Study On Social Networks And Search Engines

Semantic annotations got more attention, recently, because of the expansion in the technology adopted by social networks exponentially. Collaborative semantic annotation comes from the fact that many friends share thoughts and opinions using social networks. In these collaborative environments, there are a large number of terms. These terms come from various resources such as: users' comments, time stamps and metadata. In this study, we will solve the problem of choosing the suitable terms that can be used in searching for images in an image sharing social network.

1 Introduction

Internet is crucial for human life activities. It becomes a repository for almost all the knowledge on the planet. That involves enormous information about computers, biology, medicine…etc. The technology evolution pushed many resources to see the light as a potential source for knowledge. These resources vary from cameras, phones, books, novels and much more. The growth of resources attracts many users (amateur or researchers) to use the power of internet in order to restore specific information. This is the most precious aspect of Information Retrieval (IR) disciplinary. IR informs the user on the existence and whereabouts of documents relating to his/her request [1]. That leads to the appearance of search engines.

Search engines are one of the greatest inventions within the recent decades. As part of IR, the main objective for search engines is to facilitate knowledge access by specifying the relevant information. The user will use a dedicated user interface to enter a query and the system will get back all relevant information that has a match. Many researchers have contributed various algorithms to increase the efficiency of retrieval and shrink the consuming time. Indexing plays a significant role in these calculations.

Relatively, the area of social networks is still young. Many social networks had been activated to see the light recently i.e. the well known social network Facebook which started in 2006 as a local network before capturing the greatest share worldwide with about 500 million users. Even though the services provided by social networks were limited at the early stages, the number grew to become uncountable these days.

Social networks and search engines have similar features and options as well as some differences. One of greatest similarities ever is the annotations.

Annotations have different definitions based on the domain or scope. Generally, annotations can be defined as making a note while reading a book, a document, an online video, software code or any information in any media [2]. Many applications incorporate annotations since they simplify memorizing important aspects and ideas which makes review much easier. For instance, learning management systems (LMS) have started using annotations in order to help students attach some notes for specific

lectures [3]. The collaborative nature of these systems allows users to suggest ideas and thoughts for friends or lead discussions with colleagues and teachers [4]. The collaborative nature of annotations is referred as “Semantic Annotation.” They are helpful for expressing notions alongside with their computational representation in a formal language which plays a role in bridging the ambiguity of the natural language [5].

Semantic annotations are used in social networks for providing efficient retrieval. They will be part of objects indexing. Indexing with users’ annotations will offer a better chance that the most relevant objects will be retrieved back.

2. Semantic Annotations

Content Based Image Retrieval (CBIR) represents the realization of semantic annotations where annotations made by users will be used to index images [6]. CBIR was developed to retrieve the relevant images from a repository that contains a huge number of digital images. Semantic annotations will be beneficial for CBIR by indexing images with these terms. Queries with these terms provide accurate results and more relevant images. These terms should be filtered before indexing in order to get the optimal terms. The users’ queries that match these index terms will help in retrieving accurate images.

2.1 Social Networks

Social networks attract extensive attentions recently. Many sites have been initiated to facilitate users’ communications, such as MySpace, Facebook, Cyworld, and Bebo. According to [7], social network sites are web-based services that allow individuals to construct a public or semi-public profile within a bounded system, articulate a list of other users with whom they share a connection or have a relation with, and view and traverse their list of connections and those made by others within the system. These sites have the dynamicity to serve in a semantic manner. This opportunity seems interested for companies in addition to individuals. Big companies such as Best Buy, Deloitte, Microsoft, and IBM have their own social network which maintains collaboration among all parties [8].

The social network sites can be classified based on the purpose. There are specific sites for sharing and comments on videos (e.g. Youtube). Other sites are dedicated for images such as Flickr. Facebook and MySpace are general social network sites. Also, Twitter is social network dedicated for sending and receiving tweets (messages). There are many features and options incorporated into the state of the art social network sites. In order to investigate these features, we have to study the most commonly preferred site, Facebook. In [9], several features have been discussed such as managing personal friends, chatting, and posting on walls, sharing photos and videos and news feed. One of the greatest features ever is the adoption of annotations.

Unfortunately, in such collaborative environments (e.g. Flickr) there are a large number of terms. These terms come from various resources such as: users' comments, time stamps and metadata. Not all of these terms will retrieve relevant images. So, we propose a technique to choose the suitable terms that can be used in searching for images in an image sharing social network since these index terms represent the best terms among all available ones.

3. Related Work

Several works proposed combining annotations. In [10], semantic combinations among the set of ontology refined sense with the relationship edges to connect the terms in this ontology had been proposed to represent the semantics of this ontology. It measures the semantic similarity or difference of two ontologies from different domains simply by comparing the commonality or difference of two sets. The encouragement behind this algorithm is the measuring of the semantic similarity for the terms or concepts in natural language within the same scope.

In [11], the subspace learning under the cross-domain setting has been exploited and cross-domain discriminative locally linear embedding (CDLLE) had been proposed. It parses both local geometry and discriminative information from the training domain to the testing domain while the preserved goal is minimizing the quadratic distance between the distribution of the training samples and that of the testing samples. It will be used when training and testing samples are distributed independently and identically (come from different domains).

On the other hand, [12] had introduced a novel technique for image annotation. It is a game used to harness human brain power towards image annotation by motivating users to provide valuable information on image contents by using an entertaining manner. It aggregates image classification and previous outcomes (annotations made by other players) to filter out the user input (annotations). Therefore, it eliminates annotations made by less-rational game players. The experiments involve performance evaluation of 2 algorithms: Markov Model and Bayes.

Also one of the current researches, as proposed by [13], focuses on the effect of using context in the image annotation systems and how it will increase the quality of image management systems. The intention is to use the metadata of web images in order to automatically annotate images. The system can be used for 3 different aspects of image context which are: spatial, temporal and social context. The annotated image will be used to search for events for specific contexts. It is suitable for mobile devices with digital cameras even though it has problems with privacy and users identification.

Another research proposed by [14]. It focuses on iterative annotation process to provide the labels. The proposed model, called Iterative Multi-label Multi-Relational Classification Algorithm (IMMCA) adopts both the content and multiple relations in social networks. The intention is to learn a label propagation scheme through the

different relations of the graph and then to use this learned propagation scheme to iteratively label each unlabeled node. It handles multiple relations and to perform multi-label classification in multi-graphs. The target application is image annotation in large social media sharing web sites. The images will be assigned with labels when users and images are connected through multiple relations - authorship, friendship, or visual/textual similarities. The strengths come from the ability to deal with both content and social relations.

In [15], a method for tag spam detection in social web sites had been proposed. It is inspired by a text mining approach which could discover the relationships between Web pages as well as tags. It clusters the Web pages and their tags respectively by self-organizing map algorithm. A labeling process was then applied on the trained map to discover the relationships between Web pages and between tags. The detection of tag spams could then be achieved by examining the semantic relatedness between a tag and its tagged Web page. The method can be used to remove the spam automatically from tags and get accurate retrieval results.

4. MyFlickr

We propose a technique that can select the best term to be used in the annotations and indexing, MyFlickr. Figure1 shows the conceptual model for MyFlickr.

Figure 1: MyFlickr Conceptual Model

The connection to image sharing network will be granted by using specified API(s). In this step, we will be able to retrieve all comments made by different users. After that, a preprocess phase will be conducted to detect misspelling and correct the invalid terms after comparing with thesaurus as well as removing the remaining incorrect terms. The next phase will be the weighting of remained terms in order to identify the candidate terms to be part of final annotations.

The classification phase involves extract all terms' related information such as place. Also, the relations among terms will be built such as Kuala Lumpur is the capital of Malaysia. It is followed by expansion phase where every term will get extensions. Every term has different terms with the same meaning (i.e. synonym). The next phase is the production of annotations that will be used in the next phase to index the images and facilitate retrieval.

The last phase is the evaluation of using our novel technique for image retrieval. The expected results include a new technique that provides accurate results for the retrieval of images.