:

$$\boldsymbol{P(t_j|t_j) :=} \frac{|t_i \bigcap t_j|}{|t_j|}$$

The equation captures how often the tag $\boldsymbol{t_t}$ co-occurs with tag $\boldsymbol{t_j}$ normalised by the total frequency of tag $\boldsymbol{t_i}$. This can be interpreted as the probability of a photo being annotated with tag $\boldsymbol{t_j}$ given that it was annotated with tag $\boldsymbol{t_i}$. Many other variations of asymmetric co-occurrence measure have been proposed in the literature before to build tag (or term) hierarchies.

To illustrate the difference between symmetric and asymmetric co-occurrence measures consider the tag Eiffel Tower. For the symmetric measure we find that the most co-occurring tags are (in order): Tour Eiffel, Eiffel, Seine, La Tour Eiffel and Paris. When using the asymmetric measure the most co-occurring tags are (in order): Paris, France, Tour Eiffel, Eiffel and Europe. It shows that the Jaccard symmetric coefficient is good at identifying equivalent tags, like Tour Eiffel, Eiffel, and La Tour Eiffel, or picking up a close by landmark such as the Seine. Based on this observation, it is more likely that asymmetric tag co-occurrence will provide a more suitable diversity of candidate tags than its symmetric opponent.

The next step in the process of tag aggregation is to merge the known lists of candidate tags for each of the user-defined tags, into a single ranking. There are two aggregation methods, based on voting (a strategy that computes a score for each candidate tag) and summing (a strategy that takes the union of all candidate tag lists) [96] that can be used along with a re-ranking procedure (where tags are arranged in their order of high relatedness [55]) that promotes candidate tags containing certain properties and significance values.

To achieve this, we use three different types of tags:

- User-defined tags U refer to the set of tags that the user assigned to a photo.
- Candidate tags $\boldsymbol{C_u}$ is the ranked list with the top most co-occurring tags, for a user-defined tag $\boldsymbol{u \varepsilon U}$. We denote C to refer to the union of all candidate tags for each user-defined tag $\boldsymbol{u \varepsilon U}$.
- Recommended tags R is the ranked list of the most relevant tags produced by the tag recommendation system.

For a given set of candidate tags (C) a tag aggregation step is needed to produce the final list of recommended tags (R), whenever there is more than one user-defined tag. In this section, we define two aggregation strategies. One strategy is based on voting (a strategy that computes a score for each candidate tag), and does not take the co-occurrence values of the candidate tags into account, while the summing strategy (which takes the union of all candidate tag lists) [96] uses the co-occurrence values to produce the final ranking. In both cases, we apply the strategy to the top co-occurring (or highly related) tags in the list.

Another method of increasing the accuracy of image tags starts by expanding the example from chapter 2, where the Jaccard coefficient was employed along with WordNet to unify the tags described by the users. The 3 fields used in the example are expanded to four fields (or parameters), namely 'primary object', 'secondary object', 'action' and 'primary colour'. However, in this case, each potential tag information received will first be assessed for its value. This is also referred to as Value of Information or Value of tags.

As an example of the implementation of the 4 fields method, consider the search for a photo of a red sky at a lake. In normal circumstances, such a search may return the non-relevant image Figure(4.5), which shows a lake with red flowers but without the red sky at dusk.

Figure 4.5: Lake with red flowers

However, our enhanced method (4.3) of tagging would allow users to enter extra object names to further identify the tag. In this case, the primary object would be 'lake', the secondary object would be 'sky', the action would be 'dusk', or more precisely, 'sunset', and finally the colour would be 'red'.

In this method, before the WordNet database is queried to check if the specified words stored in the tags and returned by the search do exist, each tag returned is 'valued' against a set of pre-defined criteria. Examples of this criteria are:

- Popularity: What is the size of the tag on the Flickr tag cloud, i.e. how many times has the tag been voted for?

Figure 4.6: Lake with red sky at dusk

- Topicality: Is the tag suitable for the search topic? As an example, consider a search for an image of the city of London. The returned tags may represent London City or the novelist Jack London. In this case, the results are compared to the categories on WordNet, where London city belongs to 'noun.location'. This category is ranked higher (as it has more tags per photo) than the London Novelist category 'noun.person'

- Uniqueness: Is the tag of the photo unique and unambiguous? For example, a photo of a 'car' which is also tagged 'car' is unique and can only refer to a car, irrespective of its type.

- Redundancy: Are there too many irrelevant and redundant tags? For example, a search for a photo of a cat that returns 'cat', 'feline', 'tabby', 'fluff', 'jinx' (for a photo of a black cat) and 'cuddles' is, obviously, plagued by too many redundant tags, when 'cat' or 'feline' would suffice.

- Simplicity: How simple is a photo tag? For example, a photo of a Teapot that is tagged 'Teapot for brewing Darjeeling tea' may be too complex for search engines, as well as tag rankings algorithms (and the word Darjeeling may also be classified as spam). Ideally, the photo should be tagged as, simply, 'Teapot'.

- Spelling: Misspelled tags should, obviously, be excluded from the list of returned tags.

- Recency: For this assessment, tags are ranked by age, such that an image that has several possible tags, which were created over a long period of time, would rank the most recent tags higher than the oldest ones.

The returned list of tags is deemed to be much more accurate in terms of the search query, and this can be used to more accurately return the image in Figure(4.6), which represents exactly the criteria being searched for i.e. a lake with a red sky.

There are many other tag criteria that can be used to assess returned tag values. However, the criteria of topicality and relevance is of more importance as it answers the question "What are users tagging?" This criteria is mapped to WordNet categories, which are used to bind tags to the category with the highest ranking. Figure 4.7 shows the distribution of Flickr tags over the most common WordNet categories, which can be used to assess and classify tags. When focussing on the set of classified tags, we find that locations are tagged most frequent (28%); followed by artefacts or objects (16%), people or groups (13%), actions or events (9%), and, finally, time (7%).

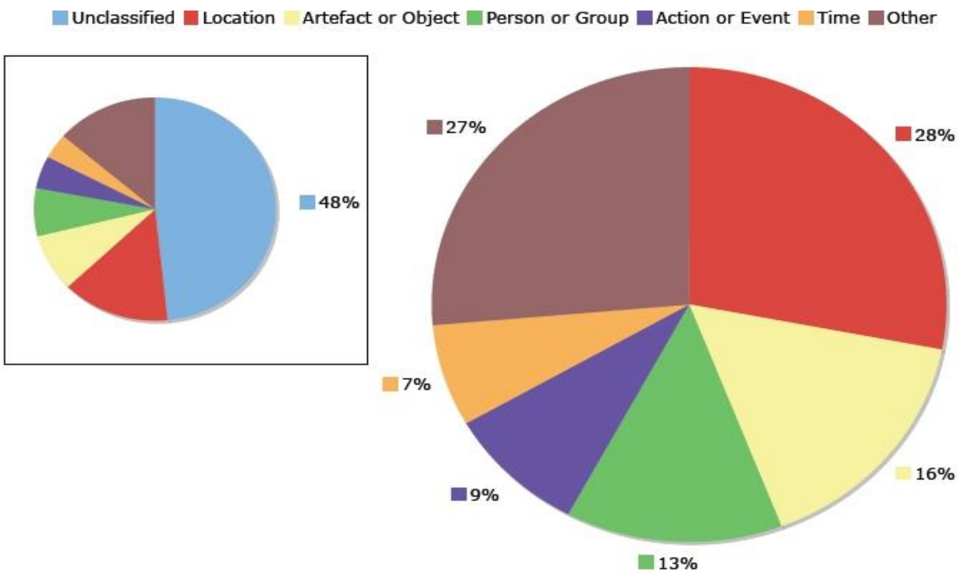


Figure 4.7: Flickr's tags Most frequent WordNet categories

From this information, we can conclude that users do not only tag the visual contents of the photo, but to a large extent provide a broader context in which the photo was taken, such as, location, time, and actions.

Another criteria that would be used to rank photo tags, is the classification of tags as defined in table 4.3, which looks at classes of photos with one tag, photos with 2-3 tags, 4-6 tags, and more than 6 tags, respectively. The table can be used to compare voting strategies (i.e. photos with a high number of user tags) against summation strategies (photos with aggregated tags). Experiments have shown that the increases in the accuracy of tag ranking is proportional to the number of a photo's user-defined tags [63]. This indicated that only 13 % of all tagged photos have a higher degree of accuracy as they contain more than six tags. The high number of tags will serve as an input into the Jaccard measure of co-occurrence, as we will discuss in the Example section.

	Tags per Photo	**Photo%**
Class I	1	31 %
Class II	2 - 3	33 %
Class III	4 - 6	23 %
Class IV	> 6	13 %

Table 4.1: **Definition of photo-tag classes and the percentage of photos in each class**

Finally, below we will introduce an example that details how to increase the accuracy of tagging by employing n-dimension of resources, i.e. as many of the above methods as possible.

4.4 Rationale

Our research work revolves around the improvements of image tagging, and for this reason, we have opted to combine many of the methods discussed in the previous sections. Users will be able to enter tags based on two searchable objects, as well as the photos action and background. This will significantly enhance the value added to the photo tags.

Once the user defined tags are saved with the photos, the returned list of tags, from a search query, will be enhanced by comparing it against a set pre-defined values (or criteria). Such an action would serve to filter out many irrelevant results. The returned list would be further enhanced by promoting the tags via the use of tag classes that utilise voting strategies.

The final filtered list of tags would then be used as recommended tags for users to choose from, as this would reduce the introduction of irrelevant tags that can be entered due to misspellings, inaccurate descriptions and attempted spamming. Users would then select one or more tags from this pre-defined list, without the ability to enter free text.

Once we get a large number of photos that have been tagged by a promoted and recommended set of tags, the set of results returned by a search query would be highly accurate. This would allow us to accurately compare similarity measures between photos using the Jaccard coefficient.

4.5 Example

In this chapter, we will use a single example that amalgamates all the n methods (or dimensions) detailed in the above sections. The example we will use is a photo of Big Ben's tower. Initially, we allow users to enter their tags into the four fields described in the previous sections; namely primary object, secondary object, action and colour. However, before the tags can be added, we use the Jaccard method to calculate co-occurrence coefficients. Both normalisation methods; symmetric and asymmetric will be used for the calculations.

For the primary object, we use the symmetric measure to find the most co-occurring tags which returns (in order): *Big Ben, Big Ben Tower, Westminster, Thames, London and England.* These recommendations will be offered to the users to populate the primary object field. Next, we use the asymmetric measure to calculate the most co-occurring tags for the secondary object which returns (in order): *London, England, Clock, Tower, Westminster, Ar-*

chitecture and Europe. It is more likely that asymmetric tag co-occurrence will provide a more suitable diversity of candidate tags than its symmetric opponent. Therefore, it is more useful for returning the secondary object's recommended list. Similarly, the co-occurring tags for action would return: *Travel, Tour, Visit and Book.* Finally, the colours returned are: *Blue, Black and White.*

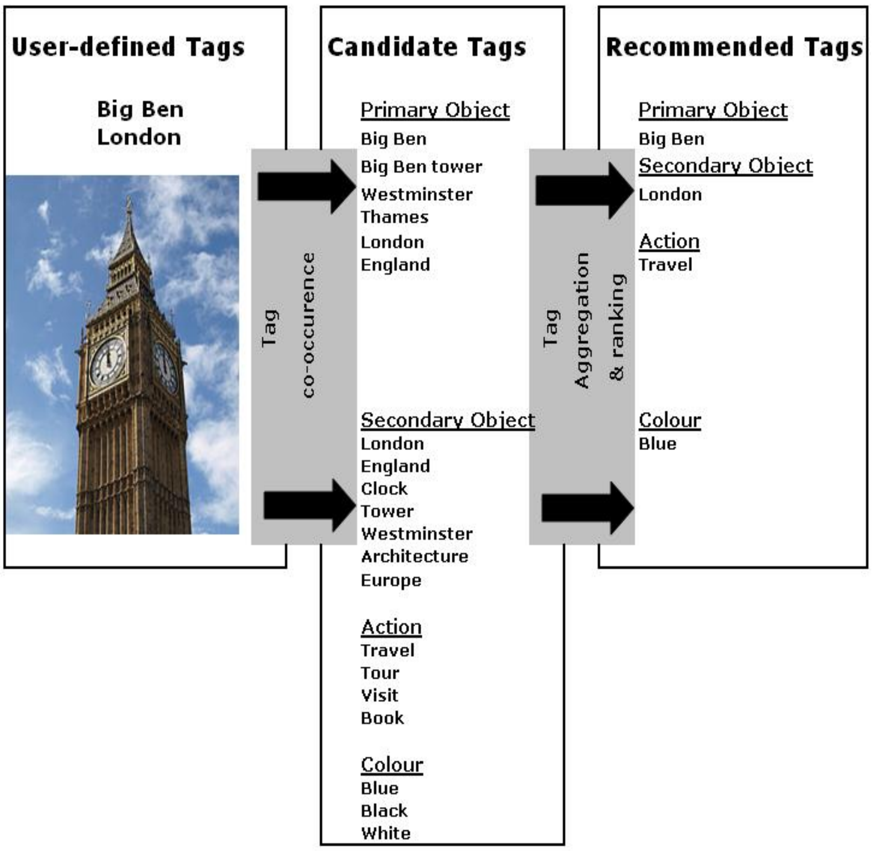


Figure 4.8: Big Ben's Tag Recommendation

In Figure 4.8, the list of tags produced by the symmetric and asymmetric measures for each of the four fields are further aggregated to produce the final list of recommended tags. We use two aggregation strategies. One strategy is based on voting, and does not take the co-occurrence values of the candidate tags into account, while the summing strategy uses the co-occurrence values to produce the final ranking. In both cases, we applied the strategy to the top co-occurring tags in the list.

The voting strategy computes a score for each candidate tag, where a vote for that candidate is cast. A list of recommended tags is obtained by sorting the candidate tags on the number of votes. The summing strategy also takes the union of all candidate tag lists, and sums over the co-occurrence values of the tags.

Figure 4.7 showed that users do not only tag the visual contents of the photo, but to a large extent provide a broader context in which the photo was taken, such as, location, time, and actions. The tags being recommended, by our above strategy, and accepted by our users can now be analysed based on vote aggregation (summing) and promotion (voting). At first, we can see that the largest (most frequent) category in the Figure is 'Unclassified' at 48%.

WordNet	Acceptance ratio %
Unclassified	39 %
Location	71 %
Artifact or Object	61 %
Person or Group	33 %
Action or Event	51 %
Time	46 %
Other	53 %

Table 4.2: **Acceptance ratio of tags of different WordNet categories**

However, when voting is taken into account, where users select one or more of the tags recommended by our strategies, we can deduce that there exists a gap between user-defined and accepted tags for those tags which can not be classified using WordNet.

Table 4.2 shows the acceptance ratio for different WordNet categories. In the Table we can see that locations, artifacts, and objects have a relatively high acceptance ratio. However, people, groups and unclassified tags (tags that do not appear in WordNet) have relatively low acceptance ratio. We conclude that our system is particularly good at recommending additional location, artifact, and object tags.

4.6 Summary

In conclusions, we assert that our strategy of recommending tags is more effective than using only user defined tags. The strategy had a more positive effect on relevance and precision, particularly when the strategy is based on voting. Additionally, the strategy was particularly good at recommending locations, artifacts and objects, both in terms of volume and acceptance ratio.

Chapter 5

VALUE OF INFORMATION TAGS - ANALYSES

Objectives

- Discussing the issues with user generated image tags.
- Discussing the benchmarks used for the development of image tagging system.
- Discussing the richness of image tag values.
- Discussing the inherent problems with image tags.
- Discussing the usability of values within image tags.
- Discussing the threshold associated with image tags.

5.1 Introduction

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

In this chapter, a thorough examination of the problems that have plagued current image tagging techniques will be carried out. Once the problems are identified, an examination of the information richness contained within images will be carried out with the view to using such information to alleviate the concerns raised by the above problems.

The properties of the images will also be mapped to a set of defined property domains that utilises the resultant products of these sets to address the solutions used to describe the image being tagged.

5.2 Problems with Tags

It is assumed that the majority of the existing image tags contain 'noise' within their contents and that personal agendas, motivations and intentions during the user tagging process may be both beneficial and harmful to improving the efficiency and accuracy of the image tag recommendation and tag search.

Semantic noise (also known as 'red herring') ([111]) can be defined as tags and textual data that are associated with an image that may be interpreted in such a way so as to distract from the actual meaning of the intended purpose of the tag.

Table 5.2 lists some of the possible issues associated with image tags, as well as the values existing within the tags, which in some cases, can be used in tag searches and recommendations:

The items listed in the 'tag values' column will be grouped in section 5.3 according to their simplicity, ease of use, availability, universality and limitation or restriction.

tag problems		**tag values**
Missing tags		Tag accuracy, precision and specificity
Missing values		Tag age
Misspelled tags		Tag topicality
Irrelevant tags		Tag reuse value
Lack of integrity		Tag complexity
Ambiguous tags		Tag simplicity
Inconstant tags		Tag acceptance and popularity
Personal choice		Tag frequency
Erroneous entry		Tag trustworthiness
Duplicated tags		Tag length (number of words)
Rule violations		Tag economic value (free vs proprietary)
		Tag language
		Tag recall-ability success
		Tag searchability

Table 5.1: **Tag problems and tags**

5.2.1 Missing tags & the Semantic Loss:

New advances in technology has facilitated the creation, upload and annotation of photos with relative ease. A user can take a snapshot with almost any mobile device and immediately upload it to Flickr (or any other image sharing social site). However, the user is not obliged to tag the uploaded image. This is known as 'Semantic Loss', the consequence of which will make it near impossible for other users to find the image by using any of the available search methods.

5.2.2 Missing tag values:

Some photos may be annotated with the correct tag, but are missing other useful tags that would help to further identify the photo during search and comparison.

Figure 5.1: Hotel by a lake

Figure 5.1 has a single tag in Flickr, namely 'hotel'. However, a more useful set of tags would also include background, colour and action, such as 'sky', 'lake' or 'sea', 'blue' and 'calm'. The extra tags would allow other users to accurately find an image of a hotel with a view of a lake or a calm sea situated in a resort with clear blue sky.

5.2.3 Misspelled tags:

User generated raw tags may contain spelling errors, context errors, slang expressions or simply a regional variation of spelling. A simple search within

Flickr tags revealed spelling errors, such as 'the boys in balck', context errors such as 'Male box' insted of 'Mail Box', slang expressions such as 'Luv' instead of 'Love' and regional variations such as 'color vs colour'. Such errors can be overcome by this research's proposed solution of tag suggestion and can also include features such as auto word completion, spell checker and word suggestion. The issue of regional spelling variation would require the tagging application to match the user's location to the applicable list of suggested tags.

In some cases, the user may genuinely be unaware of the correct spelling of a word (or a name). An example is the correct spelling of one of the Benelux counties; 'Belguim' or 'Belgium'.

5.2.4 Erroneous personal choice:

Image annotators have the choice to erroneously express their own opinion when tagging a photo, which may or may not be of use to other users when an image search is performed or when an application compares the tagged image against other images. For example, a user may choose to label a mobile phone as 'Orange', which may not be useful for other users who wish to search for this image. A proposed solution may entail suggesting extra tags that further identify the image, such as 'Mobile', 'Phone' and 'Nokia'.

5.2.5 Irrelevant Tags:

Users may over-tag their photos with too many irrelevant tags that would render many of the image search results useless. Flickr allows a user to add up to a maximum of 75 tags per photo. Many users choose to add a large number of tags, even if the end result would cause the images' comparison and search results to be rendered inaccurate. As an example, a simple search for a house photo on Flickr reveals extra images such as 'prefab', 'faade', 'site', 'wood' and 'courtyard'. Ideally, the most suitably used images would be 'house', 'construction', 'architecture' and, maybe, 'prefab'.

5.2.6 Tag Integrity:

A user's decision to add a certain type of tag to a photo must be unimpaired and completely independent and free of any pressure, coercion or influence. For example, a user annotating a photo of a music festival or a music band must be free to express his/her opinion with additional tags such as 'bad' and 'uninspiring'. The organizers of the festival and the owners of the music's band record label should not be allowed to influence public opinion by manipulating the available list of suggested tags.

5.2.7 Ambiguous tags:

Many users tag their photo with words that are common and, thus, add little information to the value of the annotation.

Figure 5.2: Photos tagged ambiguously as Paris Hilton

Examples of ambiguity include *word-sense ambiguity* as shown in Figure 5.2 such as 'Paris Hilton'. A simple search on Flickr for such an expression returns, among other search results, two completely disparate images. One image shows the celebrity model Paris Hilton, while the other shows the Paris Hilton Hotel in Las Vegas. Other types of ambiguity include:

- *Geographic ambiguity*, such as 'Cambridge', which can return results for the city in both the United Kingdom and the USA.

- *Temporal ambiguity*, such as 'FA Cup', which could be any FA cup game or any FA cup final from the early 1900's till present.

- *Language ambiguity*, such as 'mist', which means fog in the English language and dung in the German language.

5.2.8 Inconsistent tags:

User generated annotations are highly personal and subjective and may cause some issues regarding inconsistent vocabularies. For example, the tag 'Paris, which has no visual content, may have been assigned to a photo of a hotel room, a Paris restaurant or the Eiffel tower.

Even if the 'Paris' tag had visual consistency as part of a set of tags, consistency can not be guaranteed across the tags assigned by multiple users. This is due to the fact that not all users share exactly the same thoughts on visual categories.

5.2.9 Erroneous tag entry:

Users may enter the wrong text when assigning a tag to a photo without realising. An example is when a user tags two photos from two different events and erroneously assigns a 'wedding' tag to 'birthday' photo and a 'birthday' tag to a 'wedding' photo.

5.2.10 Repeated, duplicated & plural tags:

The number of tags used to annotate an image may not necessarily indicate that users are being fairly thorough in the tagging process. Users may be, simply, adding the same tag (or its plural version) more than once to the same image.

Figure 5.3: A Photo with repeated and plural tags

In Figure 5.3, an image of a tree was annotated with a repeated 'Tree' tag twice and the repeated plural 'Trees' also twice. Such an action has the effect of generating 'semantic noise', i.e. textual tags that may affect the search and comparison of the image.

Tag repetition can be defined as assigning the same tag to two or more images. The issue of tag duplication occurs when a Flickr user uploads a large amount of images in a single session and chooses to assign the same set of tags to all of the images uploaded in that particular session. This will result in a large number of images all with a high number of tags. However, the tag sets assigned may not necessarily be appropriate for all of the images which were uploaded in that session meaning the large number of tags may hinder rather than help image searches as many of the tags may be irrelevant.

5.2.11 Rules violations:

Most image sharing web sites terms and conditions stipulate that uploaded images and their tags must not encourage discrimination or hatred based on gender, ethnicity, colour or disability. This also applies to images and tags that encourages violence and crimes. In general, the combination of a published photo and its tags must strive not to be provocative, revealing, controversial or offensive.

Figure 5.4: A controversial photo of Cock Fighting

An example of a controversial photo can be seen in Figure 5.4, where some users may find such an image to be deeply offensive.

The site moderators will remove such images through a variety of content filters, however, if this process fails, then the system being developed and used to rank and recommend image tags must filter such images from the final search results.

5.2.12 User' Background:

When analysing user generated image tags, the important points to consider are the user's professional, academic and cultural background.

- Professional photographers who annotate their own photographs will generate a much higher than average proportion of generally useful tags. This group of users correspond to A1-B2 of the classification categories in Table 5.3.

- Users who are experts within their domain knowledge will also annotate their photos with domain relative tags. Such a group of users will, generally, be aligned with A1-B2 of the classification categories in Table 5.3. However, in some cases where the motivation of the group is directed solely for the benefit of the domain expert, the group will be more aligned towards C1-C5 of the classification categories in Table 5.3.

- Users with higher education and professional qualifications may use a much richer vocabulary when annotating and describing their images. The classification and tag expansion groups will be annotated with highly descriptive terms that will render them to be very useful not only to follow highly educated users, but also to the community of image users as a whole.

  This group of users will also correspond to A1-B2 of the classification categories in Table 5.3.

- Tags generated by the general public are very likely to be aligned to C1-C5 of the classification categories in Table 5.3, as most of the tags will be useful only to the individual generating the tags or a narrow group of users who will use the tags to organise and denote ownership of said group.

  Such tags are also likely to contain spelling errors and may not contain any clear relationship to their parent photos and, hence, will also conform to D1-D2 of the classification categories in Table 5.3.

- Users of different cultural and ethnic background may annotate the same photo with different tags, depending on the depth of their grammatical and vocabulary's strength.

5.3 Value of Tags

Social annotators have an opportunity to add great value to the existing value of tags within the photos being added to social photo sharing sites and, hence, improve the ability of other users to search for these photos.

To appreciate the full value of the existing and added tags, it would be best to categorize their classes, which can be further grouped by user motivation [63].

The classes specified in Table 5.3 can be used establish the rules for a definitive benchmark of an image tagging system, as will be discussed in the next sub-section. This Table can be further extended by three types and four categories:

- General types of items, people and events such as 'Storm', 'Cat', 'Roof' or 'Winter'.
- Cultural knowledge of the background of the subject or the event of the image, such as 'God', 'the American Civil War' or 'the battle of Bosworth'.

Classification Category	**Tag Motivation**	**Usefulness**
A1	Tag generically identifies what image is 'of'.	
B1a	Tag specifically identifies what image is 'of'. (place names/events)	Useful to Flickr community as a whole.
B1b	Tag specifically identifies what image is 'of' (people/animals).	
B2	Tag identifies what image is 'about'.	
C1	Refining tag	
C2	Self-reference tag	
C3	Task-organising tag	
C4	Tag which denotes ownership	Useful only to individual/group
C5	Compound tag	
D1	Misspelling	
D2	Unable to determine relationship	

Table 5.2: **tag classification**

- Items representing the image as an icon. Examples include mythical creatures such as 'Dragon', symbolic representations such as 'Youth' and emotions such as 'Sadness'.

The four categories that are perpendicular to the types above are:

- Who: The tags represent the people or the objects within the photo, such as 'Arabs', 'Trees' or 'Unicorn'.

- What: This category describes the events or actions associated with the people of objects within the photo, such as 'Birth and Death' or

'The Industrial Revolution'.

- Where: This category represents and describes the photo (or its contents) as a location, such as 'London', 'Hell', 'Heaven' or 'beach'.

- When: This category describes periodic events as well as dates and times, such as 'Today', '1966' or 'the year of the Roses'.

The above categories and types can be further represented in Table:

	General	**Specific**	Abstract
Who	Types of people.	Named people or items	Mythical beings
What	General events or state of being.	Specific events Specific events	Emotions or abstractions
Where	Types of location.	Specific location	Place symbols
When	Cyclical time.	Specific period of time	Symbolised time

Table 5.3: **14 tag categories and types**

5.3.1 Establishing a Benchmark for the proficient development of an Image tagging system

To guarantee the success of the system being developed that would improve tag recommendation and search, a set of rules that identify what makes a successful tag must be identified.

The rules can be deployed as 'constraints' to enable the cleaning and cleansing of user generated of tags, in a similar process that has widely been used in data warehousing, where the data (or tags in this instance) are loaded in a staging area to facilitate their cleansing, before being loaded into the photos.

Simple and Easy

Tags must be simple and must represent the lowest level of granularity. For example, a tag that read 'Jaguar', will need extra tags to be more accurately identified as in Figure 5.5.

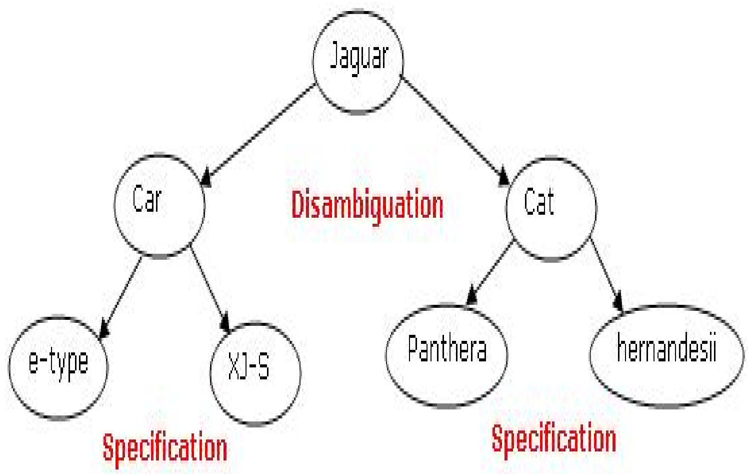


Figure 5.5: Jaguar Types

Once a user enters a photo tag that specifies a 'Jaguar', the tag recommendation mechanism would also suggest 'Cat' and 'Car'. If a user selects

'Cat', then a user has an option to further select from another recommended list of tags, such as 'panthera' and 'hernandesii'. Similarly, if a user selects 'Car', then a user has an option to further select from another recommended list of tags, such as 'e-type' and 'XJ-S'.

Highly Available - Universal

Image relevant tags must be highly available, such that they are listed higher up the tag suggestion list. Such a feature requires tags to be:

- Recent: Tags that have been used and referenced only a few times, a long time ago, may be either not listed or only listed at the bottom of the tag suggestion list and in the search engines. Examples are tags for countries that no longer exist and may no longer be used, such as 'Yugoslavia' which may have been used to tag images of 'Serbia', but is now no longer referenced. Ideally, the tag suggestion list should only list the most recently used (MRU), highest frequency tags.

- Reusable: Tags that are popular and are frequently used are highly reusable. Such tags must also have a high level of granularity and must not use generalised terms, which renders them less usable. Examples of such tags are 'Game console', which is not as reusable as the more specific tags of 'XBox', 'Nintendo' and 'PlayStation'.

- Trustworthy: Tags that have been generated and used by domain experts will have a high level of trustworthiness. Examples are tags that have been generated by professional photographers to tag their own uploaded photographs.

- Recallable: Tags that are easily remembered or have a 'catchy' name will always be among the first to be selected from a tag suggestion list.

- Searchable: In the realm of Search Engine Optimisation (SEO), "content is king" and such content must be relevant and recent. Tags that conform to these criteria will always be in the top of a list of a search engine and a tag suggestion list.

Free of Constraints

Images that are 'constrained' in any manner may dissuade users from selecting them. Hence, unconstrained tags must be:

- Impersonal: Tags must not be subjective or constrained by the personal views and preferences of the user generating them.

- Free: Tags must not be constrained by the high monetary or financial value placed on them by the users generating them. Ideally, to encourage re-usability, tags must either be free or have a low price constraint.

- Widely used language: The language used to generate the tag must be widely used by a large number of other users, who can both read and understand the context of the tag use. For example, a user who generates tags in the Flemish language, for images to be consumed by mainly Chinese users will face re-usability failure, due to the language barrier constraint.

- Topical: Finally, tag contents must not stray outside of the topic relevant to the images being tagged.

5.3.2 First experiment

The first experiment was the use of several similarity measures and combined together to arrive at a single Figure that may give an indication of how similar is one image is to another.

Five similarity measures were selected:

$\boldsymbol{Sm_1}$ = Jaccord

$\boldsymbol{Sm_2v}$ = Dice

$\boldsymbol{Sm_3}$ = Matching

$\boldsymbol{Sm_4}$ = Cosine

$\boldsymbol{Sm_5}$ = Overlap

N = Number of Tags

The Combined Similarity Measures can be expressed in the formula:

$$\boldsymbol{\Sigma_1^n \frac{(SM_1+SM_2+SM_3+SM_4+SM_5)}{N}}$$

This formula can be applied to the tags belonging to seven different objects to arrive at Table (5.4) below.

The main tag	Combined Similarity Measures
Tree	0.1847
Nice	0.0506
Apple	0.11883
Baby	0.1205
Car	0.023
Dog	0.378
Dress	0.210

Table 5.4: **Applying combined similarity measures**

In this experiment, it can be deduced that the use of the combined similarity measures by themselves is not enough, hence more meaningful measures must be applied for the results to be improved. In the next chapter, it will be demonstrated that the further weights applied to these initial results will yield improved tag comparison results.

5.4 Domains and threshold

The contents of an image can be though of as a full description written in prose, where the old adage "A picture is worth 1000 words". Alternatively, it might simply have a few keywords describing spatial, temporal, or emotional aspects [83].

To reduce the semantic gap, these few keywords can be organised into groups that can be used to work out the sets of domains for the image, then the product of the two sets can determine the proximity (or distance) of the returned image(s).

The global semantic description of an image can be represented as a logic composition of the different sets of image properties. This presentation would enable the user to pose queries in terms of natural language or by visual examples, then the system returns the semantically closest images to the query [68].

To illustrate this concept with a detailed example, consider the image in Figure 5.6, which is tagged as 'Leslie'. From the previous section that detailed the 'Problems with Tags', it can be deduced that the tag is highly ambiguous as it depicts a name that can be applied to the boy, the girl or to the pet dog.

Figure 5.6: A person or pet named Leslie

A value of a tag, generated for an image, can take a value between $/\mathbf{0}, \mathbf{1}/$, where a value that is close to one is considered to be accurately representing the image contents, while a value close to zero is less likely to be an inaccurate representation of the image contents.

Thus, in Figure 5.6, the value of the tag will be approximately 0.3, as the probability that the name 'Leslie' is earmarked for anyone of the three objects in the image, is one third of 1. In this case, a tag suggestion list or a search algorithm may ignore such an image, if the threshold for image similarity is set to 0.5

.

In contrast, each of the images in Figure are tagged as 'Leslie', and thus, the value of each of the three individual tags is 1. However, such tags may cause the wrong set of images to be retrieved as the search may specifically be looking for a girl named 'Leslie' and not a boy or a dog by the same name. To improve the returned search results, the values must be combined with the tags, as discussed below.

The concept of tag valuation can be represented by the formula:

$$\boldsymbol{T_d} \longrightarrow \boldsymbol{D_v}$$

Figure 5.7: All named Leslie

Where $\boldsymbol{T_d}$ represents a set of tags for domain d, and $\boldsymbol{D_v}$ represents a set of domain values. For example, consider a set of possible tags for the animal domain:

$\boldsymbol{T_{Animal} = \{cat, dog, horse, sheep, goat, rabbit\}}$

The set of values for any image containing one of the animals in the set would be:

$\boldsymbol{D_v = \{high, low, medium\}}$ - such that the value:

$\boldsymbol{T_{animal} \longrightarrow D_{value}}$

Thus, the image of the cat in Figure 5.8, can be represented as:

$\boldsymbol{cat \longrightarrow high}$

Figure 5.8: Cat with high tag weighting

The above example does not apply to image objects only, but also to image background and colour. Consider the example of $\boldsymbol{T_{rose}}$, which is a set of tags of roses 5.9. The set of tag values for the roses tag would be:

$$\boldsymbol{D_{vroses} = \{red, pink, purple, yellow, white, ...\}}$$

Figure 5.9: Red Rose

In this case, using natural language and visual queries, the value of:

$\boldsymbol{T_{rose}} \longrightarrow \boldsymbol{D_{rose}}$, which equates to:

$\boldsymbol{rose} \longrightarrow \boldsymbol{red}$

Further, to compare the tag similarity between two images, two sets of tag values can be multiplied to deduce the probability that the two images are similar. Consider the example where one image has the tag values:

$\boldsymbol{D_v} = \{\boldsymbol{dog}, \boldsymbol{game}, \boldsymbol{TV}\}$

Figure 5.10: Dogs playing Nintendo and Eating

While the second image has the tag values:

$\boldsymbol{D_v} = \{\boldsymbol{dog}, \boldsymbol{bowl}, \boldsymbol{food}, \boldsymbol{biscuits}\}$

A Cartesian product of the two sets of tag values can be used to work out the similarities, such that:

$$D_{dogplay} \text{ x } D_{dogfood} = \{(dog, dog), (dog, bowl), (dog, food), (dog, biscuits), (game, dog), (game, bowl), (game, food), (game, biscuits), (TV, dog), (TV, bowl), (TV, food), (TV, biscuits)\}$$

The resultant product set has value such as (dog,dog), (dog,bowl), (dog,food) and (dog,biscuits), which indicates that the two images have a high degree of similarity and will, therefore, increase the accuracy of the image search.

Additionally, other domain values can be used instead of colour. For most images, the set of domain values to be used may contain

$\{object, colour, background, fuzziness, ...\}$.

Ideally, 'Action' must also be included, but may be difficult to implement.

5.5 Conclusions

This chapter has asserted that tags that are highly reusable, simple, easy to use, recent and relevant, will have a rich set of values inherent within their contents.

Highly valuable tags must be resistant to noise, topical, trustworthy, economical, searchable, recallable, popular, unique, non-ambiguous and precise.

Further, tags must have a high level of granularity, be error free and must not violate the rules set by the assigned tagging community of the image sharing sites.

Additionally, tag values that are generated as a result of the Cartesian product of two sets of values, must equate to a value that is higher than the threshold set by the search algorithm of the similarity method.

This chapter has also asserted that existing image tags may have very rich value, which lies hitherto, unused. Ideally, such valuable tags must be extracted to a set of values, of which the resulting product can be used to work out the numerical accuracy of the searched images.

Chapter 6

WordNet-based approach

*O*bjectives:

- Provide an integrated approach to accuracy improvement based on WordNet.
- Provide a comparative analyses of search results.
- Provide a second experiment with improved results.

6.1 Introduction

In chapter 5, the initial experiment developed to ascertain the similarity between two images returned rather limited results that will need to be improved on and expanded in this chapter. The fact of the matter is that the astronomical number of current images uploaded to internet sites means that similarity search methods must be as accurate as they can be in order to return nearly acceptable results. Such "difficult to imagine" figures dictates that every effort must be made to use the most effective approaches to improve the accuracy of photo tags and, thus, the accuracy of image search.

6.2 Corpuses/images

Images are naturally rich in information, even before tags are added to them. The richness comes from the fact that visual objects and features can be categorised in what is termed as bag-of-features or Bag of Words representations for image object recognition. Such a technique has been studied by [7] [75] [15] [65]. In one technique Wu et al [39], proposed improving these techniques by limiting the loss of spatial information, where the word representing a visual image may contain multiple semantic meanings and the same semantic meaning may be represented by multiple visual words each representing a single visual image.

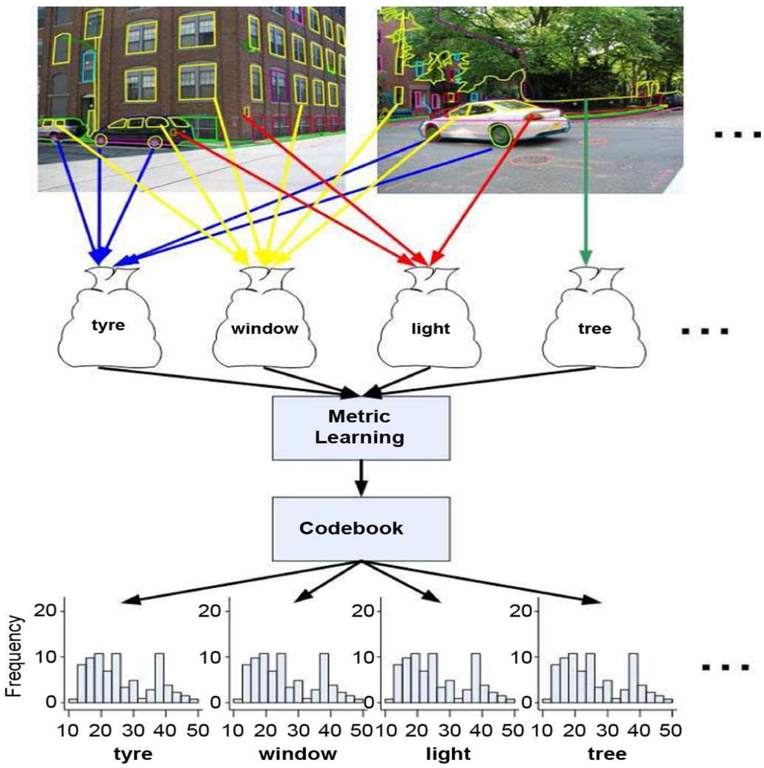


Figure 6.1: Process of building the semantics-preserving bag-of-words model

In Figure 6.1, objects within the photo are segmented, categorised and bagged by user tags. The Bag of Words model used in the process outlined in Figure 6.1 is an approach that attempts to resolve object recognition/image annotation problems by deriving it from natural language processing [58]. This model maps each visual feature (that is of interest) within the image to codewords that are grouped to generate a codebook. Models can then be employed, which considers the distance between the semantically identical features as a measurement of the semantic gap and, therefore, tries to learn a codebook.

Referring to Figure 6.1 again, the Metric Learning is a training process where segmented image objects have their features extracted to represent these objects. Following that, the frequency of the features (or tags) are calculated and used as a weighting figure for other models.

Despite the drawbacks of the Bag of Words model, as discussed by [13] [65] [90] [107] [57], the codebook learning proposed by this model has been successfully employed to overcome the limitation of semantics lost within earlier models.

Object categories that are mapped to bags of words can also be extended to every day common items, such as names, food, animals, habits, vehicles, jobs, schools, houses, habitats, colours and clothes.

6.3 Approach

6.3.1 Applied weight:

In this first approach a weight measure is applied to the results achieved from the experiments in chapter 5 (5). Recall that in chapter 3, six similarity measures where identified, of which only five were used by the initial experiment in chapter 5, namely Jaccard coefficient, Dice, Matching, Overlap coefficient and Cosine.

The applied weight measure [101] is derived form the similarities of the four objects associated with each image, namely 'primary object', 'secondary object', 'action' and 'colour'.

Let each similarity measure be $\boldsymbol{SM_n}$, such that the five utilised measure would be $\boldsymbol{SM_1}$ .. $\boldsymbol{SM_5}$. Also, let the total number of tags for the compared photos be $\boldsymbol{N}$, such that:

$\boldsymbol{SM_1}$ = Jaccord

$\boldsymbol{SM_2}$ = Dice

$\boldsymbol{SM_3}$ = Matchibg

$\boldsymbol{SM_4}$ = Cosine

$\boldsymbol{SM_5}$ = Overlap

$\boldsymbol{\beta} = \frac{\boldsymbol{1}}{\boldsymbol{N}}$

If the applied weight is $\boldsymbol{\alpha}$, then the equation for the previous chapter 5:

$$\boldsymbol{Algorithm_1} = \frac{\boldsymbol{(SM_1+SM_2+SM_3+SM_4+SM_5)}}{\boldsymbol{N}}$$

Can be modified to:

$(\beta SM_1 + \beta SM_2 + \beta SM_3 + \beta SM_4 + \beta SM_5)$

Thus, applying the weight to the equation:

$Algorithm_1 = (\beta SM_1 + \alpha\beta SM_2 + \beta SM_3 + \alpha\beta SM_4 + \beta SM_5)$

OR,

$Algorithm_1 = \frac{1}{n} * \Sigma_1^n SM_1 + \alpha SM_2 + SM_3 + \alpha SM_4 + SM_5$

This "weighting centric equation" can be used to work out the third column in the table listed in the next section. The third column formula ac be represented as:

$\Sigma_{i=1}^n W_i * SM_i$

The applied weight can take any value in the range:

$1 \leq \alpha \leq 0$

In the experiment discussed in Chapter 5, any results that yielded 0.5 or above will be considered successful, i.e. the images have adequate or high degree of similarities between them. However, results that return less then 0.5 will have the weight applied to them.

In a first example, consider the six tree images in Figure 6.2 below. Each image is labelled starting from $\boldsymbol{image_1}$ onward to $\boldsymbol{image_6}$ in a clockwise fashion. Table 6.1 displays the tags belonging to each image:

Figure 6.2: Photos of trees from Flickr

When each of the five similarity measures above were applied to yield the comparison between $\boldsymbol{image_1}$ and $\boldsymbol{image_2}$, the following results where obtained:

$\boldsymbol{SM_1}(\boldsymbol{image_1}, \boldsymbol{image_2}) = 0.5$

Image	Tag from Flickr
$\boldsymbol{image_1}$	Tree, Bushes, still, green.
$\boldsymbol{image_2}$	Tree, deer, blue, sky, grass, green, UK, Canon 350D, canon, landscape, polariser, favourite, field, fields, Dyrham Park, National Trust.
$\boldsymbol{image_3}$	Tree, green, grass
$\boldsymbol{image_4}$	Tree, lake, water
$\boldsymbol{image_5}$	Tree, lake
$\boldsymbol{image_6}$	Trees

Table 6.1: **Tags for each image in Figure 6.2**

$$\boldsymbol{SM_2}(\boldsymbol{image_1}, \boldsymbol{image_2}) = 0.136$$

$$\boldsymbol{SM_3}(\boldsymbol{image_1}, \boldsymbol{image_2}) = 3.0$$

$$\boldsymbol{SM_4}(\boldsymbol{image_1}, \boldsymbol{image_2}) = 0.306$$

$$\boldsymbol{SM_5}(\boldsymbol{image_1}, \boldsymbol{image_2}) = 0.5$$

The result of applying $\boldsymbol{Algorithm_1}$ to the above Figures yields:

$$\boldsymbol{Algorithm_1} = \frac{\mathbf{0.5+0.136+3.0+0.306+0.5}}{\mathbf{18}} = 0.24$$

However, when applying the weight measure to the results from $\boldsymbol{Algorithm_1}$, the following result is returned:

$$\boldsymbol{Algorithm_2} = \frac{\boldsymbol{SM_1 + \alpha SM_2 + SM_3 + \alpha SM_4 + SM_5}}{\boldsymbol{18}} = 0.36$$

Applying the weight measure to $\boldsymbol{Algorithm_1}$ has improved the similarity between $\boldsymbol{image_1}$ and $\boldsymbol{image_2}$ from 0.24 to 0.36, and such an approach can be applied to improve the results of the similarity measures between the other images. The improvement in results can be seen in table 6.2 where the similarity measure between each of the images is shown to improve when a weighting measures is applied to the results yielded from the first algorithm.

Compared Images	Similarity measures ($\boldsymbol{algorithm_1}$)	Similarity measures ($\boldsymbol{algorithm_2}$)
$\boldsymbol{image_1, image_2}$	0.24	0.36
$\boldsymbol{image_2, image_3}$	0.99	0.99
$\boldsymbol{image_3, image_4}$	0.89	0.89
$\boldsymbol{image_4, image_5}$	0.37	0.89
$\boldsymbol{image_5, image_6}$	0.30	0.45

Table 6.2: **Similarity measures results (with and without applied weights)**

The table shows that where any results were below 0.5, then when the weight was applied, the final result of the similarity measure was significantly improved.

In a second example, consider the six baby images in Figure 6.3. Each image is labelled $\boldsymbol{baby_1}$ onward to $\boldsymbol{baby_6}$ in a clockwise fashion. Table 6.3 displays the tags belonging to each image:

Figure 6.3: Photos of babies from Flickr

When each of the five similarity measures above were applied to yield the comparison between $\boldsymbol{baby_1}$ and $\boldsymbol{baby_1}$, the following results where obtained:

Baby Image	Tag from Flickr
$\boldsymbol{baby_1}$	Baby, India, wrap, hair, daughter.
$\boldsymbol{baby_2}$	Baby, cute, newborn, babies, basket.
$\boldsymbol{baby_3}$	Babies.
$\boldsymbol{baby_4}$	Baby.
$\boldsymbol{baby_5}$	Baby , open mouth, eyes, portrait, baby portrait, geotagged, round, head, nares, nostrils, eat, bib, babies, eyelashes, round head.
$\boldsymbol{baby_6}$	I love you, baby, girl, daughter, NADIA, canon, HELLO KITTY, precious gift.

Table 6.3: **Tags for each image in Figure 6.2**

$$\boldsymbol{SM1}(\boldsymbol{baby_1}, \boldsymbol{baby_1}) = 0.11$$

$$\boldsymbol{SM2}(\boldsymbol{baby_1}, \boldsymbol{baby_1}) = 0.2$$

$$\boldsymbol{SM3}(\boldsymbol{baby_1}, \boldsymbol{baby_1}) = 1.0$$

$$\boldsymbol{SM4}(\boldsymbol{baby_1}, \boldsymbol{baby_1}) = 0.2$$

$$\boldsymbol{SM5}(\boldsymbol{baby_1}, \boldsymbol{baby_1}) = 0.2$$

Again, applying $\boldsymbol{Algorithm_1}$ to the above five similarity measures yields:

$$\boldsymbol{Algorithm_1} = \frac{\mathbf{0.11+0.2+1.0+0.2+0.2}}{\mathbf{9}} = 0.19$$

However, when applying the weight measure to the results from $\boldsymbol{Algorithm_1}$, the following result is returned:

$$\boldsymbol{Algorithm_2} = \boldsymbol{\Sigma_{i=1}^{5} \beta * SM_i} = 0.45$$

In this second example, it can be seen that applying the weight measure to $\boldsymbol{Algorithm_1}$ has improved the similarity between $\boldsymbol{baby_1}$ and $\boldsymbol{baby_2}$ from 0.19 to 0.45, which is a significant improvement. Again, such an approach can be applied to improve the results of the similarity measures between the other baby images. The improvement in results can be seen in table 6.4 where the similarity measure between each of the baby images is shown to improve when a weighting measures is applied to the results yielded from the first algorithm.

6.3.2 Applied WordNet:

In the previous approach, applying weight measure to the results extracted from the similarity measures improved the accuracy of the similarity search. However, to add even extra meaning

Compared baby Images	Similarity measures ($\boldsymbol{algorithm_1}$)	Similarity measures ($\boldsymbol{algorithm_2}$)
$\boldsymbol{baby_1, baby_2}$	0.19	0.45
$\boldsymbol{baby_2, baby_3}$	0.14	0.42
$\boldsymbol{baby_3, baby_4}$	0.50	0.99
$\boldsymbol{baby_4, baby_5}$	0.27	0.54
$\boldsymbol{baby_5, baby_6}$	0.02	0.24
$\boldsymbol{baby_1, baby_6}$	0.36	0.56

Table 6.4: **Baby similarity measures results (with and without applied weights)**

to the tags, WordNet cognitive synonym sets (synsets) are applied on top of the weight measure. The synsets are applied to the aforementioned properties of image tags, namely Primary object, Secondary object, Action and Colour.

Recall that the Jaccard similarity measure between two images in terms of set of words can be defined as,

J = $\frac{\boldsymbol{A \bigcap B}}{\boldsymbol{A \bigcup B}}$

Where A=$\boldsymbol{WA_1}$, $\boldsymbol{WA_2}$, $\boldsymbol{WA_3}$, ...$\boldsymbol{WA_n}$ and B=$\boldsymbol{WB_1}$, $\boldsymbol{WB_2}$, $\boldsymbol{WB_3}$, ...$\boldsymbol{WB_n}$

Referring back to Figure 6.2, table 6.5 below lists the values of each of the four properties (Primary object, Secondary object, Action and Colour) for each image within the Figure.

Image	Tag Values
$\boldsymbol{image_1}$	Tree, Bushes, Still, Green.
$\boldsymbol{image_2}$	Tree, Sky, Still, Blue.
$\boldsymbol{image_3}$	Lake, Sky, Dusk, Red
$\boldsymbol{image_4}$	Mountain, Sea, Calm, Blue
$\boldsymbol{image_5}$	Tree, Sea, Windy, White
$\boldsymbol{image_6}$	Tree, Forest, Autumn, Brown

Table 6.5: **Tags values for each image's property in Figure 6.2**

To apply the Jaccard similarity measure to compare $\boldsymbol{image_1}$ and $\boldsymbol{image_2}$, the four tags for $\boldsymbol{image_1}$ are:

$\boldsymbol{image_1}$ = {Tree, Bushes, Still, Green}

and for $\boldsymbol{image_2}$:

$\boldsymbol{image_2}$ = {Tree, sky, still, blue}, hence:

$\boldsymbol{image_1} \cap \boldsymbol{image_2}$ = {Tree, still}, and:

$\boldsymbol{image_1} \cup \boldsymbol{image_2}$ = {Tree, still, pushes, green, sky, blue}

Now the Synonym for each of these tags can be extracted from WordNet and grouped in table 6.6:

Words	No. of Word Synonyms
Tree	8
Bushes	27
Still	47
Green	30
Sky	4
Blue	53

Table 6.6: **Number of word synonyms**

Now, employing Jaccard by utilising the tags and the numbers in table 6.6:

Jaccard $= \frac{\boldsymbol{Tree+Still}}{\boldsymbol{Tree+Bushes+Still+Green+Sky+Blue}}$, hence:

Jaccard $= \frac{\frac{1}{8}+\frac{1}{47}}{\frac{1}{8}+\frac{1}{27}+\frac{1}{47}+\frac{1}{30}+\frac{1}{4}+\frac{1}{53}} = 0.33$

The synonyms for each word tag are extracted from the Wordnet search browser, as shown in Figure 6.4 for the 'blue' tag and in Figure 6.5 for the 'baby' tag. The browser displays the senses and synonyms for each of the searched words.

Again, in a second example that applies WordNet on top of the similarity measures, and utilising the six baby images 6.3, table 6.5 below lists the values of each of the four properties (Primary object, Secondary object, Action and Colour) for each image in the Figure.

Image	Tag Values
$\boldsymbol{baby_1}$	baby, mother, hug, black.
$\boldsymbol{baby_2}$	baby, basket, sleep, white.
$\boldsymbol{baby_3}$	baby, floor, sleep, white.
$\boldsymbol{baby_4}$	baby, bed, smiling, pink.
$\boldsymbol{baby_5}$	baby, mouth, open, white.
$\boldsymbol{baby_6}$	baby, posing, resting, red.

Table 6.7: **Tags values for each image's property in Figure 6.3**

To apply the Jaccard similarity measure to compare the images names $\boldsymbol{baby_4}$ and $\boldsymbol{baby_5}$, the four tags for $\boldsymbol{baby_1}$ are:

$\boldsymbol{baby_1}$ = {baby, bed, smiling, pink}

And the four tags for $\boldsymbol{baby_5}$ are:

$\boldsymbol{baby_5}$ = {baby, mouth, open, white}, hence:

$\boldsymbol{baby_4} \cap \boldsymbol{baby_5}$ = {Baby}, and:

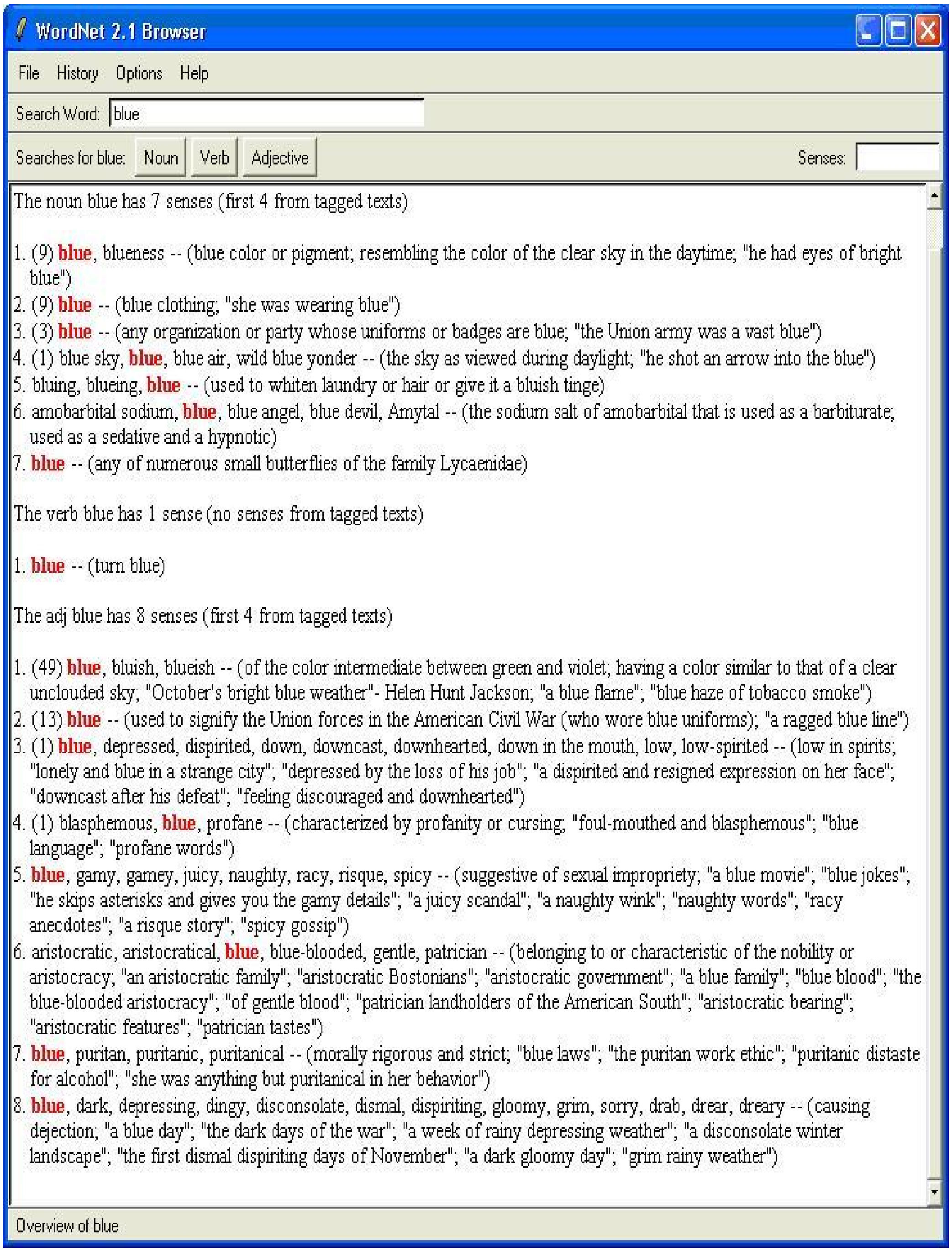


Figure 6.4: WordNet Search Browser

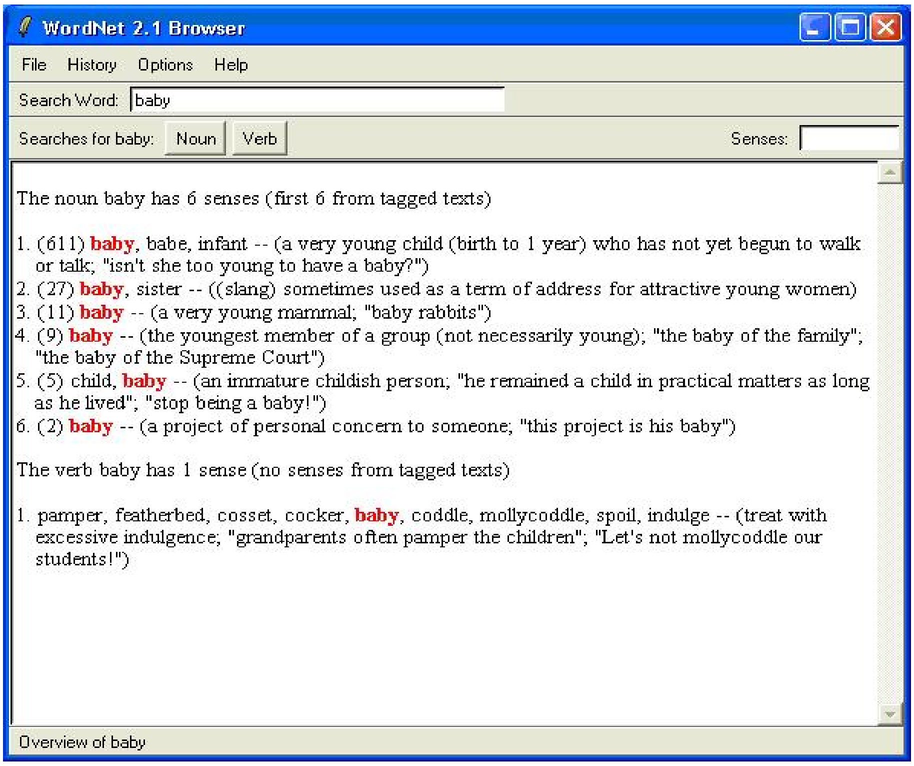


Figure 6.5: WordNet Search Browser for baby

$\boldsymbol{baby_4} \cup \boldsymbol{baby_5}$ = {Baby, bed, smiling, pink , mouth, open, white}

Now the Synonym for each of these tags can be extracted from WordNet and grouped in table 6.8:

Words	No. of Word Synonyms
Baby	13
Bed	43
Smiling	8
Pink	10
Mouth	22
Open	46
White	39

Table 6.8: **Number of baby word synonyms**

Now, employing Jaccard by utilising the tags and the numbers in table 6.8:

Jaccard $= \frac{\boldsymbol{Baby}}{\boldsymbol{Baby+Bed+Smiling+Pink+Mouth+Open+White}}$, hence:

Jaccard $= \frac{\frac{1}{13}}{\frac{1}{13}+\frac{1}{43}+\frac{1}{8}+\frac{1}{10}+\frac{1}{22}+\frac{1}{46}+\frac{1}{39}} = 0.18$

6.4 Results and analysis

The results in the previous section confirm the increased confidence in using a combination of similarity measures, then applying a weight to the results, if the value is below 0.5, and finally using a combination of WordNet synonyms to further improve the similarity result between any two compared images.

In Appendix 7.3, Table 7.3 further enforces the conclusion of this thesis, in that the application of the Combined Similarity Measure alone is neither accurate nor sufficient to yield viable similarity results.

During the research for this thesis, 130 images were analysed to return the sample results in Table 7.3. These results were further employed (7.3) to return a subset of relevant images, which were further reduced to produce a smaller list of recommended images.

To further improve the comparisons of the analysed images, the f-measure was used which relies on precision and recall values. The formulae for precision and recall can be expressed as:

$$\boldsymbol{precision} = \frac{\textit{no. of correctly recommended images}}{\textit{no. of recommended images}}$$

$$\boldsymbol{recall} = \frac{\textit{no. of correctly recommended images}}{\textit{no. of relevant images}}$$

The f-measure can be calculated from the above two formulae as follows:

$$\boldsymbol{f - measure} = \frac{\mathbf{2} \times precision \times recall}{precision + recall}$$

The f-measure was used in this research to carry out statistical analysis on the collected data. The results in scoring terms is highest around 0.9 and lowest at around 0.2 (Table 6.9). The f-measure also decreases as the scoring results increases (Figure 6.10). Hence the results of the f-measure are more promising when 60 to 90 images are analysed by using precision and recall.

To further distinguish between recommended and relevant images, consider the example images below. In this example, recommended images are required for an image of a baby and an image of a tree. By using the WordNet approach, a similarity of 0.7 was obtained for the two recommended images (6.6 and 6.7).

Figure 6.6: recommended tree image

Figure 6.7: recommended baby image

However, 'relevant' rather than 'recommended' images, usually have a similarity of $\leq$ **0.5**, where the actual similarity score for the image and image were 0.3

Figure 6.8: Relevant tree image

Figure 6.9: Relevant baby image

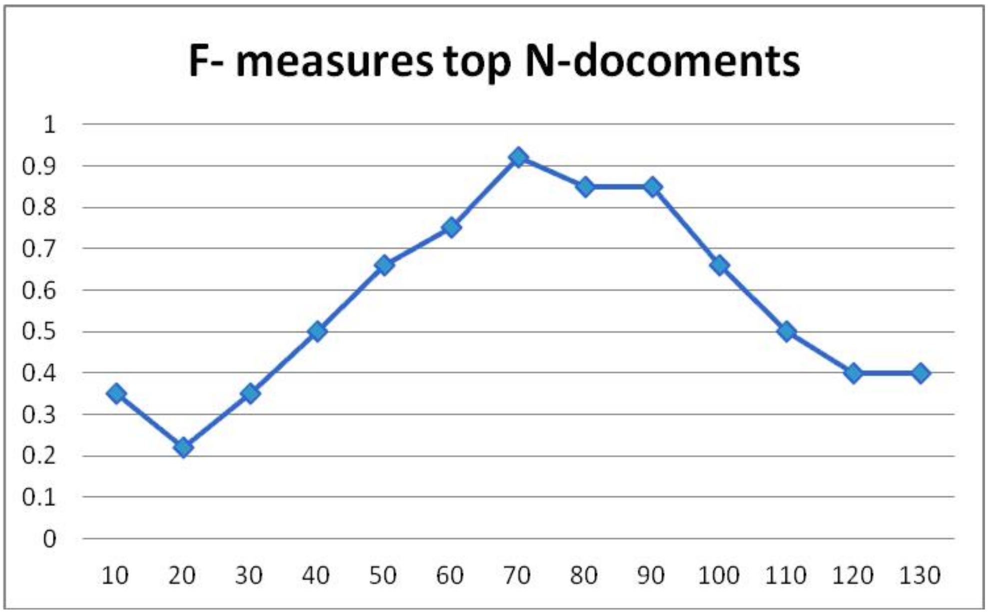


Figure 6.10: f-measure for the analysed data

No of images	F-measures
10	0.35
20	0.22
30	0.35
40	0.50
50	0.66
60	0.75
70	0.92
80	0.85
90	0.85
100	0.66
110	0.50
120	0.40
130	0.40

Table 6.9: **Table of f-measures for the analysed data**

However if we raise the threshold from 0.5 to 0.6 and 0.8 we obtain the following f-measures, respectively.

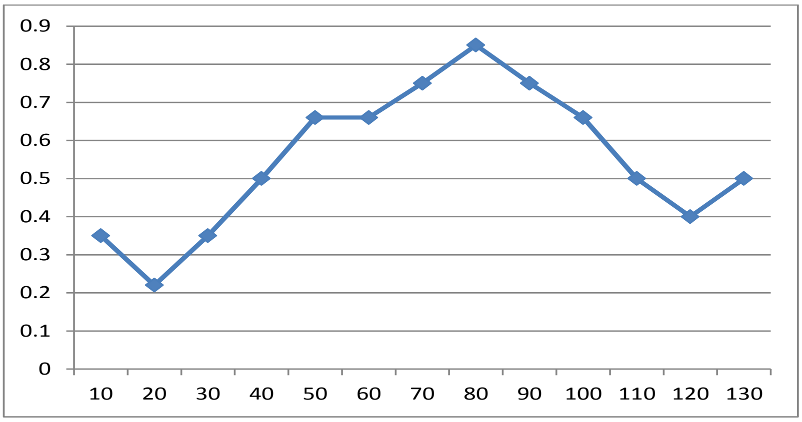


Figure 6.12: f-measure for the analysed data at 0.6 threshold

No of images	F-Measures
10	0.35
20	0.22
30	0.35
40	0.5
50	0.66
60	0.66
70	0.75
80	0.85
90	0.75
100	0.66
110	0.5
120	0.4
130	0.5

Figure 6.11: F-measure for the analysed data at 0.6 threshold

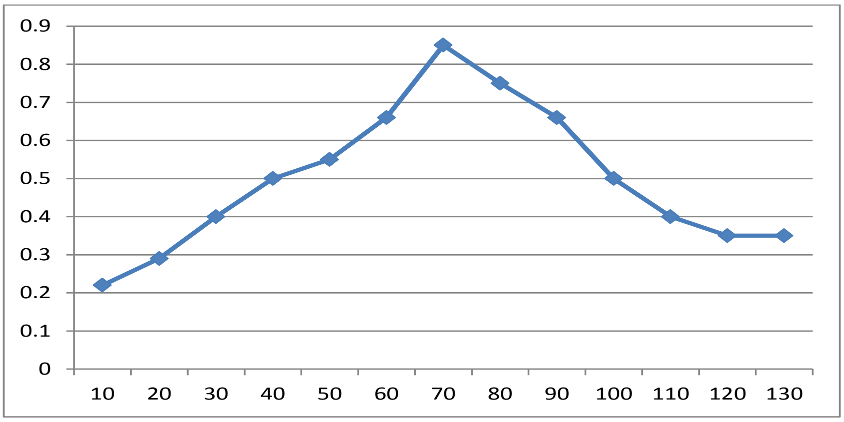


Figure 6.13: f-measure for the analysed data at 0.8 threshold

No of images	F-Measures
10	0.22
20	0.29
30	0.4
40	0.55
50	0.5
60	0.60
70	0.85
80	0.75
90	0.66
100	0.5
110	0.4
120	0.35
130	0.35

Figure 6.14: F-measure for the analysed data at 0.8 threshold

From a careful analysis of the two F-measures, Figure(6.4) and Figure (6.4), we observe the close similarity of both compared with Figure (6.10). These results conclude that the threshold of 0.5 is an optimal one and that the dataset is indeed statistically significant.

6.5 Summary

In this chapter, further confirmation was detailed about the challenging and astronomical sizes of images uploaded to social and image sharing sites. Such challenges dictate that every effort must be made to improve the accuracy of image search results.

Among the solutions discussed were the 'Bag of Words' model, as well as the application of a combination of similarity measures, followed by the application of a pre-determined weight measure and finally applying an algorithm containing WordNet's synonyms to the final results.

Many examples were analysed in this chapter, including a detailed workout of how the various algorithms were applied to each of the initial results. The final results further increased the confidence in the accuracy of the recommended algorithms.

Chapter 7

CONCLUSION AND FUTURE WORK

Objectives

- Provide a conclusion to this research.
- Provide a summary of this research's achievement.
- Discuss future research areas and direction resulting from this research.

7.1 Introduction

The enormity of the challenge faced by this research can be best explained in figure 7.1 ([79]) which is further complicated by Facebook's recent purchase and merger with Instagram [43].

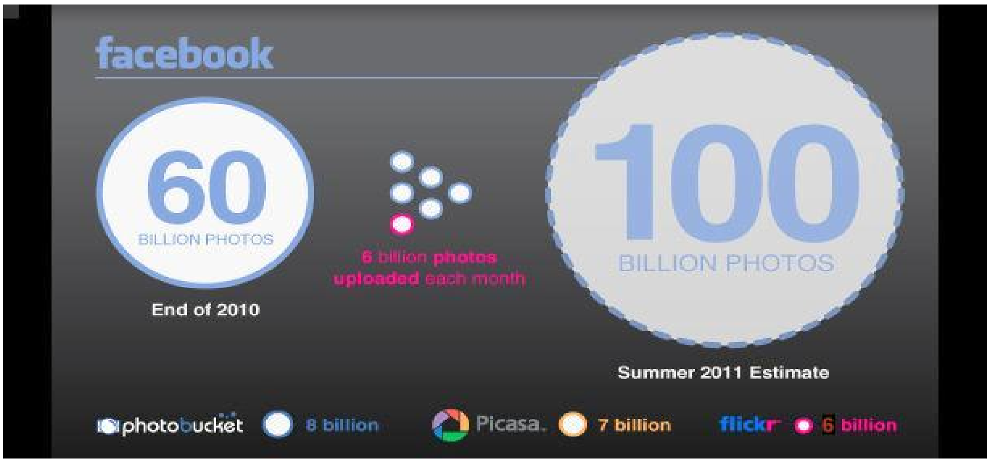


Figure 7.1: The growth of images on Facebook

This buyout by Facebook has significantly increased the numbers of uploaded images as Instagram is now growing faster than Flickr [16], and is expected to further bloat the total number of images uploaded into Facebook.

The above numbers will further highlight the importance of the dissertation's research questions, namely 'why are the current images on photo sharing web sites inaccurately tagged?' and 'How to improve the accuracy of Information Retrieval from existing images?'.

The answer to the first question can best be illustrated by a look at the purpose and motivation behind using Instagram.

Instagram allows users to take pictures, then enhance them by applying a digital filter, before sharing them on a variety of social networking such as Facebook and Twitter. Instagram will not force its users to tag their uploaded photos, neither will it enforce any rules or algorithms on users who do choose to tag their photos.

This action, or lack of it, by Instagram and other photo sharing social sites, enforces this research's motivation to solve the issues arising from the lack of tags, and where tags exist, tag ambiguity, misspelling, shorthand writing, slang and abbreviated words.

7.2 Summary of research achievement

Once the literature review for semantically enhancing tagged images was completed, the dissertation successfully presented the currently utilised similarity measures, the methods of metadata generation and the current measures of image relatedness.

We have carried out three experiments for the purpose of evaluation and validation. In the first experiment we utilised the richness and variety of information embedded within an image (also known as Value of Information), thus increase their accuracy, as explained in Figure (7.2).

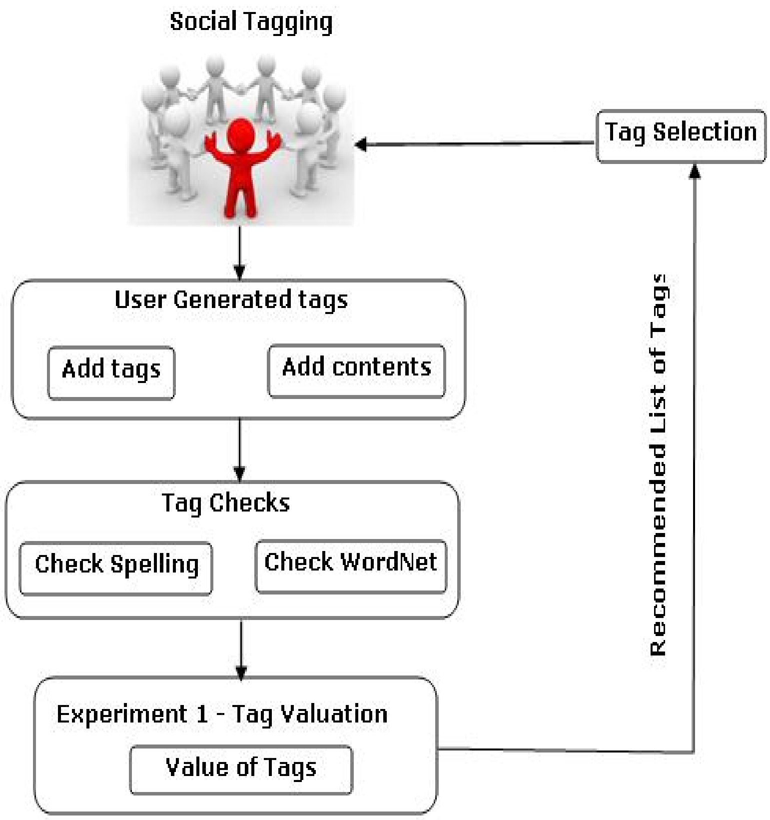
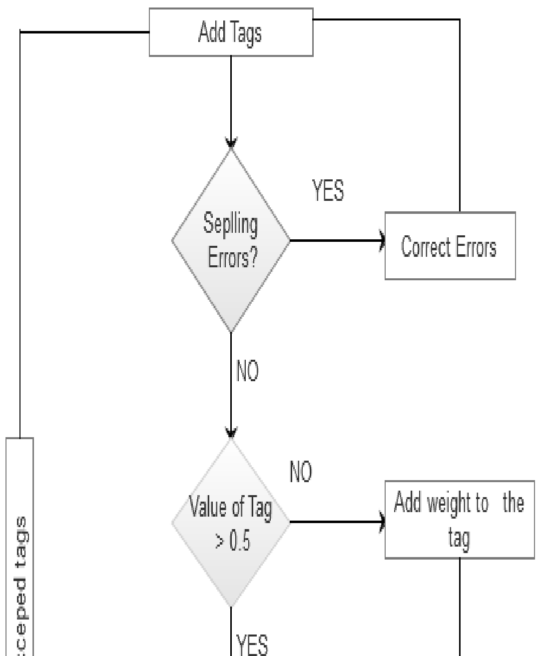


Figure 7.2: Value of Tags (Experiment 1)



The various similarity measures and weighting techniques have been employed to increase the accuracy of image tags, as depicted in Figure (7.4).

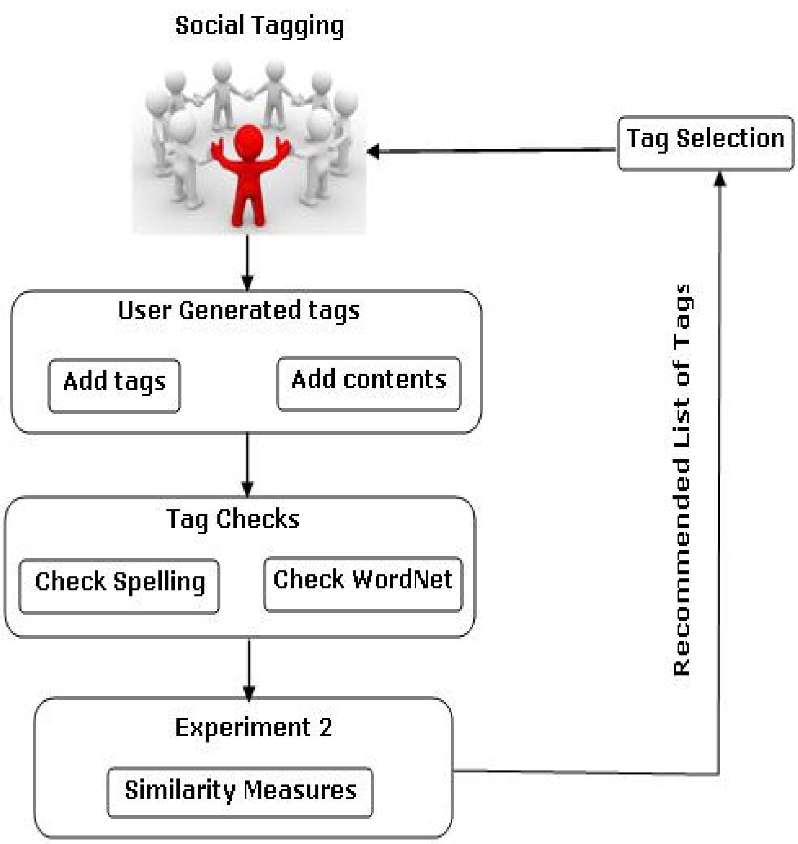
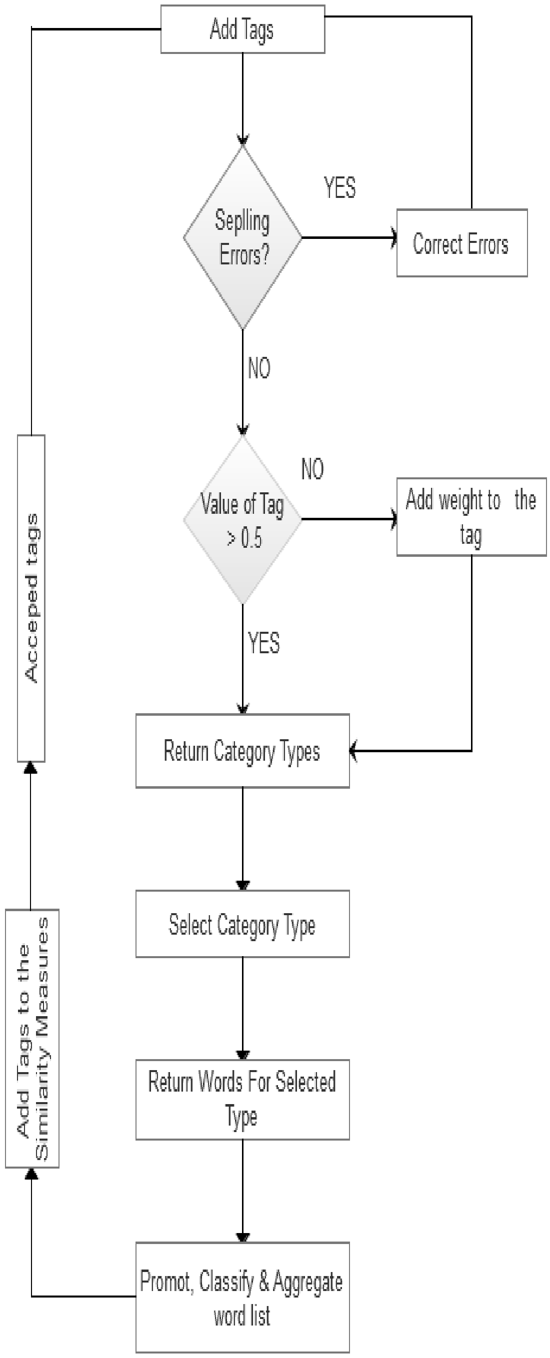


Figure 7.4: Similarity Measures (Experiment 2)


Add Tags
Seplling Errors?
YES
Correct Errors
NO
Value of Tag > 0.5
NO
Add weight to the tag
YES
Return Category Types
Select Category Type
Return Words For Selected Type
Promot, Classify & Aggregate word list
Add Tags to the Similarity Measures
Acceped tags

And finally, all the methods and resources listed above (or n-dimensions) have been used as a single unified process to improve the accuracy of the tags, as explained in Figure (7.6).

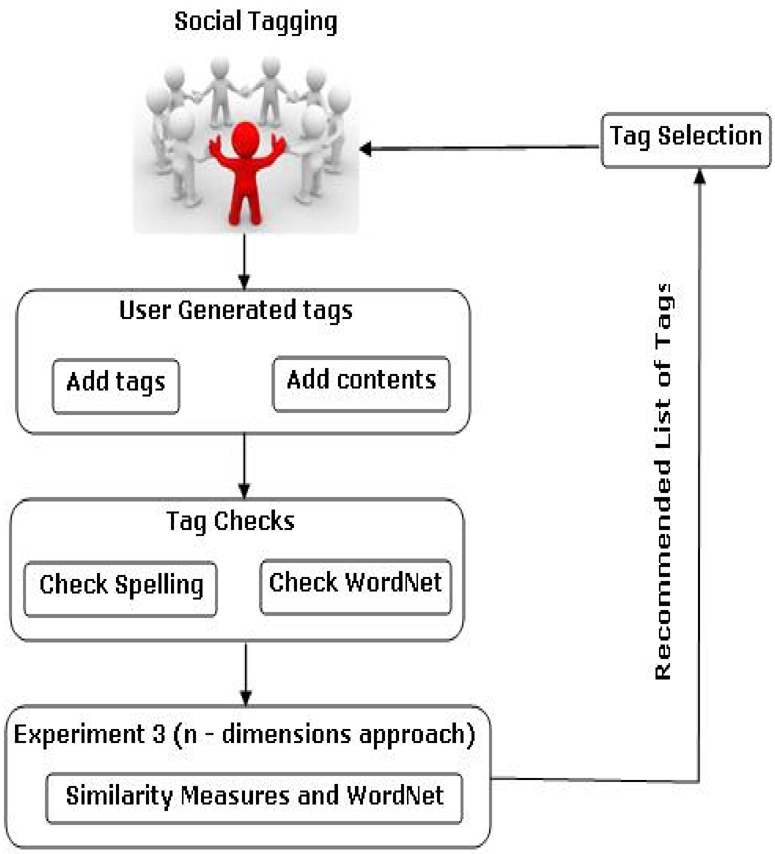
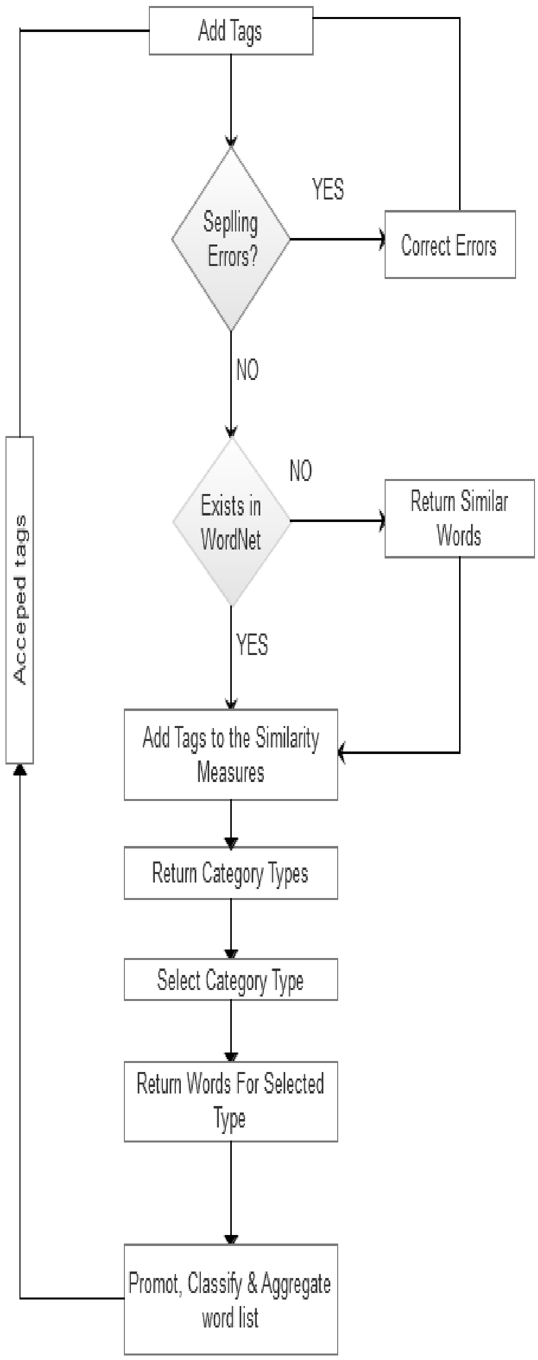


Figure 7.6: n dimensions (Experiment 3)


Add Tags
Seplling Errors?
YES
Correct Errors
NO
Exists in WordNet
NO
Return Similar Words
YES
Add Tags to the Similarity Measures
Return Category Types
Select Category Type
Return Words For Selected Type
Promot, Classify & Aggregate word list
Acceped tags

The research was successful in highlighting the various equations and methods of tag clouds, tag clustering and tag classification that are currently used in industry and in academic research methodology. Additionally, the various types of Similarity Measures that are used to measures the distance (or similarity) between the sets of attributes of two images were also outlined. These similarity measures included the Jaccard coefficient, Dice, Matching, Overlap coefficient, Mutual Information and Cosine coefficient.

Following this, a look at the main challenges of social tagging and their effects on usage statistics and on the extraction of relevance information was also completed. These included users' culture, Motivation, Domain knowledge, Semantic loss and Vocabulary, Diversity of opinion, Independence, Decentralization, Aggregation, Expertise, Reputation and Reliability.

Next, the research looked into the ways in which image tagging and text tagging differ and the logical flow of data between the components of a social tagging recommendation system. There was also a comparison between Symmetric and Asymmetric measures and the criteria governing the value of information added to tags, such as Popularity, Topicality, Uniqueness, redundancy, Simplicity, Spelling accuracy and Recency. A look at the classification of tags by voting frequency was also completed.

The success of the research continued with an answer to the first research question, namely the reason for the tag inaccuracy of existing images that are uploaded to image sharing sites. The reasons that constitute the answer include missing tags, semantic loss, missing tags, misspelling, erroneous personal choice, irrelevancy, lack of integrity, ambiguity, inconsistency, erroneous value entry, repetition, duplication, rule violation and user background. The research also stressed the benefits of establishing a benchmark for the proficient development of an image tagging system, which includes simplicity, ease of use, availability, universality and free of constraints. An initial experiment, using the collected research data, was also conducted by combining five similarity measures (Jaccord, Dice, Matching, Cosine and Overlap) to establish the relatedness of one image to another.

The second research question of how to extract information from images is also answered by a looking at domains and tag value sets, where the resultant product of tag sets can be extracted from a Cartesian product of the two sets of tag values, hence returning a set of accurately similar tags that were extracted from the images attributes.

Finally, a second experiment that uses the bulk of the image corpus to provide an integrated approach to accuracy improvement was conducted. Three approaches were proposed of which two were utilised. The initial approach of limiting the loss of spatial information by categorising the visual objects and features into 'Bags of Words' was discussed but not used.

The second approach used a calculated weight that was applied to the results for the first experiment to further improve the accuracy of tags, while the third approach applied WordNet to the results from the second approach for even more accurate results.

Finally, the research successfully provided a comparative analysis of the resulting findings from the two experiments that used the three selected approaches, namely 'combined similarity measures', 'applied weight' and 'applied WordNet'. The research concludes by recommending the use of the selected approaches to existing sites such as Flickr and Delicious.

7.3 Future Work

This research has been very successful in establishing the approaches required to increase tag accuracy in images. However, this work is by no means exhaustive and can be further extended in the following fields:

- **Video search and retrieval:**

  Much academic and research work has been done on the use of the similarity measures approach for video retrieval [94] [1] [64]. The work mainly centres around the retrieval of the most relevant videos by a concept, such as 'vehicle', which can then be ranked based on their relevance to the concept. The research work also involves the methods of automatically annotating semantic videos and using similarity based retrieval.

  This can be further extended by using the same approaches used in this research, namely the use of a combination of similarity measures then applying a WordNet weighting to further improve the accuracy of the tags for future retrieval.

- **Audio search and retrieval:**

Similarly academic research has also been done on the use of WordNet for overcoming the semantic gap in content based audio search [12] [93]. Again, the use of combined similarity measures and WordNet weighting for improving the accuracy of semantic audio tags, can be used to extend the work of this research.

- **Using colour:**

  This research has, so far, relied on the input of primary object, secondary object, action and background into the tags descriptors, to add value to an image tags. Throughout all the samples and the data and images collected for this research, an assumption has been made that all images under consideration were black and white. However, further accuracy can be established by considering the plethora of other properties within images, such as colour, texture and shape [23].

A suggested future research question would be:

'How to improve the accuracy of image tags by applying a weight based on the similarity of every available image attribute, such as action, background, colour, texture and shape'.

- **Image noise reduction:**

  Another future research extension would be reducing the noise picked up by images such as sparse light and dark disturbances. A typical source of image noise would be specks of dust inside the camera or on the camera lens.

  In such work, image processing is required to modify each pixel in the image from its original value, such that it complies with a normal distribution of noise. This will result in a noise free, uniform image that can be further processed to extract it rich and high value attributes, such as colour, background and action.

- **Algorithm utilisation:**

  Finally, the algorithm developed from the approaches in this research can be recommended to the image sharing sites, since any improvements gained in the accuracy of image tags can be translated to ease of image retrieval, and hence, extra revenue from customers who require images for commercial reasons.

  Alternatively an SDK (Software Development Kit) or an API (Application programming interface), can be developed, in further research, to enable users and commercial entities to develop applications that utilises this research's algorithm.

References

[1] Amjad Altadmri; and Amr Ahmed. Automatic semantic video annotation in wide domain videos based on similarity and commonsense knowledgebases. *In IEEE International Conference on Signal and Image Processing Applications.*, 2009.

[2] M. Ames; and M. Naaman. Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI conference on Human factors in computing systems, ACM Press*, pages 971–980, 2007.

[3] R. Miotto; L. Barrington; and G. Lanckriet. Improving auto-tagging by modeling semantic co-occurrences. *In*

Proc. Int. Society Music Information Retrieval Conference., 2010.

[4] Tim Berners-Lee. Weaving the web: The past, present, and future of the world wide web, by its inventor. *Texere Publishing Ltd.*, 1999.

[5] Tim Berners-Lee. The semantic web. *Scientific American*, 2001.

[6] Cameron Marlow; Mor Naaman; Danah Boyd; and Marc Davis. Ht06, tagging paper, taxonomy, flickr, academic article, to read. *In Proceedings of the 17th Conference on Hypertext and Hypermedia.*, page 31 40, 2006.

[7] C. Dance; J. Willamowski; L. Fan; C. Bray; and G. Csurka. Visual categorization with bags of keypoints. *In Proc. European Conference Computer Vision Int. Workshop on Statistical Learning in Computer vision.*, pages 1–22, 2004.

[8] Christian Wartena; Rogier Brussee; and Martin Wibbels. Using tag co-occurrence for recommendation. *In ISDA.*, pages 273–278, 2009.

[9] Michael Buckland. Vocabulary as a central concept in library and information science. digital libraries: Interdisciplinary concepts, challenges, and opportunities. *In Proc. Third International Conference on Conceptions of Library and Information Science*, 1999.

[10] Freddy Limpens; Michel Buffa; and Fabien Gandon Edelweiss. Bridging ontologies and folksonomies to leverage knowledge sharing on the socialweb: a brief survey. *Laboratoire d'Informatique, Signaux, et Systmes de Sophia-Antipolis.*, 2008.

[11] Stuart Butterfield. Sylloge. 2004.

[12] Pedro Cano. Content-based audio search from fingerprinting to semantic audio retrieval. *PhD thesis, University*

Pompeu Fabra, Barcelona, Spain., 2007.

[13] L. Cao and L. Fei-Fei. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes. *In Proc. EIEEE International Conference Computer Vision.*, pages 1–8, 2007.

[14] S F Cheng; W Chen; and H Sundaram. Semantic visual templates: linking visual features to semantics. *In Proceedings of the International Conference on Image Processing, ICIP98, vol. 3*, page 531 535, 1998.

[15] P. Tirilly; V. Claveau; and P. Gros. Language modeling for bag-of-visual words image categorization. *In Proc. ACM International Conference on Content-Based Image and Video Retrieval.*, pages 249–258, 2008.

[16] Media Culpa. Instagram now growing faster than flickr @ONLINE. `http://www.kullin.net/2012/01/`

`instagram-now-growing-faster-than-flickr/`, February 2013.

[17] Liu Wenyin; Susan Dumais; Yanfeng Sun; HongJiang Zhang; Mary Czerwinski and Brent Field. Semi-automatic image annotation. *Microsoft Research China.*, 2001.

[18] Carlos J. Bernardos; Antonio de la Oliva; Mara Caldern; Dirk von Hugo; and Holger Kahle. Nemo: Network mobility. bringing ubiquity to the internet access. *IEEE INFOCOM.*, 2006.

[19] T Deselaers; and V Ferrari. Visual and semantic similarity in imagenet. *In CVPR.*, 2011.

[20] C Yang; M Dong; and F Fotouhi. An interactive image annotation system. *In Proc.Intl. Semantic Web Conference 2008, volume 5318 of LNAI.*, page 6 8, 2005.

[21] Alan Mislove; Hema S. Koppula; Krishna P. Gummadi; Peter Druschel; and Bobby Bhattacharjee. Growth

of the flickr social network. *Proceedings of the first workshop on Online social networks.*, pages 25–30, 2008.

[22] Beth Emmerling. An experimental study of socialtagging behaviorand image content. *Journal of The American Society for Information Science and Technology.*, 2011.

[23] N.; Ferecatu, M.; Boujemaa and M. Crucianu. Semantic interactive image retrieval combining visual and conceptual content description. *In ACM Multimedia Systems Journal.*, 2007.

[24] Flickr. All time most popular tags ONLINE, December 2012.

[25] Flickr. Flickr gets faster drag-and-drop uploader, larger upload limits ONLINE. `http://blog.flickr.net/en/2011/08/04/6000000000/`, January 2013.

[26] Y.; R. Iyer; R.E. Schapire; Freund and Y. Singer. An efficient boosting algorithm for combining preferences. *In*

Proc. of the 15th Intl. Conference on Machine Learning., 1998.

[27] Miller G. Informavores. in: The study of information: Interdiscipiinary messages. *Wiley-Interscience*, pages 111–113, 1983.

[28] S. A. Golder; and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science 32(2)*, pages 198–208, 2006.

[29] Tye Rattenbury; Nathaniel Good; and Mor Naaman. Towards automatic extraction of event and place semantics from flickr tags. *Yahoo! Research Berkeley.*, 2007.

[30] George A. Miller; Richard Beckwith; Christiane Fellbaum; Derek Gross; and Katherine Miller. Introduction to wordnet: An on-line lexical database. *Princeton*, 1993.

[31] Thomas Gruber. Collective knowledge systems: Where the social web meets the semantic web. *In Web Semantics,*

6(1)., page 4 13, 2008.

[32] Tom Gruber. Ontology of folksonomy: A mash-up of apples and oranges. 2005.

[33] X Zhou; M Wang; Y Xiang; H Xu; and B Shi. Exploring flickr's related tags for semantic annotation of web images. *In Proceeding of the ACM International Conference on Image and Video Retrieval.*, 2010.

[34] Harry Halpin. Identity, reference, meaning, and the web. *Proceedings of the Identity, Reference, and the Web (IRW2006)*, 2006.

[35] M Aurnhammer; P Hanappe; and L Steels. Integrating collaborative tagging and emergent semantics for image retrieval. *In Collaborative Web Tagging Workshop (WWW06).*, 2006.

[36] Yusef Hassan-Montero; and Victor Herrero-Solana. Improving tag-clouds as visual information retrieval inter-

faces. *In Proceedings of Multidisciplinary Information Sciences and Technologies.*, 2006.

[37] M Heckner; M Heilemann; and C Wolff. Personal information management vs. resource sharing: Towards a model of information behaviour in social tagging systems. *In Intl AAAI Conference on Weblogs and Social Media (ICWSM), San Jose, CA, USA.*, 2009.

[38] M Sonka.; V HLavac; and R Boyle. Image processing, analysis and machine vision. *Chapman and Hall; London.*, 1993.

[39] Lei Wu; Steven C. H. Hoi; and Nenghai Yu. Semantics-preserving bag-of-words models and applications. *IEEE TRANSACTIONS ON IMAGE PROCESSING.*, 2010.

[40] Benjamin Markines; Ciro Cattuto; Filippo Menczer; Dominik Benz; Andreas Hotho; and Gerd Stumme. Evaluating similarity measures for emergent semantics of social

tagging. *In The Semantic Web ISWC 2008, Proc.Intl. Semantic Web Conference.*, pages 641–650, 2009.

[41] C Cattuto; D Benz; A Hotho; and G Stumme. Semantic grounding of tag relatedness in social bookmarking systems. in the semantic web iswc. *In Proc.Intl. Semantic Web Conference 2008, volume 5318 of LNAI.*, page 615 631, 2008.

[42] Masashi Inoue. Image retrieval: Research and use in the information explosion. *Progress in Informatics.*, 2009.

[43] Business Insider. Facebook: Users upload 300 million images a day ONLINE. `http://www.businessinsider.com/facebook-images-a-day-instagram-acquisition-2012-7`, February 2013.

[44] Varelas Ioannis. Semantic similarity methods in wordnet and their application to information retrieval on the web.

In Intelligent Systems Laboratory of the Techical niversity of Crete., 2005.

[45] D Tsai; Y Jing; Y Liu; H A Rowley; S Ioffe; and J M Rehg. Large-scale image annotation using visual synset. *In ICCV.*, 2011.

[46] Kristina M. Irvin. Comparing information retrieval effectiveness of different metadata generation methods., 2003.

[47] Jay J. Jiang; and David W. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. *In Proceedings of International Conference Research on Computational Linguistics (ROCLING X), 1997, Taiwan.*, 1997.

[48] Ruofan Wang; Shan Jiang; and Yan Zhang. Re-ranking search results using semantic similarity. *In the 8th International Conference on Fuzzy Systems and Knowledge Discovery.*, pages 1047–1051, 2011.

[49] C. Jrgensen. Image retrieval: theory and research. *The Scarecrow Press, Lanham, MA and Oxford.*, 2003.

[50] Grigory Begelman; Philipp Keller; and Frank Smadja. Automated tag clustering: Improving search and exploration in the tag space. *In Proceedings of the 15th International World Wide Web Conference, volume 6.*, 2006.

[51] Alexander Kreiser; Andreas Nauerz; Fedor Bakalov; Birgitta Konig-Ries; and Martin Welsch. A web 3.0 approach for improving tagging systems. *In Proceedings of the International Workshop on Web 3.0: Merging Semantic Web and Social Web (in Conjunction with the 20th International Conference on Hypertext and Hypermedia).*, 2009.

[52] Ivn Cantador; Ioannis Konstas; and Joemon M. Jose. Categorising social tags to improve folksonomy-based recommendations. pages 1–15, 2011.

[53] M Strohmaier; C Krner; and R Kern. Why do users tag? detecting users motivation for tagging in social tagging systems. *In Technical report, Knowledge Management Institute - Graz University of Technology.*, 2009.

[54] J Kustanowitz; and B Shneiderman. Motivating annotation for personal digital photo libraries: Lowering barriers while raising incentives. *In Tech. Report HCIL-2004-18, U. Maryland.*, 2005.

[55] Seongjae Lee; and Soosun Cho. Web image retrieval reranking with wikipedia semantics. *In International Journal of Multimedia and Ubiquitous Engineering.*, 2012.

[56] Sun-Sook Lee and Hwan-Seung Yong. Ontosonomy: Ontology-based extension of folksonomy. *In Proceedings of the 2008 IEEE InternationalWorkshop on Semantic Computing and Applications.*, 2008.

[57] J. Z. Wang; J. Li; and G. Wiederhold. Simplicity: Semantics-sensitive integrated matching for picture libraries. *IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 9.*, page 947963, 2001.

[58] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *in Int. J. Comput.Vis.*, page 91110, 2004.

[59] J Fan; D A Keim; Y Gao; H Luo; and Z Li. A novel approach to enable semantic and visual summarization for exploratory image search. *In ACM Conf. on Multimedia Information Retrieval (MIR08).*, 2008.

[60] J Fan; Y Gao; H Luo. Integrating concept ontology and multi-task learning to achieve more effective classier training for multi-level image annotation. *In IEEE Trans. on Image Processing, vol. 17, no.3*, pages 407 – 426, 2001.

[61] Y Gao; J Fan; H Luo; and S Satoh. A novel approach for ltering out junk images from google search results. *In Intl Conf. on Multimedia Modeling.*, 2008.

[62] Mathias Lux; and Gisela Dosinger. From folksonomies to ontologies: employing wisdom of the crowds to serve learning purposes. *International Journal of Knowledge and Learning.*, 2008.

[63] Angus E.; Thelwall M.; and Stuart D. General patterns of tag usage among university groups in flickr. *Online Information Review*, 2008.

[64] Ainhoa Llorente; Srdan Zagorac; Suzanne Little; Rui Hu; Anuj Kumar; Suhail Shaik; Xiang Ma; and Stefan Ruger. Semantic video annotation using background knowledge and similarity-based video retrieval. *In Proceedings of the TREC Video Retrieval Evaluation (TRECVID).*, 2008.

[65] L. Wu; M. Li; Z. Li; W.-Y. Ma; and N. Yu. Visual language modeling for image classification. *In Proc. International Workshop on Multimedia Information Retrieval.*, pages 115–124, 2007.

[66] Z Xu; Y Fu; J Mao; and D Su. Towards the semantic web: Collaborative tag suggestions. *In Collaborative Web Tagging Workshop (WWW06).*, 2006.

[67] Zhichen Xu; Yan Fu; Jianchang Mao; and Difu Su. Towards the semantic web: Collaborative tag suggestions. *In Proceedings of the Collaborative Web Tagging Workshop at the WWW.*, 2006.

[68] Najlae Idrissi; Jose Martinez; and Driss Aboutajdine. Bridging the semantic gap for texture-based image retrieval and navigation. *JOURNAL OF MULTIMEDIA*, 2009.

[69] Adam Mathes. Folksonomies - cooperative classification and communication through shared metadata. *Computer Mediated Communication*, 2004.

[70] G J Qi; X S Hua; Y Rui; J Tang; T Mei; and H J Zhang. Correlativemulti-label video annotation. *In MULTIMEDIA'07: Proceedings of the 15th International Conference on Multimedia, ACM, NewYork, USA.*, page 17 26, 2007.

[71] Peter Merholz. Metadata for the masses. 2004.

[72] Franck Michel. How many photos are uploaded to flickr every day, month, year?

[73] Peter Mika. Ontologies are us: a unified model of social networks and semantics. *n International Semantic Web Conference*, pages 522–536, 2005.

[74] K Bischoff; C S Firan; W Nejdl; and R Paiu. Can all tags be used for search? *In Proceeding of the 17th ACM*

Conference on Information and Knowledge Management (CIKM08), page 203 212, 2008.

[75] Y.-G. Jiang; C.-W. Ngo; and J. Yang. Towards optimal bag-of-features for object categorization and semantic video retrieval. *In Proc. of the 6th ACM Int. Conf. on Image and Video Retrieval.*, pages 494–501, 2007.

[76] Mohsen Kalantari; Hamed Olfat; and Abbas Rajabifard. Automatic spatial metadata enrichment: Reducing metadata creation burden through spatial folksonomies. *Centre for SDIs and Land Administration: The University of Melbourne*, 2010.

[77] Alexandre Passant and Philippe Laublet. Tracking usage in collaborative tagging communities. *In Proceedings of theWorkshop on Contextualised Attention Metadata.*, 2007.

[78] John G. Breslin; Alexandre Passant; and Denny Vrandecic. Social semantic web. *Digital Enterprise Research Institute (DERI)*, 2011.

[79] PhotoWeekly. The number of photos on facebook is exploding ONLINE. `http://www.photoweeklyonline.com/the-number-of-photos-on-facebook-is-exploding-infographic/`, February 2013.

[80] Pingdom. Internet 2011 in numbers ONLINE. `http://royal.pingdom.com/2012/01/17/internet-2011-in-numbers/`, January 2013.

[81] Pingdom. Internet 2012 in numbers ONLINE. `http://royal.pingdom.com/2013/01/16/internet-2012-in-numbers/`, January 2013.

[82] Gabriella Pigozzi; Wlodek Rabinowicz; Soroush Rafiee Rad; and Jan Sprenger. Epistemology of social decision

making. *The Tilburg Center for Logic and Philosophy of Science.*, pages 2–118, 2009.

[83] Perry Rajnovic. Attaching textual meta-information or semantic linkages to images ONLINE. `people.cs.pitt.edu/~chang/265/seminar06/raj.ppt`, January 2013.

[84] Eva Hrster; Malcolm Slaney; MarcAurelio Ranzato; and Kilian Weinberger. Unsupervised image ranking. *In LS-MMRM.*, 2009.

[85] T.; RATTENBURY and M. NAAMAN. Methods for extracting place semantics from flickr tags. *In ACM Transactions on the Web.*, 2009.

[86] P. Resnik. Using information content to evaluate semantic similarity. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 448–453, 1995.

[87] Elizeu Santos-Neto; Matei Ripeanu; and Adriana Iamnitchi. Tracking usage in collaborative tagging commu-

nities. *In Proceedings of theWorkshop on Contextualised Attention Metadata.*, 2007.

[88] Harry Halpin; Valentin Robu; and Hana Shepherd. The complex dynamics of collaborative tagging. *Proceedings of the 16th International Conference on World Wide Web*, pages 211–220, 2007.

[89] G.P. Enser; C.J. Sandom; and P.H. Lewis. Automatic annotation of images from the practitioner perspective. 2005.

[90] S. Lazebnik; C. Schmid; and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. *In Proc. CVPR.*, page 21692178, 2006.

[91] A. Hotho; R. Jaschke; C. Schmitz; and G. Stumme. Information retrieval in folksonomies: Search and ranking. pages 411–426, 2006.

[92] C Cattuto; A Baldassarri; V D P Servedio; and V Loreto. Emergent community structure in social tagging systems.

advances in complex physics. *In Proc. of the European Confeence on Complex Systems ECCS2007.*, 2007.

[93] M.; Sonnleitner R.; Hauger D.; Seyerlehner; K., Schedl and B.. Ionescu. From improved auto-taggers to improved music similarity measures. *In: Proc. Adaptive Multimedia Retrieval, Copenhagen, Denmark.*, 2012.

[94] Yusuf Aytar; Mubarak Shah; and Jiebo Luo;. Utilizing semantic word similarity measures for video retrieval. *In CVPR.*, 2008.

[95] C.E. Shannon. A mathematical theory of communication. *Technology Journal.*, pages 379–423, 1948.

[96] Brkur Sigurbjrnsson; and Roelof van Zwol. Flickr tag recommendation based on collective knowledge. *Yahoo! Research.*, 2008.

[97] Kilian Weinberger; Malcolm Slaney; and Roelof van Zwol. Resolving tag ambiguity. *ACM.*, 2008.

[98] Xirong Li; Cees G. M. Snoek; and Marcel Worring. Learning social tag relevance by neighbor voting. *In IEEE Trans. in Multimedia.*, 2009.

[99] Dominik Benz; Christian Krner; Andreas Hotho; Markus Strohmaier; and Gerd Stumme. Stop thinking, start tagging: Tag semantics emerge from collaborative verbosity. *the International World Wide Web Conference Committee (IW3C2)*, 2010.

[100] James Surowiecki. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations.* Doubleday; Anchor, New York City, 2005.

[101] G.R.J.Srinivas; Niket Tandon; and Vasudeva Varma. A weighted tag similarity measure based on a collaborative weight model. *ACM.*, 2010.

[102] Time Technology. Flickr gets faster drag-and-drop uploader, larger upload limits ONLINE. http://techland.time.com/2012/04/25/flickr-gets-faster-drag-and-drop-uploader-higher-upload-limits January 2013.

[103] Nathanael Chambers; Joel Tetreault; and James Allen. Approaches for automatically tagging affect. *American Association for Artificial Intelligence*, 2004.

[104] Internet Stats Today. 2012 facebook stats ONLINE. http://internetstatstoday.com/?p=113, January 2013.

[105] M. Uschold. Where are the semantics in the semantic web? *American Association for Artificial Intelligence*, 2003.

[106] Franco Maria Nardini; Fabrizio Silvestri; Hossein Vahabi1; Pedram Vahabi; and Ophir Frieder. On tag spell checking. *In Proc. SPIRE'10 Proceedings of the 17th inter-*

national conference on String processing and information retrieval., pages 37–42, 2010.

[107] J. Li; W. Wu; T. Wang; and Y. Zhang. One step beyond histograms: Image representation using markov stationary features. *in Proc. IEEE Conf. Computer Vision Pattern Recognition.*, pages 1–8, 2008.

[108] L. Wang; and B. S. Manjunath. A semantic representation for image retrieval., in proc. of international conference on image processing. *ICIP02.*, pages 523 – 526, 2003.

[109] Neela Sawant; Jia Li; James Z. Wang. Automatic image semantic interpretation using social action and tagging data. *The Pennsylvania State University.*, 2010.

[110] Matt Brian The Next Web. Flickr hits 6 billion total photos, but facebook does that every 2 months.

[111] wikipedia. Models of communication ONLINE. `http://en.wikipedia.org/wiki/Models_of_communication`,

January 2013.

[112] the free encyclopedia Wikipedia. Tag cloud @ONLINE. `http://en.wikipedia.org/wiki/Tag_cloud`, February 2013.

[113] K Rodden; W Basalaj; D Sinclair; K Wood. Does organization by similarity assist image browsing?. *In ACM SIGCHI.*, 2001.

[114] Manish Gupta; Rui Li; Zhijun Yin; and Jiawei Han. Survey on social tagging techniques. *SIGKDD Explor.*, 2010.

[115] H. Zhang; M. Korayem; E. You; and D. J. Crandall. Beyond co-occurrence: Discovering and visualizing tag relationships from geo-spatial and temporal similarities. *In WSDM 2012, ACM.*, pages 33–42, 2012.

[116] L. Wu; L. J. Yang; N. H. Yu; and X. S. Hua. Learning to tag. *In Proceeding of ACM International World Wide Web Conference.*, 2009.

[117] X Wu; L Zhang; and Y Yu. Exploring social annotations for the semantic web. *In Proceedings of the 15th International Conference on World Wide Web.*, pages 417 – 426, 2006.

[118] Ye Lu; Chunhui Hu; Xingquan Zhu; HongJiang Zhang; and Qiang Yang. A unified framework for semantics and feature based relevance feedback in image retrieval systems. *In Proc. of the eighth ACM international conference on Multimedia*, 2000.

[119] Jie Xiao; Wengang Zhou; and Qi Tian. Exploring tag relevance for image tag re-ranking. *In Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval.*, pages 1069–1070, 2012.

Images

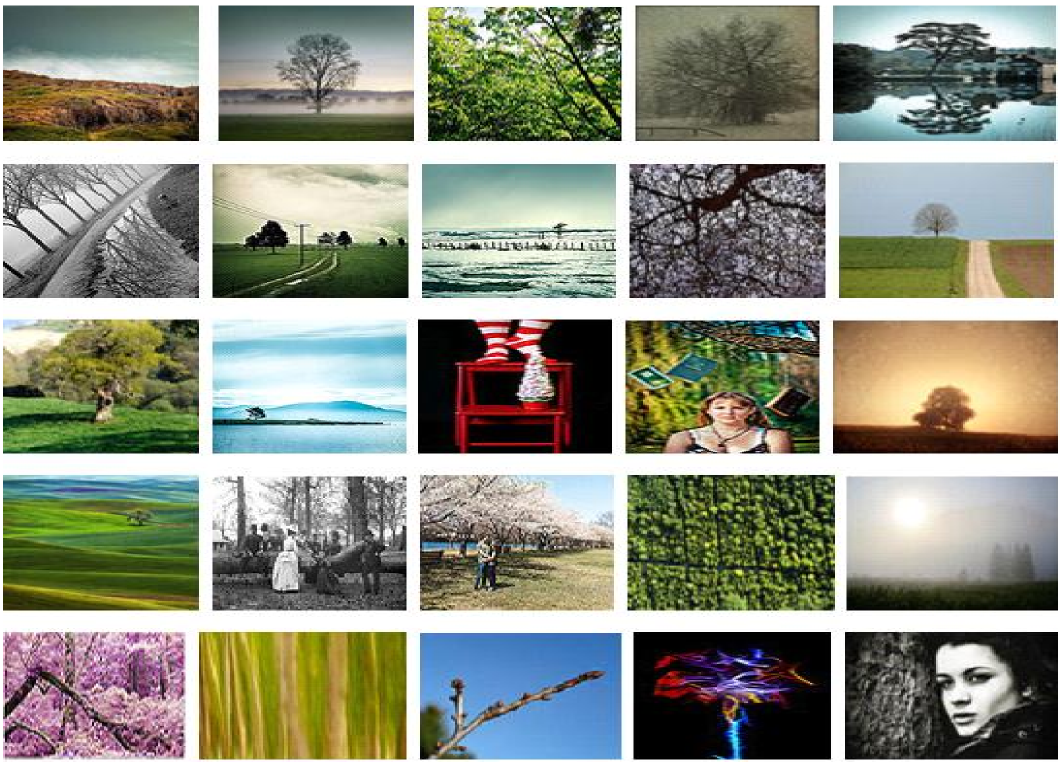


Figure 8: Example images of tree tags

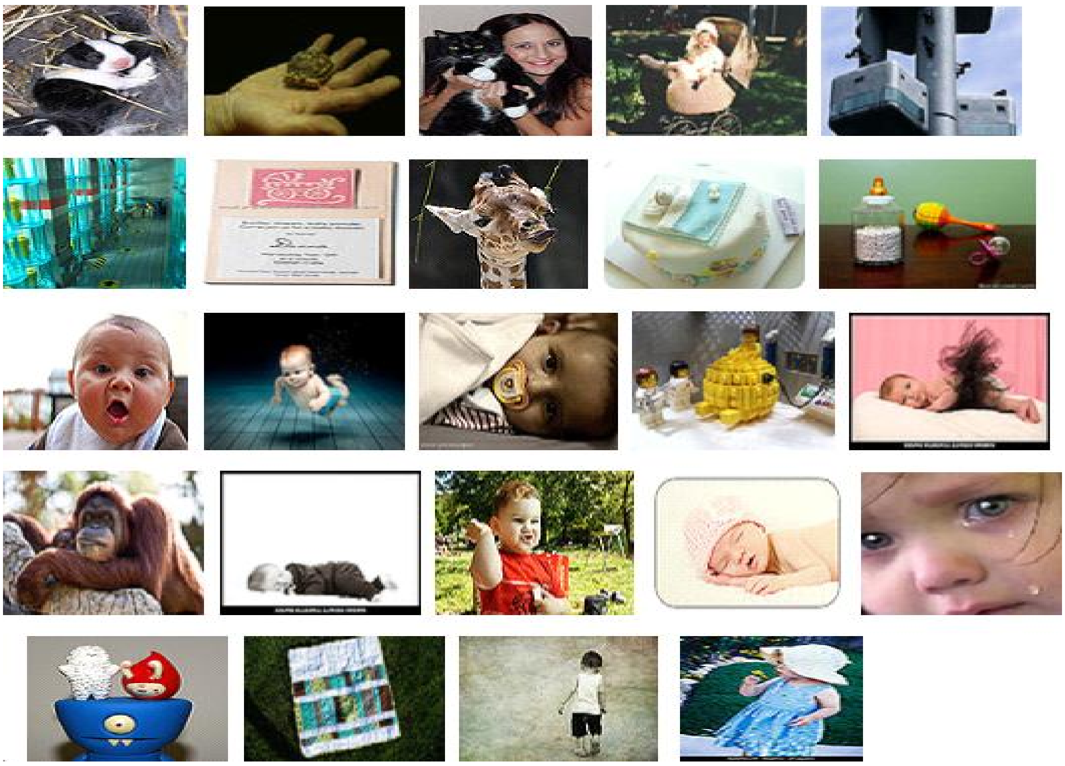


Figure 9: Example images of baby tags

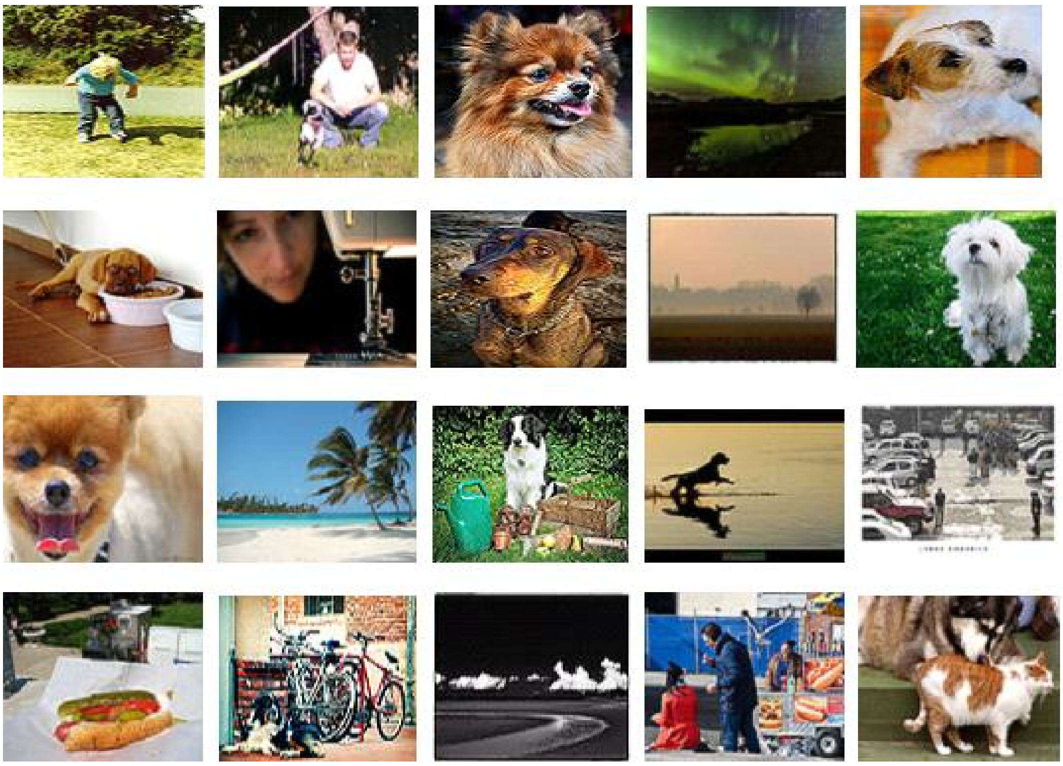


Figure 10: Example images of dog tags

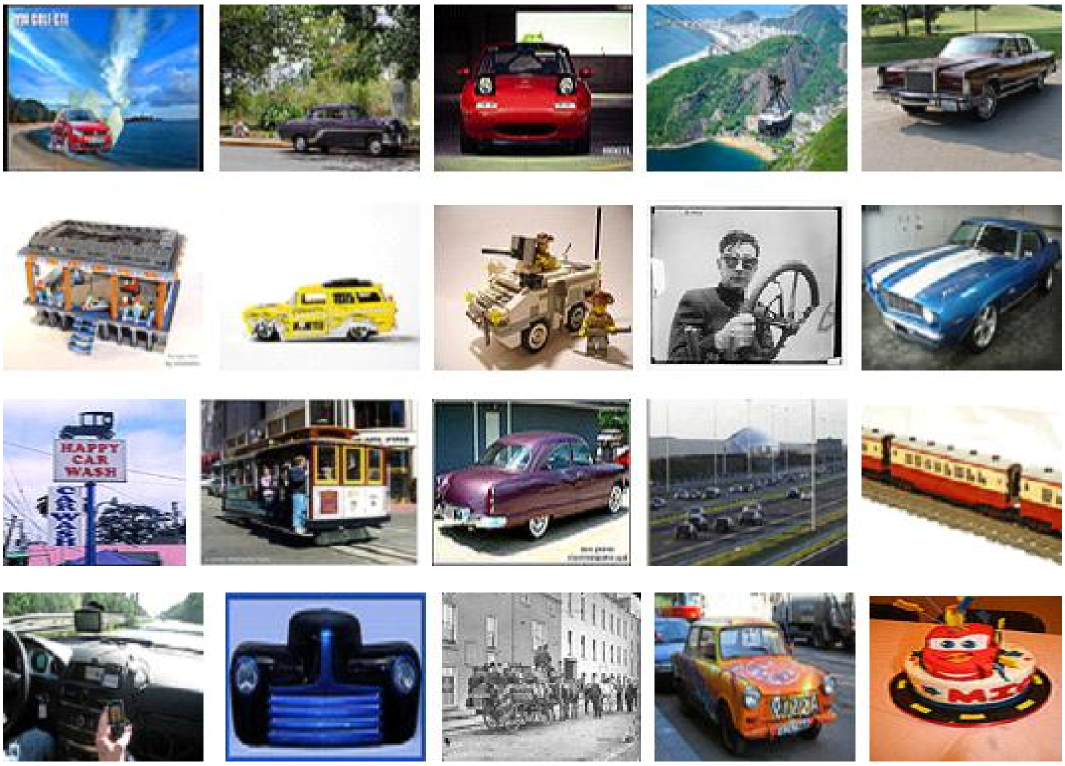


Figure 11: Example images of car tags

Sample Results

Tag	Combined Similarity Measures	Weighted Combined Similarity Measures	Combined WordNet with similarity measures
Tree	0.24	0.36	0.37
Tree	0.19	0.25	0.26
Tree	0.12	0.21	0.23
Tree	0.26	0.45	0.48
Tree	0.01	0.10	0.12
Baby	0.37	0.56	0.55
Baby	0.22	0.34	0.35
Baby	0.42	0.81	0.81
Baby	0.37	0.55	0.57
Baby	0.03	0.13	0.14
Car	0.05	0.16	0.17
Car	0.14	0.29	0.31
Car	0.22	0.42	0.39
Car	0.54	0.89	0.88
Car	0.11	0.31	0.34
Boy	0.04	0.11	0.12
Boy	0.02	0.10	0.10
Boy	0.29	0.42	0.42
Boy	0.36	0.62	0.61
Boy	0.15	0.26	0.26
Green	0.04	0.12	0.11
Green	0.26	0.45	0.45
Green	0.53	0.75	0.75
Green	0.21	0.42	0.42
Green	0.55	0.81	0.81

Table 1: **Results of some tags for the three experiments**

Relevant Images

Figure 12: Relevant image of car tags

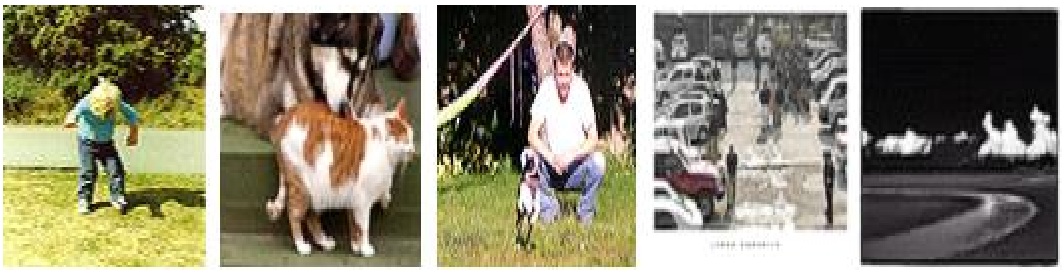


Figure 13: Relevant image of car tags

Figure 14: recommended images of dog tags

Figure 15: recommended images of car tags