UiT The Arctic University of Norway

FACULTY OF SCIENCE AND TECHNOLOGY DEPARTMENT OF COMPUTER SCIENCE

Context Centric Approach of Semantic Image Annotation and Retrieval

Najeeb, Elahi

A dissertation for the degree of Philosophiae Doctor
October 2019

Acknowledgements

First and foremost, I would like to express my deepest and sincere gratitude to my advisor, Professor Randi Karlsen. During this whole journey, I remain indebted for her not only for her continuous support for my Ph.D study but also for her understanding during the times when I was really down. I am really thankful to her for her patience, motivation, and vast knowledge. Her guidance helped me in all the time of research and writing this thesis. I could not have imagined completing this thesis without her continuous guidance and effective suggestions.

I would also like to thank my co-advisor associate professor Anders Andersen for his guidance and valuable comments during discussing and writing papers. I am also grateful to him for paying detail attention to my thesis and making helpful comments and valuable suggestions.

I would also like to thank master student Einar, Martin and Steffen Hageland for their participation in project development and other research fellows Mehmmod and Kostas Antypas for their valuable discussions and giving me this believe that I can pursue and fulfill my research goals.

Last but not the least, I would like to thank my family: my wife Sana, my mother Riffat Ara, my father and my siblings for supporting me spiritually throughout writing this thesis and my life in general.

Abstract

The invention of digital cameras and the portability offered by mobile phones empowered with digital cameras has considerably fueled the popularity of digital images. Moreover, the affordability of these devices has given the common man the opportunity to capture his world in pictures, and a number of online social network services have made it possible to conveniently share these images with others. Therefore, people are now capturing and sharing far more images than ever before. As a result, billions of searchable image data exist, with diverse semantics, visual contents, and geographically disparate locations, and is continuously growing in size. However, these collections are inherently difficult to navigate, due to their size and lack of machine understandable semantic information of the content of images.

This thesis proposes a novel approach to explore and extract context information attached with images, mainly gathered from social network sites. I first performed a user study, to understand the user behavior on social network sites. I inferred that the relationship among users have central importance.

To assist users to annotate images in social network, I use existing metadata gathered from already annotated images on social networks, to generate metadata for non-annotated images. Social network analysis techniques together with image metadata are used to automatically annotate images. As context for an image, I consider temporal and geographical values. In addition to that, I consider three basic social entities associated with images; user relationships, user activities (comments and likes) and annotations.

To retrieve the most relevant images from social network, I proposed Relation-Based Image Retrieval (RBIR). For each user I calculate their relationships with other members in the network, and a ranked list of the closest and most reputed friends is compiled by analyzing the mutual activates between two users and their overall individual reputation in the social network. Comments and likes made by highly ranked members hold more weight, and photos are ranked in accordance with the number and weight of likes and comments they receive.

To test our approach, I developed a prototype based on the Facebook platform, to annotate images and allow users to search for images among their Facebook friends. The results demonstrate that our techniques are useful for annotation and retrieving relevant images.

Contents

List of Figures

Chapter 1

Introduction

The invention of digital cameras and portability offered by mobile phone digital cameras has considerably fueled the popularity of digital images. Moreover, the affordability of these devices has given the common man the opportunity to capture his world in pictures and conveniently share them with others. For many people to express themselves by writing is a cumbersome task and an easier way to share their thoughts is by taking pictures. Thus the saying goes "A picture is worth a thousand words", suggests that complex ideas can be conveyed by a single click.

Therefore, people are now capturing and sharing far more images than ever before. It indeed confirms the Susan Sontag's vision of a world where "everything exists to end up in a photograph" [141]. As a result, billions of searchable image data exist, with diverse semantics, visual contents, and geographically disparate locations, and is continuously growing in size [34]. However, these collections are inherently difficult to navigate, due to their size and lack of machine understandable semantic information of the content of images.

Social Network Sites are online social networking services that have attracted considerable amount of attention and curiosity in recent years from all quarters. Kaplan and Haenlein [73] defined such services as consisting of a "set of Web applications, which allows the creation and exchange of user-generated contents". Thus, billions of people across the globe share their contents online using these services and photos are the most popular content. For example, according to

internet.org Flickr[1] hosts more than 5 billion pictures, while Facebook[2] user share more than 250 billion photos and continue uploading 350 million new photos each day. This makes Facebook the world's largest photo-sharing site. In this dissertation, I will be exploring both these services as source of photos.

1.1 Research Questions

The Web is increasingly populated with images that are tagged, timestamped, and surrounded with text. The recent emergence of social networks has enriched this metadata with social capital. This dissertation addresses the lack of research in the area of online social networks, particularly the images in online social networks, from an image retrieval point of view. The overall problem of finding relevant images has raised the following research questions.

How can information on online social networks support personally relevant image annotation and retrieval?

This broad research question can be narrow down and divided into the following questions.

RQ1) What is the user behavior in online social networks?

Users play the most important role in Online Social Network (OSN). Therefore, to understand the user is the first step to comprehend the OSN. Following questions are outlined to explore users in OSN.

   a) How important has online social networks become in people's daily life?

   b) Does people disclose personal information on OSN?

   c) What are the social activities users are most interested in on OSN?

   d) How does gender and age of the users play a role when making relationships/friends on an OSN?

---

[1] https://www.flickr.com/
[2] https://www.facebook.com/

RQ2) What is the meaning of context and how can social networks be explored as context of images?

RQ3) How can one reuse existing annotations of images from members of one's social network as a base for image annotation?

   a) Can an online social network context be used to automatically provide personally relevant image annotations?

RQ4) How to retrieve, rank and recommend the most relevant images on OSN?

   a) Can social network based image retrieval be used to improve the personal relevance of retrieved images?

The first research question RQ1 can be viewed as a background research within this dissertation. The research questions RQ2 and RQ3 deal with the context and features of context in social networks. The research performed in RQ1 and RQ2 was necessary to understand social networks and to identify the features of context to perform the experimental research, which was conducted in relation to research questions RQ3 and RQ4.

1.2   Hypothesis

This thesis addresses context-centric image annotation and retrieval focusing on three main areas: data semantics - how to formally specify the meaning of annotations useful to system and users; data reuse - how to integrate data from various contextual sources into information reusable for annotation; and image retrieval - once images are annotated, how to locate relevant images with respect to the query and user.

In each area, the main obstacle is the semantic gap between the low-level features content recognition can provide and the high-level abstract way people recall their memories. We therefore purpose a context-centric approach enriched with text-centric approaches. We believe that the value added by this approach can best be demonstrated when left to stand alone independent from content recognition. Our hypothesis is that the automatic integration of annotation into the image can be supported, with adequate results exclusively through context centric approach; and social networks can offer rich contextual metadata for image.

In all three areas we devise solutions (an ontology, an architecture for consolidating information from social network and an algorithm for annotation and retrieval) that are complimentary to content-recognition, rely only on context and reuse existing data and meaning from the semantic web.

1.3   Research Context

This dissertation started in 2009 when social media was just becoming popular and almost no research had been conducted in image search on social media. These research studies have been carried out as a part of a PhD program at the Department of Computer Science - UiT The Arctic University of Norway. The PhD project is a formal part of CAIM (Context-Aware Image Management) project[1]. CAIM is a research project funded by The Research Council of Norway[2] under the banner of VERDIKT program. The project is focused on research and the development of tools for context-aware image management, where image description, query formulation, retrieval from heterogeneous distributed environments, and ranking are designed for using context information. Important application domains are those requiring image capture and multimodal retrieval in mobile environments. The research work carried out in CAIM is performed with collaboration of University of Bergen, Norwegian University of Science and Technology, Munich University of Technology, University of Hawaii at Manoa and Telenor R&D in the areas of data management, image processing, information retrieval, multimedia and mobile systems. CAIM project objectives are to develop methods and techniques for:

- Dynamic context capture and management.

- Multimodal information retrieval algorithms based on visual queries (using current or system selected images), context-data and positioning information.

- Context-based ranking and presentation of multimodal information on mobile units.

- End-user applications for testing CAIM concepts and algorithms.

- Specifications and prototypes systems for next generation mobile units.

During my PhD, I also performed teaching assistant duties in course INF-3701 Advanced database systems. As part of PhD program at UiT, candidate

---

[1]http://caim.uib.no/
[2]http://www.forskningsradet.no/

must attend 30 sep (studentpoints). I attended and successfully passed 25 stp courses: INF-8320 Middleware, INF-8701 Advanced database systems, MNF-8000 Research seminar. The remaining 5 stp were achieved by attending two PhD level summer courses. I attended the IEEE 2009 Summer School on Semantic Computing. The 6-day event took place on the campus of the University of California, Berkeley, United States in July 20-25, 2009. My final report was about "Usage of RDF Linked Data". Lastly came the Web Science Doctoral Summer School, July 6-13 2011, DERI, NUI Galway (Ireland), where I gave a presentation on "Social Network Analysis and usages of NodeXL".

1.4 Methodology

There has been a long debate whether or not computer science categorically is a science. The objection that computing is not really a science since it studies man made technology, is not true. Computer science studies information processes both artificial and natural. It studies information processes which occur naturally in physical world [31, 37]. Computer science is therefor well fit within the field of natural science[37], where *hypothetical-deductive* reasoning model approach the scientist inquiry. The model begins with the formulation of a testable hypothesis, followed by the deduction of predictions and the design of experiments that either supports or refutes the hypothesis. If the experiment results are refuted, it generally leads to the reformulation the hypothesis and it is an iterative process. The final report on the core of computer science presents an intellectual framework for the discipline of computing. The report presents that research problems within the field of computer science are approached in the three following ways [31].

Theory is rooted in mathematics. Theory deals with identifying the objects, relationships among objects are hypothesis, proofs of relationships are constructed and results are interpreted.
Abstraction is rooted in experimental scientific methods. Abstractions deal with models of implementations, The models emphasize the essential features and provide the means for predicting the future behavior.

Design is rooted in the field of engineering. Design deals with the process of specifying a problem, deriving the requirements and specifications, and the system is designed and implemented. The steps are iterated if the system does not match the specifications.

The boundaries between theory, abstraction and design are not very clear, they intertwined each other and are hard to separate [31] .This is also the case in this dissertation. The work presented in this dissertation draws mainly from *design* paradigms. This dissertation is rooted in information retrieval systems, which deals with the organization of large sets of persistent, shared data for efficient query and update. In the category of theory, I devise algorithms for image annotation, and for storing and searching images. Through performance analysis of algorithm I evaluate that the requirements are satisfied. I use abstraction to derive a model for social knowledge representation and methods of processing and inference them. For the paradigm of design I assembled the requirements and specification with the help of user studies. Experimental data sets from Facebook were collected to test the prototype and steps were iterated to match the specification.

1.5 Included Publications

This section presents the list of papers published during the PhD work and documen my contributions.

Paper1: Najeeb Elahi, and Randi Karlsen. *"User behavior in online social networks and its implications: a user study."* Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics. ACM, 2012.

Description of paper: This paper presents a study on user behavior in social networks. The study discusses overall trends in making friends and explores the most popular activities among social network users. The relevance for this PhD dissertation lies in the understudying of the social network user.

Author Contributions: I had the original idea of the study. I was responsible for designing and conducting the survey.

Paper2: Najeeb Elahi., Karlsen, R. and Younas, W. (2012). *"Ontology-Based*

Image Annotation by Leveraging Social Context." International Journal of Handheld Computing Research (IJHCR), 3(3), 53-66. doi:10.4018/jhcr.2012070104.

Description of paper: This paper presents a system that can semi-automatically generate ontology-based annotations for an image on social networks by leveraging the annotations provided by the most active user. It is relevant to this dissertation because annotation is known to be a key factor in image retrieval systems.

Author Contributions: I led this study. Other authors helped in development, analysis and discussion.

Paper3: Najeeb Elahi and Randi Karlsen. 2014. *"Relation based image retrieval in online social network."* In Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication (ICUIMC '14). ACM, New York, NY, USA, , Article 26 , 7 pages. DOI=10.1145/2557977.2558019 http://doi.acm.org/10.1145/2557977.2558019.

Description of paper: In this paper we investigate how we can assist users to retrieve the most relevant images from their social network. A ranked list of the closest and most reputed friends is compiled; comments and likes made by highly ranked members hold more weight, and retrieved images are ranked in accordance.

Author Contributions: I had the original idea of the study and was responsible for the data collections. Other authors contributed with paper writing and discussion.

Paper4: Elahi Najeeb, Randi Karlsen, and Einar J. Holsbo. *"Personalized Photo Recommendation By Leveraging User Modeling On Social Network."* Proceedings of International Conference on Information Integration and Web-based Applications & Services. ACM, 2013.

Description of paper: In this paper we precisely model the user needs and interests from two social network services Facebook and Flickr, in order to recommend relevant photos. We proposed to use the Linked Open Data cloud which provides data along with metadata in such a way that it can easily be consumed by other web services.

Author Contributions: I was key person in planning the paper, performing

the review and writing the manuscript. Einar helped in ontology development and Randi Karlsen contributed by review the paper.

Paper5: Randi Karlsen, Martin Haetta Evertsen, and Najeeb Elahi. *"Metadata-based automatic image tagging."* International Journal of Metadata, Semantics and Ontologies 8.4 (2013): 298-308.

Description of paper: This was another work to annotate images. In this paper, we investigate the ability to automatically tag images based solely on image metadata, and present a novel approach to image tagging using a combination of the metadata geo-location, date/time and category keyword.

Author Contributions: As a co-author I participated in long sessions of discussion and reviewing the paper.

Paper6: Mannan, Noman Bin, Sheikh Muhammad Sarwar, and Najeeb Elahi. *"A New User Similarity Computation Method for Collaborative Filtering Using Artificial Neural Network."* Engineering Applications of Neural Networks. Springer International Publishing, 2014. 145-154.

Description of paper: In this work, we predict the rating of a particular item (movie) for a given user based on the judgment of other users, who are similar to the given user. We model similarity between two users as a function that consists of a set of adaptive weights and attempt to train a neural network to optimize the weights. Experiments and testing was done on Movielens dataset.

Author Contributions: As a co-author my contribution lies in formulating the problem. I participated in analyzing the data and writing the manuscript. I also presented the paper in the conference.

Paper7: Fernandez-Luque L, Elahi N, Grajales FJ 3rd. *"An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis."* Stud Health Technol Inform. 2009;150:292-6.

Description of paper: This paper shows results on the study of metadata, specially comments made on videos shared on social networks. It was interesting to find significant amount of comments revealing personal health information that was shared publicly on social networks.

Author Contributions: I co-led the paper with the researcher from Northern Research Institute (Norut). I participated by collecting data, refining and analyzing data. I contributed in writing the manuscript and reporting the results.

Paper8: Elahi Najeeb, Randi Karlsen, and Sigmund Akselsen. *"A context centric approach for semantic image annotation and retrieval."* Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns, 2009. COMPUTATIONWORLD'09. Computation World:. IEEE, 2009.

Description of paper: In this study, we discuss techniques to improve the quality of image retrieval and image management with the help of context information over the web. We proposed the three different aspects of image context such as spatial, temporal and most importantly social context.

Author Contributions: I was the main author and presented the hypothesis that leveraging the contextual metadata of images would yield relevant image retrieval. Other authors contributed by discussion and reviewing the text.

Paper9: Randi Karlsen, Najeeb Elahi, Anders Andersen *"Personalized Recommendation of Socially Relevant Images."* Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics (WIMS'18), Association for Computing Machinery (ACM), No. 2018, 2018, ISBN 978-1-4503-5489-9.

Description of paper: This present a social image recommender system that offers a hybrid filtering approach, combining content and knowledge-based filtering with a novel social-based filtering, that selects images of social interest to the user, by e.g. being posted by close friends or family.

Author Contributions: As a co-author my contribution lies in formulating the problem. I participated in analyzing the data and writing the manuscript. I also presented the paper in the conference.

Chapter 2

Thesis Background

2.1 Information Retrieval

Information retrieval has a long history and a mature field of research. With the invention of computers, people realized the potential of technology that can be used to organize and retrieve textual documents. Information retrieval systems were introduced as early as in 1950s when Vannevar Bush [27] published the article titled "As We May Think", where he discussed the idea of using computer for storing and accessing digital documents. This idea was later materialized in a number of research works. H.P. Luhn, proposed a new method of recording and searching information [89] and a statistical approach to automated encoding and searching of information [90]. Later G. Salton [127] proposed the Smart information retrieval system, were a few notable early research works.

For the long time, information retrieval was an activity to be preformed by professionals seeking information in a closed information system such as librarian and legal experts. However, the trend has swiftly changed in the last decade with the enormous success of the world wide web. Internet easily accessible to people, made the Web the primary source of information finding, where optimized search engine services satisfy the people's information need. The Noun "Google" became a transitive verb in English dictionary meaning "to use the Google search engine to obtain information about (someone or something) on the World Wide Web [1].

---

[1]Merriam Webster Dictionary http://www.merriam-webster.com/dictionary/google.

A Pew survey[1] found that as much as 90 percentages of people prefer Internet for getting everyday information.

Information Retrieval (IR) is a very broad term and can have different meanings, such as getting a book out of a shelf in a library is a form of information retrieval. In the academic field of study, information retrieval is defined as finding documents of an unstructured nature (usually text), that satisfies an information need, from within large collections stored on computers [93].

An information retrieval system is a system which stores, organizes, classifies and represents the information items. A **query** represents the intend of the user, and refers to the information need which user seeks from the IR system. The techniques which are used to fulfil the query are described through information retrieval models.

2.1.1 Information Retrieval Architecture

An IR system is a composition of different modules which produce different results, depending on the scope the of the system. Figure 2.1 illustrates the detailed view of a general IR system. Figure is inspired by Massimiliano work [123]. The crawler module extracts documents and relevant information from HTML code to build the document collection for the IR system. Before storing the documents in the system, documents are preprocess in order to effectively reduce the size and space. The indexer component constructs the inverted index and create the data structure in order to improve the search process. The searching component retrieves documents from the document collection by matching query words to the inverted index and the ranking component scores all the retrieved documents from the previous step to the relevance metrics. Finally the ranked list of documents are presented to the user.

Document Preprocessing Document collection and text query are preprocessed in the IR system. This involves a set of steps. The first step is **tokenization** of documents. This is the task of dividing the document (sequence of

---

[1]Pew Internet and American Life Project. The internet and daily life. Fallows, Deborah, 2004
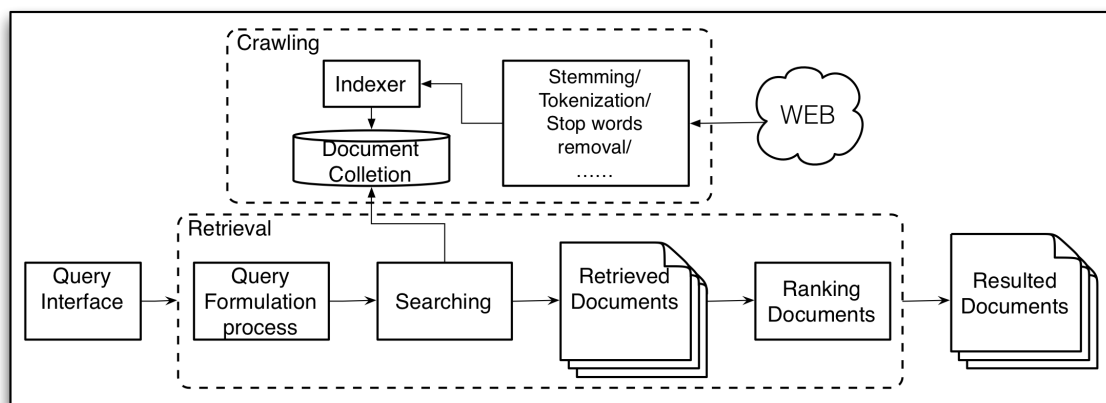
Figure 2.1: A general architecture of Information retrieval system

characters) into pieces, called **tokens**, and at the same time removing the unwanted characters such as punctuation. A token is a concept in the document that is grouped together as a useful semantic unit for processing [93]. Tokenization is not quite as straightforward as it seems. It has a number of issues, such as how to select a correct token from a sequence of characters, and which separator should be used to divide the sequence of characters.

Many tokenization issues are language-specific and requires the language of the document to be known. For instance in English, there are various uses of the apostrophe for possession and contractions. For example, assume the text *"You can't read Mr. O'Neill books"*. If you consider apostrophe as a word separator, you get two words *can* and *t* which looks intuitively wrong, but looks fine for *O* and *Neill*. Hyphenation is another issue, where hyphen (-) is used to connect the parts of compound words, such as *"co-author"* and *"a well-thought-out plan"*. In the tokenization process, hyphenation is commonly handled by applying heuristic rules. IR systems should also handle the new types of character sequences such as email address (najeeb.elahi@uit.no), web URLs (http://uit.no/startsida.html) and more. These words can be tokenized as a single token. However, the length of token would greatly expand the size of the vocabulary. In IR systems, it's considered good practice to use the same tokenizer techniques to index the document and the query [136].

Stopword removal is a process of removing the words that are considered more general and common. These words have high frequency in the document, but do not help in matching query to document. A *stopword list* is a vocabulary of unwanted words such as *"a, be, to, ect"*. Words from the stopword list are discarded during indexing, which leads to significant reduction in the number of postings that system has to store. However, eliminating words in a stopword list can harm the recall of the IR system. The typical example of which, is a search for song "Let it be". Potentially all these words are eliminated in the index as they appear in the stopword list. The retrieval process thus becomes ineffective. Because of these limitations, modern IR systems considerably reduce the size of the stopword list and web search engines are designed for "phrase search queries" and therefor abandon the use of stopword lists.

Stemming and lemmatization The goal of stemming is to reduce inflected and sometimes derived words to their common base or root form, called **stem**. For example, *listen*, *listened* and *listening* are reduced to *listen*, so that the words with variant forms can be viewed as same feature.

2.1.1.1 Relevance Ranking

There could be large number of documents that match the query keywords and number of documents would be in billion over the web and all documents can not be retrieved. To deal with this situation, there are number of relevance ranking methods to retire the most relevant documents in chronological order. TF (Term Frequency) and IDF (Inverse Document Frequency) are the most used methods [45].

2.2 Context

Human beings are very successful at conveying ideas to each other and responding accordingly. This is due to many reasons: the richness of the language they share, the common understanding of how the surrounding world works, and an implicit understanding of everyday situations. When people talk, they are capa-

ble of using implicit situational information, or context, to increase the conversational bandwidth. Unfortunately, this capability to conveying ideas does not transfer well to humans interacting with computers. Therefore, by improving the computer's access to context, we increase the productivity of communication in human-computer interaction and make it possible to produce more useful computational services [126]. To develop the specific definition of context that can be used in our application domain, we will look at how researchers have endeavoured to define context in their work.

The term context has been used in several ways in different areas of computer science, such as contextual search, context-sensitive help, multitasking context switch, context-aware information retrieval and so on [26, 129]. In fact, context is a general concept and has a loose definition. Therefore, there are numbers of definitions of context that can be found in the computer application domain [30, 129, 131]. Many of them define context in terms of characteristics of the surrounding environment that determine the behavior of user and information relevance to the user. Dey [126] defines context as :

> *"Any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves"*

Dey further elaborates the context values that generate the powerful understanding of the current situation, by using primary context such as location, entity, activity and time as an input into other sources of contextual information. Dey's work shows the significance of knowing the current situation of the user in an application domain. However, Dey leaves an important question about how information becomes relevant, unanswered [44].

In the work where Schilit and Theimer [130] first introduced the term context-aware, it refered to location, identities of people and objects in the proximity and changes to those objects. Schilit further divides context into three main categories.

- *Computing Context*: refers to the computing environment such as network

connection and capacity, communication costs, and devices accessible to user input and display.

- *User Context*: refers to the social situation of user, such as the user's profile, current activity, location, and people nearby.

- *Physical Context*: refers to for example lighting conditions, noise level and traffic conditions.

In Marc Davis' research dealing with camera phone image annotation [35], he coined the idea of Context-to-Contents inference for image retrieval that is closely in line with our need, by using context of an image to infer image content. Schilit definition is closest in spirit to the definition we desire.

"A system is context aware if it uses the context to retrieve information to the user, whereas relevancy depends on user social situation"

Investigations of different aspects of context have been central concerns in studies and theories of human information behavior, information seeking processes and information retrieval (IR), as theorists and researchers have moved away from decontextualized views of IR toward more use-centred and cognitive viewpoints. At this point, it is generally recognized that IR is an inherently interactive process, which occurs within multiple, overlapping, contexts that inform, direct or shape the nature of this interaction [32]. In other words, information seeking, use and evaluation take place within multidimensional contexts, which can be analyzed from multiple levels. One can see these developments in the theoretical models of [14, 19, 67] and others.

2.3 Social Network

Humans beings have remarkable qualities of cooperation. Political, industrial and information revolutions are the perfect examples of human cooperation for mutual benefit through forming social networks.
Fundamental questions arise to what social network are and what we (researcher)

can gain from understanding social networks such as culture, pattern of communication and people engagement with the social networks. In the following, I will analysis various social network definitions and attempt to define which will suite best in our case.

In 1954, social anthropologist J. A. Barnes [11] was the first person to coin the concept of social network. The general concept of society is in the background of social networks. Society cannot be considered as the sum of people alone, but also include relationship among people that connect one with another cultivate society [9]. Therefore, social network is the set of people tied by relationships. The modern social network paradigm was created by the sociologist Milgram. He studied the small-world problem based on six degrees of separation theory that everyone and "everything is six steps away". Milgram conducted the experiment, in which random people from Nebraska and Boston were asked to generate acquaintance chains to send a letter to a person in Massachusetts, employing the small world method. The letter could only be sent to a person with whom these people knew on a first-name basis. Afterwords, he analyzed the path of the letter. Among the letters that found the target, the average number of links was six [143].

There is no single definition that can describe social networks. Many researchers focusing on different aspects of social networks, have given different definitions. For example, Wasserman [149] and Hatala suggested that actor is a tied in social bond, where as other [121, 146] described actor is a node in a graph connected by edge.

Table 2.1 explores different concepts as described in literature that are necessary for the explanation of social networks.

2.3.1   Social Networks Sites

Social network sites are online social networking services that have attracted considerable amount of attention and curiosity in recent years from all quarters. It has emerged as a major medium of communication as it has provided a platform for sharing personal information with a vast network of friends. As of this writing,

Table 2.1: Different notions of social network

Actor	Relation	SN Definition	SN Examples	Ref(s)
An actor is a discrete individual, corporate or collective social units	A set of ties of a specific type; a tie is a linkage between a pair of actors	The finite set or sets of actors and one or more relations defined on them	Friendship among children in a classroom; all nations in the world and the formal diplomatic connections between them	[149]
Actors are also called points, nodes or agents	Relationships, edges or ties; one or more kinds of relations between pairs of actors	A set of actors that may have relationships with one another	Family; co-workers in a company; the network of neighbors; friendship among students in a classroom	[81]
People, organizations or other social entities	Relationships, such as friendship, co-working or information exchange	A set of social entities connected by a set of social relationships	Friendship among people; co-workers in a company; people who communicate with one another via computer	[72]
A node in a graph; each node represents a customer	The undirected, unweighted edges in the graph; each edge represents the connectedness between two nodes	An undirected, unweighted graph	Customer's social network which is derived from customer's interaction data from World Wide Web	[121, 146]
Actors are people or groups of people	Patterns of interaction or ties between actors	A social network is a set of people or groups of people with some pattern of contacts or interactions between them	Co-workers within a company	[84, 132]
The fundamental unit of a network, also called a site (physics)	The line connecting two vertices. Also called a bond (physics)	A set of items, which we will call vertices or sometimes nodes, with connections between them, called edges. Systems taking the form of networks abound in the world.	neural networks, metabolic networks	[109]

Table 2.2: Internet Traffic and User Engagement Report

Site	Global Rank [a]	Pageviews/user [b]	Time on Site (Minutes) [c]
google	1	18.2	17:57
facebook	2	12.2	19:54
youtube	3	6.61	16:30
yahoo	4	7.66	8:01
baidu	5	8.66	8:11
amazon	6	11.93	11:35
wikipedia.org	7	3.5	4:33
twitter	8	4.93	7:23
taobao	9	9.63	9:20
Qq.com	10	4.48	5:52

[a]An estimate of a site's popularity relative to all other sites
[b]Estimated daily unique page views per user
[c]Estimated time a visitor spends on the website per day (mm:ss)

millions of people across the globe share their contents online using these services. The social network sites play a very important role in current web applications, which accounts 4 out of top 10 sites according to statistics from Alexa[2], as shown in Table 2.2. It suggests the social network sites popularity relative to all other sites in the world, and also shows that people tend to spend more time on social network sites than most other sites.

The social network sites are very unique in a way of allowing user to show or advertise their social network to other. The flexibility provided by the social networks has given the people an opportunity to conveniently express themselves online through videos, photos, comments, and rating of online contents. Social networks also offer a platform to enrich the existing relationships and establish new relationships that would be almost impossible otherwise. Therefore, people are now using these services for building their social circle and sharing their contents far more than ever before.

[2] http://www.alexa.com on January 28th, 2015

2.4 Social Networks Analysis

The defining feature of social networks is the existence of linkage information. The concept of social network originated from the social network analysis and Mitchell defined it [100], A social network is the specific set of linkages among a defined set of persons with the additional property that the characteristics of these linkages, as a whole may be used to interpret the social behavior of the persons involved. Moreover, the definition does not just include the person as an entity of social networks but also involve objects and events or anything that can make the relationship with other entities.

Social Network Analysis (SNA) is a technique used to study the patterns of social relations among a set of social entities such as people, organizations or states [29, 149]. SNA measures the formal and informal relationship between social entities to reveal what facilitate or halt information flow. SNA is used to assess the structure of the network and identify the path and chains through which the communication happens. It differs from other methods in that its focus is on the relations between the social entities rather than the attributes of the actor (individual qualities of the actors such as age and location). Examples of connections between actors include attending the same college, giving donations to the same organizations, working in the same company etc. Examples of connection between other social entities include blogs on the same topic, videos or images covering the same event.

Social network researchers [29, 94] have identified that the relationship is the core of any kind of social network analysis.

- Social actors and their actions must be viewed as interdependent rather than independent or autonomous units.

- Relationships among actors the primary focus, while attributes of actors are secondary focus.

- The relationship between the actors are the channels for transfer or flow of resources. In other words, flow of social capital is heavily dependent on social ties.

SNA has been applied to a wide rang of disciplines, such as to successfully uncover criminals [65], in corporate partnership and professions collaboration networks to identify the missing connection between workers in organizations and finding out the influential or isolated individuals [38, 110]. In the medical field, risk network structure is analyzed in the early epidemic phase of HIV transmission in Colorado Springs to determining the relation of network structure to epidemic phase. A analysis of community-wide HIV/AIDS contact tracing records and paired partner information from other STD/HIV program records was used to augment network connections[120]. Social network analysis have been used in economics [80, 115, 121] in order to explain how market structure affects market dynamics by the adoption or diffusion of information.

2.4.1 Centrality

The concept of centrality is important to social network analysis. Centrality measures the interdependency that exists among a set of actors and identify the important actors that are located in strategic positions in the network [54, 149] . Three measures of centrality are described below

Degree Centrality identifies the most active actor that has most ties to the other actors in network.

Closeness Centrality identifies the actor that can quickly interact with all other actors in the network and has short communication paths to others.

Betweenness Centrality describes an actor as central if it lies between other actors on their geodesics, and the interaction between two nonadjacent actors depends on that central actor.

2.4.2 Limitation of Social Network Analysis

A number of limitations of social network analysis has been reported in literature. Here we briefly explain the three main limitations.

2.4.2.1 Graph-based Algorithms

The sociograms [102] is the graphical representation of social network, where people are represented by nodes and relationships by links between them. Sociograms was developed in early 1940's and most of the SNA research is based on a graph-based view of social networks and subsequently uses graph theory. The most important is the identification of sociometric features that characterize the social network.

The main problem with SNA is that its focus is to examine the existence of relationships between people and does not address different kinds of relationships. Furthermore, graph-based algorithms are most commonly used in SNA to examine the structure of social networks and uncover the informal connection between social entities. However, the problem with these graph-based classical algorithms is that they cannot represent the different aspects of human interaction without losing some knowledge [48, 49]. These graph-based representations are only concerned with syntax, they all lack semantics, and have a poor exploitation of the types of relations.

2.4.2.2 Limitation with Social Networks Data Collection

Collecting a data set that is rich enough to provide substantial social network analysis, requires immense effort. As the network captures all the social entities and their complex relations, it becomes more detailed and interesting, but leads to an exponential buildup of classification and characterization tasks. In the social network data collection process, there are many sources of uncertainty involved, and the traditional graph based algorithms do not take uncertainty into account. In traditional social network analysis, graph based algorithms are useful for determining mathematically derived facts about entities in the network. For example, the algorithm can compute the "degree centrality" for a node, which calculate the node connectivity and shows the importance of the node. However, the algorithm does not consider the certainty of the connections, or any type of metadata about the connections. Therefore, the uncertainty of the connections (links or relations) raises the questions about the validity of analysis results.

2.4.2.3 Problem with Self Promotion in Social Network

Everyone has a different way of perceiving his own social network, which makes it difficult to obtain an objective view. In many cases the self-esteem of individuals is quite high and they intend to perceive themselves centrally (important). A study showed [87] that almost nine out of ten social media using teens believe that people are over sharing (posting) on social media platforms. Sociology research has studied the phenomena of why and how people make friends, how many friends people have and how they depend on one and another for social support in social networks [83]. The online social network users have very low threshold for accepting friend requests, some accepting requests from acquaintances or even from strangers. This is to become more visible and promote themselves an imperative of their social networks. In this thesis, we encounter self-promotion problem by classifying friends into real friends.

2.5 Semantic Web Technologies

Semantic Web is an extension of the World Wide Web where both data and its explicit meaning can be effectively processed by computer programs [16]. Semantic Web is merging the existing Web technologies with knowledge representation formalisms in order to establish an infrastructure allowing data to be processed, filtered and discovered more effectively on the Web [57]. Semantic Web encloses the ideas for interoperability that go beyond the traditional programming. Ontologies define unambiguous formal semantics, which allows the common access to the information. The Web was designed with the vision that information will be useful not only for human-to-human communication, but also for machine to read and process effectively. Information on the net is developed for the human, not for the machine, which is a major obstacle for achieving machine-readable information. Instead of artificial intelligent approaches, where machines tries to act like people, semantic web brought the vision to develop languages for expressing information in a machine-readable form. Berners-Lee describes semantic layered architecture, which allows users and computer programs to write and share information in a machine-readable way. This will facilitate the development of a

new generation technologies and toolkits [16, 63].

2.5.1 Ontology

Ontology is a philosophical word used for categorizing and classifying objects in real world. In computer science, ontology was first introduced in artificial intelligent by scientist John McCarthy [95]. His idea was that for common-sense reasoning, we require common-sense knowledge and he called this ontology. In semantic web, ontologies are introduced to provide machine-understandable semantics. Ontologies used to capture the knowledge about a domain of an interest in the form of concepts and their relationships. Ontologies are being used in number of complicated application, e.g. e-commerce, knowledge management, Information integration etc. In such dynamic systems, ontology provides various functionalities, like storage or exchange of data from different ontologies and ontology-base reasoning which is gaining the main focus. According to Gruber [60], the ontology is a formal, explicit specification of a shared conceptualization. Ontology supports the need of heterogeneity, where different organizations communicate across the enterprise, sharing the formal, explicit specification of concepts. Ontologies can be categorized according to their usage. Gomez-Perez [57] presented many types of ontology; here we are considering a few of them.

General ontologies These types of ontologies describe the common knowledge that is reusable across domains.

Domain ontologies are used to formalize knowledge in precise domains.

Domain-Task ontologies describe tasks and activities in specific domain.

Application Ontologies model the knowledge of applications and cover all aspects of specific applications.

Different ontology languages are available, focusing on different aspects. In the follow session we discuss languages that are used for building ontologies.

2.5.1.1 Ontology Web Language (OWL)

The Web Ontology Language (OWL) [97] is suggested by the World Wide Web Consortium (W3C). OWL is a markup language formally derived from the DAML+OIL web ontology language [96] and mainly intended to achieve sharing, publishing and reasoning about information on the web. OWL builds on RDF [82] and RDFS[1] and provides additional vocabulary for describing concepts and properties (e.g. relations between concepts, cardinality, equality, richer typing of properties, etc). There are three species of OWL: OWL Lite, OWL DL and OWL Full and these are designed to be layered according to their increasing expressiveness.

In this work I have used OWL DL to represent my scenario.The objective in particular is to build an infrastructure based on OWL DL. Though OWL DL lacks in expressivity power compared with OWL Full, it maintains decidability and regains computational efficiency. The computational efficiency is an important feature since the mechanism has to handle scores of complex social, spatial and temporal metadata. OWL DL comprises all the OWL language constructs with restrictions and is based on Description Logics (hence the suffix DL). These are the decidable parts of First Order Logic [140] and are therefore amenable to automated reasoning. It makes sure that all its entailments are computable and the computations will be finished within a finite time. In order to achieve more expressivity and decidability, I use the Semantic Web Rule Language [114], which is designed as an extension of OWL DL, but this may come at the cost of additional complexity.

2.5.2 Social Network Semantic Web Layer

The well-known Semantic Web Layer Cake, described in literature, have been proposed by Tim Berners-Lee [16, 33]. The semantic web layer cake has reached a significant level of maturity and has been suggested as a standard architecture for semantic web applications by W3C. According to Tim Berners-Lee, every layer is seen as building block on the next layers below it. Here I propose a social network semantic web layer cake that is inspired by [16, 48]. Next, all the notions

---

[1]https://www.w3.org/TR/rdf-schema/

are explained in a simple manner illustrated in Figure 2.2.

XML is a markup language, which is intended for adding information to an existing document. RDF, RDF Schema and XML layers are very basic, but essential Internet technologies for Semantic Web. RDF and RDF Schema (RDFS) provides the framework to describe application-specific classes and properties. Classes in RDF Schema, allows RDF resources to be defined as instances of classes and sub- classes. We need ontology vocabulary on the top of the RDFS layer since RDFS vocabulary is not sufficient to handle domain knowledge. For this purpose the W3C Web Ontology workgroup has defined OWL (Ontology Web Language).

Logical reasoning is used to create the consistency and correctness of data sets, and these data sets are used for inferring important results. In other words, the logical layer enables the writing of different rules, which are further executed and evaluated by the Proof layer. The semantic reasoning is not only used to infer the hierarchy of classes of ontology or to check the validity and consistency of the OWL knowledge base, but it can also be used to deduce implicit knowledge on the bases of given explicit relationships. However, the expressivity provided by OWL is limited by a tree-like structure and implicit knowledge cannot be inferred from the indirect relation between entities. Hence, implicit knowledge is highly desired in online social networks because of their highly dynamic nature (new actors are joining, new relationships are building and the values of centrality keep changing). The semantic architecture of social network analysis makes use of RDF graphs to represent the social network. Some of the social capital (or social data) are already available with semantic meanings in form of (RDF, RDFa, ?formats), but most to the data is still not in RDF form, so that Wrappers and web 2.0 APIs are used to transform them to RDF format.

Many ontologies can be used to represent and to capture the rich social knowledge. The most popular ontology to model social network is Friend Of A Friend (FOAF [24]), as it is used to model people's identities, their relationships and their activities. The core of FOAF is to describe the user profile. For example, a person entity describes personal information properties (such as foaf:img foaf:familyName), social properties(foaf:knows, foaf:Group) and personal identi-
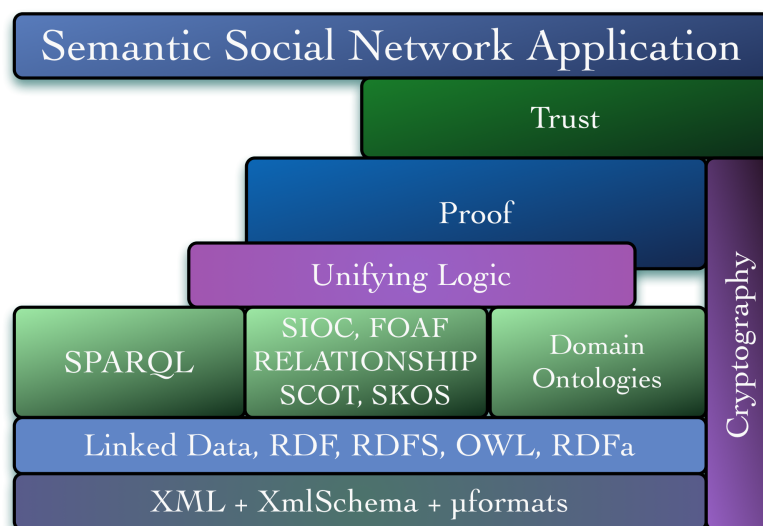
Figure 2.2: Social network semantic web architecture

ties (foaf:homepage, foaf:email). The RELATIONSHIP ontology[1] specializes the "knows" property of FOAF to deferent types of relationships in social networks. SIOC (Semantically- Interlinked Online Communities) ontology [23] is commonly used in conjunction with FOAF for representing social networking information. SIOC defines the generic concepts and properties that are needed to describe the information from online communities. The main advantage of using SIOC is that it follows the modular design approach, that is, its concepts and properties can easily be specialized and extended by other ontologies. Keyword based tags alone cannot establish the semantics of what is being marked-up, tags should have formal definition to be unambiguously classified and consequently to be able to infer new facts. Therefore, the ontology has been specifically designed for the social tagging [79] by using SCOT[2] ontology that is used to reduce the ambiguity of natural language.

---

[1]http://purl.org/vocab/relationship
[2]http://rdfs.org/scot/spec/

Chapter 3

User Behavior in Online Social Network

The work described in this chapter is related to the following questions

RQ1) What is the user behavior in online social network?

Users play the most important role in Online Social Network (OSN). Therefore, to understand the user is the first step to comprehend the OSN. Following questions are outlined to explore users in OSN.

   a) How important has online social networks become in people's daily life?

   b) Does people disclose personal information on OSN?

   c) What are the social activities users are most interested in on OSN?

   d) How does gender and age of the users play a role when making relationships/friends on an OSN?

Related Paper: Paper No. 1 [40] and Paper No. 7 [51]

A main contribution of this thesis is related to exploration of context information attached with images, mainly gathered from online social network. In the first

part of this chapter, I discuss the effects of online social networks on society in general. In particular, I take a look at how OSN has provided a platform to all, specially to females in highly conservative societies, and steps which inspire soft political revolution or change processes. The second part of this chapter is based on paper number 1 and paper 7. Paper 7 is about personal information disclosed on OSN and paper 1 describes a user study I conducted in order to learn user behavior on online social networks.

Online social networking services have attracted great interest in recent years. It has emerged as a major medium of communication as it provides a platform for sharing personal information with a vast network of friends. Every moment, millions of people across the globe share their contents online. On one hand, social networks are getting close to their assumed goal to facilitate interaction between the users around the world and therefore, bridging the social gap across the nations. On the other hand; rigorously compromising the privacy of the users [59]. A privacy breach resulted in negative repression for the users that even some time cast their life [47] in hostile situations. Therefore a cautious approach is necessary to build an individual's social network and sharing contents.

The user is the focus point of Online Social Networks (OSN), therefore it is necessary to study the user in order to design efficient algorithms to analyze social networks. In efforts to understand the OSN I investigated the following research questions

- What motivates people to join online social networks?

- What are the social activities users are most interested in online social network?

- How gender and age of the users plays a role of making relationships /friends over social network?

Online social networks are closely coupled with society. In this chapter, we will briefly discuss some of the affects OSN has made on society in the recent past. We will also present a perspective of a social network, that describes vague nature of friends' relationship and show the association of users activities with

gender, age and nationality. The results suggest that female are more conservative while accepting friends requests from stranger than males, and that the younger people are more active (and open). We also compare the social activities between two distinct nations - Norway and Pakistan. Our study revealed that despite of vast differences between these two nations, the online social activities are quite similar. The knowledge gained through this case study results is later used in the thesis to aid annotation, sorting and retrieving information (images) over social networks.

3.1 Effects of Online Social Networks on Society

It is unanimously agreed[1] that in civil societies, individuals and group of individuals have a right to their freedom of expression and freedom of the press. These freedoms are essential to build a strong civil society and are very critical in any democratic systems across the world. Social media delivers a platform to these individuals to connect with others and express their freedom. Social media is not only used as a tool for networking or for entertainment. It has inflicted almost every aspect of daily life, doing business, interacting with government, political participation and so on.

Historically the use of information as a power, was limited to the governments. Today, a blogger can impact an election, an image shared on social media can incite fear in the strongest of government [8] and all with very little capital investment and without the baggage of bureaucratic rules and nation values [105]. Many stats in Middle East control the information flow and censor the conventional media (TV, Newspaper) into government advantage [78], which make the role of social media more relevant since it is perceived more authentic and open.

In the following, I focus on the "Arab Spring" as one of the success stories of social media. Arab Spring refers to a series of protests against authoritarian regimes in Arab countries in Middle East and North Africa in 2010 and 2011. It was at the most part a non-violent moment compared to its volume and was therefore referred as a "Jasmine Revolution"[46] in the beginning. The jasmine revolution started in Tunisia, when a street vendor, Mohamed Bouaziz set himself

---

[1]http://www.un.org/en/universal-declaration-human-rights/

Table 3.1: Internet and Facebook users in Arab countries during revolution (2012 Est.)

Countries	Population	Internet Usage	Facebook User
Bahrain	1,248,348	961,228	413,200
Jordan	6,508,887	2,481,940	2,558,140
Lebanon	4,140,289	2,152,950	1,587,060
Oman	3,090,150	2,101,302	584,900
Qatar	1,951,591	1,682,271	671,720
Saudi Arabia	26,534,504	13,000,000	5,852,520
Syria	22,530,746	5,069,418	n/a
U.A.E	8,264,070	5,859,118	3,442,940
Yemen	24,771,809	3,691,000	495,440

on fire on 17 of December 2010 to protest the price hike and political repression. This event became a catalyst for the Tunisian Revolution, thanks to Facebook for playing a significant role in disseminating information and mobilizing the masses of protestors in Tunisia.

It is important to note that Mohamed Bouazizi was not the first Tunisian to set himself alight in an act of protest. Many of such cases occurred without any significant media attention [124], but what made a difference in this case, was the power of social media and that the **images** of Bouaziz were put on Facebook and everyone saw it [13]. It is also important to know the extent to which the Internet was available in Arab countries during the Arab spring. Table 3.1 , shows the degree of usage of Internet and Facebook. The Facebook user numbers is taken from Socialbakers [1], Internet user numbers is taken from Internet World Stats [2]

3.1.1 Social Media Political Effects on Authoritarian and Autocratic regimes

Social Networking Sites have changed the dynamics of political uprising in recent past, which we have witnessed in Arab spring. Zine El Abidine Ben Ali of Tunisia

---

[1] http://www.socialbakers.com/
[2] http://www.internetworldstats.com

ruled 24 years and Hosni Mubarak ruled Egypt for 30 years with iron fist. Their respective governments were toppled by young and inexperienced youth who had no prior experience in politics and certainly no arm support to the cause. Youth of both countries have challenged experienced politicians and toppled their governments. Social networking sites have provided opportunity to young activists to explore the new medium of communication to systematically diffuse information to build solidarity and identify goals for mutual benefits. The population pyramid graph in Figure 3.1 (the population data is taken from United States census bureau [1]), clearly shows the high demography of youth both male and female, where an estimated 60% population is under the age of 25 year. In those countries, youth was very much frustrated with the uncertain future and low employment opportunities. Wealth accumulation in certain hands has also played its role to bring regime change.

During this new form of political and social change, other governments were quick to learn. Many governments received the Arab spring as a wakeup call for social and political reforms. The Algerian government lifted the state of emergency nearly after two decades. Law making authority was passed to elected representatives by Sultan of Oman. Middle East oil rich states were forced to consider women empowerment, open up society, wealth distribution and better welfare facilities.

Another example where OSN has played vital role in politics comes from south Asian countries such as India and Pakistan. Aam Aadmi Party AAP (Common Man Party) of India and Pakistan Tehreek-e-Insaf PTI (Pakistan Movement for Justice) were once called a drawing room political parties, referring to the zero representation in parliament. In the presence of well-established political parties, these new parties did not have money for print or media. Hence, they started their activities on Facebook and Twitter, which appealed to thousand of followers. So much that PTI party leader Imran Khan was called by opponents "Prime Minister of republic of facebook-pakistan" suggesting that Mr.Imran Khan could become prime minister if and only if election would happen on Facebook. However, recent election held in 11 May 2013 surprised many when PTI emerged as Pakistan's second largest party, and managed to form the government in one of

---

[1]https://census.gov
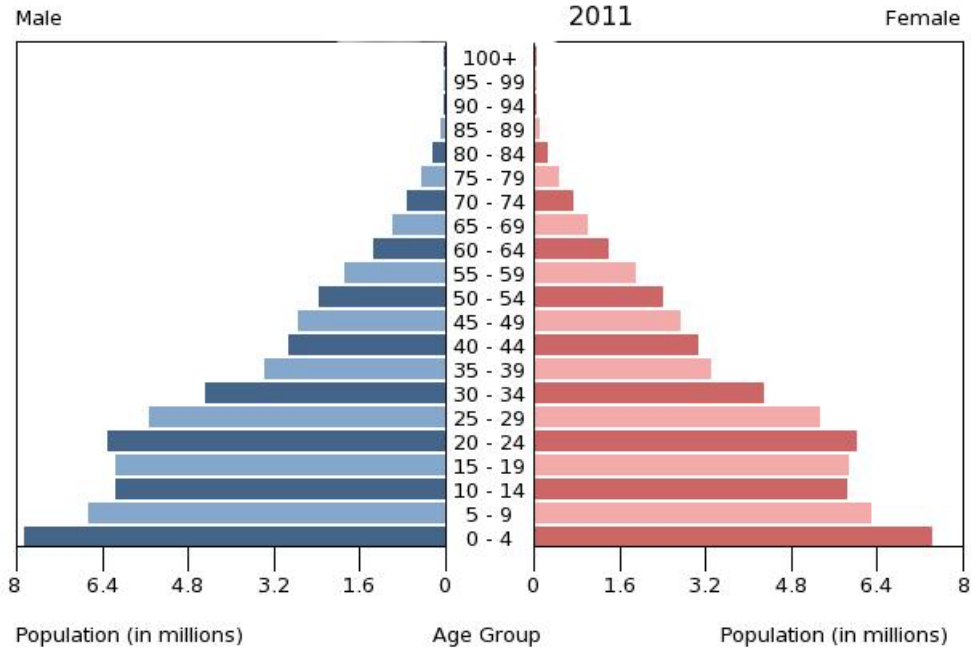
Figure 3.1: Population pyramid graph illustrates aggregated population of Egypt and Tunisia

four provinces. AAP led by Arvind Kejriwal also created history by winning 67 of Delhi's 70 seats, and formed government in Delhi. These accomplishments were attributed to their successful and well-planned social media campaign [6, 10]. The aforementioned examples show the power of OSN, which presented itself as a viable alternative to print and media.

3.1.2 Gender Equality and Social Networking Sites

Women's participation in politics on equal footing with men was denied rigorously and purposely to maintain historic supremacy of men over women in the Arab world. The term "gender" refers to the social differences and social relations between women and men. Gender equality means equal rights, responsibilities and opportunities of women and men. Social networking sites provided opportunities to women to express their political ideas and thoughts to larger audiences without any fear and restrictions. Women were thrilled to be participating in the

process while performing their domestic duties. Social networking sites have provided invaluable service to those societies where in the name of religion, culture and custom women's participation in politics were denied. Online social networks provide the opportunity to learn women's prospective that was largely absent in these countries. For instance, female Egyptian activists played a crucial role in the Egyptian revolution by effectively making use of social medial for formulating protests and connecting with other. At one point a female activist wrote [47]:

"If you can take pictures, take pictures . . . if you can use Twitter, send tweets . . . if you can blog, blog from the street. There are people demonstrating for our cause in Tunisia and Jordan, and I just found out that there are people demonstrating in Paris too. All of these people have faith in us."

The gender breakdown of Facebook users in Arab countries shows that the percentage of female users is at 33.5%, which is still lower than the global trend, where women constitute around half of the Facebook users [4]. Nevertheless Arab women take this as a window of opportunity towards women empowerment.

3.2 Steps of a Change Process

In this section, I will present the process of how online social networks have affected the political awareness and increased the participation of masses. I explain briefly how the technology inspired revolution happened in steps.

3.2.1 Voice to Voiceless People

This is the most important phase of the change process. At this stage masses were somehow aware of political repression and injustices. However, due to threats to their live, properties and not having enough skills to build common ground, online social networks answered most of the concern. Activists capitalized the advantage of modern technology to create a shared identity with defined goals without putting themselves to immediate danger.

In the initial stage, activists' friends and family made every possible effort to

shed considerable light, to attract the attention of masses to grave injustice done by the government. Primarily, mainstream media controlled by state and private channels ignored or overlooked these stories, but activists took the advantage of social media to spread information, to attract the attention and build solidarity for the common cause.

3.2.2   People Power Display

Youth of respective countries challenged the power of the mighty government by coming on the streets. OSN activities were an important driving force behind these street movements. Social activist campaigners coordinated with each other through social media sites; they uploaded images, videos, shared information and informed masses about events and activities on various localities. Other users on OSN, loyal to the same cause, contributed by endorsing and re-sharing contents uploaded by activists and by making comments which generated further conversation. These OSN discussion, counterargument and arguments in the defence of the cause, made people more engaged and committed to the cause itself. Eventually, people came out in the streets in huge numbers.

3.2.3   Global Attention

Images of atrocities committed by regimes shared through the OSN, appealed the consideration of global and regional powers, international governmental organization and non-governmental organization to act and condemn the action of the guilty party. At the end, we saw three different outcomes of the process. First, government fell to the demand of masses, as we witnessed in Tunisia and Egypt. Second, governments managed to pull off protesters and activists from the streets through force and concession, as was done in Bahrain. Third, governments used all the available and possible means of coercion to deny the basic rights of their nationals. On the other hand, people get determine for their just cause and enter into bloody civil war, Libya and Syria are the perfect example. Nevertheless, online social networks played a vital role to determine the basic human rights for the people, which was impossible to achieve otherwise.

3.3 Personal Information Disclosed on Online Social Networks

In this section, I present [51] the study about Patient Generated Contents (PGC) on social networks. Health information is by far the most personal information disclosed on social networks. This study presents the strata within this content and postulate some of the corresponding patient risks and ethical challenges associated with PGC found in YouTube video comments.

The internet is a main source of information and studies have found that most Europeans are using the Internet to search for health information [7]. Social Networks like YouTube, Facebook, and Blogger are among the most popular. The common characteristic of all these platforms is the availability of User Generated Content. This has lead to a scenario where patients are not only accessing content, but also creating new content (e.g. blogs, videos). Predominantly, this content is also used as a form of primary communication between the users, especially when commenting or evaluating (e.g likes) the content created by other patients. There are users who publish videos relating to their everyday life; including being blind, deaf or coping with chronic debilitating diseases. Many of these videos also contain comments from other users, mainly patients and relatives, which contain personal health information or medical advice. In the following section, we describe the characteristics of patient generated contents found within YouTube video comments.

3.3.1 Methodology

In this study, we queried the YouTube database for users who had published at least three videos in English about their everyday life with Multiple Sclerosis. Twenty-seven patients were found, and using the YouTube application-programming interface, user video and comment data were extracted. A total of 769 videos, 7047 comments and 2426 user profiles were collected (video creators and commentators). Using random selection, we selected 25 of these videos and analyzed their corresponding 557 comments. Comments which were not posted on health-related videos or were not written in English were excluded. A final

sample of 320 comments was analyzed and classified into the following categories:

- Personal health information: related to personal health experiences. Sub-categories: diagnosis, date of diagnosis, symptoms and medication.

- Video discussion: discussions about the videos (e.g. adding information or discussing the video topic).

- Appreciation: comments expressing gratitude for content or comments.

- Criticism: Video content criticism based on the content or quality of the information contained within the video.

- Unrelated comments: comments not covered in the other categories.

No ethical approval was necessary for the data used and gathered for this study. We used publicly accessible data that was voluntarily released by YouTube's users.

3.3.2 Results

The majority of comments met more than one category. In Figure 3.2 we see that most comments discussed the video (77%, n=247) or expressed gratitude for the author(s) (55%, n=177). A total of 22 comments met the unrelated comment criteria (e.g. **# KABOOOOM #**) and 8 criticized video authors (e.g. I'm skeptical you REALLY have MS). Overall, 70 comments (22%) contained personal health information concerning their creators or a third party (e.g. relatives). The comments with personal health information (PHI) were further classified[1]. As Figure 3.3 expresses, most of the comments contained information about medications (73%, n=51). Comments about symptoms (50%, n=35) and diagnoses (39%, n=27) were also prevalent. In one case, the information disclosed was the personal health information of a third party:

"I have been watching your videos since my daughter was diagnosed with MS on 28-12-07. She was diagnosed with an aggressive form of MS. Betaferon caused

---

[1]a comment can fall into more than one category
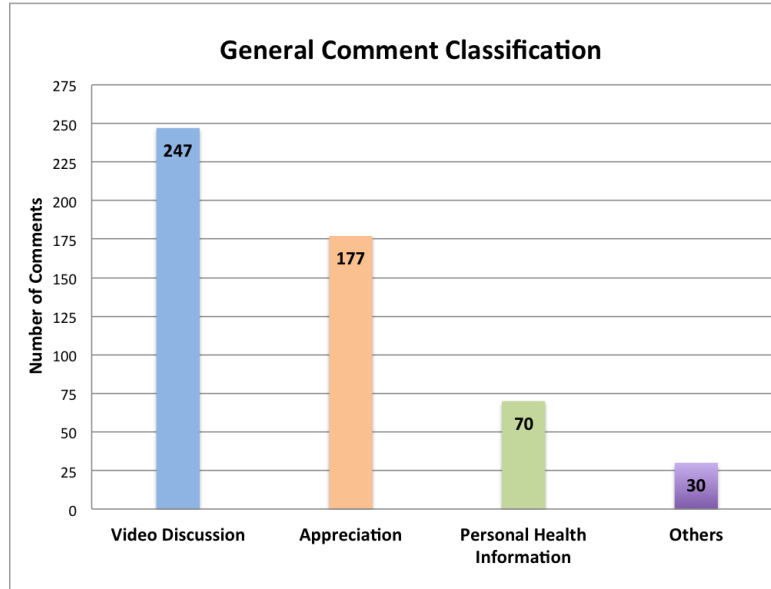
Figure 3.2: Total number of comments classified into the main categories

liver problems in a very short time. She has now had 4 infusions of tysabri and now feels she is well enough to try and go back to work. Tysabri only became available in Aust on July 1/08. This drug has given her hope that she still has a future to look forward to as she is 26 yrs old. Your improvement since your first dose gave us all hope."

Comments also encouraged other patients to take or change their medications; for example: *"im in a wheelchair is it worthwhile taking tysabri?"* The response by the video author was: *"In a word: ABSOLUTELY! It is amazing how much freedom I recovered when I started walking!"*

3.3.3 Discussion

Our results show that although the comments with personal health information had a low prevalence, the importance of the information disclosed within is extremely sensitive. Comments contained detailed user information. The public availability of these comments could constitute a risk to the user's privacy even if their comments were written under an "anonymous" alias. For example, a YouTube's nickname is often found in other OSN like Flickr or Blogger. This

Stratification of Comments with Personal Health Information
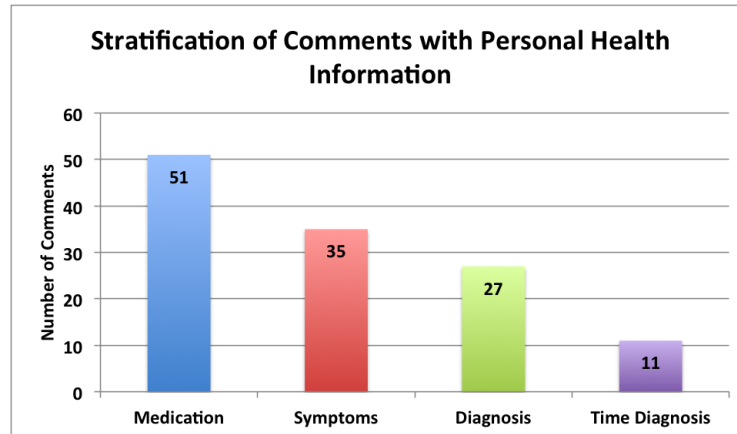
Figure 3.3: Total number of comments with personal health information by category

could lead a "curious" investigator to find other profiles in social networking sites, like Facebook, and consequently access family member information (e.g. the Facebook profile of the daughter diagnosed with MS). We also found that some of the commenters' YouTube profiles had been deleted, but their comments remained in the video database. This would make it nearly impossible to delete this public information. The public availability of this information could even have further implications that reach beyond our current ethical paradigms. For example, if the comment's author is identified, could it be used to deny health insurance coverage or job.

3.4   A User Study on Online Social Networks

In order to understand the user behavior and investigate the formation of relationships between users on OSN, I conducted a user study with tailored questionnaire. We began by investigating the overall trends in making friends in different gender. We identify the correlations between age of the participants and the number of friends and real friends they have. We explored the most popular activates among social network users and briefly examined the differences between two counties.

3.4.1 Methodology

We conducted a survey of random users of a social network. A URL invitation to participate in this survey was sent via email at the end of June 2011, and 247 people from 33 countries agreed to participated in this survey. In order to acquire diversity, we also conducted survey an in-person in Norway. To protect the individual privacy, we did not reveal the individual information. Out of 247 participants 26% were female and 74% were male. Our goal was to accumulate data from people from different gender, age, and nationality. However, people from age group over-40 were quite reluctant to participate in the survey.

3.4.2 Popular Social Networks and Motivation for Joining Social Networks

We started the survey by asking the fundamental questions, why people join social network services and which social network services they are using. An overwhelming majority turn out to be using Facebook shown in Figure 3.4 with 98%, and LinkedIn being second with 39%. This inspired me to conduct further experimental work on Facebook in this thesis. A large number of users were using multiple services at the same time. One of the reasons is that these services are built for different purposes, where Facebook primarily focused on social connections among friends and LinkedIn is used for professional networking. Another reason could be that some services are being obsolete and replaced new, such as Orkut being replaced by Facebook.

Most participants responded that they joined social network services to keep in touch with old friends and the people they already know offline. Social network services expand by allowing network members to send friend requests to non-members by email. People react to that email, which carries the link to the user profile, and are delighted to see the profile enriched with many posts, images and a whole group of friends. People like to contribute by making comments or liking the content, but it can only be done by joining the network. This finally motivates newcomer to join the network.

We have noticed that in Figure 3.5, a small number of participants responded that they joined the social network to meet new people. A large number of
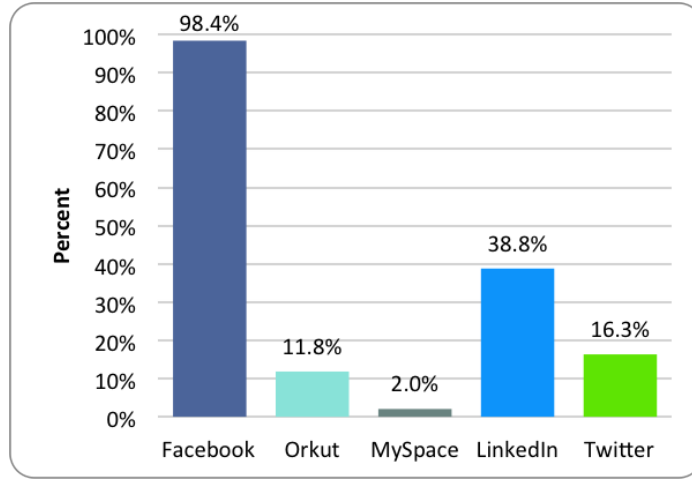
Figure 3.4: Shows the percentage of users consuming different OSN

participants responded that they joined to keep in touch with old friends. We believe that the media coverage of social networks was gigantic and that prompt many tenderfoot to notice and try new medium of networking. Later they bring whole group of friends to the network, for reason mentioned in Figure 3.5.
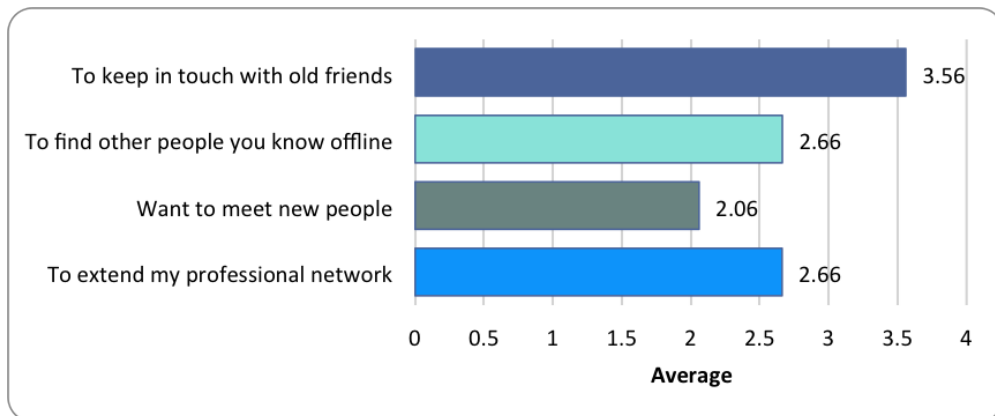
Figure 3.5: Reports the motivation behind joining social network, average satisfaction of user from scale 1 to 4

3.4.3 Significance of Social Relationships in Online Social Network

Online social networks allow anyone who claim to be of a certain age (in most cases 13 year or older depending on networks) to become member and setup a profile with a valid email address. Once the profile is created, the user can send friendship requests to other users on the same network. If a user approve a friendship request, her profile appear in the friends list. That is how the user builds their network of friends online. Once the relationship is build, user information is shared with other users. Given the vast array of information, such as birth date, location, full name and phone, concerns regarding privacy and security are regular issues [62]. Therefore, friendship or relationship built over online social networks can cause unwanted disclosure of private information.

Social relationships are usually categorized as weak or strong, but are in reality exceptionally diverse in terms of how close and inmate one perceives to be. Danah Boyd [21] reported the online social network relation in binary terms "Friend or not", he further elaborates that there is no way to determine what metric was used or what the role or weight of the relationship is. While some people are willing to indicate anyone as friends, others stick to a conservative definition. Most users tend to list anyone who they know and do not actively dislike. This often means that people are indicated as friends even though the user does not particularly know or trust the person [21]. Thus, there is no refine line that separates friends and not friends, however, in this chapter, we consider real friends, who are known offline and keep strong tie.

3.4.3.1 The Affects of Gender on Social Relationships

To find out how people make friends over social networks, we asked participants to respond to five questions, illustrated on x-axis in Figure 3.6 on a four-point scale where 1 equaled "strongly disagree" and 4 equaled "strongly agree".

Participants reported high confidence in people they already knew offline with male-mean 3.2 and female-mean 3.8. Males agree (mean = 3) to accept friend of a friend, while females agree less (mean = 2.3). When it comes to making friends with strangers, both genders showed disapproval, with the strongest disapproval
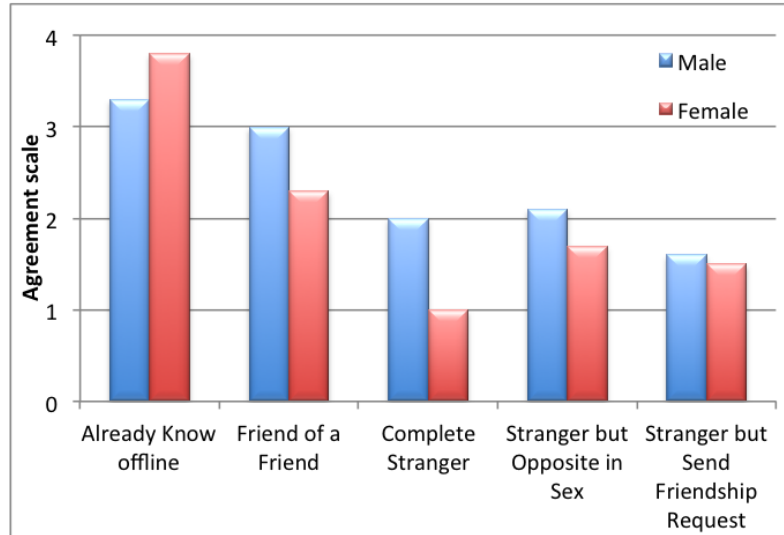
Figure 3.6: Illustrate the difference between male and female for making friends.

from females (male-mean=2, female-mean=1). It is, however, interesting to note that the approval ratting is slightly higher in case of friend request from opposite sex.

Nevertheless, females appear to be more careful while making friends as compare to males. It is noticeable that the female conservativeness in our results very much correspond with another study [59], where they found that females are substantially more conservative for reporting their phone number on the user profile. Specifically, single males reveal their phone number in even higher frequency than the single females. It also agrees with the results published in [85] where they suggests that users are largely using Facebook to learn more about people they meet offline and are less likely to use the site to initiate new connections.

3.4.3.2 Is Age a Factor in Relationships

Users were reluctant to reveal their accurate age and therefore, we formed four different age groups. The popularity of social networks among young is demonstrably high. Our survey revealed that younger people were more active on social networks. Figure 3.7 illustrates that the young people are more open to make friends online even if they are strangers. This trend declines as the age increases.

People in age group over 40 are very reluctant to accept strangers as their friends.
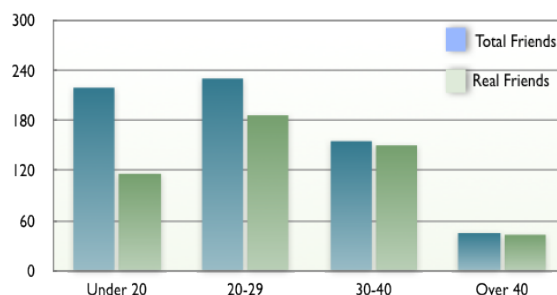
Figure 3.7: Shows the association between age group and number of total and real friends.

It is noteworthy to compare this result with the speculative theoretical limit, that you cannot have more than 150 friends on your social network, so all the people beyond this limit cannot be actual friends [39]. Our results show that the participants of age group 20-29 do exceed this theoretical limit.

Our results correlate with a research conducted about the information disclosure in online social networking profiles [112]. There, as age increases, the amount of personal information in profiles decreases. It implies that older users demonstrate mature behavior and protect their private information online as compared to the younger users.

3.4.4 Social Activities in Online Social Networks

As observed in this survey, as well as in earlier research [91], respondents are engaged in a variety of activities on OSN, as show in Figure 3.8. Users login to OSN to make comments, to check out the feed or status of friends, to see contents such as images or videos share by others and to use the email services embedded in OSN.

We attempted to determine behavior differences between different countries. People from 33 countries responded to our survey. But, due to various reasons, we chose Pakistan and Norway for comparison. One reason was that the biggest chunk of sample data was reported from Pakistan (31%) and Norway (22%).

Figure 3.8: Importance of social activities across two nations.

Another reason was that these countries are very different with respect to socioe-conomic status, political freedom and cultural differences [61, 68].

To our surprise, participants from both countries show similar behavior, which is contrary to the general viewpoint that social networking is used differently in different countries. However, the notable exception was use of email or private messages and comments. Norwegian users are more interested in making use of private messages than Pakistani users (mean 2.3 vs. 1.5) and comparatively less interested in comments on social content (mean 1.8 vs. 2.2). The importance of different social activities is illustrated in Figure 3.8.

It is noteworthy that 2.9% (5 265 880) of the Pakistan population use Face-book compare to 53.64% (2 508 380) of Norwegian population [1]. This 2.9% of Pakistani population is mostly urban class. This might be one of the reason why we see a striking resemblance between the social activities in the two coun-tries. Other reason is the platform dependency, which drives users to engage in particular activities.

---

[1]http://www.socialbakers.com/facebook-statistics

3.4.5 Discussion

We found that users have many individuals in their online social network that they do not consider a real friends. Thus, the assumption in some online social networks, such as Facebook, that if "x is friend of y" then "y is friend of x" is erroneous. Furthermore, many users do not have a very high threshold for accepting friend requests from strangers (or acquaintances), and they often accept request to become more visible and for self-promotions.

"Unreal friends" are one of the major reasons of erroneous results in SNA. Our study reveals that the number of unreal friends is inversely proportional to the users' age. We conclude that users with age above 30 have negligible "unreal friends". While analyzing the relationship the users above age of 30 can be safely neglected. An implication of having many unreal friends, is the amount of potentially not-so-interesting information received. A person using OSN often receive more than hundreds of news feeds, media contents, and comments per day, many of these are of limited value.

Another implication is user leaving online social network. Like offline social networks, people depend on one and another for social support and, similarly, cannot stand what they do not feel comfortable with. It is sometimes difficult to avoid the contents, as the publisher (friend) expects feedback. Consequently, many leave the online social network.

Contrary to the results reported in Golder's work [56], that only a small proportion of communication was through private messaging between friends, we found that the private messaging is significantly popular on social networks. This trend probably has emerged in recent years. This also implies that friends no longer feel the need to use email specific services for exchanging private messages. Researchers analyzing communications through social network between users, usually neglect the private messages.

Use of OSN raises some privacy concerns, users have little control and lack the competency to establish robust privacy settings for the profile. Prior research [91] has shown that users tend not to change default settings, even when concrete privacy invasion was experienced by users [36]. For instance, on Facebook default privacy settings, if a friend makes comments on your photo, a friend of the

friend can see the whole album. The habitual use of online social networks and its deeply integration into our daily routines, seems to outweighed the privacy concerns. Participants, aged under-20 is particularly vulnerable, since they have many unreal friends and disclose personal information for self-promotion that can be used or abused by third parties.

3.5   Conclusion

Online social networks play a vital role to support basic human rights for people, which is impossible to achieve otherwise. The aforementioned examples and the recent 2016 election in United States [116] showcase the power of OSN, that present itself as a viable alternative to the print and media.

We have examined user behavior in social networks and discussed some of its potential implications. A study on personal information disclosed on OSN has shown some of the risks users may be subject to, when disclosing personal medical information. It is particularly important to understand patient's intents and awareness on risk when disclosing personal health information. This research may also contribute to the formation of health consumer guidelines for safety on OSN, for example, not reusing web aliases in platforms where personal health information is publicly disclosed.

A study about general usage of OSN reveals that females are more reluctant to accept friend requests from strangers. We also reveal that the young users are liberal in broadening their social circle. We have compared the results from two contrasting countries - politically and socioeconomically, and the comparison shows the users' behavior is quite similar. This study is small and cannot be generalized, but it highlights the need for further research.

Chapter 4

Image Annotation by Leveraging the Online Social Network Analysis

The work described in this chapter answers the following questions

RQ2) What is the meaning of context and how can social networks be explored as context of images?

RQ3) How can one reuse existing annotations of images from members of one's social network as a base for image annotation?

    a) Can an online social network context be used to automatically provide personally relevant image annotations?

Related Paper: Paper No. 2 [43], Paper No. 5 [75] and Paper No. 8 [41]

This chapter is based on two important assertions from previous chapter 3, that viewing images on social networks is an important activity and not all friends are real friends. The second assertion is significant since I am using voluntarily available image annotation from friends to automatically annotate new images. To

avoid overhead, we rely on real friends and discard others. Paper 2 were the most central paper for this chapter, while paper 5 and paper 8 were supporting papers exploring and investigating how image metadata combined with geo-location and date/time metadata improve automatic image annotation.

In the first part of this chapter, I will discuss social networks as context of an image and in the second part discuss the graph-theoretic techniques of social network analysis. Based on these techniques, I have developed an algorithm for image annotation. The algorithm is implemented and tested on real and randomly gathered data.

4.1   Image Annotation

Annotations are observed as statements about the image created by users. Furthermore, it is not only used for describing the image content, but also to classify, organize and search images. Annotations are known to be a key factor in image retrieval systems [69]. However, manual image annotation is a time consuming and cumbersome task, yet extremely important for image management and retrieval.

This work is based on Text-Based Image Retrieval (TBIR). TBIR is achieved by matching a text query to annotations associated with images. The drawback of this approach is that manual annotation of an image is a resource-intensive task and is inherently expensive because users are unwilling to put large amount of effort to correctly annotate images. Even if human annotators are willing to annotate images, their annotation is highly subjective because humans perceive the content of an image differently, and they also make mistakes. For instance, an annotator in our lab once misidentified a building in Oslo as being in Sweden. Whereas, in automatic annotation this problem can be avoided by simply using image GPS coordinates to determine location.

4.2   Social Networks for Image Annotation

Due to the resource intensive nature of manual image annotation, a question arises about automatically acquiring relevant annotation while avoiding the unnecessary

and malicious annotations. Also, images taken by digital cameras typically include only basic metadata such as time/date and GPS coordinates, which alone are not sufficient and meaningful enough to correctly annotate images.

Hence, in order to cater both of these issues, it is imperative to find alternative ways to annotate images. Leveraging the annotation from the context of image can allow us to make meaningful connection with image contents [138, 139]. So instead of using image-processing techniques to manipulate and interpret the image [111, 117], we use context of image. Therefore, the context of an image is of central importance in my approach towards image annotation.

For the context of an image, I consider several factors like geo-location, time and relationship among actors in social networks.The purpose of my system is to automatically generate annotations for an image by leveraging the annotations provided by the most influential user in social network (also known as the central actor). Social network is a set of trusted people which are linked to each other. Their relationships are defined and this helps in acquiring relevant image annotations. More importantly images and provided annotations are trusted within a group.

In next section, I describe social networks, characteristics of a central actor and highlight the importance of relevance of annotations within social groups.

4.2.1 Relevance of Social Networks for Image Annotation

During this study, I have observed that annotations can be categorized in two types, namely: public annotations and personalized annotations.

Public Annotation   refers to annotations which are very generic in nature and relevant to masses. Annotations like "Eiffel tower" and "Oslo opera house", can be considered public annotations as they do not relate to any specific person and, hence, can be used by anyone. Table 4.1 shows more example of both annotations.

Personalized Annotation   is annotations that are relevant to a person or a specific group of persons. For example, annotations such as "birthday party in Los

Angeles" and "holidays in Tromsø" can be considered personalized annotations as they relate to specific events and people.

In a certain social network group specific events are known to the whole group and this leads the whole group to create similar annotations in most cases. We have also observed that personalized annotations can often help most users in a social network group to annotate their images. In other words, users within a certain group may find semantic agreement on personalized annotations. Furthermore, we have found personalized annotations as a salient source to aid our automatic image annotation technique for an actor within a social network group.

4.2.2 OSN Context Features for Image Annotation

Context as a broad concept is defined in chapter 2. Here I revisit this concept with illustration in Figure 4.1 . I classify context into, *Usage Context* and *Capture Context*, representing different types of image environments.

Capture Context mainly describes when and where the image was taken. This information is extracted from the digital camera in the form of EXIF (EXchangeable Image File[1]) metadata. EXIF metadata is associated with image that includes hundreds of tags, among them many are related to the image-capturing conditions. EXIF metadata can be used in several ways, for instance; focal distance, focal length and sensor size can be used to measure the size of object at the focal point and this can help us to differentiate between different objects. It may include context information such as date, time, location, sensor data, information about an area or objects in the image.

There are many types of Capture Context information that potentially can be added to an image, however not all context information is relevant to an image. The required context information will change depending on availability of context sources and the usage of the image. Hence, in this thesis, I mainly rely on spatial and temporal information.

---

[1]$http://www.digicamsoft.com/exif22/exif22/html/exif22_1.htm$
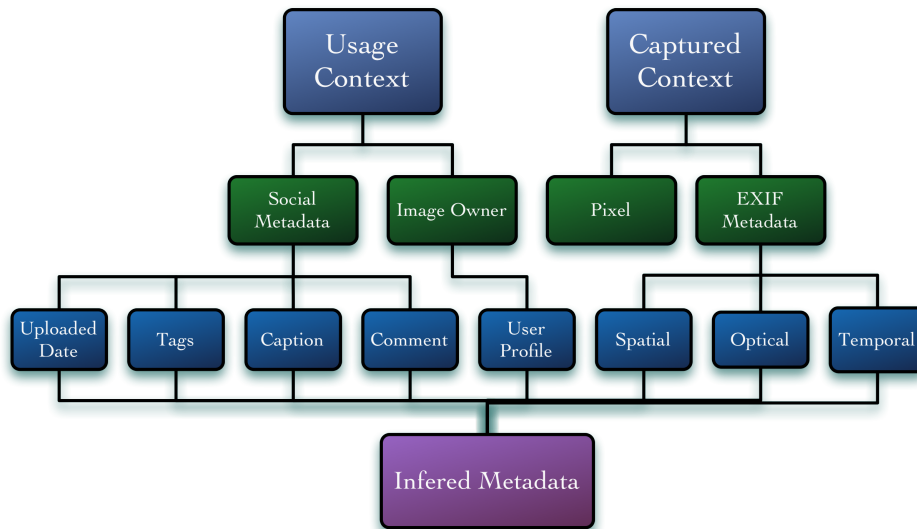
Figure 4.1: layered classification of image context

Usage Context Usage Context describes the surrounding environment as to where and how the image is being used. For instance in a document where the image is used as an illustration or in an image collection. Context information can be a textual description of a collection, a document abstract, tags describing the collection or document, or text surrounding an image.

In this thesis, I focus on social networks as Usage Context for images. I observe that the most useful feature of social network is the contributions from other members of the network in the form of comments and likes. Therefore, social network activities over images such as *comments* and *likes* emerge as a key Usage Context information. Comments are considered a collection of arbitrary keywords, which may not necessarily portray the image content. Nevertheless it is a discussion about the image. In the later part of this chapter, our results demonstrate that taking frequent keywords from comments are helpful for automatic annotation. I have also used numbers of comments as an important indicator for image retrieval in the Chapter 5.

4.2.3 Relevant Image Annotation from Social Network

Usually people with similar interests and activities happen to have mutual understanding among them. Likewise, group of like-minded people within social networks frequently has correlated activities and common interests, and this can also lead them to correlated annotation of images. Annotation of one person can be used as a basis for annotating other peoples' images. Though it is true that harvesting more annotation data from the entire group can yield more information, but it can also create noisy and unnecessary annotations. If, for example, within a social network, to annotate an image captured at "Eiffel Tower" would yields thousands of keywords (i.e. annotations), but then finding anything useful will eventually become a predicament. Such noisy annotations may not allow us to develop a robust image annotation framework for social network.

Peter A. Gloor [118] has stated that in social networks, some actors have more influence than others and can provide us with the most relevant information. Therefore, I decided to take annotations from influential actors. I call such user a central actor and by using the following three perimeters I estimate the influence of the central actor. So a central actor in my case is the one that

- Has high centrality in a given social network (explained in 4.3.2)

- Has higher number of images with annotations relevant to the target actor[1]

- Has high weighted relationship with target actor

The image annotations of this central actor, paired with analysis of geo-reference and time difference can certainly improve the validity of the results.

4.2.4 Use Case Scenario

In social networks, users actively participate by creating, sharing and communicating digital social artifacts. A significant fraction of these artifacts exist in the form of images, often with metadata. In the following, a use case scenario, illustrated in Figure 4.2, can help us understand the big picture. This use-case

---

[1]Target actor is the one whose images are going to be annotated.
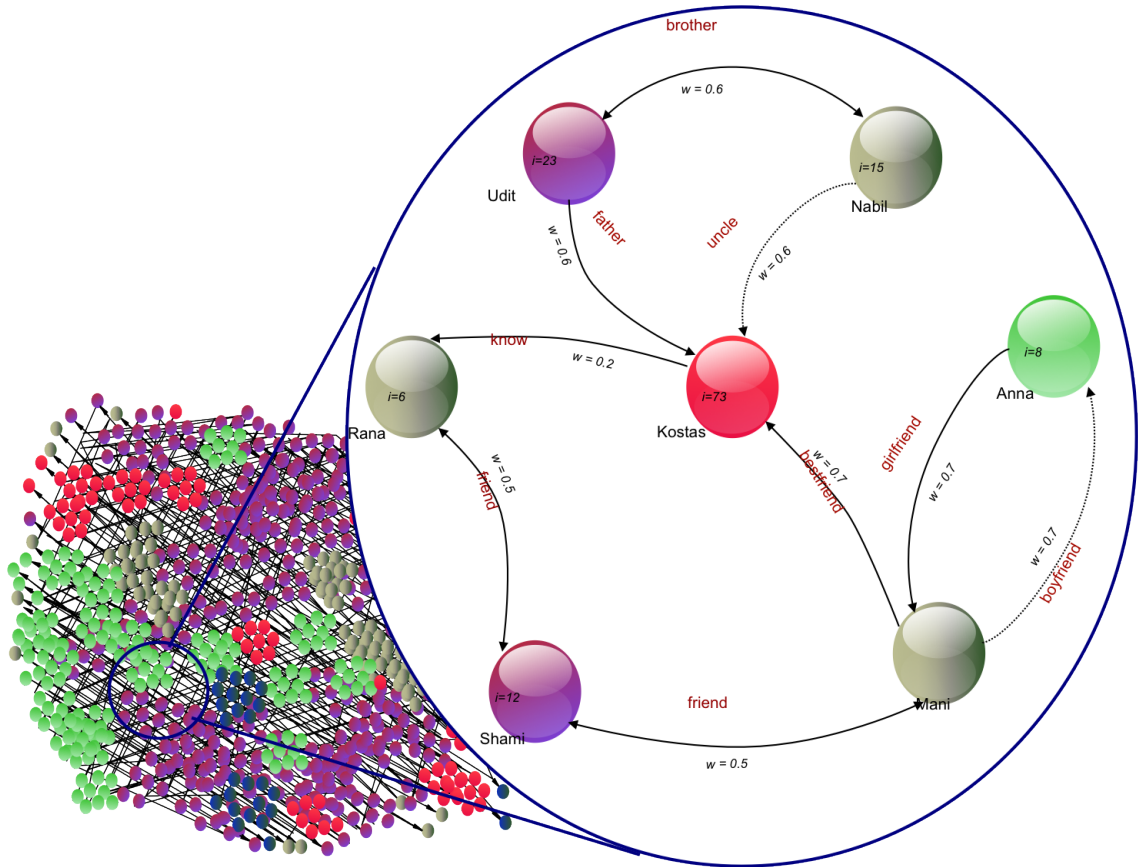
Figure 4.2: Circle illustrates the subset of a social network, "i" is the number
of images and "w" represents weights assigned to the relationships. Dotted lines
show the inferred relationships

is a story of three actors within a social group; Anna, Mani and Kostas. They
are friends and they decided to go to Paris on the occasion of 121st anniversary
of Eiffel Tower. Kostas is a very passionate photographer and he took a lot of
pictures of the Eiffel Tower, group photos, individual photos and photos of Paris
city as seen from the Eiffel Tower. After the tour ended, Kostas created an album
with a caption "Visit de Eiffel Tower" on his social network profile and also he
annotated all images with names of people depicted in the pictures, main activ-
ities, weather and etc. Mani also created an album, consisting of few pictures,
with a caption "tour de Paris". All three actors are active members of a social
network and they are connected to each other through social relationships.

Kostas is the one with most images with annotations relevant to Mani's recent activities. Our system, analyzes that Kostas is an influential member of the social network, has images with annotations which are closely related to Mani's images based on time and geo location, and Kostas and Mani are close friends. Therefore it suggests annotations to Mani. These annotations can contain items like buddy list (people annotated in Kostas' pictures), weather condition (e.g. they might be using their own particular vocabulary like "sunny", "breezy" and etc) and main activities (e.g. watching concert, partying, dancing and etc). Mani, upon receiving these annotations, is pleased to see the suggested annotations as they seem very relevant because they are indicating the same event, same activities and same group of people. Mani accepts the suggested annotations and uses these to annotate his pictures. Mani is pleased because of the fact that he did not have to manually annotate images from scratch.

4.3   Graph Theory for Social Network Analysis

In social network analysis, graph-theoretical concepts are used to understand and explain the social phenomena. A social network consists of a set of actors, who may be arbitrary entities like persons or organizations, and one or more types of relations between them. Graph theory is widely used to define the relationship between different entities in many application fields. Basic concepts of graph theory are almost the same, and then different applications have various restrictions and requirement such as, directed verses undirected graphs, weighted verses binary graphs. In this section, I begin by introducing common notations and fundamental concepts of graph theory. I closely adhere to [12, 22] however I have not included the concepts that are not used in this thesis.

4.3.1   Graph Theory Terminology

For the simplicity and in accordance with online social networks, we assume that all the networks under computation are graphs $G = (V, E)$, which consists of a set V of vertices (also mentioned as nodes of graph and actors of social network) and a set E of edges (also mentioned as ties or relationships). An edge $e \in E$

connects two vertices $a, u \in V$. The vertex a is stated as adjacent to u and e is incident between a and u. The adjacent vertices $a, u \in V$ are called neighbours.

Let $w : w(e) \rightarrow \mathbb{R}$ be a weight function on the edges $e \in E$. An incident weight function assigns each element $e \in E$ a real numeric value $w(e) \in \mathbb{R}$ and we assume that $w(e) > 0$ for weighted graphs . Weights are used to measure the strength of a link.

4.3.2 Center Actor of Social Networks

Social networks assume that actors are interconnected [76] and that connections or relationships have real influence on the way they consume each other's resources. In order to analyze the relations and discover the most important actor, we first assign the weight to the relationships $w(e) > 0$. Calculating the weight of each relationship can draw us near to the most influential actor. However, B Aleman-Meza concluded [5] that altering the weights and thresholds alone do not improve the semantic analysis of social network results. Therefore, in order to infer the influential actor and relationships among individuals in social networks, we took advantage of SNA (Social Network Analysis) techniques that provide a set of mathematical and algorithmic approaches to deduce such information [149]. Particularly we used the SNA concept centrality that provides measures of the interdependency that exists among a set of actors as centrality identifies the most important actor of the social network [48, 54, 76].

To find out the most influential actor in social networks, Wasserman [149] suggested measures that can help us identify most influential actor namely: degree centrality, closeness centrality and between centrality.

4.3.2.1 Degree Centrality

There are several measures to capture variations on the notion of a vertex's importance in a graph. To explain the graph-theoretical perspective we being with the centrality measures and explain how they can capture vertex's importance in a graph.

> `Definition` : Let $G = (V, E)$ be an undirected, weighted graph with edge strength w. The degree centrality $C_D : V \to \mathbb{R}$, of vertex $a \in V$ is defined as the sum of the strengths of its incident edges, normalized by the maximum possible degree $(V - 1)$

$$C_D(a) = \frac{\sum_{\{u,a\} \in E} w(\{u, a\})}{(V - 1)}$$

where $w(\{u, a\})$ computes the strength of edges $\{u, a\} \in E$ adjacent to the vertex $a \in V$. The degree centrality [122] identifies the most active actor that has most direct ties to the other actors in the network. Such actor grasps a structural position in a network that serves as a source for information exchange. Degree centrality is a local measure which considers only the adjacent vertexes. It is useful to recognize that every edge is a walk of length 1.

4.3.2.2 Centrality Based on Shortest Path

This centrality measure is based on the assumption that an actor is more central when the distance to other actors is short.
The *closeness centrality* [125] identifies the actor that can quickly interact with all other actors in the network and has short communication paths to others. Such actor can be reachable without going through many intermediaries.
We define a path from $a \in V$ to $u \in V$ as an arbitrary sequence of vertices and edges, starting with a and ending with u in a strongly connected graph. Let $d_g(a, u)$ denote the distance between two vertices a and u. The minimum distance between two vertices $d_g(a, a) = 0$ and $d_g(a, u) = d_g(u, a)$ by convention.

> `Definition` : Let $G = (V, E)$ be an undirected, weighted graph with edge strength w and path length l. The closeness centrality $C_C : V \to \mathbb{R}$, of vertex $a \in V$ is defined as the inverse of the sum of distances to all other vertices where $a \neq u \in V$

$$C_C(a) = \frac{1}{\sum_{u \in V} d_g(a, u)}$$

The *betweenness centrality* [53] of a node is a measure based on the number of shortest paths that a vertex lies in. It describes an actor as central if it

lies between other actors on their geodesics and the interaction between two nonadjacent actors depends on that actor. Such an actor also called a broker.

> `Definition` : Let $\sigma_{su}(a)$ denote the number of shortest paths from s to u that some $a \in V$ lies on. Let $\sigma_{su} = \sigma_{us}$ denote the number of shortest paths from $s \in V$ to $u \in V$, where $\sigma_{ss} = 1$ by convention. The following are standard measures of centrality:

$$C_B(a) = \sum_{s \neq a \neq u \in V} \frac{\sigma_{su}(a)}{\sigma_{su}}$$

Thus, $C_B(a)$ is an index of the degree to which s and u need a in order to communicate along the shortest path linking them together.

Each graph-theoretic method is defined for graphs of specific characteristics and we can assign weights to the different centralities according to their importance. We calculate the actor centrality (total centrality C_T) as a sum of all the centralizes normalized by 3.

$$C_T(a) = \frac{C_D(a) + C_c(a) + C_B(a)}{3}$$

4.4   Images Close in Time and Space

During testing, I observed that image annotations of central/relevant actor, corresponding with analysis of geo-reference and time difference improves the validity of the results.

Time is usually used effectively to classify and manage digital images. The simplest geo-reference annotation take placed, first, by finding out images in a certain radius of associated GPS coordinates of target image and then finding out the common keywords in these recently found images. If two images shared by two different people, have short time difference and are geographically close according to GPS coordinates, it is most likely they capture the same event and have strong correlations between annotations.

Several functions are assumed, $getActors(SN)$ returns the actors list from the social network.

The functions $verifyTimeLimit(pic.time)$ and $verifyCoordinatesLimit(pic.coords)$

Algorithm 1: Relevant Actor

Require: $relevantActor$ is a list of actors and SN is a social network of user

$relaventActor \leftarrow empty\ list$

$relevantActorPic \leftarrow empty\ list$

for all $actor \in getActor(SN)$ **do**

$\quad pics \leftarrow getPic(actor)$

$\quad$**for all** $pic \in pics$ **do**

$\quad\quad$**if** $verifyTimeLimit(pic.time)\ and\ verifyCoordinatesLimit(pic.coords)$ **then**

$\quad\quad\quad relaventActor.add(actor)$

$\quad\quad\quad relevantActorPic.add(actor, pic)$

$\quad\quad$**end if**

$\quad$**end for**

end for

for all $actor \in getActor(relaventActor)$ **do**

$\quad$**if** $verifyRelation(actor)$ **then**

$\quad\quad$**return** $actor.relevantActorPic$

$\quad$**end if**

end for
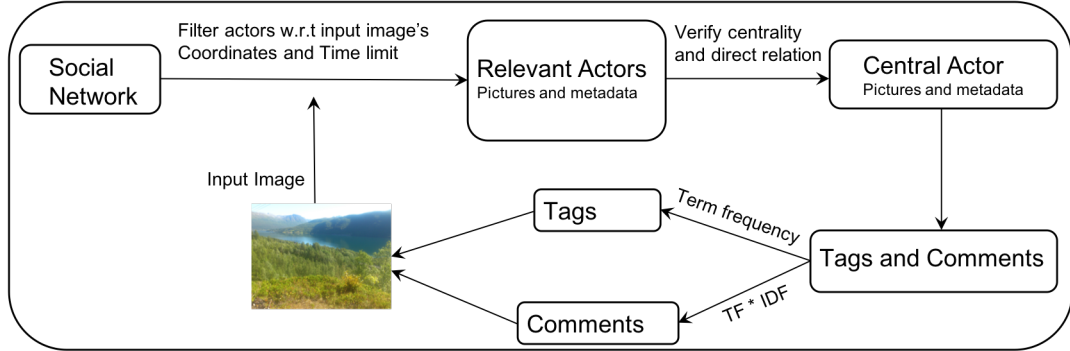
Figure 4.3: Algorithm illustration of Image Annotation

verify and return True if the pictures time and coordinates are sufficiently close to a specified time and GPS coordinate position, respectively. The function $verifyRelation(actor)$ returns True, if the given actor has centrality and direct relationship values above a given threshold. The $relevantActorPic.add(actor, pic)$ is a add function with the actors as keys and an array for each actor containing a list of pictures associated with that actor.

4.4.1 Annotation from Tags and Comments

Aforementioned algorithm returns the set of $releventActorPic$ along with associated tags[1] and comments. Following figure 4.3 illustrates annotation process in steps. As an input, an image with social network profile of a user is given. On the basis of image coordinates and time difference, relevant actors and their images along with metadata are selected. Among relevant actors, an actor who has high centrality and high weight of direct relation is selected. Tags and comments are extracted from the already selected images of central actor. I use the two different techniques TF and TF*IDF are used to extract the annotation for a given image. I extract two top ranked tags from each technique.

Term Frequency for Tags

TF (Term Frequency) weight [128] is used which is common practice in IR (Information Retrieval) to represent the relevance of term in a document. Terms

---

[1]Album name and people name were also considered as a tags.

that appear often in a document should get high weights. In our model, we use the term frequency to find out the relevance of keywords which are used as a tag for other pictures (pictures close in time or space). So the high frequency tags were selected for annotation. The intuition behind is that, tags that appear often in a closely (time or space) related pictures should get high weights.

$TF_{(d,t)} = log(1 + n(d,t)/n(d))$

Where as $n(d,t)$ is the number of occurrences of term t in all the associated tags with $releventActorPic$, and $n(d)$ is the total number of associated tags with $releventActorPic$

TF*IDF for Comments

Since the comments, unlike tags, are sequence of meaningful keywords, that can be dealt as a document. Therefore, in my approach, all the comments associated with $releventActorPic$ are consider as a paragraphs of a single document. In order to infer the suitable tag from comments, both term frequency and inverse document frequency measures are used. IDF of a word is the measure of how significant that term is in the whole corpus. For a term t in a document d, the weight $W_{(t,d)}$ of term t in document d is given by:

$W_{t,d} = TF_{t,d} * log(N/n_t)$

Where as $W_{t,d}$ is the weight assigned to term t in document d and $TF_{t,d}$ number of occurrence of term t in document d. N represent the number of documents in entire corpus and n_t is a number of documents with term t.

$W_{t,d}$ weights were represented in Vector Space Model (VSM) and In this approach, annotation can be viewed as a ranking problem, where the aim is to rank the correct keyword (Tag) at first position. Above-mentioned technique does have some limitation that are discussed in the limitation section 7.3.

4.5   Implementation

I make use of a semantic web framework to represent the social network and to deal with the diversity of relationships, illustrations shown in figure 4.4.
The Extraction Manager gathers a user's Facebook data and converts it to Resource Description Framework (RDF [1]) triples. These extracted triples are then

---

[1] https://www.w3.org/RDF/
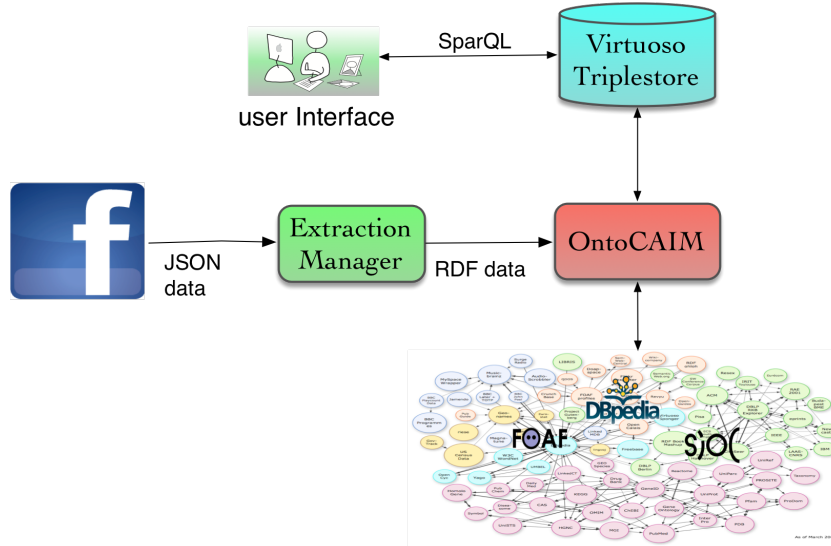
Figure 4.4: Functional Architecture of Image Annotation

later stored in a Virtuoso[50] triple store. The Protégé framework [106] was used to construct and mange OntoCAIM ontology. OntoCAIM ontology illustrated in 4.5 is developed which not only encompasses Social Network Analysis functionality but also defines mechanism to annotate the images with an underlying ontology. The first milestone was to convert Facebook data from JSON (JavaScript Object Notation) to RDF format. We build a simple application that uses Python Facebook API[1] to retrieve user information and then translates it into RDF. Once the data was stored in RDF formate in Virtuoso triple store, SPARQL queries were made to extract the results. To better understand the semantic of tags, I detected the name entities and annotated them with DBpedia entities such as places, organizations or people, for instance the tag of "Oslo" is linked to unique URI *http://dbpedia.org/resource/Oslo* and the meaning of the concept "Oslo" is well defined. Annotation was realized via the name entity recognition services DBpedia Spotlight [98]. GeoNames information could be found in order to further enrich the social tags, once you have determined, via Spotlight, that Oslo is a place, since most GeoNames instance e.g. gn: Feature are linked with DBpedia resource through *owl:sameAs* property.

---

[1] http://django-facebook.readthedocs.org/en/latest/index.html

4.5.1 The OntoCAIM Ontology

OntoCAIM ontology specifies the domain knowledge for image annotation in social network and is formally represented in OWL (Web Ontology Language). I decided to use OWL DL which is based on description logic and is also the decidable part of First-Order Logic (FOL). Though OWL DL lacks expressivity when compared with OWL Full, but it maintains decidability and computational efficiency. Computational efficiency is an important feature since the ontology has to deal with scores of social relations.

One part of OntoCAIM ontology covers annotations that are properties of the image itself e.g title, creator, image format, image size whereas the other part captures relationship among users. OntoCAIM ontology import many other ontologies (e.g. FOAF, EXIF & SIOC) to represent the social network and annotations. In the following figure 4.5, the rectangles represent the OWL concepts and arrows between them could represent any of the three properties depending upon the shape of arrow i.e. data property or object property or sub-class/super-class axioms. Hence, concepts and properties without prefix in the figure belong to OntoCAIM ontology whereas ones with prefix are imported from corresponding ontologies (e.g. FOAF, EXIF & SIOC).

Many ontologies can be used to represent the user. The most popular ontology to model people is FOAF [25]; it is used to model people's identity, their relationships and their activities. The core of FOAF is used to describe the user profile. For example, a Person entity describes personal information properties (such as foaf:img foaf:familyName), social properties(foaf:knows foaf:Group) and personal identities (foaf:homepage foaf:mbox). Our system largely uses classes and properties from FOAF, but it is also necessary to model some additional concepts.

We created a concept Actor as a subclass of foaf:Person ($Actor \subseteq foaf : Person$) and property hasImage as a sub property of foaf:img ($hasImage \subseteq foaf : img$). Actor is a specialized person who must get some values from image through the property $\exists hasimage.Image$ shows in figure 4.6. Enforcing an Actor to get values from the image property increases the efficiency in terms of reasoning; so instead of examining all the persons reasoner will only consider those,
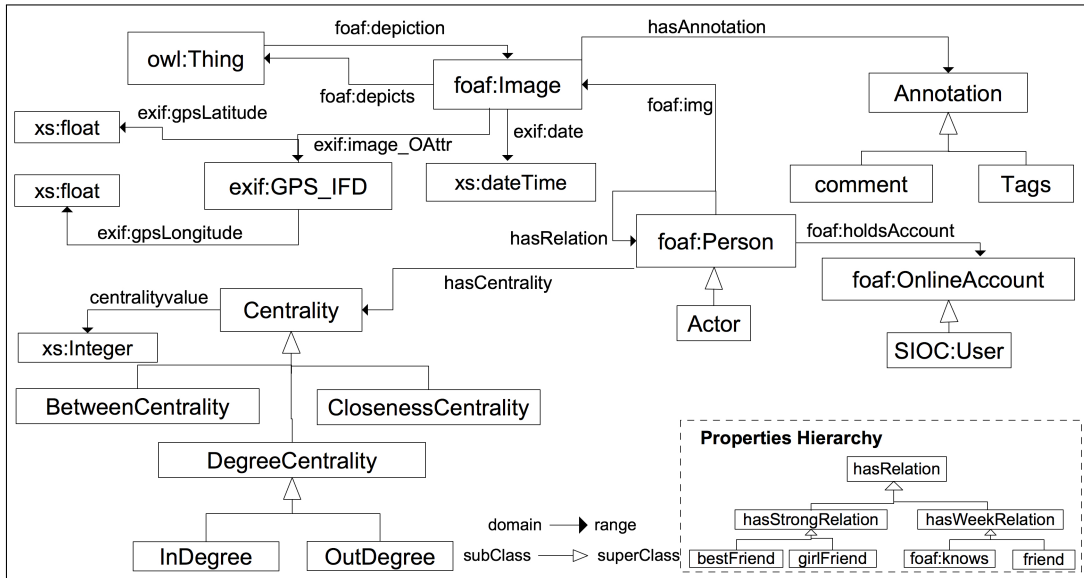
Figure 4.5: Illustrates the main concepts of Formulation of Ontology through classes and properties

```
<!-- http://www.owl-ontologies.com/Ontology1271829425.owl#Actor_ -->

  <owl:Class rdf:about="#Actor_">
    <rdfs:subClassOf rdf:resource="&FOAF;Person"/>
    <rdfs:subClassOf>
      <owl:Restriction>
        <owl:onProperty rdf:resource="#hasImage"/>
        <owl:someValuesFrom rdf:resource="#AnnotatedImages"/>
      </owl:Restriction>
    </rdfs:subClassOf>
  </owl:Class>
```

Figure 4.6: Specialized Actor

in an effort to find the desired annotation, that have some images. The concept Centrality captures the centrality of person through property centralityvalue and is determined through is-a relationship with

$\{BetweenCentrality, ClosenessCentrality, DegreeCentrality\} \subseteq Centrality.$

Table 4.1: System collected annotation Sample

	Automatically generated Keywords
Public Keywords	Winter, Snow, Summer, Tree, Hill, mountains, hike, peak, Garden, Arctic
Personalized Keywords	fjellhies, Studentship, Delicious, Palov, mustache, kjempefint, bilde, midnattsol, norlys, hasu

4.6 Results

In this section, I describe how my approach to automatic image annotation was evaluated. I performed a experiment over a data set, which was volunteered by four distinct social network users by allowing the system to crawl their Facebook profile. For each user we collected the basic information, their friend list, images along with image metadata, which included, if available, associated keywords, description, date/time, GPS and comments made over images. If the image time was not available, I used the time when the image was uploaded to Facebook. After cleansing the collected data we run our experiments over 18000 images with 82000 keywords and 22100 comments associated with images.

We involved user to take part in the experiment and assess the performance of automatic image annotation task. For a given un-annotated image, the task is for the user, to annotate image that is then compare to the automatically generated annotation. Each assessor was requested to annotate 7 randomly chosen images from his/her own social network. For each selected image, assessor annotates with 4 keywords of his/her choice and then the same image was later annotated by system. We presented the 4 top keywords which system automatically generated for given image.

I compute the precision on each image. let K represents the keyword that annotate the image, where K_C is the correct annotation, K_U represents the annotation created by user and K_{SYS} is the annotation which is automatically generated by the system. Precision was calculated as $precision = \frac{|K_C|}{|K_{SYS}|}$. I also calculated the ratio between user created annotation and automatically generated annotation by $UserSystemRatio = \frac{|K_U|}{|K_{SYS}|}$

Table 4.2: Annotation Results

	Image	Annotation	Precision	$\frac{|K_U|}{|K_{SYS}|}$
1		*User*: Beautiful, View, Paris, Eiffel-Tower *System*: EiffelTower, KTHgroup, View, EuroTrip	1	0.5
2		*User*: Winter, Snow, Tree, Swing *System*: Swing, SwedishWinter, Tree, Garden	1	0.5
3		*User*: Summer, Flower, Stroller, Baby *System*: Outfit, Hashu, Stockholm, Baby,	0.75	0.5
4		*User*: Hill, Snow, Friends, Troms *System*: Arctic, Fjellheisen, Geoscientists, Red	0.75	0
5		*User*: Summer, Lake, Grilling, BBQ *System*: Eating, Respect, Handsome, Ketchup	0.25	0

Table 4.3: Annotation Results

Image No.	Number of System Tags	Number of Personalized Tags	Number of Public Tags	Precision
1	4	2	2	1
2	4	2	2	1
3	4	2	2	0.75
4	4	2	2	0.75
5	4	0	3	0.25
6	4	1	3	0.50
7	4	1	3	0.50
.	.	.	.	.
.	.	.	.	.
18	2	0	0	0
Arithmetic Mean	3.2	1.6	1.7	0.63

4.6.1 Discussion

Table 4.2 shows the annotation results. There are two type of annotations in an annotation column, one made by $User$ and the other one generated by $System$. Precision was calculated as $precision = \frac{|K_C|}{|K_{SYS}|}$. In order to realize the difference between two annotations, I calculated the ratio between user created annotation and automatically generated annotation by system $UserSystemRatio = \frac{|K_U|}{|K_{SYS}|}$. One image can be annotated with a number of keywords, therefore in many cases an image was correctly annotated by both $User$ and $System$ but with different keywords.

The first image in table 4.2 presents the panorama view of Paris from Eiffel Tower and it was annotated with keywords "View" and "EiffelTower" both by the $User$ and $System$. Other two annotation "KTHgroup" and "EuroTrip" picked by the $system$, were also correct and very important. Since these two keywords categorized as Private keywords, which repeatedly appear in discussion on social network related to the image. The image owner revealed to us that the image was captured when a group of students from KTH University went on Euro trip.

The second image in table 4.2 shows a beautiful scene of a garden in the winter, where a tree along with swing are covered with heavy snow. *User* and *System* correctly annotated this image with keywords "Swing" and "Tree". *System* offered two other keywords "Garden" and "SwedishWinter". The keyword SwedishWinter was obtained from discussion about the image on Social network, which image owner acknowledged as the important descriptor of image.

The third image in table 4.2 displays a young lady with blue outfit carrying a baby in stroller. The image was annotated with "Baby" both by *User* and *System*. *System* also annotated image correctly with "Stockholm" and "Hashu". "Hashu" is the nickname of the baby Hashum and it appears repetitively in discussion on Social network. The image owner recognized the importance of "Hashu", however she thinks that, "Outfit" is unnecessary to describe this image.

The forth image in table 4.2 displays four young boys posing in front of alpine skiing resort. Whereas the *System* precision was 0.75 and none of the annotation was same with *User*. Arctic, Fjellheisen and Geoscientists were the correct keywords to describe this picture, however Red was unnecessary to highlight the image.

The last image in table 4.2 is the perfect example where contribution for social network leads to the wrong annotation. *System* precision for this image is only 0.25, which means only one out of four, is correct. Respect, Handsome and Ketchup were detected wrong keywords to describe this picture, Eating is the only keyword which somewhat suit this image.

Table 4.3 shows the results of arbitrarily chosen 18 images. The table explains the total number of tags system generates for a given image, out of them, how many are personalized tags and how many are private tags. Last column of the table explains the precision for each image. Arithmetic mean of system tags is 3.2, which indicate that some of the images could not receive even 4 keywords for annotation. Images were annotated more with Public tags 1.7 than Personalized tags 1.5, however the margin is nearly negligible, which highlights the importance of personalized tags. Overall precision for images is 0.63 that also considered a good precision for randomly selected images.

Chapter 5

Relation Based Image Retrieval In Online Social Network

> The work described in this chapter answered the following questions
>
> **RQ4)** How to retrieve, rank and recommend the most relevant image on OSN?
>
> > a) Can social network based image retrieval be used to improve the personal relevance of retrieved images.
>
> **Related Paper:** Paper No. 3 [108] , Paper No. 4 [42] and Paper No. 9 [74]

Paper 3 were the most central paper for this chapter, while paper 4 and paper 9 were supporting papers. Exploring and investigating, how to rank a list of the closest and most reputed friends and precisely modelling the user needs and interests from social network. Communication is the foundation of true friendship. As the Alex says "compromise, communication and consistency are needed in all relationships, not just in romantic ones". Agreements and disagreements are less important than a consistent communication in every day life. It develops a gradual understanding between people and bring them close to each other. The same phenomenon reflects in online social networks. We can measure the level of

friendship by measuring the communication between people.

The social network is formed with a number of users connected through different relationships [1]. The term "relationship" is rooted on associations and various levels of trust between two people. Social theories[58, 76] have discussed the importance of strength of a relationships and that it has real influence on the way users consume each others resources. These relationships are the key to many features such as personalization, security, recommendations and most importantly search for social artifacts, for instance photos and videos.

We propose a Relation-Based Image Retrieval (RBIR) approach that makes use of social information already available in a social network, to improve the user's search experience. We consider three basic social entities associated with photos: user relationships, user activities (comments and likes) and tags, and derive scoring functions for ranking relationships and photos. Relationship score measures the strength of a relationship and is derived from the time and effort invested to maintain contact between two participants. Photo score is derived by considering the quality and worth of both comments and likes on each photo. The worth of comments and likes are determined by analyzing the already ranked relationships through which they are coming from, for instance, comments coming from highly ranked friends weight more than other comments.

Social image retrieval is enabled by first selecting potentially relevant images by matching keywords in the text query against annotations related to images in the social network. The socially most relevant images are subsequently identified through a social ranking algorithm using relationship and photo scores. To test our approach, we developed a prototype on Facebook real-world dataset and results demonstrate that our techniques are useful for retrieving relevant images.

5.1 System Architecture

Our Architecture for Relation Based Image Retrieval is illustrated in Figure 5.1. We generated the ranked list of friends by analyzing the users mutual activities in social networks. Socially rich photos, commented and liked by highly ranked

---

[1]To date, most of social networks such as Facebook, orkut and Myspace offer only positive relationships between two users, such as close friend, family member.

friends were consider more relevant. The *Data Collector* collects social network data about a user, such as her friends, their photos and comments, likes and tags associated with photos from Facebook. For the first time user, *Data Collector* dispatches the social network information to the *Friendship and General user Score Calculator* to calculate two different scores, relationship score and general user score, which are then store in a local database (Local DB). The *Data Collector* later receives updated data from Facebook and enriches the already stored social network information. The *Friendship and General user Score Calculator* recalculates the scores once a week since changes are not very frequently observed and to avoid the unnecessary computation load.
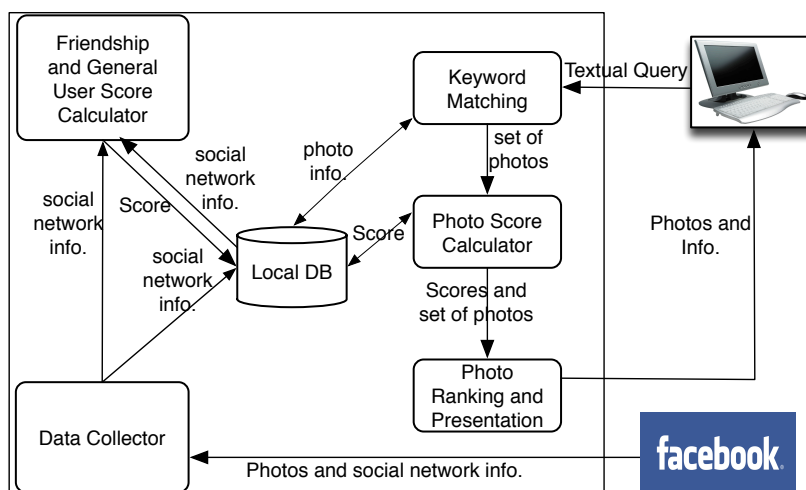
Figure 5.1: Architecture of the system

When a textual query is submitted to the system, *Keyword Matching* performs similarity measure between search string and keywords associated with photos and selected a set of photos from the *Local DB* that are send to *Photo Score Calculator*. *Photo Score Calculator* calculates the social relevance of the photos on the bases of relationship and general user score between query initiator and photos owner, which are already stored in *Local DB*. *Photo Ranking and Presentation* receives the scores and set of photos, and ranks them according to the social relevance for the query initiator. Ranked photo are finally send for presentation.

5.2 Social Network Model

Social networks are online platforms that allow users to create their personal profile, and also help them to articulate various types of connections with other members of the network. Facebook[1], Orkut[2], Flickr[3] and Myspace[4] are some examples of social networks.

Following are the four entities of social networks that have an important role in this thesis.

- **Users:** People who are members of the network, perform some activities, upload photos, produce tags and search.

- **Photos:** Basic content items that are published by users and accessible for others to view, to make comments on or like. Social networks enable their users to publish many other contents, but in this thesis we only consider photos.

- **Tags:** Keywords used to describe the photo content

- **Activity:** In general terms, activity is anything that one user does on another user's guestbook, such as browsing through photos, leaving comments, reading a public diary, writing testimonial and sharing contents for their established friends. However, to reduce the complexity, we only consider the comments and likes over photos made by other users as an activity.

5.3 Ranking of Users and Photos

The proposed solution uses three different scoring functions for ranking the users and photos. *relationship-score*, which is a score regarding the friendship between two users, *general user-score* that is a global and general score for a specific user, and *photo-score* which indicates the importance of a photo to a user.

We use the following notations for formally defining the scoring functions. Each

---

[1]https://www.facebook.com/
[2]http://www.orkut.com/
[3]http://www.flickr.com/
[4]https://myspace.com/

user has a set of photos, denoted P, i.e. $P = \{p_1, ..., p_n\}$. Each photo $p \in P$ has a set of likes and comments given by other users in the social network. The likes given to photo p are represented as a set $L(p)$, while the set of comments to p are represented as the set $C(p)$:

$$L(p) = \{l_1, ..., l_n\} \, C(p) = \{c_1, ..., c_m\}$$

Each like and comment has a property *uid*, which identifies the user providing the like or comment. A photo can receive only one like per user, while a user can comment a number of times to the same photo, for instance in a discussion between several users over a specific photo. In our approach, we want to give more value to those photos where many users participate in discussions. Thus, the first comment from a user U to a photo p is given more weight then subsequent comments from U. For each comment, a property *num* represents the number this comment has in a possible succession of comments from user U to photo p.

5.3.1 Relationship Score

The relationship-score aims to categorize the real binding between users. It is difficult to determine which relationship is more significant than others, because they are established by authoritative permission (single click), rather than through regular interaction [56]. In the beginning, social network users populate their friend list with people they know offline. As time passes by, they start inviting new users with which they share common interests and accepting friendship requests made by others. It is also observed that the user often accepts friend requests from acquaintances because of various social reasons [40]. As a result, the user get access to large sets of photo artifacts that are potentially not relevant to her/him.

To identify the closer relationships between people, we consider the activities between two specific users, and assume that users that have a close social relationship will have a higher likelihood to comment and like each others photos. Friends liking photos represents a vote of confidence, while comments from the social circle may invoke further discussion and generate interest in the photos. We did not analyze whether it is positive or negative information given in the

comment, but either way, it shows the commenter was interested in the photo, hence she takes the time to make a comment. A relationship score between two users, U_i and U_j, is thus calculated based on the comment and like activities happening between the two users. The relationship score $F_{(U_i,U_j)}$ is described in Formula (5.1) .

$$F_{(U_i,U_j)} = \frac{\left|L_{(U_i,U_j)}\right| + \left|C_{(U_i,U_j)}\right|}{\left|P^{U_i}\right|} + \frac{\left|L_{(U_j,U_i)}\right| + \left|C_{(U_j,U_i)}\right|}{\left|P^{U_j}\right|} \qquad (5.1)$$

Here $\left|L_{(U_i,U_j)}\right|$ represents the number of all likes from user U_j to photos owned by user U_i , $\left|C_{(U_i,U_j)}\right|$ the number of all comments from user U_j to photos owned by user U_i, and $\left|P^{U_i}\right|$ the number of photos owned by U_i.

The sets $L_{(U_i,U_j)}$ and $C_{(U_i,U_j)}$ are described as follows:

$$L_{(U_i,U_j)} = \left\{l \mid l \in L\left(P^{U_i}\right) \wedge l.uid = U_j\right\}$$

$$C_{(U_i,U_j)} = \left\{c \mid c \in C\left(P^{U_i}\right) \wedge c.uid = U_j\right\}$$

Here $L\left(P^{U_i}\right)$ and $C\left(P^{U_i}\right)$ represents the set of all likes and comments on all photos owned by U_i. The properties $l.uid$ and $c.uid$ describes the user giving the like and comment respectively. The sets $L\left(P^{U_i}\right)$ and $C\left(P^{U_i}\right)$ are described as follows:

$$L\left(P^{U_i}\right) = \bigcup_{k=1}^{n} L\left(p_k^{U_i}\right)$$

$$C\left(P^{U_i}\right) = \bigcup_{k=1}^{n} C\left(p_k^{U_i}\right)$$

In these formulas $p_k^{U_i}$ represents photo k owned by user U_i, and n is the number of photos owned by user U_i.

5.3.2 General User Score

The characteristic of each of the user's friends is an aspect, which plays an important role in online and physical social interaction [28] . We define a general

user-score as a score that describes a user's general popularity among his friends with respect to photos. To estimate this score we assume that users with a high general popularity will have high activity on his photos. Moreover, a large number of friends in the user's social circle also indicate his popularity. The general popularity-score for the user is described in Formula (5.2).

$$G_{U_i} = \frac{\left|L\left(P^{U_i}\right)\right| + \left|C\left(P^{U_i}\right)\right|}{\left|P^{U_i}\right|} + log_2\left(N + 1\right) \qquad (5.2)$$

The general user score is achieved by calculating the number of comments and likes that are received over photos and dividing by the number photos. N is the number of friends in U_i's social network, and log_2 is used to gradually decrease the impact of a large number of friends. Here $\left|L\left(P^{U_i}\right)\right|$ and $\left|C\left(P^{U_i}\right)\right|$ represent the number of all likes and comments (respectively) given to photos owned by U_i. $\left|L\left(P^{U_i}\right)\right|$ and $\left|C\left(P^{U_i}\right)\right|$ were described in section 5.3.1.

5.3.3 Photo Score

The photo-score is a score that describes how relevant a photo is for a specific user. The score is calculated by determining the strength of relationship between the user searching for photos and a photo owner, combining the quantity and weights of the likes and comments on the photo. The weights of the likes and comments are derived from general user-score and relationship-score.

Assume now that user U_j is searching for photos. The photo-score $PS_j\left(p^{U_i}\right)$, described in Formula (5.3), represents the photo score for photo p owned by U_i with respect to user U_j. $PS_j\left(p^{U_i}\right)$ is calculated as a combination of three different scores: relationship-score, $F_{(U_i,U_j)}$, which defines the direct binding between query initiator U_j and photo owner U_i, likes-score $S_L\left(p^{U_i}\right)$ and comments-score $S_C\left(p^{U_i}\right)$, defines the amount and worth of the likes and comments on photo of p^{U_i}. Time ΔT represents the time in months since the photo was uploaded, and is used to give preference to the newly uploaded photos.

$$PS_j\left(p^{U_i}\right) = h * F_{(U_i,U_j)} + S_L\left(p^{U_i}\right) + S_C\left(p^{U_i}\right) - \frac{\Delta T}{2} \qquad (5.3)$$

Formula (5.4) describes the Like Score, $S_L\left(p^{U_i}\right)$, for photo p owned by U_i,

and is calculated as a summarization of a combined relationship and general user score over all n users liking photo p. The value of the Like Score is determined by the strength of the relationship between photo owner and users liking p, and the general reputation of users liking p.

$$S_L\left(p^{U_i}\right) = \sum_{U_a \in L_u\left(p^{U_i}\right)} \left(F_{(U_i, U_a)} + G_{U_a}\right) \tag{5.4}$$

In formula (5.4) $L_u\left(p^{U_i}\right)$ represents the set of users liking photo p^{U_i}. Assume the set of all users u

$$L_u\left(p^{U_i}\right) = \{U_x \in u \mid (\exists l)\left[l \in L\left(p^{U_i}\right) \wedge l.uid = U_x\right]\}$$

where $L\left(p^{U_i}\right)$ is the set of all likes on photo p owned by U_i.

Formula (5.5) describes the Comment Score, $S_C\left(p^{U_i}\right)$, for photo p owned by U_i. The score is a summarization of a combined friendship and general user score.

$$S_C\left(p^{U_i}\right) = \sum_{U_a \in C_u\left(p^{U_i}\right)} \begin{cases} s\left(F_{(U_i, U_j)} + G_{U_a}\right) + \\ t\left(F_{(U_i, U_j)} + G_{U_a}\right) * \left(\left|C_{U_a}\left(p^{U_i}\right)\right| - 1\right) \end{cases} \tag{5.5}$$

In formula (5.5) $C_u\left(p^{U_i}\right)$ represents the set of users commenting photo p^{U_i}. Assume the set of all users u

$$C_u\left(p^{U_i}\right) = \{U_x \in u \mid (\exists c)\left[c \in C\left(p^{U_i}\right) \wedge c.uid = U_x\right]\}$$

where $C\left(p^{U_i}\right)$ is the set of all comments on photo p owned by U_i.

In some cases photos can have long discussions between two or more users, and since we want photos with many different users commenting ranked higher than the photos with just a discussion between a few people, the first comment to a photo p by user U_a is given higher weight then subsequent comments from U_a. The first part of the summarization equation in Formula (5.5), is for the first comment to p from user U_a, second part is for all the other comments from U_a. The construct $\left(\left|C_{U_a}\left(p^{U_i}\right)\right| - 1\right)$ holds the number of comments from user U_a to photo p, minus the first comment. The set of all comments from user U_a to a

photo p is described as:

$$C_{U_a}\left(p^{U_i}\right) = \left\{c \mid c \in C\left(p^{U_i}\right) \wedge c.uid = U_a\right\}$$

5.4 Evaluation

In this section, we describe how our approach to relation-based image retrieval was evaluated. We discuss how the data was collected and system testing was done.

5.4.1 Data Collection and Privacy Issues

The application gathered data from Facebook about users, their friends and photos. The information collected was the identifier of the user and all his friends, all the photos they have uploaded, all likes and comments on the photos, as well as the title, caption, location and tags on the photo. The photo itself is not downloaded; instead the URL to a static version of the picture (on Facebook-servers) was stored in the database.

We gathered all the test data in the period from 15 Feb 2013 to 04 Mar 2013. When the participants accepted to use our system, their data was downloaded to the database and calculations of relationship and general user score were made. Our data set contained a total of 193,869 photos belonging to 1,088 unique users. These photos had total of 227,111 comments, 288,704 likes and 886,817 tags. Figure 5.2 shows the details of how the data increased as the number of user increased.

The data were collected by approaching 6 core users that contributed with all photo information and 5 of these core users were used as assessor when evaluating the system. As we download much data about the user and his/her friends, we can assume that some of the data might be in some sense private and not wished to be posted public for everyone. The photos as well might only be meant for his/her friends on Facebook and they do not wish them to be publicly shared. For this reason we had to ensure that users testing the system were not able to see any more data than they could already by using Facebook.
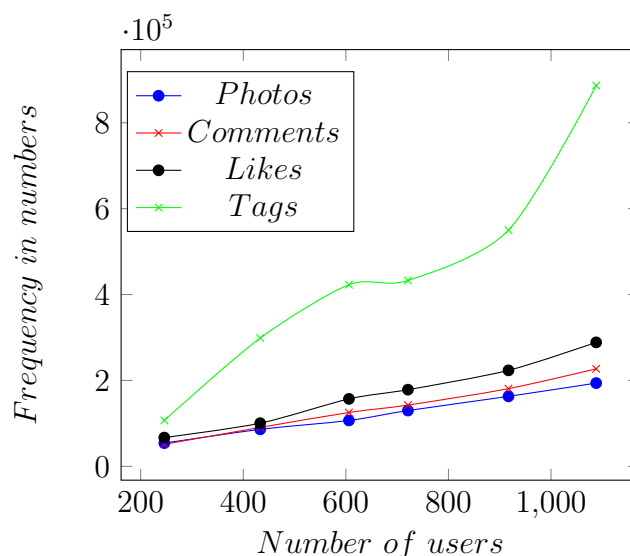
Figure 5.2: Statistics of collected data from Facebook

5.4.2 Evaluation Method

We evaluate our relation-based image retrieval system, in the following denoted *Social Ranking*, by comparing with a baseline system *Without Ranking*, where relation-based ranking were not used. The baseline system retrieves images based on keyword matching. We used the integrated match against function in SQL, which returns a relevance score; that is, a similarity measure between the search string and keywords attached with the photos. Relevance scores are non-negative floating-point numbers. Zero relevance means no similarity.

It was important that we evaluated our system on the same social dataset. Hence, we developed a baseline system Without Ranking, which does not consider the relationship and photos score but execute the textual query on the system with same environment such as network user account, data set and computational power. Each query issued by the users was executed by both systems, and the top 10 recommended photos from each system were presented to the user.

5.4.3 Testing

The notion of relevance for images in a social network is very much subjective and depends on the social context of the query initiator. Every query initiator

obtains a different result for the same query since the scoring model explore the friendship and photo scores, which are user specific.

Every assessor was asked to make 7 queries, 4 queries about Events and 3 about Places in their respective social network. They were presented with a few example queries but they were free to make their own queries within the categories. Some of the queries made by assessors were *wedding*, *birthday*, *Oslo* and *snow in tromso*. Most of the queries we received were single word text queries.

For each textual query, the assessors were displayed a two column result page illustrated in Figure 5.2 that shows the top-10 results for both strategies *Social Ranking* and *Without Ranking*. The assessors were requested to identify for each of the strategies, the relevant images for each query and also mention the ranking position of the images. To help the assessors, resulted images were presented with keywords, description, photo owner and link to Facebook to find additional context when needed.

5.4.4 Evaluation Metrics

For the evaluation metrics, we adopted two metrics to evaluate the different aspects of our systems

- **Precision at K (P@K)**
  Precision at K is to measure whether the users were receiving relevant documents at the top of the ranking or not. We reported the precision at k=5 and k=10. It measures the precision when 5 and 10 photos have been seen.

- **Event and Place Category**
  We divided textual queries into two categories Event and Place. Event queries were made to search for photos, which depicted events, such as *football match* and *snow skating*. Place queries were made to search for the photos that were captured at or showed a specific place, such as *Oslo* or *Stockholm*.

5.5   Results

This section presents the results of testing our algorithm *Social Ranking* and comparison with the baseline *Without Ranking*. Baseline is the text query over the same data set and other parameters for performance test.

Table 1 shows the average precision score for images socially relevant to the query initiator. The table displays separate scores for P@5 and P@10.

Table 5.1: Average precision of relevant images

	P@5	P@10
Social Ranking	0.711	0.671
Without Ranking	0.379	0.354

For the precision at 5 (P@5), we measure a precision of 0.711 for *Social Ranking*, which indicates that on average 71% of recommended photos were found relevant in top 5 photos. We measured the precision 0.379 for *Without Ranking* that shows only 38% photos were relevant in top 5. Expanding to the top 10 recommended photos (P@10) we see the performance goes slightly down for both *Social Ranking* and *Without Ranking* to 67% and 35% respectively and *Social Ranking* again produce better results but with less margin compare to the (P@5). We can thus safely argue that the *Social Ranking* strategy performs very well for both precision metrics.

Table 5.2: Average precision of relevant images

	Event	Place
Social Ranking	0.694	0.841
Without Ranking	0.347	0.466

Table 2 shows the average precision scores for images socially relevant to the query initiator. The table illustrates the separate scores for the two search query categories; Events and Places. We calculated the precision for top 10 photos against queries from both categories. We measured a precision of 0.694 and 0.347 for *Social Ranking* and *Without Ranking* respectively for the Event category. While the precision for the Place category was 0.841 and 0.466. Thus, queries

for Place perform slightly better than queries for Event. When looking at the relative difference in performance between the two strategies, *Social Ranking* and *Without Ranking*, we observe that for all metrics the *Social Ranking* strategy outperforms the *Without Ranking* strategy.

5.6 Personalized Recommendation of Socially Relevant Images

Recommender systems are software tools and techniques that provide suggestions for items that are most likely of interest to a particular user. Based on a user profile, relevance of items are determined using filtering techniques such as content-based, collaborative, demographic and hybrid filtering [3, 17].

To recommend images of personal interest to the user, we present a framework image recommender system in paper 9 [74]. The image recommender system offers a hybrid filtering approach, including content-based, knowledge-based and social-based filtering, that selects images of social interest to the user, by for example being posted by close friends or family. To enable this, we generate a user profile based on user activity on social media, that reflects user interests and social context. Images are recommended by filtering available images according to a combination of two relevance measures, interest relevance and social relevance. Our system handles both cross-source user profiling and image recommendation across social media, currently focusing on images from Facebook and Flickr.

Figure 5.3 shows the components of our image recommendation system. The lower part of the figure (below the dotted line) shows user profile construction based on user data collected from Facebook and Flickr. This includes an Extraction Manager collecting and processing user data, semantic enrichment based on information from WordNet[1] and DBpedia[2], before the user profile is generated and stored. To recommend images through a filtering process, the user profile is compared with textual representations of images. Feature extraction collects and process image metadata to identify representative terms describing the im-

---

[1] https://wordnet.princeton.edu
[2] http://wiki.dbpedia.org
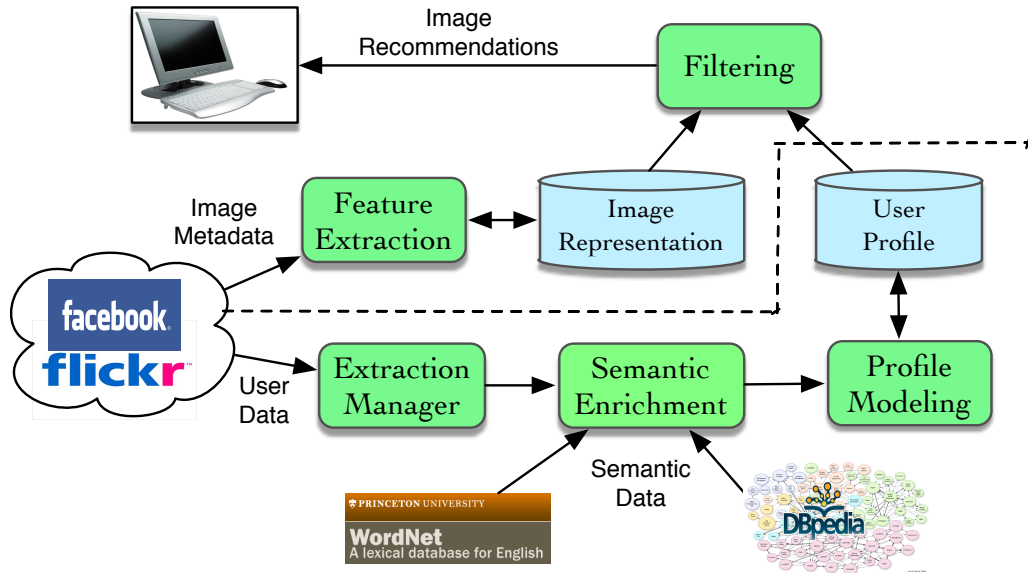
Figure 5.3: Framework of the recommended system

age. This is illustrated in the upper part of Figure 5.3. User profiling and the recommendation process, including feature extraction and filtering of images, are described in the following sections.

5.7 Discussion

The results presented in the previous section indicate that using *Social Ranking* for relevant photo retrieval can be highly beneficial. We examine the performance of *Social Ranking* with comparison to the baseline and over all satisfaction of users with our system. Figure 5.4 illustrates the top 3 photos as a result of Event query *wedding.* The assessor describes that the socially ranked photos on the left column are very much relevant, where the first photo depicts her cousin getting married and the family had heavily commented and liked the photo. The second ranked photo is her close friend in marriage ceremony and the photo is commented and liked by many other close friends. In the same manner the last photo is relevant too.

In contrast, *Without Ranking* showed in the right column of Figure 5.4, did not

perform well for the same query. The assessor reported that she knew the top ranked photo, while the remaining two photos were not interesting at all. However, these photos have been commented and liked by many, but not by friends of the assessor. In other words, these photos are maybe very popular in other social circles but not for this particular user.

The second result, illustrated in Figure 5.5, shows the top 5 photos as result

Figure 5.4: First three photos for wedding query

of the Place query *Oslo*. The assessor explained that the socially ranked photos were very relevant. The topmost photo depicts one of his close friends portraying in a famous place in *Oslo*, and many of the assessor's close friends have liked and commented on the picture, which make it more interested for the assessor. In the same way, assessor found the second, fourth and fifth picture highly relevant, while the third is not that much relevant and therefore does not justify the third place in the photo ranking.

On the right side of Figure 5.5, we see the result of query matching using classic
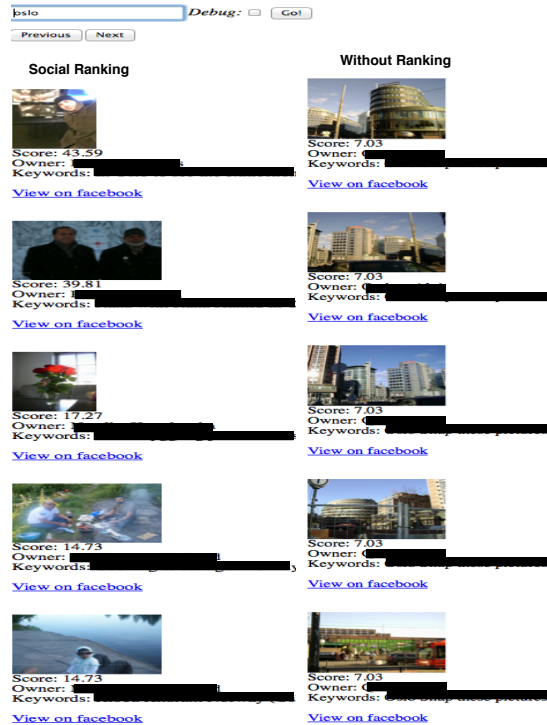
Figure 5.5: First five photos for Oslo query

image retrieval without relation-based ranking. There the query *Oslo* retrieved photos of city buildings and famous places but fail to offer the social importance for the query initiator. Moreover, all the socially ranked photos came from different albums, uploaded by different users, whereas *Without Ranking* retrieved the photos from one person as a result of best textual matching.

The higher effectiveness of *Social Ranking* over the *Without Ranking* can be attributed to the fact that people share the likeness of photos with their social group. Friends liking photos consider a vote of confidence and comments from the close social circle generate high interest in the photos. Sometime, discussion over the photos has more importance for the user than the photo itself.

To use our system, the first time user grants permissions to the application, which acquires a Facebook-authentication token that is used to get access the users profile and download data. The execution time for the initial download of information from the Facebook-server is currently rather high. There are three main factors: Due to the large amount of data we are collecting about a user, the

execution time of each query against the Facebook-servers was high. For some queries it could take up to 8 seconds to get a response. The second issue was that Facebook had a limit on the amount of data you could download for each query. So to retrieve all the data we had to split up the queries, instead of getting all the information with just 1 query. This led to the last issue: Facebook has a limitation on the amount of queries that can be executed. For normal applications, like ours, this limit was 6000 queries per 10 minutes and if the user had a large amount of friends on Facebook we would often need to execute as much as 10,000 queries.

Nevertheless, once the data is downloaded, we calculate the friendship and general user scores locally, which do not require long execution time. Furthermore, we recalculate the scores once a week since the changes are not very often observed. Hence the system is not computational expansive and run seamlessly.

Chapter 6

RelatedWork

The accomplishment of online social media and amount of information partici-
pants voluntarily reveal, have attracted the attention of researchers that focus on
harvesting online social media information for use in information retrieval (rec-
ommender) systems [17, 20, 142]. Exploiting social connections, social behaviors
(such as shares, clicks, and likes between users), tags, posts and status message
have been proven beneficial for applications that need to grasp user preferences
and demands. A variety of such information has therefore been used for user
profiling [1, 2, 66, 99].

6.1   User Behavior in Online Social Network

Since the beginning of online social network revolution, researchers have raised
the issue of user's privacy and discussed some of its potential implications [47,
59]. Researchers have explored the content quality of YouTube videos and the
disclosure of personal health information within social network public profiles [18,
77] . The analysis of comments is important due the prevalence and incidence of
social network platforms; with issues related to privacy and ethics. Additionally,
automatic comment analysis, through the use of crawlers and algorithms, may
be used to harvest more information about the users for personalization purposes
(i.e. user targeted advertising) [134]. The overarching objectives of our study
is to learn the user behavior on social network platforms and to increase the
understanding of the social network phenomenon. In particular, it is focused

on Patient Generated Content (PGC) found within YouTube video comments created by patients and the disclosure of personal health information within.

6.2   Image Annotation in Online Social Network

Recent research [103, 104] on image annotation has mainly focused on Global Positioning System (GPS) and timestamps of image while ignoring the context of the image. While developing context-based annotation tool PhotoCompas [107], author obtains ground-truth from previously annotated photos to generate label suggestions for people that may be depicted in photos that are yet not annotated. They make use of timestamps and GPS coordinates as a context and identify the patterns of re-occurrence and co-occurrence of different people. However, they do not investigate the granularity of relationship among the people. In order to understand the semantics of an image, the most common approach to bridge the "semantic gap" is to use the multimedia ontological infrastructure [139]. In semantic annotation of image collections [86] author elaborates the use of ontology for image annotation and demonstrate the advance search strategies through the concept instead of keywords. However, their approach, for the most part, relies on manual annotation. Another ontology-based system called CONFOTO [113] attempted to annotate conference photos through collaborative tagging with Resource Description Framework. Photocopain [145] is a semi-automatic image annotation system, which acquires various type of metadata from the context of the image such as GPS coordinates, EXIF metadata, user's calendar and Flickr tags. Photocopain created service-based interoperable annotation system by using Semantic web standards. They have also used extraction techniques for image analysis. Our sources for acquiring metadata and use of technologies are very much like Photocopain, except that that did not significantly investigate the social context of an image.

Image annotations are the focus in a number of publications, which describe different approaches to collecting tags from previously tagged images. Much work is done on tagging based on content analysis of images, where machine learning techniques are used to develop image annotation systems that map low-level

visual features to high-level concepts [92, 144, 147, 150].

A number of systems annotate query images by selecting tags from related images gathered from online image collections based on a combination of geographic position and visual similarity [70, 71, 103, 119, 133, 137]. The general technique is first to collect a set of images within a certain radius of the query image, narrow down the set by using visual similarity techniques, and finally collect tags from the remaining images.

The work of Jones [70] and Popescu and Moellic [119] describe annotation of landmark images, where identification of landmark and subsequent annotation relies on a reference image corpus collected from Flickr [119] and from Flickr and Panoramio [70]. In the work of Joshi [71] location-driven tags are suggested for geo-tagged images by collecting tag clouds from three different sources; a point-of-interest database, Flickr photos and personal photos. Tags are ranked according to distance and visual similarity to the query image. SpiritTagger[1] [103] is a tag suggestion tool that combines geographical context with content-based image analysis to collect geographically relevant tags from images in Flickr. To maximise the number of images that are visually similar to the query image [133] suggests an iterative visual matching approach, while in work of Silva and Martins [137] the most relevant tags are suggested by using supervised learning to rank or unsupervised rank aggregation methods that combine different estimators of tag relevance.

Approaches for annotating images based on a keyword and the image itself have been presented [55, 148]. The methods require at least one accurate keyword and combine keyword search with image content analysis to retrieve similar images from which tags can be selected.

Related to our approach, authors such as Liu et al [88], work on automatic landmark mining where landmarks are automatically recognised by mining image metadata, such as tags and geo-location, from large-scale social image collections. The user first submits a location name to the system, which subsequently crawls Flickr and selects photos and related metadata. The selected images are clustered according to visual features and different metadata types (such as time stamp, tags and location), and candidate landmarks mined from the clusters.

---

[1]https://vision.ece.ucsb.edu/abstract/513

Much of the referred works that provide image tagging are based on content analysis of images [70, 71, 103, 119, 133, 137], in some cases in combination with geographic position or user provided keyword(s) [55, 70, 71, 88, 103, 119, 133, 137, 148]. Our work contrasts with the referred works as we do not include analysis of visual features in images, but rather provide automatic image annotation based solely on image metadata. We chose this approach in order to avoid the semantic gap problem and investigate the effectiveness of using image metadata as basis for the image annotation process.

6.3 Ranking and Image Retrieval In Online Social Network

Social searching receives enormous recognition, for instance through the use of Google's PageRank algorithm, which determines the importance of web pages by analyzing the link structure on the web. The social approach gives the opportunity for more personalized and in consequence more relevant answers to queries asked by individuals. FolkRank [64] is another algorithm for assigning the authority scores to elements in the network and consequently improves the result ranking. FolkRank is inspired by the renowned PageRank [101] that exploit the structure of folksonomy.

The usefulness of social information to recommend the most relevant tags for images in Flickr is described in [135]. They have investigated the user behavior of tagging images in Flickr and derived the scoring function from analyzing the collective social knowledge. In [52] authors make use of social information, particularly the wide range of social messages from the audience, posted on Twitter about the TV programs in Japan and infer the TV viewing rates as oppose to the conventional survey-based TV rating.

Focusing on an approach that retrieves documents on the base of social information, the work presented in [15] is most aligned with our approach. They exploited the social relations to rank the resulted contents items from deli.cio.us and Flickr. They presented a framework to cast the different entities of networks

into a unified graph model representing the mutual relationship of users, content and tags, and described the scoring functions for each entity (users, content and tags) and relations between them. However, they simplify the friendship between users, by only measuring the shortest path connecting the two users. They also did not consider the likes and comments made over the social contents.

We believe in general, that both the social information existing in social networks as well as user ranking is needed to determine the social relevance of content items. However, both will benefit more if we combine them effectively. In our proposed approach, we ranked users based on the social activities on the each other photos (comments and likes). Moreover, we predominantly consider the activities associated with our target item photos.

Chapter 7

Conclusions and Future Work

The main focus of this thesis has been annotating and retrieving images in online social network. In the methods presented in this thesis, the information surrounding the images has been an important feature. This includes, textual tags, comments, timestamp and social network.
In the following I summarizes the main contribution of the thesis.

7.1 Contributions of the Thesis

The main contributions of the thesis are specifically related to image annotation and retrieval in online social networks. Therefore user behavior of using social network and user relationship with other user in the network was important properties. I proposed the methods to use aforementioned properties to retrieve the most relevant images.

7.1.1 User Behavior in Online Social Network

We have examined the user behavior in Online Social Networks (OSN) and discussed some of its potential implications. A study about personal information disclosed on OSN has shown some of the risks users may be subject to, when disclosing personal medical information. This research may also contribute to the formation of health consumer guidelines for safety on OSN, for example, not

reusing web aliases in platforms where personal health information is publicly disclosed.

A study about general usage of OSN was also conducted. It uncovers that most of the participant join the OSN to keep in touch with old friends and the people they already know offline. For many, online social network was the digital replica of their network in society. The study also discusses overall trends in making friends on OSN. It reveals that young people are more open to make friends online even if the other person is a stranger. This trend declines as the age increases. Given the vast array of private information shared with OSN users, sharing with strangers, can bring dangerous consequences. While inquiring about the social activities users are most interested in OSN, we found that users login on to OSN to check out feeds, to see images and videos shared by other members and to make comments.

7.1.2 Image Annotation by Leveraging the Online Social Network Analysis

I am using voluntarily available image's annotation from friends to automatically annotate new images. One of the important assertions while observing the user behavior in online social network was that, where as friends are an important feature of social network, not all the friends are real friends. Hence, I categorise the friends into real and not real friends. I rely only on real friends to improve the annotation efficiency and avoid the overhead.

During this study, I observed that annotations could be categorised in two types, namely: public annotations and personalized annotations. Public annotations refer to annotations, which are very generic in nature and relevant to masses such as "Eiffel tower" and "Oslo opera house". Personalized annotations are annotations that are relevant to a specific group, for example, annotations such as "birthday party in Los Angeles" and "besøk Tromsø(visiting Tromsø)". These can be considered personalized annotations as they relate to specific events and people.

I presented social network as a context of an image and developed an algorithm

for image annotation based on the graph-theoretic techniques of social network analysis. Its practical implications are validated by experiments on real and randomly generated data. We involved users to take part in the experiment and assess the performance of the automatic image annotation. For each selected image, an assessor annotates with 4 keywords of his/her choice and then the same image was later annotated by the system, developed as part of this work.

I compute the precision of automatically generated annotations on each image. Images were annotated more with public tags 1.7 than personalized tags 1.5, however the margin is nearly negligible, which highlights the importance of personalized tags. Overall precision for images is 0.63 that also considered a good precision for randomly selected images.

7.1.3 Relation Based Image Retrieval In Online Social Network

Social network provide a platform to connect users through different relationships. These relationships have real influence on the way that users consume each other's resources. I propose an Relation Based Image Retrieval (RBIR) approach to make use of social information already available in social networks, to improve user's image search experience.

I consider three basic social entities associated with photos, i.e. user relationships, user activities (comments and likes) and tags. I derive scoring functions for ranking relationships and photos. Relationship score measures the strength of a relationship and is derived from the time and effort invested to maintain contact between two participants. Photo score is derived by considering the quality and worth of both comments and likes on each photo. The worth of comments and likes are determined by analyzing the already ranked relationships through which they are coming from, for instance, comments coming from highly ranked friends weight more than other comments.

RBIR is enabled by first selecting potentially relevant images by matching keywords in the text query against annotations related to images in the social network. The socially most relevant images are subsequently identified through

a social ranking algorithm using relationship and photo scores. To test our approach, we developed a prototype on Facebook real-world dataset and results demonstrate that our techniques are useful for retrieving relevant images. I examined the performance of my solution with comparison to the baseline and over all satisfaction among users of our system. The results show that using Social Ranking for relevant photo retrieval is highly useful.

7.2 Answers to Research Questions

The following elaborate on how the research questions in section 1.1 have been answered.

RQ1) What is the user behavior in online social network?

To answer this question we have included two user studies, which were conducted to observe the user behavior on Online Social Networks (OSN). Chapter 3 has presented the methods and results of these studies. I present the study in section 3.3 about Patient Generated Contents (PGC) on social network, which is by far the most personal information disclosed on social network. The study concluded that patient unintentionally discloses very personal information publically. The public availability of such information could even have further implications that reach beyond our current ethical paradigms.

Other study in section 3.4 about general usage of OSN reveals that young users are not reluctant to accept friend's requests from strangers. Therefore exposing personal information to unknowns, which could be harmful. However, female users were relatively more carful. This study also shows that images attract significant interest amid user-generated contents and relationships among users on OSN have a central importance.

RQ2) What is the meaning of context and how can social networks be explored as context of image?

Context is a general concept and has a loose definition. We defined the context as, "A system is context centric if it uses the context to retrieve

image to the user, whereas relevancy depends on user social network". I classify context into, Usage Context and Capture Context, representing different types of image environments, Figure 4.1 illustrates the context features.

Capture Context mainly describes when and where the image was taken and includes metadata extracted from the digital camera, such as temporal and geographical metadata. Usage Context describes the surrounding environment as to where and how the image is being used. I consider that the image is being used on online social network. I observe that the most useful feature of social network is the contributions from other members of the network in the form of comments and likes. Therefore, social network activities over images such as comments and likes emerge as a key Usage Context features.

RQ3) How can one reuse existing annotation of images from members of one's social network as a base for image annotation?
A group of like-minded people within online social network frequently has correlated activities and common interests that lead them to produce similar annotation of images. In other words, users within a certain OSN group may find semantic agreement on image annotations. Therefore, I used already available image annotation within one's OSN to annotate images, with no annotation. I have observed that annotations can be categorized in two types, namely: public annotations and personalized annotations, as explained in section 4.2.1.

In social network some actors have more influence than others, as they provide relevant information to many others. Therefore, I decided to take reference annotation from influential actors (central actors). I use the graph-theoretical concepts to deduce the central actors 4.3. An algorithm 4.4 was implemented to find relevant image annotations. Section 4.6 presents the results. Images were annotated with more Public tags 1.7 than Personalized tags 1.5, however the margin is nearly negligible, which highlights the importance of Personalized tags. Overall precision for annotatins is 0.63 that also considered a good precision for randomly selected images.

RQ4) How to retrieve, rank and recommend the most relevant image on OSN? Chapter 5 presented a Relation-Based Image Retrieval (RBIR) approach that makes use of social information already available in a social network, to retrieve relevant images. We consider three basic social entities associated with image, i.e. user relationships, user activities (comments and likes) and tags, and derive scoring functions for ranking relationships and images. My RBIR approach shows how the ranked relationships and images help to retrieve relevant images.

7.3 Limitations

A main problem I had to face was to ask people for their personal and detailed online social information therefore it was very difficult to persuade people to participate in my studies. Another main limitation was to deal with rapid changes occurring to social network platforms, specially Facebook. Facebook has introduced many features related to user content sharing and rating which was unknown at the time this thesis was conducted.

A problem for the user behavior survey was to persuade people to participate in the survey. It may undermine the generalizability of results reported. Regardless the uniqueness and the novelty of the findings in the survey, 247 participants do not allow any generalization. Moreover, female participants were more reluctant to reveal their private information. Therefore, it only accumulates the 26% of the participants. There may be a relationship between likelihood of participation in the web survey and online social network users, especially in developing nations. People more likely to participate in online surveys from developing nations are the tiny urban educated class, which has access to the Internet. There is also a chance of likeminded participants groups.

To evaluate image annotation and retrieval, I download considerable data about the user and his/her friends from Facebook, using the token given by the user. One can assume that some of the data might be in some sense private - and not wished to be posted public for everyone. The photos as well might only be meant for his/her friends on Facebook - and they do not wish them to be publicly

shared. For this reason, we implemented that people should not be able to see any more data on our website then they already could on Facebook with their account.

However, there is a problem in the current implementation of the system. On Facebook you can set different relationships. For instance you can have relationships like *family*, *friends* and *acquaintances*. If the user then selected to share a photo with only family and we had 2 persons on the website, a normal friend and a family member of the user, the photo would be visible and searchable by both, even though only one of them should be able to view the photo. This is something that should be changed before making such an application available to public.

By default, the user is asked to authorize the app to access basic information that is available publicly or by default on Facebook. If your app needs more than this basic information to function, you must request specific permissions from the user. This is accomplished by adding a scope parameter to the OAuth Dialog request followed by comma separated list of the required permissions.

7.4 Future Work

My proposed ranking of photos in this work could be extended to ranking of any item *i.e. post or video* on social network. In social networks implicit feedback usually comes in the form of *like, comment* and *share*. A *like* for an item express endorsement by the user, while *comment* illustrates opinion for the item. Generally, if the opinion about an item is positive *like* is accompanied by a detailed expression through *comment*. If the negative opinion is expressed in the *comment*, it is usually without *like*. Sharing is in social network is much more like educating or disseminating information for others which is useful or interesting. Sharing can be consider a recommendation and the highest form of endorsement. As there are three possible ways to express opinion about an item in social network, we can form actually seven types of expression sets, which include *like, share, comment*, (*like, share*), (*share, comment*), (*like, comment*) and (*like, comment, share*). The conversion from *like, comment* and *share* to a rating, it would be on a scale from

one to seven.

Table 7.1: Social opinion to item rating conversion table

social opinion set	proposed rating
comment	1
like	2
like, comment	3
share	4
share, comment	5
like, share	6
like, comment, share	7

In Table 7.1 shows the social opinion to rating conversion table for a specific item. Since sharing is considered as a form of recommendation for an item, any set containing *share* has been given higher rating than the *like* or *comment* alone or their combination. However, *comment* alone has been given the lowest rating as without *like* a *comment* would have a negative feeling. Recently OSN are providing more option to rate an item such as *like, love, haha, wow, sad, angry*. This detailed feedback provides an opportunity to further classify user preference about any item.

Paper I

User behavior in Online Social Networks and its Implications: A user study

Najeeb Elahi
Department of Computer Science,
University of Tromsø, Norway.
+47 99 88 7869

najeeb.elahi@uit.no

Randi Karlsen
Department of Computer Science,
University of Tromsø, Norway.
+47 7764 4040

randi@cs.uit.no

ABSTRACT

Online social networking service has attracted great interest in recent years. It has emerged as a major medium of communication as it has provided a platform for sharing personal information with vast network of friends. In this paper we present a different perspective of a social network, one that vague nature of friends relationship and show the association of users activities with gender, age and nationality. Results suggest that female are more conservative while accepting friends requests from stranger than males, and that the younger people are more active (and open). We also compare the social activities between two distinct nations – Norway and Pakistan. Our study revealed that despite of vast differences between these two nations, the online social activities are quite similar. The knowledge gained through such results can aid in filtering, sorting and recommending information over social networks.

General Terms

Documentation, Human Factors, Verification.

Keywords

Social networks, user behavior, social network analysis, user study of online social network.

1. INTRODUCTION

Every moment, millions of people across the globe share their contents online. The flexibility provided by the social networks has given the people an opportunity to conveniently express themselves online through videos, photos, comments, and rating on online contents. Social networks also offer a platform to enrich the existing relationships and establish new relationships that would be almost impossible otherwise. Therefore, people are now using these services for building their social circle and sharing their contents far more than ever before.

On the one hand, social networks are getting close to

their assumed goal to facilitate interaction between the users around the world and therefore, bringing the social gap across the nations. On the other hand; rigorously compromising the privacy of the users [1].

Social relationships are usually categorized as weak or strong, but in reality are exceptionally diverse in terms of how close and inmate one perceives to be. Danah Boyd [2] reported the online social network relation in binary terms "Friend or not" he further elaborates that "there is no way to determine what metric was used or what the role or weight of the relationship is. While some people are willing to indicate anyone as Friends, and others stick to a conservative definition, most users tend to list anyone who they know and do not actively dislike. This often means that people are indicated as Friends even though the user does not particularly know or trust the person" [2]. Thus, there is no refine line that separates friends and not friends, however, in our paper, we consider real friends, who are known offline and keep strong tie.

Social Network Analysis (SNA) primarily focuses on the patterns of social relationships among the set of social entities such as people or organization [3, 4]. SNA assumes that relationship is important. Social network researchers Wasserman and Faust [3] have identified that the relationship is the heart of any kind of social network analysis and thus, the social actors and their actions must be viewed as interdependent rather than independent or autonomous units. SNA have been used in different application domains for a variety of purposes: It has been used to successfully uncover criminals [5], identifying missing connection between workers in organizations and to find out the influential or disconnected individuals.

In order to investigate the formation of a relationship on a social networks, we conducted a user study based upon the following questions.

- How gender affects the social relationships.

- What is the affect of age on social relationships.

- How do social activities vary across nations.

Social network analysis is relatively new field of research and consequently faces many impediments. In this paper we have elaborated some of the implications of social network analysis. However, some of the implications are related to human behavior,

which will be resolved over time by human learning and technical improvement in online social network services.

The paper is organized as follows. The next section discusses the methodology of survey. Section-3 describes the survey results and briefly compare with the other research work. Section-4 explains limitation to our work. In Section-5, we discuss the survey results and their implications. Finally, paper concludes with the main features of the work and some comments on the future work.

2. METHODOLOGY

We conducted a survey of random users of a social network. A URL invitation to participate in this survey was sent via email at the end of June 2011, 247 people from 33 countries participated in this survey. In order to acquire diversity, we also conducted survey an in-person in Norway. To protect the individual privacy, we cannot reveal the individual information. Out of 247 participants 26% were female and 74% were male. Our goal was to accumulate data from people from different gender, age, and nationality. However, people from age group over-40 were quite reluctant to participate in the survey.

3. RESULTS

We begin by discussing the overall trends in making friends in different gender. We identify the correlations between age of the participants and number of friends and real friends they have. We explore the most popular activities among social network users and examine the difference between two countries.

3.1 The Affects of Gender on Social Relationships

To find out how people make friends over social networks, we asked participants to respond to five questions illustrated on x-axis in fig.1 on a four-point scale where 1 equaled "strongly disagree" and 4 equaled "strongly agree".

Participants reported high confidence in people they *already knew offline* with male-mean 3.2 and female-mean 3.8. Males agree (mean = 3) to accept *friend of a friend* but females agree less (mean = 2.3). When it came to making friends with strangers, both genders showed disapproval, with the strong disapproval came from females (male-mean=2, female-mean=1). It is, however, interesting to note that the approval rating is slightly higher in case of *friend request from opposite sex*.
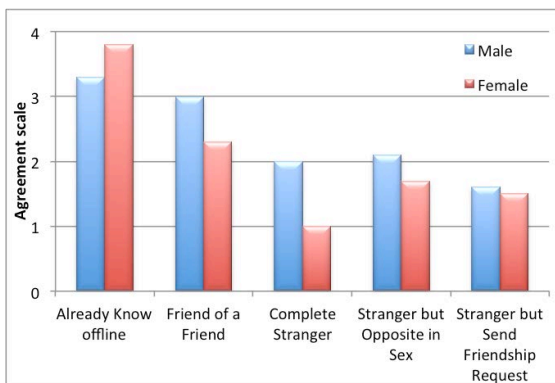

Fig 1: Illustrate the difference between male and female for making friends.

Nevertheless, females appear to be more careful while making friends as compare to males and the average attitude of

females is tend towards disapproval. It is noticeable that the female conservativeness in our results very much correspond with [1], where they found the female are substantially more conservative for reporting their phone number on profile. Specifically, single male reveal their phone number in even higher frequency than the single female. It also agrees the results published in [6] where they suggests that users are largely using Facebook to learn more about people they meet offline and are less likely to employ site to initiate new connections.

3.2 Is Age a Factor in Relationships

Users were reluctant to reveal their accurate age and therefore, we formed four different age groups. The popularity of social networks among young is demonstrably high. Our survey revealed that younger people were more active on social networks. Fig. 2 illustrates that the young people are more open to make friends online even if they are strangers. This trend declines as the age increases. People in age group over 40 are very reluctant to accept strangers in their friend circle.
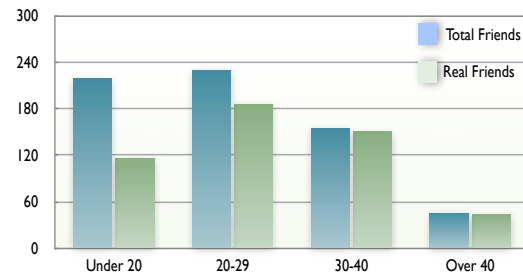
Fig 2: Shows the association between age group and number of total and real friends.

It is noteworthy to compare this result with the speculative theoretical limit, that you cannot have more than 150 friends on your social network, so all the people beyond this limit cannot be an actual friend [7], whereas our results show that the participants of age group 20-29 do exceed this theoretical limit.

3.3 Variations of Social Activities Across Nations

Finally, we attempted to determine behavior differences between different countries. People from 33 countries responded to our survey. But, due to various reasons, we chose Pakistan and Norway for comparison. One reason was that the biggest chunk of sample data was reported from Pakistan (31%) and Norway (22%), other main reason was that these countries are very different with respect to socioeconomic status, political freedom and cultural differences [8, 9].

To our surprise, participants' from both countries shows similar behavior, which is contrary to the general viewpoint that social networking is used differently in different countries. However, the notable exception was use of email or private messages and comments. Norwegian users are more interested in making use of private messages than Pakistani users (mean 2.3 vs. 1.5) and comparatively less interested in comments of social contents (mean 1.8 vs. 2.2) illustrated in Fig. 3.
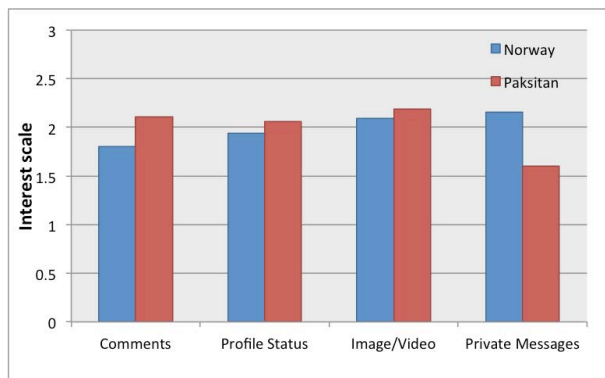
Fig 3: Importance of social activities on social network

It is noteworthy that 2.9% (5 265 880) of Pakistan population use Facebook compare to 53.64% (2 508 380) of Norwegian population[1]. This 2.9% of Pakistani population is mostly urban class this is one of the reasons why see stickling resemblances between the social activities of two countries.

4. LIMITATIONS

The primary limitation of this study is that, it may undermine the generalizability of results reported here. Regardless the uniqueness and the novelty of the findings in this survey, 247 participants do not allow any generalization. Moreover, Female participants were more reluctant to reveal their private information. Therefore it only accumulates the 26% of the participants.

There may be a relationship between likelihood of participation in the web survey and online social network users, especially in developing nations. People more likely to participate in online surveys from developing nations are the tiny urban educated class, which has access to the Internet. There is also a chance of like-minded participants groups, however, this is unlikely as we sent survey to many different people and also partly conducted in-person.

5. DISCUSSION

Users have many individuals in their online extended network that can hardly be termed as a real friend for various reasons. Thus, the assumption in some online social networks such as Facebook, that if "x is friend of y" then "y is friend of x" is erroneous. Furthermore, users usually do not have very high threshold for accepting friend requests from strangers (or acquaintances) - to become more visible and for self-promotions.

"Unreal friends" are one of the major reasons of erroneous results in SNA. Our study reveals that the number of unreal friends is inversely proportional to the users' age. We conclude that users with age above 30 have negligible "unreal friends". While analyzing the relationship the users above age of 30 can be safely neglected. Furthermore, the implication of having "unreal friends", is that online profiles receive more than hundreds of news feeds, media contents, and comments per day; many of these are of limited value. Despite the many advantages user gain when they interact through social networks with group of strangers, enough poor-quality and similar contents prove too difficult to be worthwhile.

Like offline social networks, people depend on one and another for social support in online social networks and, similarly, cannot stand what they do not feel comfortable with. Due to social pressure they sometimes cannot avoid the contents, as the publisher (friend) expects feedback. Consequently, many leave the online social network.

Contrary to the results reported in Golder work [10], that only a small proportion of communication was through private messaging between friends, we have found that the private messaging is significantly popular in the social networks. This trend probably has been changed in the recent years. This also implies that the friends no longer feel the need to use the email specific services for exchanging private messages. Researcher analyzing communications through social network between users usually neglect the private messages.

The one issue is privacy; users have little control and lack the competency to establish robust privacy settings for the profile. Prior research [11] has shown that users tend not to change default settings, even when the concrete privacy invasion was experienced by users [12]. For instance, on Facebook default privacy settings, if a friend makes comment on your photo, a friend of friend can see the whole album. The habitual use of online social network and its deeply integration into our daily routines outweighed the privacy concern. Participants' age under-20 is particularly vulnerable since they have many unreal friends and disclose the personal information for self-promotion that can be use or abuse by third parties.

6. CONCLUSION AND FUTURE WORK

Our study has examined the user behavior in social networks and discussed some of its potential implications. The results reveal that females are more reluctant to accept the friend's requests from strangers. We also reveal that the young users are liberal in broadening their social circle. We have compared the results from two contrasting countries – politically and socioeconomically, the comparison shows the users' behavior is quite similar. This study is small and cannot be generalized, but it highlights the need for further research.

Currently we are developing an automated semantically enabled classification system based on ontology, which will be tested in a bigger set of social network users and contents. These techniques will be further improved and integrated in our project for delivering personalized social contents. We will look for other ways to measure the strength of relationships and categorize it in different kinds of relationships. The inference capabilities of semantic web technology allow us to avoid explicit definition of some indirect relationships that will help us to find out content relevant for the user, which is not explicitly stated, and also more characteristics about the content.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Gross, R. and A. Acquisti, Information revelation and privacy in online social networks, in Proceedings of the 2005 ACM workshop on Privacy in the electronic society2005, ACM: Alexandria, VA, USA. p. 71-80.

---

[1] http://www.socialbakers.com/facebook-statistics

[2] Boyd, D. Friendster and publicly articulated social networking. in CHI '04: CHI '04 extended abstracts on Human factors in computing systems. 2004. ACM Press.

[3] Wasserman, S., Social network analysis: methods and applications1994: Cambridge University Press.

[4] Kate Ehrlich, I.C., Inside Social Network Analysis, 2005, IBM. p. 05-10.

[5] Hougham, V. Sociological Skills Used in the Capture of Saddam Hussein. 2005 October 17, 2008].

[6] Lampe, C., N. Ellison, and C. Steinfield, A face(book) in the crowd: social Searching vs. social browsing, in Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work2006, ACM: Banff, Alberta, Canada. p. 167-170.

[7] Dunbar, R., Grooming, Gossip, and the Evolution of Language.1998: Harvard University Press, Cambridge, MA.

[8] Jeong, Y. and R. Mahmood, Reading the world's mind: Political, socioeconomic and cultural approaches to understanding worldwide Internet search queries. International Communication Gazette, 2011. **73**(3): p. 233-251.

[9] Hasan, A., The Unplanned Revolution: Observations on the Processes of Socio-economic Change in Pakistan 2009: Oxford University Press.

[10] Golder, S., D. Wilkinson, and B. Huberman, Rhythms of social interaction: messaging within a massive online network. 2006.

[11] Mackay, W.E., Triggers and barriers to customizing software, in Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology1991, ACM: New Orleans, Louisiana, United States. p. 153-160.

[12] Debatin, B., et al., Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. Journal of Computer-Mediated Communication, 2009. **15**: p. 83-108.

Paper II

Ontology-Based Image Annotation by Leveraging Social Context

Najeeb Elahi, University of Tromsø, Norway

Randi Karlsen, University of Tromsø, Norway

Waqas Younas, Center for Advanced Studies in Engineering, Pakistan

ABSTRACT

Manual image annotation is an extensive and a cumbersome task, yet extremely important for image management and retrieval. The purpose of the authors' system is to semi-automatically generate ontology-based annotations for a social network by leveraging the annotations provided by the most active user (i.e., the central actor). Context of an image is of central importance in their approach towards semantic semi-automatic annotation. For context of an image, the authors consider several factors like geo-reference, time and relationship among actors in social networks and instead of using image-processing techniques to manipulate and interpret the image, their system leverages the context, which is automatically available along with the image and have also extended Social Network Analysis techniques by considering the granularity of relationships among actors under consideration. The authors use a semantic web framework to represent the social network and to deal with the diversity of relationships. OntoCAIM ontology is developed which not only encompasses Social Network Analysis functionality but also defines mechanism to annotate the images with an underlying ontology.

Keywords: Context, Image Annotation, Semantic Web, Semi-Automatically Generating Ontology-Based Annotations, Social Network Analysis

INTRODUCTION

Annotations are observed as statements about the image created by user, furthermore, not only they can be used for describing the image's content but also to classify, organize and search images. Annotation is known to be a key factor in image retrieval systems (Kahan & Koivunen, 2001). The most common approach used for image retrieval is Text-Based Image Retrieval (TBIR). TBIR retrieval is achieved by matching text to annotation associated with the image. The drawback of TBIR is that manual annotation of an image is a resource-intensive task and it is inherently expensive because users are unwilling to put large amount of effort to correctly annotate images. Even if human annotators are willing to annotate images, their annotation is highly subjective because different human perceive different things in the same image

and it is often found erroneous. For instance, the annotator in our lab once misidentified a building in Oslo as being in Sweden, whereas, in semi automatic annotations this problem can be avoided by plainly using GPS coordinates of image.

Therefore, it is imperative to find alternative ways to annotate images. Leveraging the annotation from the context of image can allow us to make meaningful connection with image contents (Sinha & Jain, 2008; Smeulders, Worring, Santini, Gupta, & Jain, 2000). Context of an image is of central importance in our approach towards semantic semi-automatic annotation. For the context of an image, we consider several factors like geo-reference, time and relationship among actors in social networks.

Due to resource intensive nature of manual image annotation question arises about acquiring most relevant annotation while avoiding the unnecessary and malicious annotations? Also, images taken by digital cameras typically include only basic metadata such as time and date, which alone is not sufficient and meaningful enough to correctly annotate images. Hence, in order to cater both of these issues we will be using a social network because, in a social network set of trusted people are linked to each other and their relationships are defined and this helps in acquiring relevant image annotations as they are filtered and recommended and more importantly images and provided annotations are trusted within a group. As group within social network involves like-minded users so this implies that users usually create similar annotations. We may also find semantic agreement among users over a set of images.

Our proposed system makes use of the existing metadata gathered from already annotated images in social networks, to generate more semantic metadata for non-annotated images. Since our approach for the most part relies on existing annotation provided by actors in social network, therefore, it is necessary to choose a correct actor (central actor). Hence, first we infer the central actor from the social network and then we will take her image annotations as a ground-truth for the target images. The

reference annotation is further investigated and paired with the analysis of Geo-reference and time difference as it improves the validity of results.

We took advantage from Social Network Analysis (SNA) techniques that provide the set of mathematical and algorithmic approaches to deduce central actor in social network. Resource Description Framework is used in SNA to examine the structure of social network and uncover the informal connection between social entities.

OntoCAIM ontology is developed that comprises of functionalities SNA and annotate the images with formal semantic, so that they can be integrated with other resources and consumed by machine.

The article starts with discussion of sources of image annotation, briefly describes the importance of social networks for image annotation and illustrates with a simple use case scenario. Next are the importance of ontology for our work and the functional architecture of system. The design of OntoCAIM ontology is followed by the rules developed to enhance the expressivity of ontology and the overview of related works. Finally, article concludes with the main features of the work and some comments on the future work.

SOURCES OF IMAGE ANNOTATION

In our application domain the context is the source of annotation, image's context that we are considering happens to be different from the context used in other application domains (Chen & Kotz, 2000; Schilit, Adams, & Want, 1994; Schmidt et al., 1999). The primary entity is an image and to determine the image's content/behavior we have gathered the surrounding environment of image from two different sources; namely Captured Context and Usage Context. Both of these contexts are illustrated in Figure 1 and using these contexts we will infer new metadata.

Figure 1. Layered classification of image context and infer metadata
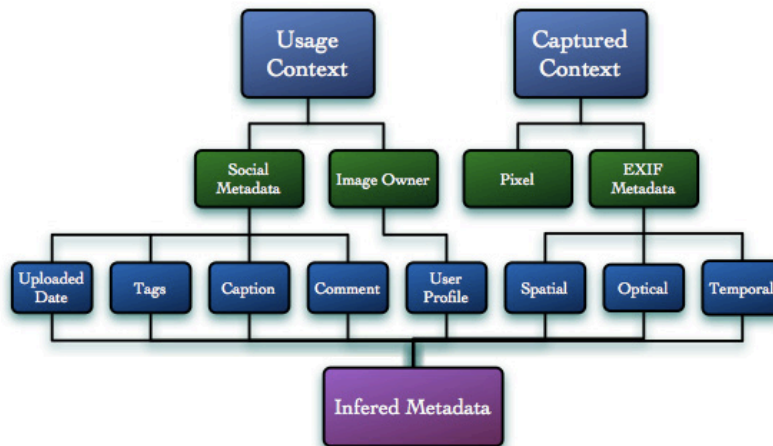
Captured Context

Captured Context describes when and where the image was taken and this information is extracted from digital camera in the form of EXIF (EXchangeable Image File) metadata associated with image that includes hundreds of tags, among them many are related to the image-capturing conditions. The EXIF metadata can be used in several ways for instance; focal distance, focal length and sensor size can be used to measure the size of object at the focal point and this can help us differentiate between different objects, provided that they are of the same color and shape.

Usage Context

Usage Context describes the surrounding environment as to where and how the image is being used. For our purposes we consider social network sites as a potential application domain, primarily because of their well defined structure (with apt granularity) and community activities over images.

SOCIAL NETWORK FOR IMAGE ANNOTATION

Collaborative nature of social network makes them a suitable candidate for applying our semi-automation image annotation technique. In this section, we describe social networks, characteristics of a central actor and more importantly also highlight the importance of relevance of annotations within social groups.

Social Network

Social networks are online platforms that allow users to create their personal profile, to establish various types of connections with other members of network and to publish variety of contents such as comments, images, videos, private messages and etc. Social network sites (Facebook, Flickr, MySpace, etc.) have revolutionized the way users share and upload images, annotate them with captions/tags/ comments/discriptions and the most importantly how their allowed friends can contribute in annotating images.

Social Network Analysis

Social network analysis primarily focuses on the analysis of patterns of social relationships among the set of social entities such as people, organization or states (Kate Ehrlich, 2005; Wasserman, 1994). Social networks analyses have been used in different application domains for variety of purposes. They have predominantly been used to successfully uncover the criminals (Hougham, 2005), identifying the missing connection between workers in organizations and

finding out the influential individuals (Ehrlich, 2005).

However, our primary goal in using a social network in this work is to find out the person who has most relevant images with annotations in a given network. Characteristics of such an individual are elaborated in next section.

Relevance of Social Group Specific Annotation

We have observed that annotations can be categorized in two types, namely: public annotations and private annotations. Private annotations are annotations that are relevant to a specific group whereas, public annotations are relevant to masses. For example, annotations such as "my birthday party in Los Angeles" and "besøk til Tromsø i 1981" can be considered as private annotations as they relate to specific events and people. Annotations like "Eiffel tower," "Oslo opera house" can be considered as public annotations as they do not relate to any specific person and, hence, can be used generally. In a certain social network group specific events are known to the whole group and this leads the whole group to create similar annotations in most cases. We have also observed that private annotations can often help most users in a social network group to annotate their images. In other words, users within a certain group may find semantic agreement on private annotations. Furthermore, we have found private annotations as a salient source to aid our semi-automatic image annotation technique for an actor within a social network group.

Relevant Image Annotation from Social Network

Usually people with similar interests and activities happen to have rapport between them. Likewise, group of like-minded people within social network frequently has correlated activities and common interests, and this can also lead them to correlated annotation of images. Annotation of one person can be used as ground-truth for other people in a group. Though it is true that harvesting more reference annotation data from the entire group can yield more ground-truth, but it can also create noisy and unnecessary annotations. If, for example, within a social network search for "Eiffel Tower" yields thousands of pictures but then finding anything useful will eventually become a predicament. Such noisy annotations may not allow us to develop a robust image annotation framework for social network.

Gloor (2009) has stated that in social network some actors have more influence than others and they can provide us most relevant information. However, finding many actors with more influence within a group can be computationally expensive and also it can lead our system to many undesired information paths. Therefore, in order to cater these issues we've relied on finding only one most influential central actor.

The way we choose center actor depends on following three conditions, this definition builds on contemporary definition of central actor. So a central actor in our case is the one that

- Has high centrality in a given social network (see section *Center Actor of Social Network*).
- Has more images with annotations relevant to the target actor (*target actor is the one whose images are going to be annotated*).
- Has high weighted relationship with target actor.

The image annotations of this central actor, paired with analysis of geo-reference and time difference can certainly improve the validity of the results.

USE CASE SCENARIO

In social networks, users actively participate by creating, sharing and communicating digital social artifacts. A significant fraction of these artifacts exist in the form of images, often with metadata. Following is a use case scenario that can help us understand the big picture.

This use-case is a story of three 3 actors within a social group, namely: Anna, Mani and

Kostas. They are friends and they decided to go to Paris on the occasion of 121st anniversary of Eiffel Tower. Kostas is a very passionate photographer and he took a lot of pictures of Eiffel Tower, lot of group photos, individual photos and photos of Paris city as seen from the Eiffel Tower.

After tour ended, Kostas created an album with a caption "Visit de Eiffel Tower" on his social network profile and also he annotated all images with names of people depicted in pictures (buddy list), main activities, weather and etc. Mani also created an album, consisting of few pictures, with a caption "tour de Paris." All of these three actors are active member of social network and they are connected with each other with a set of social relationships.

Kostas is one with most images with annotations relevant to Mani's recent activities. Our system, then, analyzes Kostas' annotations and suggests annotations to Mani. These annotations can contain items like buddy list (people annotated in Kostas' pictures), weather (e.g., they might be using their own particular vocabulary like "sunny," "breezy," etc.) condition, and main activities (e.g., watching concert, partying, dancing and etc.). Mani, upon receiving these annotations, is pleased to see the suggested annotations as they seem very relevant because they are indicative of the same event, same activities, same group of people and etc. Mani accepts the suggested annotations and he uses suggested annotations to annotate his pictures. Mani is pleased because of the fact that he did not have to manually annotate images from scratch.

DIMENSIONS OF IMAGE ANNOTATION

We are considering three dimensions in our work when it comes to image annotation. Firstly, we describe how to find the central actor in a social network and then we describe how to find and relate the geo-reference metadata and time difference information to our approach. Finally, we describe one possible scenario that sheds some light on these three image annotation dimensions.

Center Actor of Social Network

Social network assumes that actors are interconnected (Ehrlich, 2005) and that connections or relationships have real influence on the way they consume each other's resources. In order to analyse the relations and discover the most important actor we first assign the weight to the relationships. Calculating the weight of each relationship can draw us near to the most important actor. However, Aleman-Meza et al. concluded (2006) that altering the weights and thresholds alone do not improve the semantic analysis of social network results.

Therefore, in order to infer the most important actor and relationships among individuals in social network, we took advantage of SNA techniques that provide the set of mathematical and algorithmic approaches to deduce such information. Particularly we used SNA concept *centrality* that provides measures of interdependency that exists among set of actors as centrality identifies the most important actor of the social network (Erétéo et al., 2008; Freeman, 1978).

To find out the most important actor in social network, Wasserman (1994) suggested measures that can help us identify most important actor namely: degree centrality, closeness centrality and between centrality. The *degree centrality* identifies the most active actor that has most ties to the other actors in the network. The *closeness centrality* identifies the actor that can quickly interact with all other actors in the network and has short communication path to others i.e., he can communicate through shortest of edges. The *betweenness centrality* describes an actor as central if it lies between other actors on their geodesics and the interaction between two nonadjacent actors depends on that central actor.

The main problem with SNA is that, its focus is to examine the existence of relationships between people and it does not address different kind of relationships (Koelle et al., 2006). To address this problem, we assigned

weights to few edges/relations as per the nature of interaction that identifies the intensity of relationships. For instance, actor A is related to actor B with relation girlfriend and although A is not central actor as a whole, but for actor B, A can be inferred as a central actor by taking weight into account.

Since our purpose is to use the annotations provided by most important actor (discovered using above techniques) as ground-truth, we analyze the number of images and associated annotations by that central actor and this also contributes in the process of selecting central actor together with weight and centrality. In this paper we did not analyze the comments made on images or albums to avoid the Natural Language Processing complexity; however we counted the number of comments that show the popularity of images.

In a nutshell, we are calculating the central actor by using traditional SNA techniques, that is, by adding weights and by analyzing the provided annotation.

Time Difference

Time is usually used effectively to classify and manage digital images. If two images have a short time difference (threshold = 5 min but this can vary) and captured by same camera then most likely they are similar and encompass the same semantics too, because one camera cannot be located at different places in such a short time. Furthermore, a single user could carry more than one or different image capturing devices such as Phone Camera, Digital Camera, PDA etc., hence, we assume that user has already registered these devices with the system beforehand as it will help system to identify user. However, we can avoid the pain of registering devices with our system as we can extract image capturing device's code from the EXIF metadata. We classify all the images taken by single user with different devices in a certain time frame. We can use this time-difference to annotate a group of images with other single/group images. For instance, we can relate 8 images captured by Kostas'

different devices in a small period of time to Mani's single picture; because this (Mani's) single picture has a short time difference with 8 of the images captured by Kostas.

Geo-Referenced Annotation

The simplest geo-reference annotation can take place, first, by finding out images in a certain radius (threshold is 50 meters though it can vary) of associated GPS coordinates of target image and then finding out the common labels/keywords in these recently found images. Since keywords can have arbitrary meaning, clear identification of every important object/concept is needed. Suppose, we have a building labeled just as a "bank", which alone is a vague term because bank can be financial or can be used to define river of a bank. There has to be a mechanism for system to differentiate financial bank from the river of the bank so in order to do that it's imperative to clearly identify objects. In order to clearly identify correct concept of a bank, WordNet ontology is used.

Realization of Image Annotation Dimensions

Following is one of the possible semantic image annotation scenario in which we try to define our approach.

Mani and his friends are part of a social network. Mani went to Paris to celebrate 121st celebrations of Eiffel Tower. Mani is a social network user and he has uploaded a picture with GPS coordinates 2.294447,48.858306. Furthermore, this picture was taken on 2009-06-05T16:07:30 (the time and coordinates were extracted from image's EXIF metadata).

Next, system finds all users whose pictures, when compared to Mani's, are within 50 meters radius and also within 14 hours of time difference. All of the actors whose pictures fulfill this criterion will be included in our system. Let's call this subset of actors 'M'. For example, let's say Mani has a picture P and another actor, named A, in his social network has a picture P'. If coordinates of P' are within 50 meters radius of P coordinates and also if time difference of

Figure 2. Circle illustrates the subset 'M' of Social graph, "i" is number of images and "w" represents weights assigned to the relationships. Dotted lines show the semantically inferred relationships.
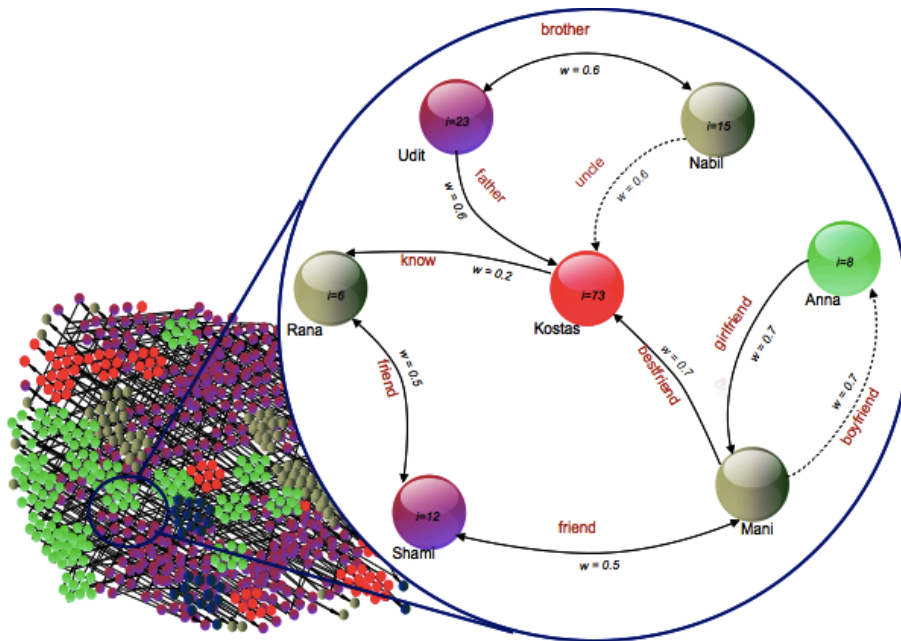
both P and P' is less than 14 hours then we're going to include A in subset M.

System will then have to find the central actor within subset M. In our pursuit of finding the central actor, we find that highly related actor in Mani's network is Kostas. Social Network Analysis reveals that, not only Kostas is linked to many people but he also has huge number of images with annotations that have short time difference and also taken within 50 meters of Mani's coordinates. Further, he has highly weighted relationship with Mani, thus making him central actor of the subset M (Figure 2).

Our system's aim is not to generate automatic annotation but instead to suggest the most relevant annotation to the user. So the system will analyze Kostas' annotations and semantically inferred annotations will be suggested to Mani. The newly inferred annotation is optionally visible to Mani and he can accept the suggestions given to him by this newly inferred annotation and his accepted suggestions go to the corresponding elements in the ontology.

Finally the Kosta's reference annotation is unambiguously classified and associated with Mani's album by using the reasoning capability of semantic technology.

LEVERAGING SEMANTIC TECHNOLOGIES TO ANNOTATE IMAGES

Keyword based annotation alone cannot establish the semantics of what is being marked-up, annotation should have formal definition to be unambiguously classified and consequently to be able to infer new facts. Our proposed system facilitates ontology-based annotation, which provides shared and common vocabulary, including the set of concepts, their properties, mutual relations and constraints. Machine can interpret and use the formal meaning for reasoning and validating the annotation.

Furthermore, Graph-based algorithms are most commonly used in SNA to examine the structure of social network and uncover the informal connection between social entities. However, problem with these graph-based classical algorithms is that they cannot represent the

Figure 3. The functional architecture of social contextual semantic image annotation
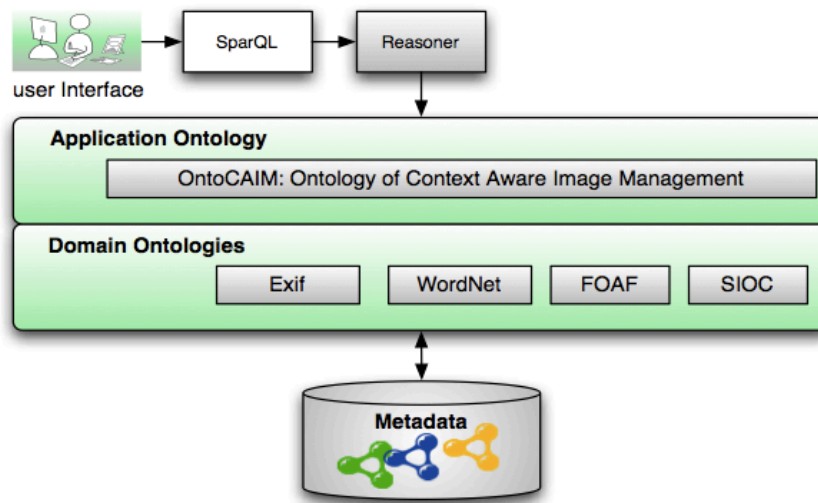
different aspects of human interaction without losing some knowledge (Guillaume Erétéo, Buffa, Gandon, & Corby, 2009). Therefore, we have built our proposed solution on a semantic web framework to deal with the problem of preserving, representing and sharing a rich social knowledge with a directed typed graph model Resource Description Framework (RDF) (W3C, 2004). Next section will detail the functional architecture of our semantic framework.

FUNCTIONAL ARCHITECTURE

Our implementation is based on the semantic web technologies that ensure the required fine-grained structure of social network and is sufficient to support annotations of all resources. It provides various technologies to formally specify domain knowledge to facilitate automation. Ontology is the backbone of semantic web. It captures the domain knowledge in the form of concepts and their relationships. We used some existing ontologies in social network domain as it is recommended to make better system integration. Our ontology OntoCAIM follows this recommendation by reusing the FOAF, SIOC (Beruetta, Brickley, Decker, Fernández, Görn, Harth et al., 2010), EXIF (MK, 2003), and

WordNet (W3C, 2006) ontologies to describe the social network and image resources.

Many ontologies can be used to represent the usage context of an image. The most popular ontology to model social network is FOAF (Brickley & Miller, 2010), as it is used to model people's identities, their relationships and their activities. The core of FOAF is to describe the user profile. For example, a *Person* entity describes personal information properties (such as foaf:img foaf:familyName), social properties (foaf:knows foaf:Group) and personal identities (*foaf:homepage foaf:email*). SIOC (Semantically-Interlinked Online Communities) ontology is commonly used with the conjunction of FOAF for representing the social networking information. SIOC defines the generic concepts and properties that are needed to describe the information from online communities. The main advantage of using SIOC is that it follows the modular design approach, that is, its concepts and properties can easily be specialized and extended by other ontologies.

EXIF is a well-known standard of representing basic image metadata used in digital cameras and in digital image management applications. To annotate the captured context of an image we take advantage of EXIF ontology. This ontology makes use of IDF (Image File Description) which exposes the

Figure 4. Main concepts of formulation of ontology through classes and properties
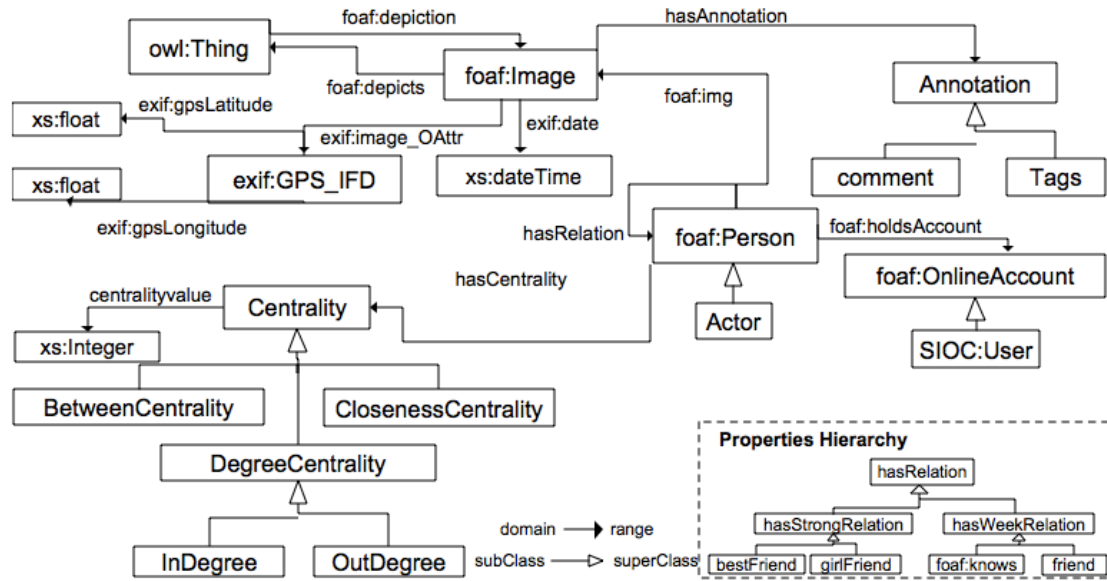
structure of data description (represented as exif:Exif_IFD), and have other concepts that are needed to represent the Geo-reference and time of image(such as exif:GPS_IFD and ExifData along with properties gpsLatitudeRef, gpsLongitudeRef, gpsAltitudeRef, dateTimeOriginal, dateTimeDigitized etc.). WordNet is another ontology that is used to reduce the ambiguity of natural language. WordNet not only offers the synonyms and antonyms of the words but it also provides rich semantics with relationship hierarchy (holonymy or part of relationship).

THE ONTOCAIM ONTOLOGY

Our system is grounded upon OntoCAIM ontology, which specifies domain knowledge for image annotation in social network and is formally represented in OWL (Web Ontology Language) (Bechhofer, van Harmelen, Hendler, Horrocks, McGuinness, Patel-Schneider, & Stein, 2004). We decided to use OWL DL which is based on description logic and is also the decidable part of First-Order Logic (FOL). Though OWL DL lacks expressivity when compared with OWL Full, but it maintains decidability and computational efficiency. Computational efficiency is

an important feature since the ontology has to deal with scores of social relations. One part of OntoCAIM ontology covers annotations that are properties of the image itself e.g., title, creator, image format, image size whereas the other part captures relationship with other objects depicted in the image.

In the following Figure 4, the rectangles represent the OWL concepts and arrows between them could represent any of the three properties depending upon the shape of arrow i.e., data property or object property or subclass/super-class axioms. Hence, concepts and properties without prefix in Figure 4 belong to OntoCAIM ontology whereas ones with prefix are imported from corresponding ontologies (e.g., FOAF, EXIF & SIOC).

Person is the fundamental concept of social network. We created a concept *Actor* as a subclass of *foaf:Person (Actor ⊆ foaf:Person)* and property *hasImage* as a sub property of *foaf:img (hasImage ⊆ foaf:img)*. Actor is a specialized person who must get some values from image through the property ∃ *hasimage.Image Actor*. Enforcing an Actor to get values from the image property increases the efficiency in terms of reasoning; so instead of examining all the persons reasoner will only consider those, in

an effort to find the desired annotation, that have some images. The concept *Centrality* captures the centrality of person through property *centralityvalue* and is determined through *is-a* relationship with {*BetweenCentrality, ClosenessCentrality, DegreeCentrality*} ⊆ *Centrality.*

As described in FOAF specification (Brickley & Miller, 2010) the property *foaf:knows* is a rather vague in design. More fine-grained properties are required to define the relationship among social actors. We introduced the hierarchy of relationships that implement the semantics of weight carried by relationships and further we categorized them according to weight, {*foaf:knows, friend*} ⊆ *hasWeakRelation*, {*bestfriend, girlfriend, boyfrined*}⊆ *hasStrongRelation*, {*hasStrongRelation, hasWeakRelation*}⊆*hasRelation.* In OWL, sub properties (*e.g., bestfriend, girlfriend*) are specialized form of super property (e.g., *hasstrongrelation*) and if the *sub property* links two individuals this implies that the two individuals are related by the super property. Only weak relation *Friend* is stated as a symmetric property in ontology, other relations have no such property restriction. For example, Mani can be declared as bestfriend of Kostas but Kostas can declare other type of relationship with Mani, or not have any relation with him at all. Properties *exif:gpsLatitude, exif:gpslongitude, exif:date and cetralityvalue* are datatype properties which takes xml primitive datatypes as a range (xs:float, xs:dateTime and xs:Integer), whereas remaining properties are the object properties which take classes as range.

ENHANCING EXPRESSIVITY OF ONTOCAIM

The semantic reasoning is not only used to infer the hierarchy of classes of ontology or to check the validity and consistency of the OWL knowledge base, but it can also be use to deduce the implicit knowledge on the bases of given explicit facts. However, the expressivity provided by OWL is limited by tree-like structure (Motik, Sattler, & Studer, 2005) and implicit knowledge cannot be inferred from the indirect relation between entities. Hence, implicit knowledge is highly desired in online social networks because of their super dynamic nature (new actors are joining, new relationships are building and the values of centrality keeps changing). To be able to infer indirect implicit knowledge, we use SWRL (Horrocks, Patel-Schneider, Boley, Tabet, Grosof, & Dean, 2004) (Semantic Web Rule Language), which add additional expressive power and maintain decidability to an extent when used in conjunction with OWL DL.

Formulation of Semantic Web Rules

SWRL is designed on the description logic foundations just like OWL DL, therefore, it is fully compatible with the formal semantic of OWL DL. SWRL has been proposed to extend the OWL axioms with rule axioms. The rules are expressed in terms of OWL classes, properties and individuals. Rules are defined as a set of antecedent (body) and consequent (head) parts containing conjunctions of atoms. If all the atoms in the antecedent hold, then conditions specified in the consequent must also hold. Atoms can be formed as

$$\text{Atom} \leftarrow C(x) \mid P(x,y) \mid Q(x,z) \mid \text{builtIn}(r,z_1,...,z_n)$$

Where C = OWL class, P = OWL Object property, Q= Data type property, x,y= object variable or individual z = data variable or value and r = Built-in relation

- C(x) hold if x is an instance of OWL description or data range.
- P(x,y) is true if x is related to y through P.
- Q(x,z) holds if x is related to z through Q.
- builtIn relation r holds if the given arguments z satisfy.

The following rules are formulated over the ontology that is treated as axioms of the

Table 1. Semantic web rules

Rule1	$Actor(?actor_A) \wedge Actor(?actor_B) \wedge hasStrongRelation(?actor_A, ?actor_B) \wedge hasCentrality(?actor_B, ?centrality_B) \wedge centralityValue(?centrality_B, ?centralityValue_B) \wedge swrlb:greaterThanOrEqual(?centralityValue_B, 5) \Rightarrow hasCentralActor(?actor_A, ?actor_B)$
Rule2	$Actor(?actor_A) \wedge Actor(?actor_B) \wedge hasCentralActor(?actor_A, ?actor_B) \wedge hasImage(?actor_A, ?image_A) \wedge hasImage(?actor_B, ?image_B) \wedge hasAnnotation(?image_B, ?annotation_B) \Rightarrow optionalAnnotation(?image_A, ?annotation_B)$
Rule3	$girlFriend(?person_A, ?person_B) \wedge FOAF:gender(?person_B, "Male") \Rightarrow boyFriend(?person_B, ?person_A)$
Rule4	$boyFriend(?person_A, ?person_B) \wedge FOAF:gender(?person_B, "Female") \Rightarrow girlFriend(?person_B, ?person_A)$

OntoCAIM ontology and this set of rules can be reused (Table 1).

Rule 1 compares each instance of class Actor and the one that has strong relation and high centrality will become central actor per rule1. Rule 2 takes central actor's image annotations and proposes newly inferred annotation to the target actor. Target actor can accept the newly inferred annotation and ontology knowledgebase is updated.

Rule 3 and 4 are developed to work like inverse property of OWL considering the OWL data property foaf:gender value. We had to take this into account because girlfriend cannot be inverse property of boyfriend, for instance, girl can be a girlfriend of a girl.

Varity of rule engines can be used to reason with SWRL rules; we chose the Jess rule engine for performing inference because of its great compatibility with Protégé OWL platform which is used to develop OntoCAIM ontology. At the beginning, Jess engine converts the relevant OWL knowledge and SWRL rules to Jess knowledge. Then the engine executes the rules and at the end the inferred facts exported back to the OntoCAIM knowledge base.

RELATED WORK

Recent research (Moxley, Kleban, & Manjunath, 2008; Moxley, Kleban, Xu, & Manjunath, 2009) on image annotation has mainly focused on Global Positioning System (GPS) and time-stamps of image while ignoring the context of the image.

While developing context-based annotation tool PhotoCompas (Naaman, Yeh, Garcia-Molina, & Paepcke, 2005), author obtains ground-truth from previously annotated photos to generate label suggestions for people that may be depicted in photos that are yet not annotated. They make use of timestamps and GPS coordinates as a context and identify the patterns of re-occurrence and co-occurrence of different people. However they do not investigate the granularity of relationship among the people.

In order to understand the semantics of an image, the most common approach to bridge the "semantic gap" is to use the multimedia ontological infrastructure (Smeulders et al., 2000). In semantic annotation of image collections (Laura, Guus, Jan, & Bob, 2003) author elaborates the use of ontology for image annotation and demonstrate the advance search strategies through the concept instead of keywords. However their approach, for the most part, relies on manual annotation. Another ontology based system called CONFOTO (Nowack, 2006) attempted to annotate conference photos through collaborative tagging with Resource Description Framework.

Photocopain (Tuffield et al., 2006) is a semi-automatic image annotation system, which acquires various type of metadata from the context of the image such as GPS coordinates, EXIF metadata, user's calendar and Flickr tags. Photocopain created service-based interoper-

able annotation system by using Semantic web standards. They have also used extraction techniques for image analysis. Our sources for acquiring metadata and use of technologies are very much like Photocopain, except that that did not significantly investigate the social context of an image.

CONCLUSION AND FUTURE WORK

In this paper we have proposed semantic image annotation with the help of social context. We have observed that inferring annotations from central actor's images can greatly reduce the overhead for manual and meaningful annotations of large set of images. We also demonstrated that ontologies together with the rules can provide an underlying platform to analyze the social network and then, infer the related annotation for the images from social network. We have used popular ontologies for establishing social network and for image annotation purpose. In the process, we identified some major stumbling blocks in applications that leverage semantics, such as reasoning limitations when it comes to mathematical operations with instances. Reasoning in OWL DL is based on what is known as the open world assumptions (OWA). When using OWA we cannot assume something does not exist until it is explicitly stated that it does not exist. Therefore, we were unable to count the number of image instansce that were crucial to find a social actor with large number of images. These limitations remain even after adding the rules, because SWRL is an extension of OWL DL. As a future work, we will implement Jess reasoner programming API to cope with aforementioned limitations and also we are concentrating on the practical realization of functional architecture, which has been illustrated in Figure 3. It includes linking OntoCAIM RDF triple store with SPARQL end point through an interface. Thus, it can be accessed by making SPARQL queries.

REFERENCES

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A. P., et al. (2006). Semantic analytics on social networks: Experiences in addressing the problem of conflict of interest detection. In *Proceedings of the 15th International Conference on World Wide Web* (pp. 407-416).

Bechhofer, S., van Harmelen, F., Hendler, J., Horrocks, I., McGuinness, D. L., Patel-Schneider, P. F., & Stein, L. A. (2004). *OWL Web ontology language reference*. Retrieved from http://www.w3.org/TR/owl-ref/

Beruetta, D., Brickley, D., Decker, S., Fernández, S., Görn, C., & Harth, A. …Polleres, A. (2010). *SIOC core ontology specification*. Retrieved from http://rdfs.org/sioc/spec/

Brickley, D., & Miller, L. (2010). *FOAF vocabulary specification. 98*. Retrieved from http://xmlns.com/foaf/spec/

Chen, G., & Kotz, D. (2000). *A survey of context-aware mobile computing research (Tech. Rep.)*. Hanover, NH: Dartmouth College.

Ehrlich, I. C. K. (2005). *Inside social network analysis* (No. 05-10). Armonk, NY: IBM.

Erétéo, G., Buffa, M., Gandon, F., & Corby, O. (2009). Analysis of a real online social network using semantic Web frameworks. In A. Bernstein, D. Karger, T. Heath, L. Feigenbaum, D. Maynard, E. Motta, & K. Thirunarayan (Eds.), *Proceedings of the International Conference on the Semantic Web* (LNCS 5823, pp. 180-195).

Erétéo, G., Buffa, M., Gandon, F., Grohan, P., Leitzelman, M., & Sander, P. (2008). A state of the art on social network analysis and its applications on a semantic web. In *Proceedings of the Social Data on the Web Workshop at the 7th International Semantic Web Conference*.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, *1*(3), 215–239. doi:10.1016/0378-8733(78)90021-7

Horrocks, I., Patel-Schneider, P. F., Boley, H., Tabet, S., Grosof, B., & Dean, M. (2004). *SWRL: A semantic Web rule language combining owl and RuleML*. Retrieved from http://www.w3.org/submission/swrl/

Hougham, V. (2005). *Sociological skills used in the capture of Saddam Hussein*. Retrieved October 17, 2008, from http://www.asanet.org/footnotes/julyaugust05/fn3.html

Kahan, J., & Koivunen, M.-R. (2001). Annotea: An open RDF infrastructure for shared Web annotations. In *Proceedings of the 10th International Conference on World Wide Web* (pp. 623-632).

Koelle, D., Pfautz, J., Farry, M., Cox, Z., Catto, G., & Campolongo, J. (2006). Applications of Bayesian belief networks in social network analysis. In *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA.

Laura, H., Guus, S., Jan, W., & Bob, W. (2003). Semantic annotation of image collections. In *Proceedings of the K-Cap Workshop on Knowledge Markup and Semantic Annotation*, Sanibel, FL.

MK. (2003). *Exif data description vocabulary*. Retrieved from http://www.kanzaki.com/ns/exif

Motik, B., Sattler, U., & Studer, R. (2005). Query answering for OWL-DL with rules. *Web Semantics*, *3*(1), 41–60. doi:10.1016/j.websem.2005.05.001

Moxley, E., Kleban, J., & Manjunath, B. S. (Spirittagger: A geo-aware tag suggestion tool mined from flickr. 2008). In *Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval*.

Moxley, E., Kleban, J., Xu, J., & Manjunath, B. S. (2009). Not all tags are created equal: Learning Flickr tag semantics for global annotation. In *Proceedings of the IEEE International Conference on Multimedia and Expo* (pp. 1452-1455).

Naaman, M., Yeh, R. B., Garcia-Molina, H., & Paepcke, A. (2005). Leveraging context to resolve identity in photo albums. In *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 178-187).

Nowack, B. (2006). CONFOTO: Browsing and annotating conference photos on the Semantic web. *Web Semantics: Science. Services and Agents on the World Wide Web*, *4*(4), 263–266. doi:10.1016/j.websem.2006.09.001

Peter, A. G. (2009). Web Science 2.0: Identifying trends through semantic social network analysis. In *Proceedings of the International Conference on Computational Science and Engineering* (Vol. 4).

Schilit, B., Adams, N., & Want, R. (1994). *Context-aware computing applications*. Palo Alto, CA: Xerox.

Schmidt, A., Aidoo, K. A., Takaluoma, A., Tuomela, U., Laerhoven, K. V., & Velde, W. V. d. (1999). Advanced interaction in context. In *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing* (pp. 89-101).

Sinha, P., & Jain, R. (2008). Classification and annotation of digital photos using optical context data. In *Proceedings of the International Conference on Content-Based Image and Video Retrieval* (pp. 309-318).

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(12), 1349–1380. doi:10.1109/34.895972

Tuffield, M., Harris, S., Dupplaw, D. P., Chakravarthy, A., Brewster, C., Gibbins, N., et al. (2006). Image annotation with Photocopain. In *Proceedings of the First International Workshop on Semantic Web Annotations for Multimedia* (Vol. 4).

W3C. (2004). *Resource description framework (RDF)*. Retrieved from http://www.w3.org/RDF/

W3C. (2006). *WordNet RDF/OWL files*. Retrieved from http://www.w3.org/2006/03/wn/wn20/

Wasserman, S. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511815478

Najeeb Elahi is a PhD fellow at the Department of Computer Science at the University of Tromso (UiT). Najeeb has obtained Master of Science in Software Engineering of Distributed Systems (SEDS) from Royal Institute of Technology (KTH), Stockholm Sweden and has worked in UniK research center in Oslo and Digital Enterprise Research Institute, Galway, Ireland. His research interests include Multimedia, Distributed Systems, Semantic Web, Ontology Design and Context Centric Approach for Image Management.

Randi Karlsen is an associate professor at Department of Computer Science, University of Tromsø, Norway. She has a PhD from University of Tromsø, and has been involved in a number of research projects for instance "Context-Aware Image Management" and "Advanced Transaction Management". Her main research interests include Interoperable systems, Information retrieval, Semantic Web, Data management in new applications, Transaction services in middleware, and Data management and transaction services in mobile environments.

Waqas Younas is graduate student in the Department of Electrical and Computer Engineering at Center for Advanced Studies in Engineering, Islamabad. His academic interests are varied but primarily he is interested in large-scale systems design, parallel computing and semantic web.

Paper III

Relation Based Image Retrival In Online Social Network

Najeeb Elahi
Department of Computer Science
The Arctic University of Norway
Tromsø-9037, Norway
najeeb.elahi@uit.no

Randi Karlsen
Department of Computer Science
The Arctic University of Norway
Tromsø-9037, Norway
randi@cs.uit.no

ABSTRACT

Online social networks allow users to share their photos with friend, family and the community at large. Social networks are formed with a number of users connected through different relationships, and the strength of these relationships has an influence on the way users react on each other photos. In this paper we investigate how we can assist users to retrieve the most relevant images from their social network. We propose Relation-Based Image Retrieval (RBIR), where social relationships are of central importance. For each user we calculate their relationships with other members in the network, and a ranked list of the closest and most reputed friends is compiled by analyzing the mutual activates between two users and their overall individual reputation in the social network. Comments and likes made by highly ranked members hold more weight, and photos are ranked in accordance with the number and weight of likes and comments they receive. To test our approach, we developed a prototype based on the Facebook platform, allowing users to search for images among their Facebook friends. The results demonstrate that our techniques are useful for retrieving relevant images.

Categories and Subject Descriptors

H.3.3 [**Social Interaction**]: Information Management—*Information Search and Retrieval*

General Terms

Image Retrival

Keywords

Photo Search, Social Photo Search, Social Ranking

1. INTRODUCTION

Online social network services have provided a delicate way of sharing personal photos among friends. People are attracted towards the photos that are somewhat related to them such as depicting known people, cities, featuring interesting events and most interesting relevant discussions over the photos from like-minded group of people. As social network sites are rapidly growing, the users are flooded with information, such as posts, video and photos contents. Easy access to such content emerging from vast networks of loosely defined acquaintance and complete strangers, leads to information overload and raise cautions of relevance of contents. Unlike all other contents in online social network arena, image contents are the most sustainable and therefore most explored [1].

The most common approaches used for image retrieval are Content-Based Image Retrieval (CBIR) and Text-Based Image Retrieval (TBIR)[3]. TBIR retrieval is achieved by matching text to annotation associated with the images, while CBIR uses an input image, which is matched to the visual content of the image archives. A drawback of TBIR is that manual image annotation is resource-intensive and highly subjective, while bridging the semantic gap [16] between low-level visual features and high-level semantic concepts is still a challenging task for CBIR. Furthermore, these conventional image retrieval techniques fall short to deal with online social image search as they disregard the social aspects and only focus on content quality and tags attached to photos and therefor may yield socially irrelevant retrieval results. This demands a new approach that in addition to the content and tags should exploit other social entities such as user relationships, comments and likes over photo.

The social network is formed with a number of users connected through different relationships [2]. The term Relationship is rooted on associations and various levels of trust between two people. Social theories[10, 7] have discussed the importance of strength of the relationships and that it has real influence on the way users consume each others resources. These relationships are the key to many features such as personalization, security, recommendations and most importantly search of social artifacts for instance, photos and videos.

We propose a Relation-Based Image Retrieval (RBIR) approach that makes use of social information already available in social network, to improve the user's search experience. We consider three basic social entities associated with

---

[1]Pixable reported in Feb, 2011 that average number of photos per user in social network Facebook is 97,000 http://blog.pixable.com/2011/02/14/facebook-photo-trends-infographic/.

[2]To date, most of social networks such as Facebook, orkut and Myspace offer only positive relationships between two users, such as close friend, family member.

photos, i.e. user relationships, user activities (comments and likes) and tags, and derive scoring functions for ranking relationships and photos. Relationship score measures the strength of a relationship and is derived from the time and effort invested to maintain contact between two participants. Photo score is derived by both considering the quality and worth of comments and likes on each photo. The worth of comments and likes are determined by analyzing the already ranked relationships through which they are coming from, for instance, comments coming from highly ranked friend weight more than other comments.

Social image retrieval is enabled by first selecting potentially relevant images by matching keywords in the text query against annotations related to images in the social network. The socially most relevant images are subsequently identified through a social ranking algorithm using relationship and photo scores. To test our approach, we developed a prototype on Facebook real-world dataset and results demonstrate that our techniques are useful for retrieving relevant images.

2. SYSTEM ARCHITECTURE

Our Architecture for Relation Based Image Retrieval is illustrated in Figure 1. We generated the ranked list of friends by analyzing the users mutual activities in social networks. Socially rich photos, commented and liked by highly ranked friends were consider more relevant. The *Data Collector* collects social network data about a user, such as her friends, their photos and comments, likes and tags associated with photos from the Facebook. For the first time user, *Data Collector* dispatches the social network information to the *Friendship and General user Score Calculator* to calculate the two different scores, relationship score and general user score, which are then store in a local database (Local DB). The *Data Collector* later receives updated data from Facebook and enriches the already stored social network information. The *Friendship and General user Score Calculator* recalculates the scores once a week since changes are not very frequently observed and to avoid the unnecessary computation load.
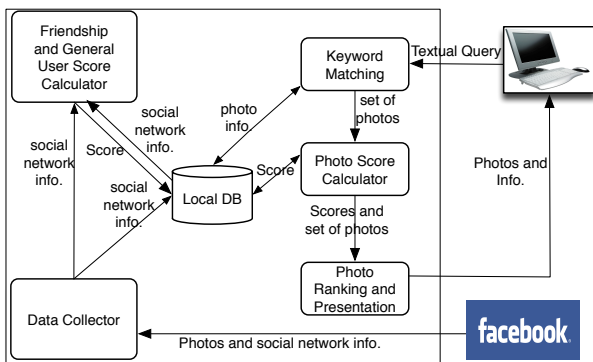

Figure 1: Architecture of the system

When a textual query is submitted to the system, *Keyword Matching* performs similarity measure between search string and keywords associated with photos and albums, and selected a set of photos from the *Local DB* that are send to *Photo Score Calculator*. *Photo Score Calculator* calculates

the social relevance of the photos on the bases of relationship and general user score between query initiator and photos owner, which are already stored in *Local DB*. *Photo Ranking and Presentation* receives the scores and set of photos, and ranks them according to the social relevance for the query initiator. Ranked photo are finally send for presentation.

3. SOCIAL NETWORK MODEL

Social networks are online platforms that allow users to create their personal profile, and also help them to articulate various types of connections with other members of the network. Facebook[3], Orkut[4], Flickr[5] and Myspace[6] are some examples of social networks.

Following are the four entities of social networks that have an important role in modeling of our social ranking algorithm.

- **Users:** people who are the member of the network, perform some activities, upload photos, produce tags and search.

- **Photos:** basic content items that are published by users and accessible for others to view, to make comments on or like. Social networks enable their users to publish many other contents, but in this paper we only consider photos.

- **Tags:** keywords used to describe the photo contents

- **Activity:** In general terms, activity is anything that one user does on another user's guestbook, such as browsing through photos, leaving comments, reading a public diary, writing testimonial and sharing contents for their established friends. However, to reduce the complexity, we only consider the comments and likes over photos made by other users as an activity.

4. RANKING OF USERS AND PHOTOS

Our proposed solution uses three different scoring functions for ranking the users and photos. *relationship-score*, which is a score regarding the friendship between two users, *general user-score* that is a global and general score for a specific user, and *photo-score* which indicates the importance of a photo to a user.

We use the following notations for formally defining the scoring functions. Each user has a set of photos, denoted P, i.e. $P = \{p_1, ..., p_n\}$. Each photo $p \in P$ has a set of likes and comments given by other users in the social network. The likes given to photo p are represented as a set $L(p)$, while the set of comments to p are represented as the set $C(p)$:

$$L(p) = \{l_1, ..., l_n\} \, C(p) = \{c_1, ..., c_m\}$$

Each like and comment has a property *uid*, which identifies the user providing the like or comment. A photo can receive only one like per user, while a user can comment a

number of times to the same photo, for instance in a discussion between several users over a specific photo. In our approach, we want to give more value to those photos where many users participate in discussions. Thus, the first comment from a user U to a photo p is given more weight then subsequent comments from U. For each comment, a property num represents the number this comment has in a possible succession of comments from user U to photo p.

4.1 Relationship Score

The Relationship-score aims to categorize the real binding between users. It is difficult to determine which relationship is more significant than others, because they are established by authoritative permission (single click), rather than through regular interaction [6] . In the beginning, social network user populates their friend list with people they know offline. As time passes by, they start inviting new users with which they share common interests and accepting friendship requests made by others. It is also observed that the user often accepting friends requests from acquaintances because of various social reasons [4] . As a result, a user get access to large sets of photo artifacts that are potentially not relevant to her/him.

To identify the closer relationships between people, we consider the activities between two specific users, and assume that users that have a close social relationship will have a higher likelihood to comment and like each others photos. Friends liking photos consider a vote of confidence and comments from the social circle invoke further discussion and generate interest in the photos. We did not analyze whether it is positive or negative information given in the comment, but either way, it shows the commenter was interested in the photo, hence she takes the time to make a comment. A relationship score between two users, U_i and U_j , is thus calculated based on the comment and like activities happening between the two users. The relationship score is described in Formula (1) .

$$F_{(U_i,U_j)} = \frac{\left|L_{(U_i,U_j)}\right| + \left|C_{(U_i,U_j)}\right|}{\left|P^{U_i}\right|} + \frac{\left|L_{(U_j,U_i)}\right| + \left|C_{(U_j,U_i)}\right|}{\left|P^{U_j}\right|} \quad (1)$$

Here $\left|L_{(U_i,U_j)}\right|$ represents the number of all likes from user U_j to photos owned by user U_i , $\left|C_{(U_i,U_j)}\right|$ the number of all comments from user U_j to photos owned by user U_i, and $\left|P^{U_i}\right|$ the number of photos owned by U_i. The sets $L_{(U_i,U_j)}$ and $C_{(U_i,U_j)}$ are described as follows:

$$L_{(U_i,U_j)} = \left\{l \mid l \in L\left(P^{U_i}\right) \wedge l.uid = U_j\right\}$$

$$C_{(U_i,U_j)} = \left\{c \mid c \in C\left(P^{U_i}\right) \wedge c.uid = U_j\right\}$$

4.2 General User Score

The characteristic of each of the user's friends is an aspect, which plays an important role in online and physical social interaction [2] . We define a general user-score as a score that describes a user's general popularity among his friends with respect to photos. To estimate this score we assume that

users with a high general popularity will have high activity on his photos. Moreover, a large number of friends in the user's social circle also indicate his popularity. The general popularity-score for the user is described in Formula (2).

$$G_U = \frac{\left|L(P^U)\right| + \left|C(P^U)\right|}{\left|P^{U_i}\right|} + log_2\left(N + 1\right) \quad (2)$$

The general user score is achieved by calculating the number of comments and likes that are received over photos and dividing by the number photos. N is the number of friends in U's social network, and log_2 is used to gradually decrease the impact of a large number of friends. Here $\left|L\left(P^U\right)\right|$ and $\left|C\left(P^U\right)\right|$ represent the number of all likes and comments (respectively) given to photos owned by U. The sets $L\left(P^U\right)$ and $C\left(P^U\right)$ are described as follows:

$$L\left(P^i\right) = \bigcup_{k=1}^{n} L\left(P^{U_k}\right)$$

$$C\left(P^U\right) = \bigcup_{k=1}^{n} C\left(P^{U_k}\right)$$

4.3 Photo Score

The photo-score is a score that describes how relevant a photo is for a specific user. The score is calculated by determining the strength of relationship between the user searching for photos and a photo owner, combining the quantity and weights of the likes and comments on the photo. The weights of the likes and comments are derived from general user-score and relationship-score.

The photo-score $S_p\left(P^{U_i}\right)$, described in Formula (3), represents the photo score for photo p owned by U_i, and is calculated as a combination of three different scores: relationship-score, $F_{(U_i,U_j)}$, which defines the direct binding between query initiator U_j and photo owner U_i, and likes-score, $S_L\left(P\right)$, and comments-score, $S_C\left(P\right)$, defines the amount and worth of the likes and comments respectively. Time ΔT represents the time in months since the photo was uploaded, and is used to give preference to the newly uploaded photos.

$$S_P\left(P^{U_i}\right) = h * F_{(U_i,U_j)} + S_L\left(P^{U_i}\right) + S_C\left(P^{U_i}\right) - \frac{\Delta T}{2} \quad (3)$$

Formula (4) describes the Like Score, $S_L\left(P^{U_i}\right)$, for photo p owned by U_i, and is calculated as a summarization of a combined relationship and general user score over all n users liking photo p. The value of the Like Score is determined by the strength of the relationship between photo owner and users liking p, and the general reputation of users liking p.

$$S_L\left(P^{U_i}\right) = u\sum_{j=1}^{n}\left(vF_{(U_i,U_j)} + G_{U_j}\right) \quad (4)$$

Formula (5) describes the Comment Score, $S_C\left(P^{U_i}\right)$, for photo p owned by U_i. The score is a summarization of a combined friendship and general user score over all m users commenting to photo p.

$$S_C\left(P^{U_i}\right) = \sum_{j=1}^{m} \left\{ \begin{array}{l} \left(sF_{(U_i,U_j)} + G_{U_j}\right) + \\ \left(tF_{(U_i,U_j)} + G_{U_j}\right) * \left(\left|C_{U_j}\left(P^{U_i}\right)\right| - 1\right) \end{array} \right. \quad (5)$$

In some cases photos can have long discussions between two or more users, and since we want photos with many different users commenting ranked higher than the photos with just a discussion between a few people, the first comment to a photo p by user U is given higher weight then subsequent comments from U. The first part of the summarization equation in Formula (5), is for the first comment to p from user U_j, second part is for all the other comments from U_j. The construct $\left(\left|C_{U_j}\left(P^{U_i}\right)\right| - 1\right)$ holds the number of comments from user U_j to photo p, minus the first comment. The set of all comments from user U_j to a photo p is described as:

$$C_{U_j}(P) = \{c \mid c \in C(P) \wedge c.uid = U_j.uid\}$$

5. EVALUATION
In this section, we describe how our approach to relation-based image retrieval was evaluated. We discuss how the data was collected and system testing was done.

5.1 Data Collection and Privacy Issues
The application gathered data from Facebook about users, their friends and photos. The information collected was the identifier of the user and all his friends, all the photos they have uploaded, all likes and comments on the photos, as well as the title, caption, location and tags on the photo. The photo itself is not downloaded; instead the URL to a static version of the picture (on Facebook-servers) was stored in the database.
We gathered all the test data in the period from 15 Feb 2013 to 04 Mar 2013. When the participants accepted to use our system, their data was downloaded to the database and calculations of relationship and general user score were made. Our data set contained a total of 193,869 photos belonging to 1,088 unique users. These photos had total of 227,111 comments, 288,704 likes and 886,817 tags. Figure 2 shows the details of how the data increased as the number of user increased.
The data were collected by approaching 6 core users that contributed with all photo information and 5 of these core users were used as assessor when evaluating the system. As we download much data about the user and his/her friends, we can assume that some of the data might be in some sense private and not wished to be posted public for everyone. The photos as well might only be meant for his/her friends on Facebook and they do not wish them to be publicly shared. For this reason we had to ensure that users testing the system were not able to see any more data than they could already using Facebook.

5.2 Evaluation method
We evaluate our relation-based image retrieval system, in the following denoted *Social Ranking*, by comparing with a baseline system *Without Ranking*, where relation-based ranking
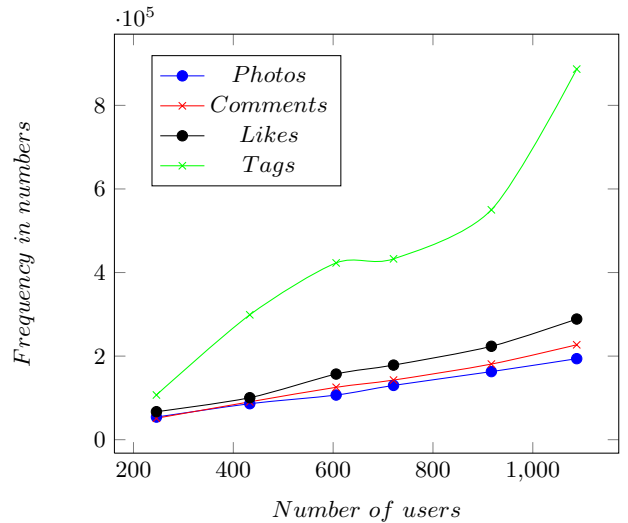

Figure 2: Statistics of collected data from Facebook

were not used. The baseline system retrieves images based on keyword matching. We used the integrated match against function in SQL, which returns a relevance score; that is, a similarity measure between the search string and keywords attached with the photos. Relevance scores are non-negative floating-point numbers. Zero relevance means no similarity. It was important that we evaluated our system on the same social dataset. Hence, we developed a baseline system Without Ranking, which does not consider the relationship and photos score but execute the textual query on the system with same environment such as network user account, data set and computational power. Each query issued by the users was executed by both systems, and the top 10 recommended photos from each system were presented to the user.

5.3 Testing
The notion of relevance for images in a social network is very much subjective and depends on the social context of the query initiator. Every query initiator obtains a different result for the same query since the scoring model explore the friendship and photo scores, which are user specific.
Every assessor was asked to make 7 quires, 4 queries about Events and 3 about Places in their respective social network. They were presented with a few example quires but they were free to make their own queries within the categories. Some of the quires made by assessors were *wedding*, *birthday*, *Oslo* and *snow in tromso*. Most of the quires we received were single word text quires.
For each textual query, the assessors were displayed a two column result page illustrated in Figure 2 that shows the top-10 results for both strategies *Social Ranking* and *Without Ranking*. The assessors were requested to identify for each of the strategies, the relevant images for each query and also mention the ranking position of the images. To help the assessors, resulted images were presented with keywords, description, photo owner and link to Facebook to find additional context when needed.

5.4 Evaluation Metrics

For the evaluation metrics, we adopted two metrics to evaluate the different aspects of our systems

- **Precision at K (P@K)**
  Precision at K is to measure whether the users were receiving relevant documents at the top of the ranking or not. We reported the precision at k=5 and k=10. It measures the precision when 5 and 10 photos have been seen.

- **Event and Place Category**
  We divided textual queries into two categories Event and Place. Event queries were made to search for photos, which depicted events, such as *football match* and *snow skating*. Place queries were made to search for the photos that were captured at or showed a specific place, such as *Oslo* or *Stockholm*.

6. RESULTS

This section presents the results of testing our algorithm *Social Ranking* and comparison with the baseline *Without Ranking*. Baseline is the text query over the same data set and other parameters for performance test.

Table 1 shows the average precision score for images socially relevant to the query initiator. The table displays separate scores for P@5 and P@10.

Table 1: Average precision of relevant images

	P@5	P@10
Social Ranking	0.711	0.671
Without Ranking	0.379	0.354

For the precision at 5 (P@5), we measure a precision of 0.711 for *Social Ranking*, which indicates that on average 71% of recommended photos were found relevant in top 5 photos. We measured the precision 0.379 for *Without Ranking* that shows only 38% photos were relevant in top 5. Expanding to the top 10 recommended photos (P@10) we see the performance goes slightly down for both *Social Ranking* and *Without Ranking* to 67% and 35% respectively and *Social Ranking* again produce better results but with less margin compare to the (P@5). We can thus safely argue that the *Social Ranking* strategy performs very well for both precision metrics.

Table 2: Average precision of relevant images

	Event	Place
Social Ranking	0.694	0.841
Without Ranking	0.347	0.466

Table 2 shows the average precision scores for images socially relevant to the query initiator. The table illustrates the separate scores for the two search query categories; Events and Places. We calculated the precision for top 10 photos against queries from both categories. We measured a precision of 0.694 and 0.347 for *Social Ranking* and *Without Ranking* respectively for the Event category. While the precision for the Place category was 0.841 and 0.466. Thus, queries for

Place perform slightly better than queries for Event. When looking at the relative difference in performance between the two strategies, *Social Ranking* and *Without Ranking*, we observe that for all metrics the *Social Ranking* strategy outperforms the *Without Ranking* strategy.

7. DISCUSSION

The results presented in the previous section indicate that using *Social Ranking* for relevant photo retrieval can be highly beneficial. We examine the performance of *Social Ranking* with comparison to the baseline and over all satisfaction of users with our system. Figure 3 illustrates the top 3 photos as a result of Event query *wedding*. The assessor describes that the socially ranked photos on the left column are very much relevant, where the first photo depicts her cousin getting married and the family had heavily commented and liked the photo. The second ranked photo is her close friend in marriage ceremony and the photo is commented and liked by many other close friends. In the same manner the last photo is relevant too.

In contrast, *Without Ranking* showed in the right column of Figure 3, did not perform well for the same query. The assessor reported that she knew the top ranked photo, while the remaining two photos were not interesting at all. However, these photos have been commented and liked by many, but not by friends of the assessor. In other words, these photos are maybe very popular in other social circles but not for this particular user.

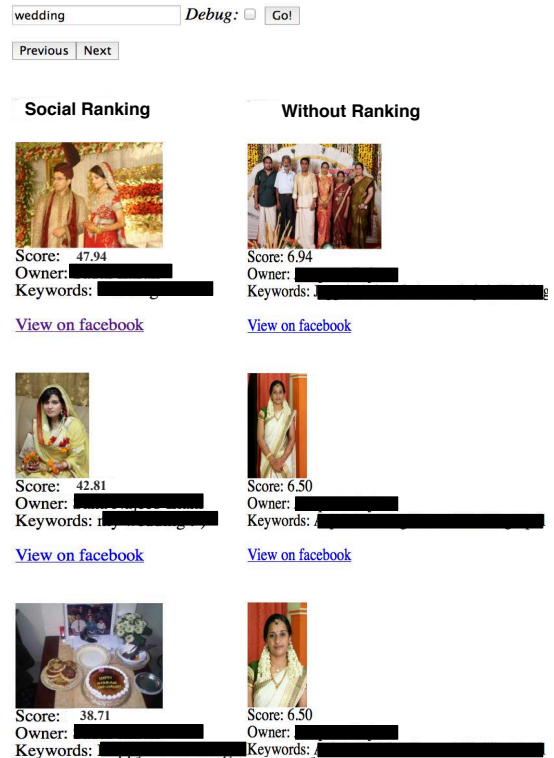The second result, illustrated in Figure 4, shows the top

Figure 3: First three photos for wedding query

5 photos as result of the Place query *Oslo*. The assessor explained that the socially ranked photos were very rele-
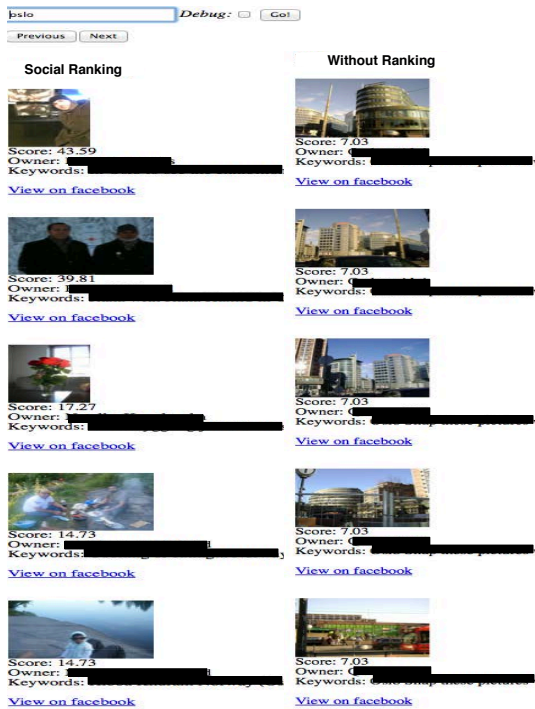
Figure 4: First five photos for Oslo query

vant. The topmost photo depicts one of his close friends portraying in a famous place in *Oslo*, and many of the assessor's close friends have liked and commented on the picture, which make it more interested for the assessor. In the same way, assessor found the second, fourth and fifth picture highly relevant, while the third is not that much relevant and therefore does not justify the third place in the photo ranking.

On the right side of Figure 4, we see the result of query matching using classic image retrieval without relation-based ranking. There the query *Oslo* retrieved photos of city buildings and famous places but fail to offer the social importance for the query initiator. Moreover, all the socially ranked photos came from different albums, uploaded by different users, whereas *Without Ranking* retrieved the photos from one person as a result of best textual matching.

The higher effectiveness of *Social Ranking* over the *Without Ranking* can be attributed to the fact that people share the likeness of photos with their social group. Friends liking photos consider a vote of confidence and comments from the close social circle generate high interest in the photos. Sometime, discussion over the photos has more importance for the user than the photo itself.

To use our system, the first time user grants permissions to the application, which acquires a Facebook-authentication token that is used to get access the users profile and download data. The execution time for the initial download of information from the Facebook-server is currently rather high. There are three main factors: Due to the large amount of data we are collecting about a user, the execution time of each query against the Facebook-servers was high. For some quires it could take up to 8 seconds to get a response. The second issue was that Facebook had a limit on the amount

of data you could download for each query. So to retrieve all the data we had to split up the queries, instead of getting all the information with just 1 query. This led to the last issue: Facebook has a limitation on the amount of queries that can be executed. For normal applications, like ours, this limit was 6000 queries per 10 minutes and if the user had a large amount of friends on Facebook we would often need to execute as much as 10,000 queries.

Nevertheless, once the data is downloaded, we calculate the friendship and general user scores locally, which do not require long execution time. Furthermore, we recalculate the scores once a week since the changes are not very often observed. Hence the system is not computational expansive and run seamlessly.

8. RELATED WORK

The accomplishment of online social networks and amount of information participants voluntarily reveal has attracted the attention of researchers [12, 8] that focus on harvesting online social network profiles for use in recommender systems.

Social searching receives enormous recognition, for instance through the use of Google's PageRank algorithm, which determines the importance of web pages by analyzing the link structure on the web. The social approach gives the opportunity for more personalized and in consequence more relevant answers to queries asked by individuals. FolkRank [9] is another algorithm for assigning the authority scores to elements in the network and consequently improves the result ranking. FolkRank is inspired by the renowned PageRank [14] that exploit the structure of folksonomy.

The usefulness of social information to recommend the most relevant tags for images in Flickr is described in [15]. They have investigated the user behavior of tagging images in Flickr and derived the scoring function from analyzing the collective social knowledge. In [5] authors make use of social information, particularly the wide range of social messages from the audience, posted on Twitter about the TV programs in Japan and infer the TV viewing rates as oppose to the conventional survey based TV rating.

Focusing on an approach that retrieves documents on the base of social information, the work presented in [1] is most aligned with our approach. They exploited the social relations to rank the resulted contents items from deli.cio.us and Flickr. They presented a framework to cast the different entities of networks into a unified graph model representing the mutual relationship of users, content and tags, and described the scoring functions for each entity (users, content and tags) and relations between them. However, they simplify the friendship between users, by only measuring the shortest path connecting the two users. They also did not consider the likes and comments made over the social contents.

We believe in general, that both the social information existing in social networks as well as user ranking is needed to determine the social relevance of content items. However, both will benefit more if we combine them effectively. In our proposed approach, we ranked users based on the social activities on the each other photos (comments and likes). Moreover, we predominantly consider the activities associated with our target item photos.

9. CONCLUSIONS

In this paper, we described our approach to retrieve socially relevant photos. Social relationships of user in online social network is the base of the proposed Relation-Based Image Retrieval (RBIR) approach. We generated the ranked list of friends by analyzing their mutual activities. Comments and likes made by highly ranked friends hold more weight, and photos are ranked in accordance with the number and weight of likes and comments they receive. We developed a prototype based on the Facebook platform, to test our approach. The results show that our techniques are useful for retrieving relevant images.

We have observed that the single word text quires outperform the longer text quires. An interesting future direction could build on the idea in [17] to develop a classifier that employs Wikipedia to expand semantic sense for query text via the WordNet ontology. Furthermore, We observed that there is a gap between textual query and underneath social graph. We plan to use semantic web technologies to structure the social network data. DBpedia [13, 11] will be used for name disambiguation of tags and keywords for easier search.

10. ACKNOWLEDGMENT

11. REFERENCES

[1] Bender, M., Crecelius, T., Kacimi, M., Michel, S., Neumann, T., Parreira, J. X., Schenkel, R., and Weikum, G. Exploiting social relations for query expansion and result ranking. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on* (2008), pp. 501–506.

[2] Caropreso, E., and Chen, S.-J. Effects of Personality on Small Group Communication and Task Engagement in an. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education* (2005), no. 1, pp. 1921–1930.

[3] Datta, R., Joshi, D., Li, J., and Wang, J. Z. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys 40*, 2 (Apr. 2008), 1–60.

[4] Elahi, N., and Karlsen, R. User behavior in online social networks and its implications. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12* (New York, New York, USA, June 2012), ACM Press, p. 1.

[5] Firan, C. S., Nejdl, W., and Paiu, R. The Benefit of Using Tag-Based Profiles. In *Proceedings of the 2007 Latin American Web Conference* (Oct. 2007),

IEEE Computer Society, pp. 154–154.

[6] Golder, S., Wilkinson, D., and Huberman, B. Rhythms of social interaction: messaging within a massive online network. *Communities and Technologies 2007* (2007), 41–66.

[7] Granovetter, M. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory 1* (1983), 201.

[8] Guy, I., Zwerdling, N., Ronen, I., Carmel, D., and Uziel, E. Social media recommendation based on people and tags. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (Geneva, Switzerland, 2010), ACM, pp. 194–201.

[9] Hotho, A., Jäschke, R., Schmitz, C., and Stumme, G. FolkRank: A Ranking Algorithm for Folksonomies. In *Proc. FGIR 2006* (2006), pp. 111–114.

[10] Kate Ehrlich, I. C., and Report, I. B. M. T. Inside Social Network Analysis, 2005.

[11] Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P. N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., and Bizer, C. DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia. *Semantic Web Journal* (2013).

[12] Liu, H., and Maes, P. Interestmap: Harvesting social network profiles for recommendations. In *In Proceedings of the Beyond Personalization 2005 Workshop* (2005).

[13] Mendes, P., Jakob, M., García-Silva, A., and Bizer, C. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems* (New York, NY, USA, 2011), pp. 1–8.

[14] Mizzaro, S., and Vassena, L. A social approach to context-aware retrieval. *World Wide Web 14*, 4 (Feb. 2011), 377–405.

[15] Sigurbjörnsson, B., and van Zwol, R. Flickr tag recommendation based on collective knowledge. In *Proceeding of the 17th international conference on World Wide Web - WWW '08* (New York, New York, USA, Apr. 2008), ACM Press, p. 327.

[16] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell. 22*, 12 (2000), 1349–1380.

[17] Wakamiya, S., Lee, R., and Sumiya, K. Towards better TV viewing rates. In *Proceedings of the 5th International Confernece on Ubiquitous Information Management and Communication - ICUIMC '11* (New York, New York, USA, Feb. 2011), ACM Press, p. 1.

Paper IV

Personalized Photo Recommendation By Leveraging User Modeling On Social Network

Najeeb Elahi
Department of Computer
Science. The Arctic University
of Norway
najeeb.elahi@uit.no

Randi Karlsen
Department of Computer
Science. The Arctic University
of Norway
randi@cs.uit.no

Einar J. Holsbø
Department of Computer
Science. The Arctic University
of Norway
einar.j.holsbo@uit.no

ABSTRACT

An online social network is a digital representation of the set of human beings on the Internet. Social network services generate large amount of usage data; for example, Facebook defines detailed user profiles, and provides a platform for sharing information with a vast network of friends, and Flickr offers sophisticated ways for sharing and searching for photos. In this paper, We propose cross-domain user profile modeling that acquires background knowledge from Linked Open Data and measures user interests. We infer the user's preferences by analyzing her Facebook profile, and expand it by linking it to Flickr in order to recommend socially relevant photos.

Categories and Subject Descriptors

H.3.3 [**Recommendation Systems**]: Information Storage and Retrieval—*Social Interaction*

General Terms

Image Retrieval

Keywords

Social Networks, Semantic Annotation, User Profile Modeling

1. INTRODUCTION

The provision of personalized content demands precise modeling of user interests and needs. To effectively model a user's interests, appropriate data and background knowledge about the user is required. Online social networks process and manage large amount of data [12]; and services such as Facebook[1] and Flickr[2] have emerged as a major medium of communication as they provide a platform for sharing personal information with a vast network of friends.

---

[1]https://www.facebook.com/
[2]http://www.flickr.com/

To explore personalization of metadata-based photo retrieval, we examine two online social networks as sources of photos, photo metadata, and profile information for a given user. We examine Facebook, which defines detailed user profiles, and allows the user to share information with a large group of friends. The users explicitly fill out their Facebook profiles with user name, email, likes, favorite groups and other information. In order to address certain holes in the information that Facebook provides (e.g., Facebook aggressively compresses any photo uploaded to their servers and strips all $exif$ information), we look into augmenting our recommendation engine with Flickr data. Flickr has a special focus on facilitating photo sharing both publically and with a network of friends. They provide photos with more detailed metadata, and at a much higher resolution than photos on Facebook, but at the same time they provide much less information about the user. Thus the two social networks can be used to complement each other.

However, social networks lack standards to uniformly process and manage social information [12]. Semantic web technologies can be used to structure the social data so that it can be integrated and be consumed across the platform [4]. Tim Berners-Lee [13] warns about the increase of islands on the Web, and proposed the Linked Data [5] which refers to a set of best practices for publishing and connecting structured data on the Web. Linked Open Data cloud provides data along with metadata in such a way that can easily consume by other services. Linked Open Data is leading to the creation of a global data space, the Web of Data, which contains billions of assertions from a number of applications[3].

We model the user's profile by aggregating user data across different social networks in order to provide personalization. We have analyzed the nature of user profile traces distributed on both social networks Flickr and Facebook, to infer the preference of the user towards photos. Linked Open Data is used to acquire background knowledge for name disambiguation of tags and topics, to provide semantic enrichment of social tags. For instance, the GeoNames [4] Ontology makes it possible to add geospatial semantic information to the photos, and DBpedia is a rich source of different types of sematic information.

2. USER PROFILE FRAMEWORK

Our framework for photo recommendation is illustrated in Figure 1. Following is the detailed description of the system.
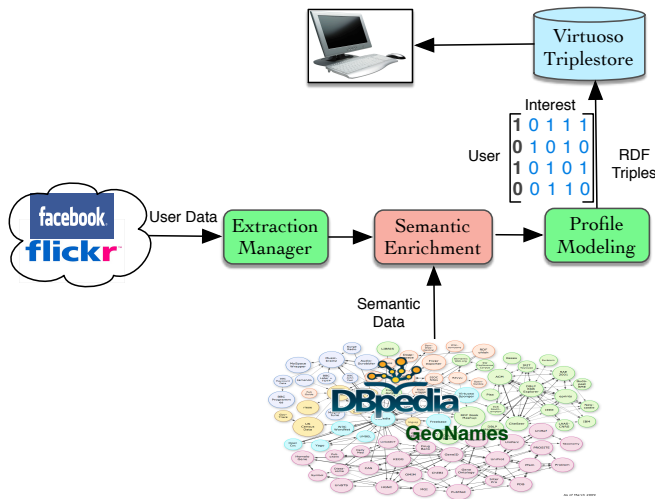
---

[3]http://www.w3.org/2001/sw/sweo/public/UseCases/
[4]http://www.geonames.org/ontology/documentation.html

Figure 1: Framework of the system

2.1 User Data

We collect social network data about a user, such as her friends, their photos and comments, likes, as well as the title, caption, location, upload date, and tags associated with photos from Facebook. The photo itself is not downloaded; instead, a URL to a static version of the picture (on Facebook-servers) is stored. We also collect similar data about the user from Flickr.

Note that a tag in Facebook is not like a social tag in another system, such as Flickr. Flickr tags are labels to describe and organize photos, which makes the photos easier to find later. Facebook tags, on the other hand, are associated with photos are only supposed to be used if the photo depicts someone. i.e. a tag in Facebook is a way of saying "this person is in this photo". Interestingly, a trend has emerged where people are tagged in a photo not because they are in it, but because the owner of a photo wants the tagged user to see the photo (a user gets notified whenever she is tagged in a photo). Either way, it suggests the importance of a photo to the people tagged. We assume, that a photo tagged with a person actually has the same importance as depicting the user.

2.2 Extraction Manager

The Extraction Manager gathers a user's Facebook data and converts it to Resource Description Framework (RDF[5]) triples. These extracted triples are then later stored in a Virtuoso[7] triple store. Once the Facebook data has been gathered, the manager attempts to find public Flickr accounts corresponding to these users. It would be preferable for the users to register their Flickr accounts with the system, but an automated harvester has the advantage that it might locate a user's friends as well. The Flickr harvester searches Flickr for users based on provided nickname, and/or email-address.

In Virtuoso it is possible to create an inverted index for triples with text as their object, which again enables one to do binary text searches in the indexed triples. This forms a

---

basis for image retrieval, where possible candidates can be retrieved by this binary search, and then ranked.

2.3 Semantic Enrichment

We extract the photo tags and user interests from their profiles and annotate these with DBpedia[6] resources. DBpedia is used as an interlinking hub in the Web of Data and enables access to the rich data sources in the Linked Open Data cloud. To better understand the semantic of tags and interests, we detected and annotated them with DBpedia entities such as places, organizations or people, for instance the tag of "Oslo" is linked to unique URI $http://dbpedia.org/resource/Oslo$ and the meaning of the concept "Oslo" is well defined. Annotation was realized via the name entity recognition services DBpedia Spotlight [10]. GeoNames information could be found in order to further enrich the social tags, once you have determined, via Spotlight, that Oslo is a place, since most GeoNames instance e.g. $gn : Feature$ are linked with DBpedia resource through $owl : sameAs$ [7] property.

3. ONTOLOGY MODELING

Many ontologies can be used to represent the user. The most popular ontology to model people is FOAF [6]; it is used to model people's identity, their relationships and their activities. The core of FOAF is used to describe the user profile. For example, a Person entity describes personal information properties (such as *foaf:img foaf:familyName*), social properties(*foaf:knows foaf:Group*) and personal identities (*foaf:homepage foaf:mbox*). Our system largely uses classes and properties from FOAF, but it is also necessary to model some additional concepts.
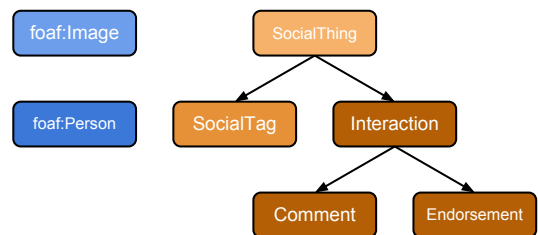
Figure 2: Ontology classes

Figure 2 shows the classes used in this system. In addition to foaf:Person and foaf:Image, we have developed the SocialThing class to our ontology. The SocialThing class describes things that are ubiquitous in social networks, but FOAF does not cover. FOAF primarily aims to describe a person, not a person's activity in a social network. We use user interactions for calculating a score of how close two friends are (more interactions is assumed to imply a "closer" relationship). Interactions encompass commenting on another user's photos, and endorsing another user's photos. An endorsement in Facebook is a Like, an endorsement in Flickr is the act of adding a photo to one's favorites, and similarly in other social networks. A social tag is when users are allowed to annotate photos with their own keywords. The ability to add social tags is limited on Facebook, but is central on Flickr.

Figure 3 shows the properties that enable us to integrate the
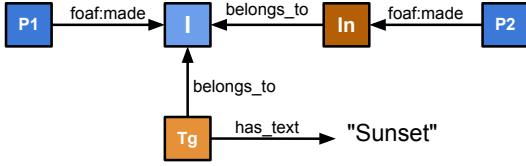
Figure 3: Ontology properties

*SocialThing*s with FOAF classes, and how the *SocialThing*s are used. A *foaf:person*, P1, will have *foaf:made* a photo, I. Another person, P2, has interacted with P1 via the image, i.e., made a comment on or endorsed the image. The interaction *belongs_to* the *foaf:Image*. Similarly, a *SocialTag*, Tg, also *belongs_to* the *Image*, but we do not keep track of who made it. The tag Tg *has_text* "Sunset" is annotated with DBpedia resource. The interaction *Comment* could conceivably also have text, but we do not collect the text of comments, as it is not relevant to our use. Finally, a photo can be *taken_near* somewhere, and a person location can be described as *based_near* somewhere.

In addition to the above properties, a number of FOAF properties are used to model the user. When available, *foaf:mbox* and *foaf:homepage* are registered, a user obviously *foaf:knows* the people listed as friends, a user's likes and interests are recorded as *foaf:topic_interest*, and their current location is recorded as *foaf:based_near*. If someone is Facebook-tagged in an image, i.e., someone claims the person is actually in the image, this image *foaf:depicts* the user. Where these things are available, they are also recorded for a user's friends. As we will see, most of the above mentioned properties can be used directly or indirectly in recommending a photo for a user.

4. USER PROFILE MODELING

Users register different types of profiles on social networks. In most cases, they are offered to fill online form to make a profile to the system. These profiles can be viewed as (field, value) pairs[3]. Let p be the profile where
$p(u) = \{(f, v) \forall f \in F \mid F$ is the set of fields, and v is corresponding value to $f\}$

F defines the range of fields that models the characteristics of user, u. In this paper, we are dealing with photo recommendations therefore we focus on these characteristics $F_u = \{Friends, Interests, Photos\}$. To develop an understanding about the user, we aggregate the user data from two different sources.

for the given user u_i
Let $p_{aggr}(u_i) = \{\}$ to be aggregated with more values.
for all $(f, v) \in p(u_i)$ **do**
  if $v \neq \{\}$ **then**
    $(f, v) \rightarrow_{align} (f', v')$
    add (f', v') *to* $p_{aggr}(u_i)$
  end if
end for
return $p_{aggr}(u_i)$

In our system we take the Facebook profile (f, v) and map it with Flickr (f', v') pair. As profiles are sparse in information, there are not too many features to use for profile matching; nickname and full name are the most commonly

available features. Also freely available are the photo tags the user uses. These will obviously not be the most reliable features, but may give some hint, everything else being equal. We make use of these features to identify the user in Flickr. A user's email, where available, is a particularly attractive feature, as it is pretty much a guaranteed match as compared to nickname, which can provide a false positive, or real name, which many users do not supply at all unless forced to. The problem is that people are often wary of displaying their emails to outsiders, and as such, it is often impossible to use this feature in practice.

Given a set of user profiles from Facebook B, and a set of user profiles from Flickr Q, we calculate similarity scores $sim(b, q) \forall b \in B, q \in Q$ following the approach of [14]. If $sim(b, q) > t$ for some threshold t, we deem the two user profiles to belong to the same person. The similarity function (1) compares the profiles field-by-field from the available information, considering some fields more important than others.

$$sim(b, q) = \sum_{f \in F} w[cmp(v_b, v_q)] \qquad (1)$$

Let b, q be two different profiles we want to compare. Let the set of possible fields be denoted F, and the values for some $f \in F$ be denoted v_b and v_q for profile b and q respectively. The comparison function cmp depends on f and w is the weight assigned to field f. In our system, the possible fields are, in order of importance, email, full name, nickname, and social tag set. Email and full name are fuzzy-matched as in [14], nickname is substring matched and tag sets are compared by simply counting the number of overlapping tags used in the two profiles.

5. PHOTO RECOMMENDATIONS

Our photo recommendations are based on the user's top ranked friends and their interests, which we infer by analyzing her aggregated profile. Let the aggregated user profile $p(u)$ consist of the ranked friend list $R(u)$, and the user's interests $I(u)$:

Ranked Friends: We made the list of ranked friend in our previous work [11] for each user u, we calculate their relationships with other members in the network, and a ranked list $R(u)$ of the closest and most reputed friends is compiled by analyzing their mutual interactions (i.e, likes and comments on photos) between two users, and their overall individual reputation in the social network. We now investigate the *foaf:depiction* property to find the appearance in each other photos. We assume that for users that appear in each other photos in social network, there is a high likelihood that they are interested in each other's photos.

User Interests and other metadata: A user's interest is here described through the pages that the user has Liked on Facebook (this includes things marked as interests, books they like, etc.), the user's current location, and whether a photo depicts the user or any of her friends. The user's five top social tag on Flickr is also added to her interests.

The available photo metadata on Facebook, e.g., caption, album name, is for simplicity treated as *SocialTags*. The exception is photo location, which, where available, is treated as a *taken_near* property.

In order to recommend an image for a user, we want to take into account her interests, but we also consider what

her closest friends are interested in. Thus we use the set of ranked friends to figure out which are the more popular interests in the user's immediate circle. Let $I(u)$ denote the interests of user u, and $I(R)$ denote the top n interests in the set of u's ranked friends, for some suitable n. The user's group interest set can be defined as $I_g(u) = I(u) \cup I(R)$. Consider the vector space with all possible interests and social tags as its dimensions. Let a photo, o be represented as a vector $\bar{o}$ in this space by its social tags, and let $I_g(u)$ be another vector $\bar{i_g}$. We define a recommendation threshold t_r, and define a recommendation function (2), where $dist(\bar{o}, \bar{i_g})$ is the euclidean distance between $\bar{o}$ and $\bar{i_g}$.

$$rec(o, u) = \begin{cases} \text{if } dist(\bar{o}, \bar{i_g}) < t_r, & True \\ \text{else.} & False \end{cases} \quad (2)$$

It is possible to augment $\bar{o}$ with the I_g of all users who have interacted with o, as the social tags are bound to be pretty sparse.

6. RELATED WORK

The accomplishment of online social networks and amount of information participants voluntarily reveal has attracted the attention of researchers [9, 8] that focus on harvesting online social network profiles for use in recommender systems. Exploiting posts and status message on Google+, Facebook and Twitter by people have been proven beneficial for the applications that need to grasp the demands of users [1, 2]. In [15] authors make use of social information, particularly the wide range of social messages from the audience, posted on Twitter about the TV programs in Japan and infer the TV viewing rates as oppose to the conventional survey based TV rating. Focusing on an approach that retrieves documents on the base of social information, the work presented in [8] is most aligned with our approach. They exploited the usefulness of top friends and top tags in social network to recommend the most relevant item. However, they simplify the recommendation by only measuring the top friends and tags and unlike our work, do not consider social activities such as like and comments.

7. CONCLUSIONS

In this paper, we describe our approach to recommend socially relevant photos. We propose cross-domain user profile modeling that acquires background knowledge from Linked Open Data and measures user interests. We infer the user's preferences by analyzing her profiles on Facebook and Flickr in order to recommend the socially most relevant photos. One can imagine a scenario where other social networks could be brought in to provide further information of the user and her network, including Twitter and Instagram. All of the above mentioned networks provide almost similar models for interacting with photo, e.g. a user's friends can endorse an image by "liking" it on Facebook, "hearting" it on Instagram.

8. REFERENCES

[1] ABEL, F., GAO, Q., HOUBEN, G.-J., AND TAO, K. Analyzing user modeling on twitter for personalized news recommendations. *Proceedings of the 19th international conference on User modeling, adaption, and persona* (July 2011), 1–12.

[2] ABEL, F., HAUFF, C., HOUBEN, G.-J., AND TAO, K. *Leveraging user modeling on the social web with linked data*, vol. 7387 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, July 2012.

[3] ABEL, F., HERDER, E., HOUBEN, G.-J., HENZE, N., AND KRAUSE, D. Cross-system user modeling and personalization on the Social Web. *User Modeling and User-Adapted Interaction 23*, 2-3 (Nov. 2012), 169–209.

[4] BERNERS-LEE, T., HENDLER, J., LASSILA, O., AND OTHERS. The semantic web. *Scientific american 284*, 5 (2001), 28–37.

[5] BIZER, C., HEATH, T., AND BERNERS-LEE, T. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems 5*, 3 (2009), 1–22.

[6] BRICKLEY, D., AND MILLER, L. FOAF Vocabulary Specification 0.98, 2010.

[7] ERLING, O., AND MIKHAILOV, I. RDF Support in the Virtuoso DBMS. *Proceedings of the 1st Conference on Social Semantic Web CSSW 221* (2007).

[8] GUY, I., ZWERDLING, N., RONEN, I., CARMEL, D., AND UZIEL, E. Social media recommendation based on people and tags. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (Geneva, Switzerland, 2010), ACM, pp. 194–201.

[9] LIU, H., AND MAES, P. Interestmap: Harvesting social network profiles for recommendations. In *Proceedings of the Beyond Personalization 2005 Workshop* (2005).

[10] MENDES, P. N., JAKOB, M., GARCÍA-SILVA, A., AND BIZER, C. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems* (New York, NY, USA, 2011), I-Semantics '11, ACM, pp. 1–8.

[11] NAJEEB, E., AND RANDI, K. Relation Based Image Retrieval In Online Social Network. In *IMCOM* (January 2014), ACM.

[12] SAN MARTÍN, M., AND GUTIERREZ, C. *Representing, Querying and Transforming Social Networks with RDF/SPARQL*, vol. 5554 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, May 2009.

[13] TIM BERNERS-LEE. The next Web of open, linked data. Recorded at TED2009, February 2009 in Long Beach, California., February 2009.

[14] VOSECKY, J., HONG, D., AND SHEN, V. Y. User identification across multiple social networks. In *2009 First International Conference on Networked Digital Technologies* (July 2009), IEEE, pp. 360–365.

[15] WAKAMIYA, S., LEE, R., AND SUMIYA, K. Towards better TV viewing rates. In *Proceedings of the 5th International Confernece on Ubiquitous Information Management and Communication - ICUIMC '11* (Feb. 2011), ACM Press.

Paper V

Metadata-based automatic image tagging

Randi Karlsen*, Martin Hætta Evertsen and Najeeb Elahi

Department of Computer Science,
UiT the Arctic University of Norway,
9037 Tromsø, Norway
Email: randi@cs.uit.no
Email: mhe@iktsenteret.no
Email: najeeb.elahi@uit.no
*Corresponding author

Abstract: Today, keyword-based search is the most common technique for searching images, and requires the availability of tags and annotations that reflect image content. While there are many techniques for automatic annotation based on image content analysis, less work has utilised image metadata as a basis for image tagging. We investigate the ability to automatically tag images based solely on image metadata, and present a novel approach to image tagging using a combination of the metadata *geo-location*, *date/time* and *category keyword*. We aim to collect relevant tags for a target image by selecting tags from related images on a community image collection, such as Flickr. Contributions of our approach are the use of dynamic techniques that adapt tag selection to the availability of relevant images and tags in the community collection. Through testing we demonstrate the usefulness of the techniques and the use of metadata information in the image tagging process.

Keywords: automatic image tagging; image metadata; metadata-based tagging; image retrieval; textual search; tag selection; dynamic techniques; content relevant tags.

Biographical notes: Randi Karlsen is an Associate Professor at the Department of Computer Science, UiT the Arctic University of Norway. She has a PhD in Computer Science and has been involved in a number of research projects. Her main research interests include interoperable systems, information retrieval, multimedia, context awareness, metadata management, health informatics and transaction management. She is the author of more than 60 publications in scientific books, journals and conferences, and has been a member of programme committees of over 30 national and international scientific conferences.

Martin Hætta Evertsen is currently a System Developer at the Norwegian Centre for ICT in Education. He received his MSc from the Department of Computer Science, Faculty of Science and Technology, UiT the Arctic University of Norway, in 2010. His research interests include automatic image tagging, image retrieval, metadata-based tagging and context awareness.

Najeeb Elahi is a PhD fellow at the Department of Computer Science at UiT the Arctic University of Norway. He obtained his Master in Science in Software Engineering of Distributed Systems from Royal Institute of Technology (KTH), Stockholm, Sweden. He has worked at the Unik Research Center in Oslo and Digital Enterprise Research Institute, Galway, Ireland. His research interests include multimedia, distributed systems, semantic web, ontology design and context-centric approach for image management.

1 Introduction

Digital images are captured in an ever-growing rate, and an important challenge is to manage them so that relevant images can be located in an effective manner. With the huge number of available images, supplementary information (such as tags and annotations) is important, not only for supporting image retrieval, but also for providing users with information about what an image depicts.

The two predominant approaches for image retrieval are: Content-Based Image Retrieval (CBIR) and Text-Based Image Retrieval (TBIR) (Baeza-Yates and Ribeiro-Neto, 2011; Datta et al., 2008). CBIR approaches focus on identifying and comparing low-level visual image features (such as colour, texture and shape) for the purpose of organising and locating images. An example image or some image features are used as query to CBIR systems, which return a set of visually similar images. TBIR approaches

allow textual search and locate relevant images by matching keywords in the query against textual image metadata, such as tags and annotations.

Today, keyword-based search is the most common technique for searching images (Sergieh et al., 2012; Pagare and Shinde, 2012. TBIR allows users to formulate high-level semantic queries, and is often more accurate and efficient in identifying relevant images compared to CBIR (Datta et al., 2008; Wu et al., 2012; Zhang et al., 2012). However, to be effective, a necessary basis for TBIR is the availability of image annotations that reflect image content. Manual annotations are usually subjective and the annotation task is time consuming so that users are not likely to devote enough time to tag all their images (Whittaker et al., 2010; Zhang et al., 2012). Another aspect is that people do not necessarily know or remember the names of all depicted objects (for instance attractions visited during a holiday).

Many tools for automatic and semi-automatic image annotation use CBIR techniques for linking visual features to keywords (Wang et al., 2012; Zhang et al., 2012). Despite the achievements in CBIR, bridging the semantic gap between low-level visual features and high-level semantic concepts is still a challenging task (Datta et al., 2008; Smeulders et al., 2000). In addition, CBIR techniques often suffer from low efficiency and scalability caused by the high dimensionality of visual features (Zhang et al., 2012; Wang et al., 2006a). To leverage these problems, some resent approaches for automatic image annotation combine content analysis techniques with the use of image metadata, where for instance GPS position of an image is used to focus tag collection to a relevant area (Jones et al., 2010; Joshi et al., 2011; Sergieh et al., 2012; Moxley et al., 2008; Popescu and Moëllic, 2009; Silva and Martins, 2011).

Digital photos can have several types of camera-generated metadata, such as the time/date of image capture, camera settings and exposure information, and for a growing number of cameras, the GPS location of image capture. In addition, image collections, such as Flickr,[1] encourage users to annotate images with semantic tags to describe image content. These data offer a potentially rich source for automatic annotation of images.

In our work, we have investigated the ability to automatically tag images based on image metadata alone. Our approach uses context metadata to first select related, previously tagged images from a source image collection, and subsequently collect the most frequently used tags from those images. We apply some simple, yet efficient, techniques that are based on text matching and document retrieval techniques. Important in our approach are some dynamic techniques that adapt image and tag collection to the availability of relevant images and tags in the source image collection.

We have tested our approach in two implementations, named as LoCaTagr[2] and LoTagr,[3] which use different sets of image metadata as input to the tagging process. LoCaTagr image tagging based on the metadata set {*GPS coordinates*, *date/time*, *category keyword*}, while LoTagr automatically tags image based on *GPS coordinates* only. This paper describes testing and evaluation of the two systems, and compare performance to other relevant systems.

2 Related work

Automatic and semi-automatic image annotations are the focus in a number of publications, which describe different approaches to collecting tags from previously tagged images. Much work is done on tagging based on content analysis of images, where machine learning techniques are used to develop image annotation systems that map low-level visual features to high-level concepts (Makadia et al., 2010; Tsai and Hung, 2008; Wang et al., 2012; Zhang et al., 2012).

A number of systems annotate query images by selecting tags from related images gathered from online image collections based on a combination of geographic position and visual similarity (Jones et al., 2010; Joshi et al., 2011; Sergieh et al., 2012; Moxley et al., 2008; Popescu and Moëllic, 2009; Silva and Martins, 2011). The general technique is first to collect a set of images within a certain radius of the query image, narrow down the set by using visual similarity techniques, and finally collect tags from the remaining images.

The work of Jones et al. (2010) and Popescu and Moëllic (2009) describe annotation of landmark images, where identification of landmark and subsequent annotation rely on a reference image corpus collected from Flickr ((Popescu and Moëllic, 2009)) and from Flickr and Panoramio (Jones et al., 2010). In the work of (Joshi et al. (2011) location-driven tags are suggested for geo-tagged images by collecting tag clouds from three different sources; a point-of-interest database, Flickr photos and personal photos. Tags are ranked according to distance and visual similarity to the query image. SpiritTagger[4] (Moxley et al., 2008) is a tag suggestion tool that combines geographical context with content-based image analysis to collect geographically relevant tags from images in Flickr. To maximise the number of images that are visually similar to the query image (Sergieh et al., 2012) suggests an iterative visual matching approach, while in work of Silva and Martins (2011) the most relevant tags are suggested by using supervised learning to rank or unsupervised rank aggregation methods that combine different estimators of tag relevance.

Approaches for annotating images based on a keyword and the image itself have been presented (Gao et al., 2010; Wang et al., 2006b). The methods require at least one accurate keyword, and combine keyword search with image content analysis to retrieve similar images from which tags can be selected.

Expansion of user provided image tags has been described (Kucuktunc et al., 2008; Rae et al., 2010; Sigurbjörnsson and Zwol, 2008; Wu et al., 2012). In the work of Kucuktunc et al. (2008) recommended tags are selected based on visual similarity between images, while in the work of Rae et al., (2010) and Sigurbjörnsson and Zwol (2008) tag selection is based on cooccurrence analysis of tags. Personalised tag recommendations are provided by Sigurbjörnsson and Zwol (2008) using tags from personal images and images from social contacts and fellow group members. Tag completion is addressed by Wu et al. (2012), where the goal is to automatically fill in missing tags for given images, based on observed tags and visual similarity.

Automatic image tagging in mobile phone applications is described by Ahern et al. (2006), Naaman and Nair (2008) and Qin et al. (2011). ZoneTag (Ahern et al., 2006; Naaman and Nair, 2008) suggests tags to photos, based on information such as location, previously used tags, tags from social contacts and temporal information, while TagSense (Qin et al., 2011) focuses on identifying people in an image by using sensors that detect for instance movement and direction.

Related to our approach, authors such as Liu et al., (2012), work on automatic landmark mining where landmarks are automatically recognised by mining image metadata, such as tags and geo-location, from large-scale social image collections. The user first submits a location name to the system, which subsequently crawls Flickr and selects photos and related metadata. The selected images are clustered according to visual features and different metadata types (such as time stamp, tags and location), and candidate landmarks mined from the clusters.

Much of the referred works that provide image tagging are based on content analysis of images (Jones et al., 2010; Joshi et al., 2011; Sergieh et al., 2012; Moxley et al., 2008; Popescu and Moëllic, 2009; Silva and Martins, 2011), in some cases in combination with geographic position or user provided keyword(s) (Jones et al., 2010; Joshi et al., 2011; Sergieh et al., 2012; Moxley et al., 2008; Popescu and Moëllic, 2009; Silva and Martins, 2011; Gao et al., 2010; Wang et al., 2006b; Liu et al., 2012; Kucuktunc et al., 2008; Wu et al., 2012). Our work contrasts with the referred works as we do not include analysis of visual features in images, but rather provide automatic image tagging based solely on image metadata. We chose this approach in order to avoid the semantic gap problem, and investigate the effectiveness of using image metadata as basis for the image tagging process.

Our work differs from tag expansion approaches (Rae et al., 2010; Sigurbjörnsson and Zwol, 2008) in that we do not require a query image to have a set of associated tags. This makes cooccurrence analysis of tags not applicable on our case. The works described by Ahern et al. (2006) and Naaman and Nair (2008) focus on suggesting tags to images based on location information. Our approach differs in that we base image tagging on more metadata, and use different techniques and sources for collecting tags, not giving priority to the user's own tagging history and tags used by the user's social contacts.

Our approach is unique in that we use a combination of the metadata *category, GPS location* and *date/time* as basis for automatic image tagging, together with novel techniques for dynamically selecting source images from Flickr and for selecting tags from the selected set of source images.

3   Metadata-based automatic image tagging

3.1   Architecture

Our architecture for metadata-based image tagging is depicted in Figure 1. There a query image is submitted to the

context-based tagger, where the *Image Selector* determines search parameters, based on the image metadata, and composes a query that is issued to Flickr. The image selector issues a Flickr photo search query using image metadata, as described in Sections 3.2 and 3.3, as parameters. The image itself is not part of the query. When a list of photos is returned from Flickr, the image selector component executes an algorithm for *dynamic image selection*, which is described in Section 3.4. This algorithm may require repeated Flickr photo searches until the requirements in the algorithm are satisfied, and a set of images has been selected as the source for tagging the query image.

Figure 1   Architecture for metadata-based image tagging (see online version for colours)
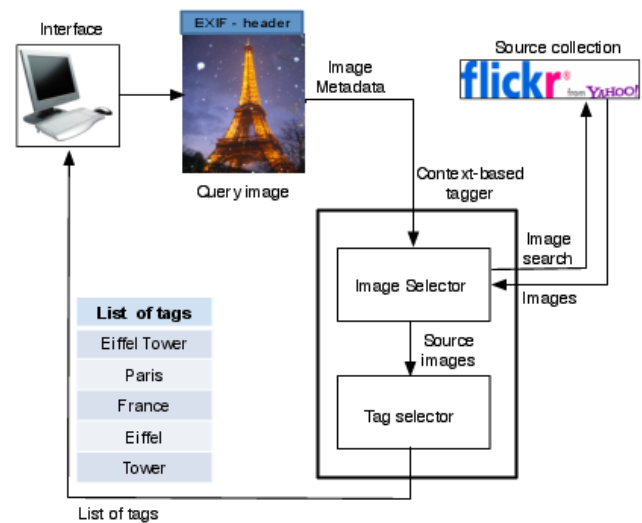
The *tag selector* processes the selected source images, collects the most frequently used tags and assigns them to the query image. The tag selector component executes an algorithm for *dynamic tag selection* that is described in Section 3.5.

Based on this architecture, we have implemented two systems: LoTagr and LoCaTagr, where the only difference is the set of metadata used in the Flickr query. LoCaTagr composes Flickr queries based on the metadata set $\{G, D, c\}$, while LoTagr uses the set $\{G\}$. Here G is a pair of numbers, $G = \{lat, lon\}$ representing latitude and longitude coordinates, D represents date/time of image capture and c is a keyword representing the image category.

Any image collection containing geo-positioned and tagged images can in principle be used as source collection. In our implementation, we chose Flickr because of its huge amount of geo-positioned and tagged images and the availability of API functions that simplify the selection of source images.

3.2   Image type and metadata use

The LoTagr and LoCaTagr systems are both designed for automatic tagging of object and event images. An *object image* typically depicts a thing, for example a building, statue or a natural thing (such as a river or mountain), while

an *event image* depicts some sort of happening, such as a football match or concert. As events happen within a limited time period, event images are in LoCaTagr handled differently from object images in that date/time values are used as parameter in Flickr image search.

Images are collected from Flickr by using the Flickr API function flickr.photos.search.[5] Input parameters, *P*, to the function depend on system and image type, as shown in Table 1.

Table 1   Parameters in Flickr search

	Object images	*Event images*
LoTagr	P={G, r}	P={G, r}
LoCaTagr	P={G, r, c, ST}	P={D1, D2, G, r, c, ST}

The search parameters in Table 1 determine *area of interest* by combining GPS coordinates, *G*, and radius, *r*. *Period of interest* is determined by *D*1 and *D*2, representing start-date/time and end-date/time, respectively. An *image category* keyword is represented by *c*, while *ST* represents a set of *synonym terms* of *c*. The synonyms are collected from STANDS_4 Web Services – Synonyms API v1[6] prior to the Flickr search.

In response to a flickr.photos.search request, Flickr returns a list of images in XML format that, for each image, includes *image id*, *owner id* and *image tags*. For LoCaTagr, Flickr image data is returned if (a) the image is positioned within the area of interest, (b) the category keyword, *c*, or terms from the set *ST* is found within the image title, description or tags, and, for event images and (c) the image is taken within the period of interest. In LoTagr, there is no distinction between object and event images. Every Flickr search uses area of interest as parameter, and only condition (a) needs to be satisfied.

3.3   Image category

Category can be defined as 'any of several fundamental and distinct classes to which entities or concepts belong'.[7] An *image category* keyword is in LoCaTagr, a term that represents a general description of the image content, such as 'tower', 'statue' or 'concert', and is used for narrowing down the search for relevant Flickr images. Among the systems described in this paper, LoCaTagr is the only system using category as image metadata.

A category, *c*, always belongs to a *type*, reflecting the type of image. Currently, the set of image types $T = \{Object, ShortEvent, LongEvent\}$ is predefined in LoCaTagr. Categories of type *Object* can for instance be 'tower', 'church' or 'statue'. *ShortEvent* describes an event that lasts for a short period (up to a day), such as a 'concert' or 'football match', while a *LongEvent* can last several days, for example a 'tournament' or 'festival'.

A category is associated with exactly one of the image types in *T*. The association between category and image type is necessary in order to avoid ambiguity regarding the meaning of a category. The category 'football' can for instance in one case refer to football as an object, and in a different case to the event of a football match. For this case, two categories may be defined by the user, 'football' as an *Object* and 'football' as a *ShortEvent*.

LoCaTagr allows the user to choose an image category, for an image or a set of images, from an extensible image category list, *L*. To add a new category, *c*, to *L*, the user enters the name of the category and selects an image type *t*, such that $t \in T$. The type *t* of *c* is used to automatically determine parameters for the LoCaTagr Flickr search. For example, when the category keyword 'football' of type *ShortEvent* is assigned to an image, the LoCaTagr system identifies the image as an event, and collects source images from Flickr using a full set of parameters, i.e. $P = \{D1, D2, G, r, c, ST\}$.

The reason for using a generic category keyword is that the user can select the keyword without having detailed knowledge of image content. This opposed to specific tags that name the object or event, and where detailed knowledge is required.

3.4   Dynamic image selection

The image selector processes the images received from Flickr and selects a *Source set* of images from which tags are collected. Since images with very few tags provide little valuable information, images are selected so that the *Source set* does not include images with less than two tags.

FlickrResults in formula (1) represents the set of images returned from Flickr using parameters *P* as input to flickr.photo.search. The content of *Source set*, *S*, is described in formula (2), where we assume a function *Tag*(*I*) that returns all tags associated with image *I*.

$$FlickrResults = flickr.photo.search(P) \qquad (1)$$

$$S = \left\{ I \in FlickrResults \,\middle|\, \left(|Tag(I)| \geq 2 \right) \right\} \qquad (2)$$

We use a dynamic image selection method that selects many Flickr images if there are many relevant images available close to the query image and fewer images if there are few relevant images available.

The technique initially searches Flickr with a very small radius, *r*, and a fairly high amount of required images, *ReqIm*. In the current implementation, the initial values are $r = 0.001$ km and $ReqIm = 50$. The set *S* is large enough if *ReqIm* or more images are selected. If a sufficient number of images is selected, tags can be collected; otherwise the radius is doubled and the number of required images is decreased with one-tenth of its last value, as shown in equation (3).

While $|S| < ReqIm$ AND $r < 32$ km
    $r = r*2$
    $ReqIm = ReqIm - (ReqIm/10)$ $\qquad (3)$
    Get new *FlickrResults* (1) and
        Compute *S* (newSourceset)(2)

The image search to Flickr continues until S has reached the required number of images. The maximum allowed radius in flickr.photo.search is 32 km, and the search process will therefore end when the radius reaches 32 km, even if no images are found.

3.5   Dynamic tag selection

The tag selector collects the most frequently used tags from images in the source set, and assigns them to the query image. Formula (4) calculates tag frequency for term t in source set S, using $|S(t)|$ and $|S|$, which represent the number of images in S, from unique owners, that use term t and the number of images in S, respectively.

$$TagFreq(t,S) = \frac{|S(t)|}{|S|} \tag{4}$$

When calculating tag frequency, each unique tag is counted only once per image owner. The rationale for this is that some people upload many images with the same set of tags, and it is not desirable to use the same tags from the same user more than once. The reason is the risk of including irrelevant tags, such as name of photographer, name of unknown people in the images, camera name, and tags that are only relevant for the specific user (e.g. honeymoon).

Formula (5) specifies the set of images from S having tag t from unique owners. Here, we assume a function $UniqueOwner(I,t)$ that determines whether owner of image I has already contributed with tag t in the source set.

$$S(t) = \{I \in S | t \in Tag(I) \wedge UniqueOwner(I,t)\} \tag{5}$$

The set of selected tags from source set S is specified in formula (6). Our approach is to discard tags with low frequencies, determined by the value of *TagThreshold*. When the number of image in S is low, i.e. less than *NumLow*, the minimum time a tag must appear is set to *TagMin* (instead of using a tag frequency value). This prevents noisy tags from appearing.

$$SelectedTags(Q) = \{t | (t \notin FilterList) \wedge$$
$$((TagFreq(t,S) \geq TagThreshold \wedge |S| \geq NumLow) \vee \tag{6}$$
$$t(|S| < NumLow \wedge |S(t)| \geq TagMin))\}$$

The value of *TagThreshold* is in our current implementations set to 0.2, meaning that a tag must used by at least 20% of the images in S in order to be selected as a tag for query image Q. The value of *NumLow* is set to 15 and the value of *TagMin* to 3. During testing we found these values to be effective for discarding noisy tags.

Formula (6) refers to the set *FilterList* which contains tags that are unwanted for the query image. Some tags are simply of no interest to any query image, for instance the tags 'geotagged', 'latitude', 'flickr' and camera name, and are thus filtered out.

Our tests show that when the number of images in the source set is very low, precision of the collected tags is also low. To tag a query image, we therefore require a minimum number of images, *ReqImMin*, in S, see formula (7). In the current testing, *ReqImMin* is set to 5.

$$\text{If } |S| < ReqImMin \text{ then } SelectedTags(Q) = \phi \tag{7}$$

3.6   Tag relevancy

When testing our approach, we manually examined the collected tags for each image and distinguished between tags that were *content relevant*, *location name*, *unsure* and *irrelevant*. Examples of the different types of tags are shown in Figure 2, where we see one of the query images and the tags provided by LoCaTagr and LoTagr. The category keyword used by LoCaTagr was 'tower'.

Figure 2   Query image (of John Hancock Centre in Chicago) and collected tags (see online version for colours)

'Content relevant' tags (displayed in green in Figure 2) are relevant with respect to what the user sees in the image, while tags classified as 'location name' (blue) give correct location names. Incorrect tags and tags that are not relevant to image content or location are classified as 'irrelevant' (orange), while tags that may or may not be relevant to image content are classified as 'unsure' (violet). Often, some will regard unsure tags as relevant whereas others will not. Typical examples are tags relevant for the location of the image but not visible on the image. Other examples are the two tags 'hancock' and 'city', as shown in Figure 2.

As exact location names can be easily obtained for geo-positioned images, through for instance the Flickr function flickr.places.findByLatLon,[8] our main interest here is to evaluate the ability to select content relevant tags. All location names are therefore disregarded in the calculation of precision, even though they are relevant to the image. To take into account, the possible relevance of unsure tags, we use two different precision scores for each image, Q, where the first ($prec1$) is stricter than the second ($prec1$), see Formulas (8) and (9). Here, $CRelev(Q)$, $LName(Q)$ and $Unsure(Q)$ represent

the sets of content relevant, location name and unsure tags, respectively, for image Q, while *SelectedTags(Q)* represents the set of all collected tags for Q.

$$prec1(Q) = \frac{|CRelev(Q)|}{|SelectedTags(Q)| - |LName(Q)|} \qquad (8)$$

$$prec2(Q) = \frac{|CRelev(Q)| + |Unsure(Q)|}{|SelectedTags(Q)| - |LName(Q)|} \qquad (9)$$

The two formulas in equation (10) calculate the arithmetic mean for each precision score over n test images.

$$Prec1 = \frac{1}{n}\sum_{i=1}^{n} prec1(Q_i)$$
$$Prec2 = \frac{1}{n}\sum_{i=1}^{n} prec2(Q_i) \qquad (10)$$

4 Testing of LoTagr and LoCaTagr

This section presents testing of LoTagr and LoCaTagr and compare performance to SimpleTagr, SpiritTagger and Google search by image. We evaluate the ability of the systems to collect relevant tags, and the accuracy of LoCaTagr and LoTagr in terms of their ability to collect a tag that name the object or event depicted in an image.

4.1 Method

We tested LoTagr and LoCaTagr using 110 queries, where each query was represented by an image with associated metadata. The set of query images, *TestIM*, contained 78 images depicting an object and 32 depicting from some events, and the content varied from very famous objects and events, such as the Eiffel Tower and Carnival in Rio, to less famous objects/events. The categories we selected for the query images were for instance 'bridge', 'church' and 'tower' for object images, and 'concert', 'festival' and 'football' for event images.

Performance of LoTagr and LoCaTagr was compared to SimpleTagr and SpiritTagger (described in Section 4.2), by using each query image in *TestIm* as query to each of the systems. For each image, we thus received four different tag sets. We manually examined all tag sets and calculated precision scores as described in Section 3.6.

All systems collect tags from relevant images available on Flickr at the time of request. To ensure that the systems were tested on an approximately identical state of Flickr, each image was tagged by all four systems within a short-time period (i.e. a couple of minutes).

To evaluate the system's ability to name the depicted object or event, we compared LoCaTagr and LoTagr to SimpleTagr and Google search by image. Also in this testing, images in *TestIm* were used as queries, and tagging was done by all four systems within a short-time period.

4.2 Comparison systems

To test the effectiveness of using dynamic image and tag selection techniques, we implemented *SimpleTagr*, which is a simplified version of LoTagr (and LoCaTagr). Like LoTagr, SimpleTagr collects a source set of images from Flickr based on GPS coordinates only. However, no dynamic techniques are used, and the system simply takes the up to 250 images that are closest to the query image and returns all tags that appear in at least 20% of the images found.

SpiritTagger (Moxley et al., 2008) is a tag suggestion tool that assembles visually relevant images from Flickr, weighted by geographic distance from the target image. With the goal to capture the spirit of a location, a set of 20 geographically representative and frequently used tags is collected and suggested to the user.

Google Search by Image[9] (GSbI) uses computer vision techniques and matching with other images on internet, to try to generate a *Best Guess* text description for a query image. The best guess is typically the name of the object depicted in the image (such as 'Eiffel Tower' or 'Big Ben') or a location name.

SimpleTagr uses, in the same manner as LoTagr and LoCaTagr, only image metadata as input to the tagging process, while SpiritTagger and GSbI use visual features and content-based image analysis as a basis for recommending tags and providing a best guess text description. SpiritTagger and GSbI were chosen as comparison systems since they both have objectives related to our work. SpiritTagger collects relevant tags from nearby images on Flickr; while GSbI seeks the internet to find a content relevant description (i.e. tag/Best Guess) for the image.

4.3 Collected tags and precision scores

In this section, the numbers of collected tags and precision scores of LoTagr and LoCaTagr are compared to SimpleTagr and SpiritTagger. Of the 110 test images, LoCaTagr was able to tag 91 images; 71 object and 20 event images. For the remaining images, the image source set, collected from Flickr, was too small to provide a useful basis for tag selection in LoCaTagr. The following test results are based on the 91 images that were tagged by all four systems.

The average number of tags collected by each of the systems is shown in Table 2. We distinguish between four different tag types: content relevant, unsure, location name and irrelevant, and present separate results for object and event images. Figure 3 shows the relative proportion of the collected tag types for each system.

From Table 2 we see that LoCaTagr collected the highest number of content relevant tags for both object and event images. For object images, LoCaTagr also had the highest total number of tags compared to LoTagr and SimpleTagr. When comparing LoCaTagr and LoTagr, LoCaTagr increased the total number of collected tags for object images from 6.4 to 11.2 (on average), an increase of 75%, while the number of content relevant tags increased with 108%. For event images, LoCaTagr increased content

relevant tags with 16% while the total number of tags was equal to LoTagr. We also notice that LoTagr had the fewest irrelevant tags for object images, while LoCaTagr had fewest irrelevant tags for event images.

Figure 3   Relative proportion of different tag types as provided by the tested systems
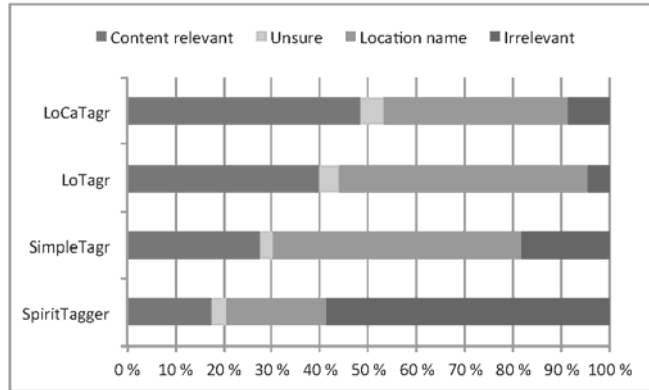
(a) Object images
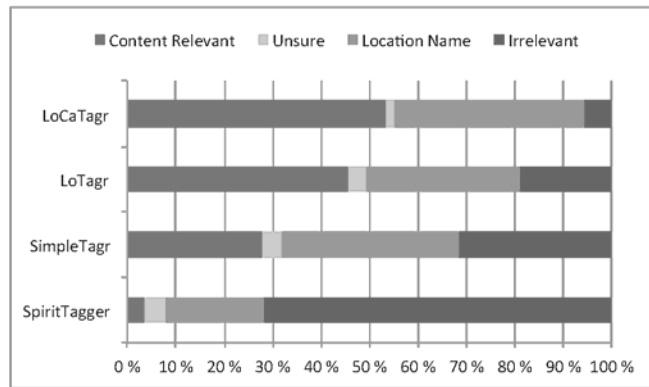
(b) Event images

Table 2   Average number of tags. Distinguishing between object and event images

	CRelev	Unsure	LName	Irrelevant	Total
Object Image					
LoCaTagr	5.4	0.5	4.3	1.0	11.2
LoTagr	2.6	0.2	3.3	0.3	6.4
SimpleTagr	1.6	0.2	3.0	1.1	5.9
SpiritTagger	3.5	0.6	4.1	11.8	20
Event images					
LoCaTagr	4.4	0.2	3.2	0.4	8.2
LoTagr	3.8	0.3	2.6	1.5	8.2
SimpleTagr	2.6	0.4	3.4	2.9	9.3
SpiritTagger	0.7	0.9	4.0	14.4	20

When comparing image types in Figure 3, the relative proportion of content relevant tags was approximately the same for object and event images. Also, LoCaTagr collected fewer irrelevant tags for event images compared to object images, while LoTagr, SimpleTagr and SpiritTagger had a

relative increase in irrelevant tags compared to the results from object images.

Table 3 shows the average precision scores with respect to content relevant tags. Table 3 displays separate scores for object images, event images and combined scores for all images in the test. From these results, we see that both LoCaTagr and LoTagr had good precision scores for content relevant tags, while the scores were noticeably lower for SimpleTagr and SpiritTagger.

Table 3   Average precision of content relevant tags

	Object images		*Event images*		*All images*	
	Prec1	*Prec2*	*Prec1*	*Prec2*	*Prec1*	*Prec2*
LoCaTagr	0.82	0.88	0.75	0.76	0.81	0.86
LoTagr	0.73	0.82	0.75	0.80	0.74	0.82
SimpleTagr	0.44	0.48	0.57	0.59	0.47	0.51
SpiritTagger	0.22	0.26	0.05	0.10	0.18	0.22

SpiritTagger had, compared with the other systems, the lowest precision scores and the highest number of irrelevant tags. One reason for this is that SpiritTagger always suggests exactly 20 tags whereas the other systems use a more dynamic approach, selecting only tags having a certain frequency. Note also that the goal of SpiritTagger is to collect tags that capture the spirit of the location. In our testing, we were only counting content relevant and location relevant (i.e. location names) tags, and were not considering tags that may be relevant to 'the spirit of the location'.

4.4   Tag accuracy

To evaluate the accuracy of the collected tags, we identified the number of test images that received a tag that named the depicted object or event. In this test, we used all 110 images in the set *TestIm* and compared LoCaTagr and LoTagr with SimpleTagr and Google Search by Image (GSbI).

When testing GSbI, we found that the strength of GSbI lies in the ability to guess image content when identical images are found on the internet. GSbI also provided good results when the query image depicted some distinct and well-known structure with a lot of similar images available on internet. For each image, GSbI provided zero or one guess. There were very few wrong guesses, which indicate that if a guess is presented, it is highly probable that it is a correct guess.

When comparing LoCaTagr, LoTagr, SimpleTagr and GSbI we measured the fraction of images that in LoCaTagr, LoTagr and SimpleTagr received a tag that named the object or event and in GSbI were given a correct guess. The results are presented in Table 4, where we firstly distinguish between object and event images, secondly between public and private images, and finally give the score over all images. A *public* image is here an image where an identical copy is found on the internet, while for a *private* image, an identical image is not publicly available. Among the 110 test images, we had 38 public images and 72 private images.

Table 4 Fraction of images with a tag naming the object or event

	Object-Event view		Public-Private view		All images
	Object	*Event*	*Public*	*Private*	
LoCaTagr	79%	47%	76%	67%	70%
LoTagr	65%	44%	74%	51%	59%
SimpleTagr	40%	34%	50%	31%	37%
GSbI	50%	9%	68%	22%	38%

From Table 4, we see that for all types of images, LoCaTagr collected the highest number of naming tags, closely followed by LoTagr. All systems followed the trend that it is easier to find content relevant information for object images compared to event images.

LoCaTagr, LoTagr and GSbI all provided good results for public images. But, for GSbI we see a clear decrease in performance for private images, LoCaTagr has only a slightly lower performance for those images. This difference is understandable knowing the different techniques used by the two systems, where LoCaTagr is not relying on finding visually similar images on the internet.

4.5 Effects of dynamic selection techniques

The only difference in the design of SimpleTagr and LoTagr is that SimpleTagr lacks the dynamic image and tag selection methods implemented in LoTagr (and LoCaTagr). From Table 3 we see that the average precision scores for all images increased from 0.47 to 0.74 for Prec1 and 0.51 to 0.82 for Prec2 when using the dynamic selection methods. Table 4 shows the fraction of images that received a tag naming the object or event increased from 37% to 59% when using the dynamic selection methods.

The improvements in average precision scores and improved amount of naming tags are all statistically significant with a confidence level of 0.95, and show that the dynamic methods for images and tag selection are effective.

4.6 Effects of category keyword

To evaluate the effect of using category as parameter for tag collecting, we compare LoCaTagr to LoTagr. For object images we see that LoCaTagr had the best precision scores for both Prec1 and Prec2 measures. In Table 4, we also see that the fraction of images that were given a tag naming the object or event increased from 59% to 70% when using category. These are not statistically significant improvements (using confidence level 0.95), but may indicate that using category information when selecting source images might be useful.

The precision scores for event images (see Table 3) do not show a similar trend. Here we see that LoCaTagr and LoTagr have identical Prec1 scores, and the Prec2 score for LoTagr is slightly better than the LoCaTagr score. However,

note that the set of event images is small, and we believe that more tests are needed in order to make some general conclusions for this group of images.

5 Discussion

5.1 Usefulness of category

The performance results in Tables 3 and 4 do not show statistically significant improvements that can support a conclusion stating the usefulness of a category keyword in the tagging process. The usefulness of category can, however, be seen when inspecting the test results of specific images. The first example is an image of a small attraction located close to a much more famous attraction. The image was taken at an aquarium in Paris, called Cineaqua, which is located close to the Eiffel Tower. For LoCaTagr we attached the category 'aquarium' to the image, and the system returned five content relevant tags and no irrelevant tags. Among the relevant tags were 'cineaqua', 'aquarium', 'fish' and 'blue'. LoTagr and SimpleTagr did not find any relevant tags, while SpiritTagger was only able to recognise the colour blue.

The problem for the three last systems is that there are so many images from the bigger attraction that tags from the smaller attraction are not selected. Thus, location alone is not sufficient to find relevant tags for this type of images, whereas combining location with image category information may give good results.

In the second example, we had an event image of a U2 concert taken place at Camp Nou, which is the football stadium of FC Barcelona. The image depicts thus an activity that is rather unusual at this specific location. For LoCaTagr we used the category 'concert', and the system returned nine content relevant tags and no irrelevant tags. Among the relevant tags were 'U2', 'Camp Nou', 'bono' and 'edge' (name of band members) and '360' (name of concert tour). The only content relevant tag collected by LoTagr and SimpleTagr was the name of the stadium. Otherwise, both systems collected a number of tags concerning football, including 'football', 'fc barcelona', 'barca' and 'soccer'. Also this example indicates the usefulness of the category keyword.

5.2 Availability of source images

As described in Section 3.5, LoCaTagr and LoTagr will not tag a query image if the number of images in the source set is lower than a required minimum. For LoCaTagr this happened for 7 (9%) of the object images and 12 (37%) of the event images. This illustrates in particular the specific challenge of finding a sufficient number of source images for event images.

Contrary to object images, the source set for an event image is in LoCaTagr restricted to a specific time period. Collecting a sufficient number of source images can be difficult since the availability of images from the specific area and time period may be scarce. However, as the

popularity of image geo-tagging grows, images of future events may increasingly be geo-tagged, which will improve the ability to tag event images.

For object images, the amount of potential source images can expand over time, as new images of the object are taken and added to Flickr. This may in fact contribute to improve the performance of both LoCaTagr and LoTagr over time.

The problem of not being able to tag an image because of a too small source set was only experienced for the LoCaTagr system. Since LoTagr also shows good performance, we believe that this system can be used as an alternative approach in those cases when LoCaTagr cannot provide a result.

5.3   Combining approaches

During testing we found that LoCaTagr and LoTagr are both useful, but under somewhat different circumstances. By combining LoCaTagr and LoTagr into a single system, we can provide an automatic tagging system where LoCaTagr can be the preferred approach for focused tagging when image category is available, while LoTagr is the alternative if the user is not willing to provide a category keyword. LoTagr is also the alternative if LoCaTagr cannot tag an image because of a too small source image set.

LoCaTagr and LoTagr (together with SimpleTagr) have the advantage of using GPS coordinates as input to the tagging process, and thereby the possibility to narrow down a search for source images to the specific area where the image was taken. However, as image geo-position is required input to the systems, our approach cannot be applied on images that lack GPS coordinates.

GSbI, on the other hand, is very useful when similar or identical images are found on the internet and can provide good best guess keywords for images that are not geo-positioned. A best guess represents an alternative to the category keyword and can narrow the search for source images. Combining our solutions with a best guess approach may well give improved image tagging in that GPS coordinates need not be an absolute requirement for the query image, and that a specific best guess keyword may give a very relevant set of source images, that again may improve the quality of the collected tags.

6   Evaluation

The goal of this work has been to investigate the ability to automatically tag images based only on image metadata. We implemented and tested two systems: LoCaTagr and LoTagr, which both base image tagging on image metadata, and additionally use dynamic techniques for (a) selecting a set of source images from Flickr and (b) retrieving relevant tags from the source images.

Based on the testing described in this paper, we conclude that both LoCaTagr and LoTagr are beneficial for automatic image tagging. The systems are capable of automatically collecting a number of relevant image tags, and at the same time avoiding too many irrelevant tags. The proportion of irrelevant tags is only 8.2% for LoCaTagr and 8.5% for LoTagr. We have also shown that the systems are capable of automatically providing tags that name the object or event depicted in an image. For LoCaTagr this applies to 70% of the image and for LoTagr to 59%. A tag that names the object/event of an image is specifically useful in that it can be used as basis for accurately obtaining more information of what the image depicts.

The availably of *GPS position* is crucial for tagging the image using our approach. GPS coordinates must thus be added (either automatically by the camera or manually by the user) to the image, before it can be tagged using either LoCaTagr or LoTagr. This may currently limit the usefulness of the systems. However, as the number of devices that automatically add GPS position to images increases, this requirement may in the future be easy to fulfil.

In addition to showing the usefulness of basing automatic image tagging on image metadata, an important contribution of our work is also the development of the dynamic techniques for image and tag selection. Testing shows that these techniques are effective for collecting a high amount of relevant tag.

Our test results indicate that additionally using a *category keyword* and *date/time* in the image tagging process, as is done in LoCaTagr, may improve performance. The test results for LoCaTagr are on average better then LoTagr, with a higher number of collected tags, lower percentage of irrelevant tags and a higher proportion of image where a tag naming the object or event is given. This improved performance comes, however, with the cost that the user must manually provide a category keyword before using LoCaTagr to automatically collect image tags.

With a reasonable amount of images, we believe that manually adding a category keyword is manageable for the user. Both in terms of time spent (the system is designed so that attaching a category to an image is easily done) and since the user only needs to describe what he/she sees in the image (e.g. a 'bridge', 'tower' or 'church'), and does not need to know the name of the object or event. However, with a large number of images, the user may not be willing to devote enough time to provide a category for each image. In this case, we find the LoTagr system to be a good alternative to LoCaTagr, since LoTagr does not require any user provided metadata, and still performs well with respect to amount and relevancy of the automatically provided tags.

7   Conclusion

We have described a novel approach for automatic image tagging based on image metadata, where frequently used tags are collected from related images available on Flickr. The approach base image tagging on the metadata *GPS coordinates*, *date/time* and *category* information, and use dynamic techniques that adapt image and tag selection to the availability of relevant images and tags on Flickr.

We implemented two prototype systems, LoTagr and LoCaTagr, that were tested and compared with three other systems: SimpleTagr, SpiritTagger and Google search by image. From the testing, we conclude that applying image metadata as the basis for automatic image tagging is useful, and that the dynamic techniques for selecting source images and tags are effective.

In this work we investigated automatic image tagging based solely on image metadata. This can be considered both an alternative and a complementary approach to the multitude of techniques doing image tagging based on image content analysis. Our systems show good performance, but we also see possibilities for improvements by combining the metadata approach with other tagging approaches. This will be investigated as part of our future work. In future work, we also want to improve our image tagging approach by investigating using a broader range of metadata in the tagging process, and we seek to design systems that are flexible with respect to the required set of input metadata.

References

Ahern, S., Davis, M., Eckles, D., King, S., Naaman, M., Nair, R., Spasojevic, M. and Yang, J. (2006) 'ZoneTag: designing context-aware mobile media capture to increase participation', *Proceeding of PICS 2006 in Conjunction with Ubicomp 2006*, Irvine, CA.

Baeza-Yates, R. and Ribeiro-Neto, B. (2011) *Modern Information Retrieval, the Concepts and Technology behind Search*, 2nd ed., Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Datta, R., Joshi, D., Li, J. and Wang, J. (2008) 'Image retrieval: Ideas, influences, and trends of the new age', *Computing Surveys (CSUR)*, Vol. 40, No. 2.

Gao, S., Wang, Z., Chia, L-T. and Tsang, I.W-H. (2010) 'Automatic image tagging via category label and web data', *Proceedings of the International Conference on Multimedia in MM'10*, ACM, New York, NY, USA, pp.1115–1118.

Jones, G.J.F., Byrne, D., Hughes, M., O'Connor, N.E. and Salway, A. (2010) 'Automated annotation of landmark images using community contributed datasets and web resources', *Proceedings of the 5th International Conference on Semantic and Digital Media Technologies in SAMT'10*, Springer-Verlag, Berlin, Heidelberg., pp.111–126.

Joshi, D., Luo, J., Yu, J., Lei, P. and Gallagher, A. (2011) 'Using geotags to derive rich tag-clouds for image annotation', *Social Media Modeling and Computing*, , Springer, London, pp.239–256.

Kucuktunc, O., Sevil, S.G. Tosun Burak A., Zitouni, H., Duygulu, P. and Can, F. (2008) 'Tag suggestr: automatic photo tag expansion using visual information for photo sharing websites', *SAMT'08: Proceedings of the 3rd International Conference on Semantic and Digital Media Technologies*, Springer-Verlag, Berlin, Heidelberg, pp.61–73.

Liu, Z., Yan, H. and Han, H. (2012) 'Mining large-scale social images with rich metadata and its application', *Journal of Software*, Vol. 7, No. 4, pp.749–756.

Makadia, A., Pavlovic, V. and Kumar, S. (2010) 'Baselines for image annotation', *International Journal of Computer Vision*, Vol. 90, No. 1, pp.88–105.

Moxley, E., Kleban, J. and Manjunath, B.S. (2008) 'Spirittagger: a geo-aware tag suggestion tool mined from flickr', *MIR'08: Proceeding of the 1st ACM International Conference on Multimedia Information Retrieval*, ACM, New York, pp.24–30.

Naaman, M. and Nair, R. (2008) 'ZoneTag's collaborative tag suggestions: what is this person doing in my phone?', *IEEE Multimedia*, Vol. 15, pp.34–40.

Pagare, R. and Shinde, A. (2012) 'A study on image annotation techniques', *International Journal of Computer Applications*, Vol. 37, No. 6, pp.42–45.

Popescu, A. and Moëllic, P-A. (2009) 'MonuAnno: automatic annotation of georeferenced landmarks images', *CIVR'09: Proceeding of the ACM International Conference on Image and Video Retrieval*, ACM, New York, USA.

Qin, C., Bao, X., Choudhury, R.R. and Nelakuditi, S. (2011) 'TagSense: a smartphone-based approach to automatic image tagging', *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services in MobiSys'11*, ACM, New York, NY, USA, pp.1–14.

Rae, A., Sigurbjörnsson, B. and van Zwol, R. (2010) 'Improving tag recommendation using social networks', *Adaptivity, Personalization and Fusion of Heterogeneous Information in RIAO'10*, Le centre de hautes etudes internationales d'informatique documentaire, Paris, France, pp.92–99.

Sergieh, H.M., Gianini, G., Döller, M., Kosch, H., Egyed-Zsigmond, E. and Pinon, J-M. (2012) 'Geo-based automatic image annotation', *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval in ICMR'12*, ACM, New York, USA.

Sigurbjörnsson, B. and Zwol, R. (2008) 'Flickr tag recommendation based on collective knowledge', *WWW '08: Proceeding of the 17th International Conference on World Wide Web*, pp.327–336.

Silva, Ana and Martins, B. (2011) 'Tag recommendation for georeferenced photos', *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks in LBSN '11*, ACM, New York, NY, USA, pp.57–64.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A. and Jain, R. (2000) 'Content-based image retrieval at the end of the early years', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, pp.1349–1380.

Tsai, C-F. and Hung, C. (2008) 'Automatically annotating images with keywords: a review of image annotation systems', *Recent Patents on Computer Science*, Bentham Science Publishers, pp.55–68.

Wang, C., Jing, F. Zhang, L. and Zhang, H-J. (2006a) 'Scalable search-based image annotation of personal images', *Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval in MIR '06*, ACM, New York, NY, USA, pp.269–278.

Wang, M., Ni, B., Hua, X-S. and Chua, T-S. (2012) 'Assistive tagging: a survey of multimedia tagging with human-computer joint exploration', *ACM Computing Surveys*, Vol. 44, No. 4.

Wang, X.J., Zhang, L., Jing, F. and Ma, W.Y. (2006b) 'Annosearch: image auto-annotation by search', *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp.1483–1490.

Whittaker, S., Bergman, O. and Clough, P. (2010) 'Easy on that trigger dad: a study of long term family photo retrieval', *Personal Ubiquitous Computing*, Vol. 14, No. 1, pp.31–43.

Wu, L., Jin, R. and Jain, A. (2012) 'Tag completion for image retrieval', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.716–727.

Zhang, D., Islam, M.M. and Lu, G. (2012) 'A review on automatic image annotation techniques', *Pattern Recognition*, Vol. 45, No. 1, pp.346–362.

Notes

1   http://www.flickr.com/

2   location, category and time-based automatic image tagging system using Flickr

3   location-based automatic image tagging system using Flickr

4   http://vision.ece.ucsb.edu/multimedia/spirittagger.shtml

5   http://www.flickr.com/services/api/flickr.photos.search.html

6   http://www.abbreviations.com/synonyms_api.asp

7   http://www.merriam-webster.com/

8   http://www.flickr.com/services/api/flickr.places.findByLatLon.html

9   http://www.google.com/insidesearch/searchbyimage.html

Paper VI

A New User Similarity Computation Method for Collaborative Filtering Using Artificial Neural Network

Noman Bin Mannan[1], Sheikh Muhammad Sarwar[1], and Najeeb Elahi[2]

[1] Institute of Information Technology, University of Dhaka,
Dhaka, Bangladesh
nomanbinmannan@gmail.com, smsarwar@du.ac.bd
[2] UiT The Arctic University of Norway
najeeb.elahi@uit.no

Abstract. A User-User Collaborative Filtering (CF) algorithm predicts the rating of a particular item for a given user based on the judgment of other users, who are similar to the given user. Hence, measuring similarity between two users turns out to be a crucial and challenging task as the similarity function is the core component of the item rating prediction function for a particular user. In this paper, we investigate the effectiveness of a multilayer feed-forward artificial neural network as a similarity measurement function. We model similarity between two users as a function that consists of a set of adaptive weights and attempt to train a neural network to optimize the weights. Specifically, our contribution lies in designing an error function for the neural network, which optimizes the network and sets weights in such a way that enables the neural network to produce a reasonable similarity value between two users as its output. Through experimentation on Movielens dataset, we conclude that neural network, as a similarity function, gains more accuracy and coverage compared to the Genetic Algorithm (GA) based similarity architecture proposed by Bobadilla et al.

Keywords: Collaborative filtering, Recommender System, Similarity measures, Artificial Neural Network.

1   Introduction

Recommender system (RS) makes custom-made recommendations to its users for products, services or information by applying various knowledge discovery techniques. The general objective of a RS is to predict rating of items of which the user has no knowledge. Different filtering algorithms are used in RS and a filtering algorithm is the core component of a recommender system. Most common filtering algorithms are demographic filtering [1] and content based filtering [2]. Demographic filtering is established using an intuition that users with common personal attributes like sex, age, region etc. will also have common personal preferences. Content based filtering recommends items to the user according to

the content of the previously preferred items. This algorithm analyzes user's past behavior and recommends items according to user's preference history.

In recent years, collaborative filtering (CF) has been the mostly used filtering algorithm [8] [9] in RS. This filtering algorithm is based on the assumption that rating prediction for an unknown item for a given user should be influenced by a neighborhood of users with which the given user is similar. The neighborhood of similar users usually rate an important number of items in a similar way as the given user. For example "A Beautiful Mind" movie could be highly rated for an individual based on the positive ratings of a group of similar people about this movie who also rated this movie very highly. This recommendation will often provide the user of the service with inspiring positive information from the collective knowledge of all other users of the service.

The importance of recommender systems is increasing day by day. In recent years, RS have played an important role in reducing the unnecessary information overload on those websites, where users have the option of voting for their preferences on a series of products or services. Most well-known example of RS will be movie recommendation websites (i.e. IMDb) not only in aspect to users but also for researchers [3]. There are also many other application fields of RS such as e-commerce [4], e-learning [5], digital libraries [6] and so on. RS have great rule in future and it's importance is increasing day by day.

Artificial neural network (ANN) is a machine learning tool, which can be used for generating a series of nonlinear result values for real-valued and vector-valued functions over continuous and discrete-valued attributes. ANNs are also strong to noise in the training data [7]. The contribution of this paper is the introduction and applicability of ANN as a new similarity function for collaborative filtering. We train the neural network to enable it to produce a similarity value between two users. Through experimentation we show that ANN performs well than Genetic Algorithm [13] in terms of Mean Absolute Error (MAE) and Coverage.

2   Background and Related Work

Similarity computation between users is the main task in collaborative filtering algorithms. For a User-User CF algorithm, similarity, $sim_{x,y}$, between the users x and y who have both rated the same items is calculated first. To calculate this similarity different metrics [10] [11] [12] are used.

Generally, for computing correlation-based similarity, similarity $sim_{x,y}$ between two users x and y is calculated using Pearson correlation or other correlation-based similarities. Other correlation-based similarities are constrained Pearson correlation, Spearman rank correlation and Kendalls τ correlation. In constrained Pearson correlation, mid point is used in place of mean rate which is the main difference with Pearson correlation. Spearman rank correlation is similar to Pearson correlation, except that the ratings are ranks. Kendalls τ correlation is similar to the Spearman rank correlation, but instead of using ranks themselves, only the relative ranks are used to calculate the correlation [17] [18]. Vector-cosine based similarity metric use user as a vector of ratings and

measure the rating vectors cosign angle [15]. There exists other useful similarity measures based on conditional probability [19] [20]. The goal of all the similarity measures is to produce appropriate similarity values between two users depending on their item rating vector.

2.1 Predicted Rating Computation and Mean Absolute Error (MAE)

To calculate the predicted rating p_x^i for user x of an item i, the following Deviation From Mean (DFM) as aggregation approach is used:

$$p_x^i = \bar{r_x} + \frac{\sum_{n \epsilon k_x}[sim_w(x, n) \times (r_n^i - \bar{r_n})]}{\sum_{n \epsilon k_x} sim_w(x, n)} \tag{1}$$

Where $\bar{r_x}$ is the average of ratings made by the given user x and $\bar{r_n}$, r_n^i is the average of ratings and rating of the neighbor for that item respectively made by the neighbor n. After calculating every possible prediction according to the similarity function sim_w, the mean absolute error (MAE) of the RS is measured as following:

$$MAE = \frac{1}{U} \sum_{u \epsilon U} \frac{\sum_{i \epsilon I_u} |p_u^i - r_u^i|}{l_u} \tag{2}$$

When running the similarity function, U and I_u represent respectively the number of training users and the number of training items rated by the user u.

2.2 Similarity Method Using Genetic Algorithm(GA Method)

The main goal of a CF based RS is to obtain better rating prediction for an unknown item by applying a similarity function that improves the accuracy [8] [11] of prediction of CF based RS. For this purpose, Bobadilla et al. proposed a genetic algorithm based similarity method.

First they generate some vector values of a user subject to another user for obtaining similarity between each pair of users. Then vector values are passed to a similarity function which is associated with some weight vectors. Weight vectors for optimal similarity function are obtained by genetic algorithm.

Genetic algorithm (GA) based similarity metric [13] is described below:

Vector Values. The pre-processing stage of the GA based method involves the computation of a vector between two users. Later, the vector is used to asses the similarity between the users. In order to understand the vector computation let us consider a RS with a set of U users, $(1, ..., U)$, and a set of I items $(1, ..., I)$. Users rate those items with a discrete range of possible values $(m, ..., M)$, where value m represents a scenario where the user is completely unsatisfied and value M indicates a situation where the user is completely satisfied.

The ratings made by a particular user x can be represented by a list, $r_x = r_x^{(1)}, r_x^{(2)}, ..., r_x^{(l)}$, where I is the number of items in the RS and r_x^i represents the rating that the user x has made over the item i. If an item is not rated by the user, mark $\bullet$ is used and therefore the expression $r_x^i = \bullet$ states that the user x has not rated the item i yet.

To compare both user x and y, their rating lists, r_x, r_y are compiled to another vector:

$$v_{x,y} = (v_{x,y}^{(0)}, ..., v_{x,y}^{(M-m)}) \tag{3}$$

whose dimension is the number of the possible ratings that a user can make over an item. Each component $v_{x,y}^{(i)}$ of the vector $v_{x,y}$, represents the ratio of items, j, rated by both users (that is to say, $r_x^{(j)} \neq \bullet$ and $r_y^{(j)} \neq \bullet$) and over which the absolute difference between the ratings of both users is $i(|r_x^i - r_y^i| = i)$, to the number of items rated by both users. That is to say, $v_{x,y}^{(i)} = a/b$ where b is the number of items rated by both users, and a is the number of items rated by both users over which the absolute difference in the ratings of both users is i.

In such a way the vector $v_{x,y}$ is produced from the rating lists of user x and user y.

Similarity Function for GA Based Architecture. The resultant vector produced from two users is used to compute similarity between two users. For similarity calculation between two users using the vector, GA method considers the following equation:

$$sim_w(x,y) = \frac{1}{M - m + 1} \sum_{i=0}^{M-m} w^{(i)} v_{x,y}^{(i)} \tag{4}$$

Here, GA method introduces a weight vector $w = w^{(0)}, ..., w^{(M-m)}$, whose components lie in the range $[-1, 1]$ (that is to say, $w^{(i)} \in [-1, 1]$). The rationale for assigning different weights to different components is to indicate relative importance of different components. As the first scalar component of the vector denotes the number of movies on which two users completely agree with each other, it would possibly have higher value that other scalar components.

Genetic Algorithm (GA) Method. In order to find an optimal similarity function, sim_w, genetic algorithm has been used to search for an optimal weight vector w, which is associated with the optimal similarity function sim_w (Eq. 4). In this context, genetic algorithm performs a supervised learning task [14], whose fitness function or evaluation function is the Mean Absolute Error (MAE) of the RS. The population of the genetic algorithm is the set of different vectors of weights, w. This method stops to generate population when the output of the population evaluation function (MAE) of the RS is lower then a threshold value, γ.

3   New Similarity Computation Method Using ANN

In this section we present our rationale for using ANN as a similarity function. Moreover, we show that how we have modeled the similarity function using an ANN. Specifically, we design the objective function or cost function in such a way that the ANN being trained with substantial amount of instances produce satisfactory similarity values.

3.1   Rationale Behind Choosing Neural Network

In [13], genetic algorithm has been used to find optimal weight vector for the equation in 4. Genetic algorithms perform well with chromosomes represented as binary string, and even though there are methods in existing literature to represent chromosome as a floating point vector, genetic algorithms sometimes perform poorly when used for floating point weight adjustment. Moreover, performance of a genetic algorithm is mostly dependent on the set of initial population, which if not carefully chosen according to a good heuristic can lead to non-optimal solutions. In this specific problem, a five-length floating point weight vector is found with genetic algorithm, which produces minimum MAE. A short note on MAE is given in section 2.1. We argue that we can have a large set of weight vectors to train if we use neural network, and we can make the weight vector as large as we can by increasing number of hidden layers and nodes in each hidden layer. Furthermore, ANN is naturally designed to handled floating point values and hence is perfectly suitable for finding acceptable similarity value between two users, if provided with vector values 2.2 between two users.

3.2   Modeling Similarity Function as ANN

As a similarity function, we have used a multilayer feed forward neural network with one input layer, one hidden layer and one output layer. The neural network is depicted in section 3.3 (Fig. 1). Vector value between two users is calculated using equation 3, and the resulting vector is modeled as the input of the network. In order to model similarity using ANN, we have used five input nodes in input layer and three hidden nodes in hidden layer. We obtain the expected similarity (Eq.7) from the output node and execute error back propagation algorithm until the expected similarity (minimum error function is set) is obtained from the output node. After training all users for all training item we expect to find an optimal neural network which, if give vector values between two users can produce satisfactory similarity value.

3.3   Neural Network Cost Function

The design of our cost function emerges from equation 1, which is used for predicting rating for an item in collaborative filtering for a given user based his similarity with other users. If we consider that rating for a given user (u_x) is
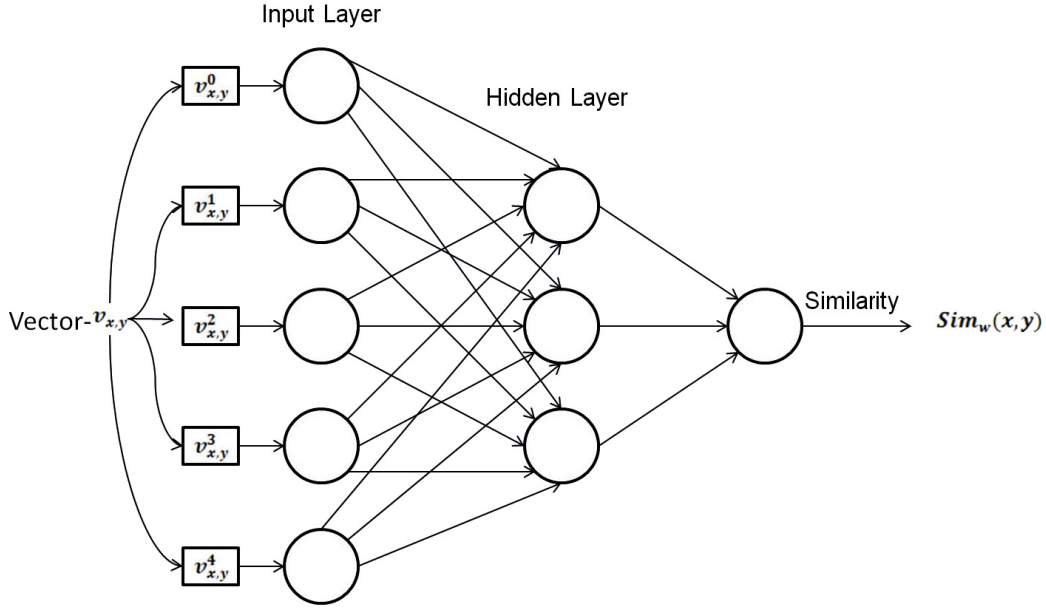
Fig. 1. Design of the similarity function as a multilayer feed-froward neural network

predicted using his similarity with another single user (u_n), the equation takes the following form:

$$p_x^i = \bar{r_x} + sim_w(x, n) \times (r_n^i - \bar{r_n}) \tag{5}$$

In the above equation, $\bar{r_x}$ is the average rating of u_x and $\bar{r_n}$ is the average rating of u_n. p_x^i is the predicted rating for u_x for item i, while r_n^i is the original rating of u_n for item i. Now, if we know the original rating r_x^i for u_x, we can write the above equation as below by slightly modifying it:

$$r_x^i = \bar{r_x} + sim_w(x, n) \times (r_n^i - \bar{r_n}) \tag{6}$$

From the above equation we can find the similarity as below:

$$sim_w(x, n) = \frac{r_x^i - \bar{r_x}}{r_n^i - \bar{r_n}} \tag{7}$$

So, if rating vectors of r_x and r_n are available we can always asses the similarity for each rating pair (r_x^i, r_n^i). The similarity $sim_w(x, n)$ would lie in the range of [-1, 1].

For example, let us consider a rating database shown in table 1 for set of 6 users 1,2,.,6 and 6 items 1,2,.,6. Items are represented in the columns and users are shown in the rows. From the table we can obtain the rating vector for each of the users. Now, if we want to asses the similarity between the first user (u_1) and the third user (u_3) for movie 6, then the result will be as following:

$$sim_w(1, 3) = \frac{4 - 3}{3 - 2.33} = 1.49 \cong 1 \tag{8}$$

Table 1. Rating Table

	1	2	3	4	5	6	Avg. Rating
1	4	5	0	3	2	4	3
2	1	0	4	0	5	2	2
3	2	0	5	0	4	3	2.33
4	4	1	1	3	4	5	3
5	2	0	3	5	0	0	1.67
6	5	1	0	4	2	0	2

According to our design, we model ground truth similarity for neural network between two users for each item. By careful examination, it can be seen that the ground truth similarity between two users will change for each item. So, for each item if they have similar preference the neural network will have to go through less correction. So if, we have the rating vector for two users we can compute the vector values as shown in section 2.2, use them as input values for neural network and train the network to produce correct similarity value for each of the commonly rated items in the rating vector.

4   Experimental Results

4.1   Procedure

Data. We have run experiments with data from the MovieLens database developed by GroupLens and Internet Movie Database (IMDb). This database contain rating values for a set of movies by a set of users. We selected the first 1000 users as collaborative users that had rated more than 40 movies. The target users were selected from the users who's id was over 1000 (so that the collaborative group and the test group of users are disjoint) and had also rated approximately 30 movies.

Table 2. Descriptive information of the database used in the experiments

Dataset	Movielens
Users	6040
Movies	3952
Ratings	1000209
Min and Max Rating	1-5

Parameter Settings for Traning and Testing. For the training we have used 3952 collaborative users from the Movielens database. We train all users (3952) for a set of movies. After that we ran several experiments working with different parameters to find the best combination. After a set of experiments we obtained better result for a neural network with three hidden layers. The maximum number of epochs to train the network was five hundred and cross validation was used to test the neural networks performance.

Prediction and Recommendation Result. In this section, we show the results obtained using the dataset specified in Table. 2. The results of our proposed ANN method are compared with the ones obtained using GA method and traditional metrics on RS collaborative filtering: Pearson correlation, Cosine and Mean Squared Differences. The comparative results are shown in Fig. 2, in terms of Mean Absolute Error (MAE) and in terms of coverage they are shown in Fig. 3.
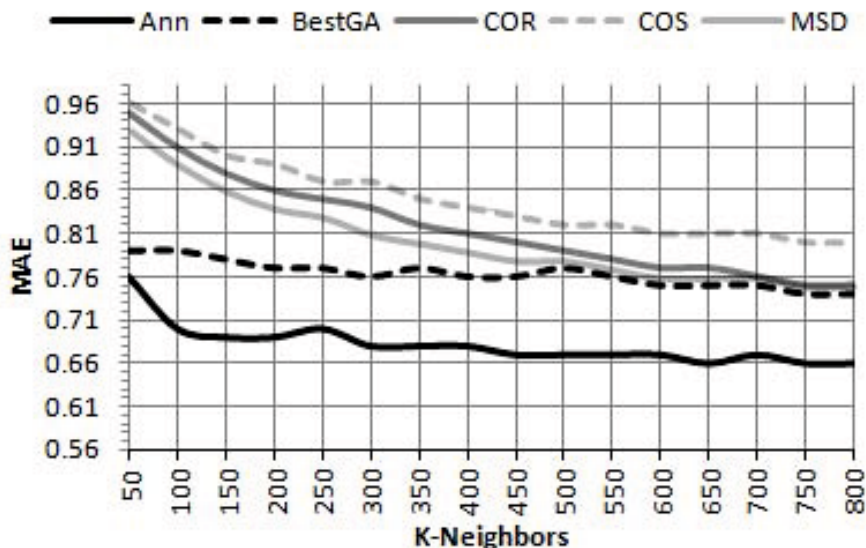

Fig. 2. Comparative resutls for GA, traditional metrics and proposed ANNs similarity method on Movielens dataset in terms of Mean Absolute Error

Fig 2 informs about the MAE obtained for Movielens when applying Pearson correlation (COR), cosine (COS), Mean Squared Differences (MSD), GA-method and the proposed ANNs method. The ANNs leads to fewer errors, particularly in the most used values of K. The black dashed and continuous lines represent respectively the best GA method and ANNs result.

Fig 3 informs about the coverage obtained. As may be seen, ANN method can improve the coverage for any value of K (the number of neighbors for each user) in relation to GA method and other traditional metric used.

The constant K is related to the number of neighbors for each given user and it varies between 50 and 800. These values enable us to view the trends of the graphics obtained from our ANN method compared to GA method and other traditional metric.

Graphic 2 shows that the best results in MAE with the ANN method are obtained when using a medium value in K, Graphic 3 shows that the best results with the ANN method are obtained in coverage using medium values in K. In this way, we should use intermediate values in K ($K \in \{300, ..., 400\}$) for obtaining the most satisfactory results both in MAE and in coverage. As our method provides high values in the quality measures applied on the MovieLens (mostly used database in RS) database, we can conclude that the proposed metric will work on a variety of Recommender Systems.
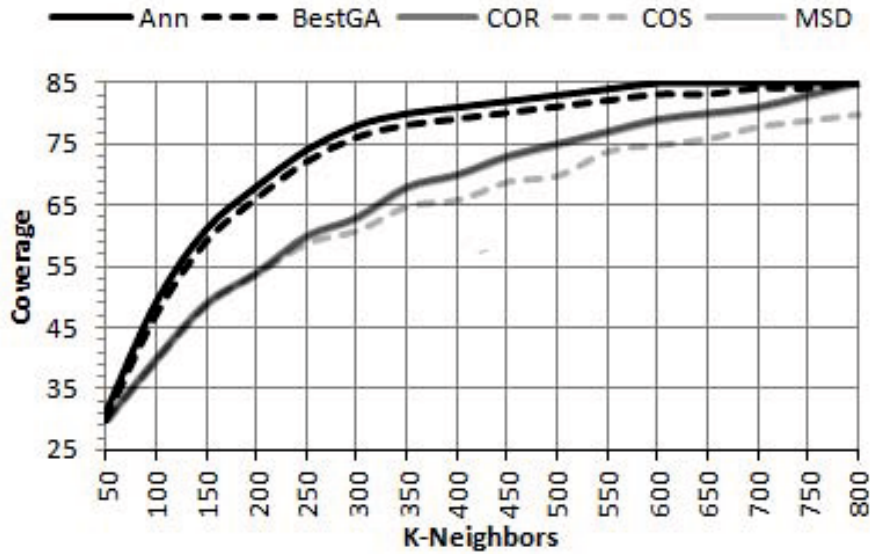
Fig. 3. Comparative resutls for GA, traditional metrics and proposed ANNs similarity method on Movielens dataset in terms of Coverage

5 Conclusion

The main contribution of this paper is the creation of a new similarity computation method using artificial neural network. When compared to GA method and other architectures, our similarity method architecture was able to reduce mean absolute error (MAE) convincingly. To be able to further evaluate our work it would be interesting to use singular value decomposition over our data. This would greatly reduce the amount of missing values in our dataset and most likely further increase the accuracy and coverage of our results. The MovieLens database have been extensively used in recommender systems research. Nevertheless, it would be useful to test the ANN based architecture with other datasets like eachmovie, film affinity, netflix and also other domains. Finally, even though applying neural network would slow down the process of recommendation generation by a few milliseconds if compared with the genetic algorithm (which gives results using a linear combination of weights), the MAE would be much less than the genetic algorithm based approach. Our experimental result clearly depicts this fact.

References

1. Krulwich, B.: Lifestyle finder: Intelligent user profiling using large-scale demographic data. AI Magazine 18(2), 37 (1997)
2. Lang, K.: Newsweeder: Learning to filter netnews. In: Proceedings of the Twelfth International Conference on Machine Learning, Citeseer (1995)
3. Miller, B.N., Konstan, J.A., Riedl, J.: Pocketlens: Toward a personal recommender system. ACM Transactions on Information Systems (TOIS) 22(3), 437–476 (2004)

4. Wei, K., Huang, J., Fu, S.: A survey of e-commerce recommender systems. In: 2007 International Conference on Service Systems and Service Management, pp. 1–5. IEEE (2007)
5. Bobadilla, J., Serradilla, F., Hernando, A.: Collaborative filtering adapted to recommender systems of e-learning. Knowledge-Based Systems 22(4), 261–265 (2009)
6. Porcel, C., Herrera-Viedma, E.: Dealing with incomplete information in a fuzzy linguistic recommender system to disseminate information in university digital libraries. Knowledge-Based Systems 23(1), 32–39 (2010)
7. Mitchell, T.M.: Machine learning, vol. 45. McGraw Hill, Burr Ridge (1997)
8. Herlocker, J.L., Konstan, J.A., Riedl, J.T., Terveen, L.G.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22(1), 5–53 (2004)
9. Manolopoulus, Y., Nanopoulus, A., Papadopoulus, A.N., Symeonidis, P.: Collaborative recommender systems: combining effectiveness and efficiency. Exp. Syst. Appl. 34(4), 2995–3013 (2008)
10. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans. Knowl. Data Eng. 17(6), 734–749 (2005)
11. Bobadilla, J., Serradilla, F., Bernal, J.: A new collaborative filtering metric that improves the behavior of recommender systems. Knowl. Based Syst. 23(6), 520–528 (2010)
12. Ingoo, H., Kyong, J.O., Tae, H.R.: The collaborative filtering recommendation based on SOM cluster-indexing CBR. Exp. Syst. Appl. 25, 413–423 (2003)
13. Bobadilla, J., Ortega, F., Hernando, A., Alcalá, J.: Improving collaborative filtering recommender system results and performance using genetic algorithms. Knowl. Based Syst. 24(8), 1310–1316 (2011)
14. Goldberg, D.E.: Genetic algorithms in search, optimization, and machine learning (1989) ISBN: 0-201-15767-5
15. Salton, G., McGill, M.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)
16. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: Grouplens: an open architecture for collaborative filtering of netnews. In: Proceedings of the ACM Conference on Computer Supported Cooperative Work, New York, NY, USA, pp. 175–186 (1994)
17. Goldberg, K., Roeder, T., Gupta, D., Perkins, C.: Eigentaste: a constant time collaborative filtering algorithm. Information Retrieval 4(2), 133–151 (2001)
18. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22(1), 5–53 (2004)
19. Karypis, G.: Evaluation of item-based top-N recommendation algorithms. In: Proceedings of the International Conference on Information and Knowledge Management (CIKM 2001), Atlanta, Ga, USA, pp. 247–254 (November 2001)
20. Deshpande, M., Karypis, G.: Item-based top-N recommendation algorithms. ACM Transactions on Information Systems 22(1), 143–177 (2004)

Paper VII

An Analysis of Personal Medical Information Disclosed in YouTube Videos Created by Patients with Multiple Sclerosis

Luis FERNANDEZ-LUQUE [a,1], Najeed ELAHI [b], Francisco J. GRAJALES III [c]

[a] *Northern Research Institute (Norut), Tromsø, Norway*
[b] *Norwegian Center of Telemedicine, University Hospital of North Norway, Tromsø, Norway*
[c] *eHealth Strategy Office, Faculty of Medicine, The University of British Columbia, Vancouver, Canada*

Abstract. The Internet has become one of the main sources of health information. Today, content generation is no longer limited to the healthcare professionals of the late nineties; Web 2.0 services and platforms have empowered patients to create and interact with various forms of Patient-Generated Content (PGC); these include: videos, blogs, and social networking pages, among others. This investigation evaluated the characteristics of PGC found within YouTube video comments. We selected a random sample of 25 out of 769 Multiple Sclerosis patient-generated videos and analyzed their corresponding 557 comments for health information. 320 comments met the inclusion criteria and 70 contained personal health information (PHI). Comments with PHI were sub-characterized for the type of medical information (i.e., diagnosis, date of diagnosis, medication, among others). In this descriptive study, we present the strata within this content and postulate some of the corresponding patient risks and ethical challenges associated with PGC found in YouTube video comments.

Keywords. Web 2.0, privacy, YouTube, health information, patient-generated content

1. Introduction

The Internet is one of the main sources of health information. Recent studies have found that most Europeans are using the Internet to search for health information [1]. Web 2.0 platforms like YouTube, Facebook, and Blogger are among the most popular. The common characteristic of all these platforms is the availability of User-Generated Content. This has lead to a new scenario where patients are not only accessing content, but also creating new content (e.g., blogs, videos); predominantly, this content is also used as a form of primary communication between the users, especially when commenting or "evaluating" the content created by other patients. For example, there are users who publish videos relating to their everyday life; including being blind, deaf or coping with chronic debilitating diseases. Many of these videos also contain comments from other users, mainly patients and relatives, which contain personal

---

[1] Corresponding Author: Northern Research Institute, Postboks 6434 Forskningsparken, 9294 Tromsø, Norway; E-mail: luis.luque@norut.no.

health information or medical advice. Although we have published research on the usage of the web services by health consumers elsewhere [2], there is paucity in the literature with regards to the characteristics of user comments posted within Patient-Generated Content (PGC). Other researchers have explored the content quality of YouTube videos and the disclosure of personal health information within social network public profiles [3–5]. The analysis of comments is important due the prevalence and incidence of Web 2.0 platforms and the PGC within; with issues related to privacy and ethics. Additionally, automatic comment analysis, through the use of crawlers and algorithms, may be used to harvest more information about the users for personalization purposes (i.e., user-targeted advertising) [6]. In our project MyHealthService [7] we are investigating the use of PGC for the enhancement and personalization of health education.

The overarching objective of our study is to increase the understanding of the Patient 2.0 phenomenon. In particular, it is focused on PGC found within YouTube video comments created by patients with Multiple Sclerosis and the disclosure of personal health information within.

2. Method

During the first week in December of 2008, we queried the YouTube database for users who had published at least three videos in English about their everyday life with Multiple Sclerosis. Twenty-seven patients were found and using the YouTube application-programming interface, user, video and comment data were extracted. A total of 769 videos, 7,047 comments and 2,426 user profiles were collected (video creators and comments). Using random selection, we selected 25 of these videos and analyzed their corresponding 557 comments. Comments which were not posted on health-related videos or were not written in English were excluded. A final sample of 320 comments was analyzed and classified into the following strata:

- Personal health information: related to personal health experiences. Subcategories: diagnosis, date of diagnosis, symptoms and medication.
- Video discussion: discussions about the videos (e.g., adding information or discussing the video topic).
- Appreciation: comments expressing gratitude for content or comments.
- Criticism: Video content criticism based on the content or quality of the information contained within the video.
- Unrelated comments: comments not covered in the other categories.

No ethical approval was necessary for the data used and gathered for this study; we used publicly accessible data that was voluntarily released by YouTube's users.

3. Results

The majority of comments met more than one category definition (Figure 1) and most of them discussed the video (77%, n=247) or expressed gratitude for the author(s) (55%, n=177). A total of 22 comments met the unrelated comment criteria (e.g., **#~KABOOOOM~#**) and eight criticized video authors (e.g., I'm skeptical you

REALLY have MS). Overall, 70 comments (22%) contained personal health information concerning their creators or a third party (e.g., relatives).
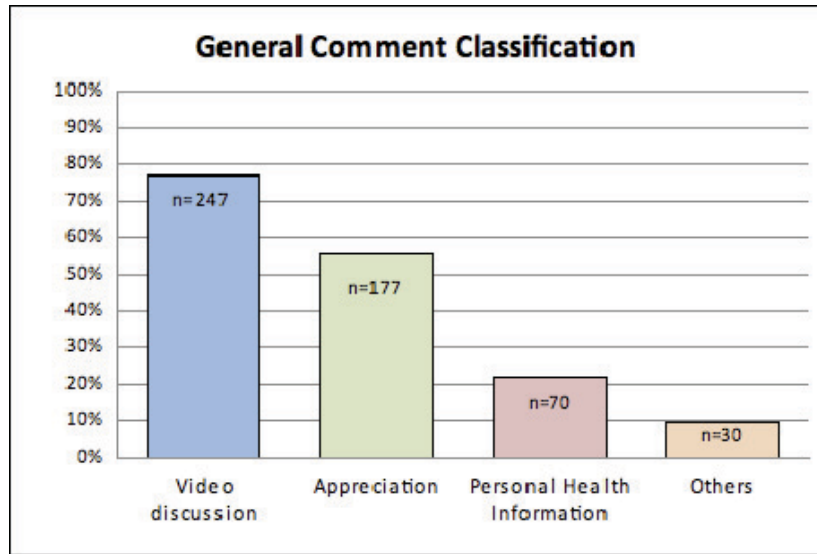
Figure 1. Total number of comments classified into the main categories

The comments with personal health information (PHI) were further stratified. As Figure 2 denotes, almost half of the comments contained information about medications (73%, n=51). Comments about symptoms (50%, n=35) and diagnoses (39%, n=27) were also prevalent. In one case, the information disclosed was the PHI of a third party:

> *I have been watching your videos since my daughter was diagnosed with MS on 28-12-07. She was diagnosed with an aggressive form of MS. Betaferon caused liver problems in a very short time. She has now had 4 infusions of tysabri and now feels she is well enough to try and go back to work. Tysabri only became available in Aust on July 1/08. This drug has given her hope that she still has a future to look forward to as she is 26 yrs old. Your improvement since your first dose gave us all hope.*
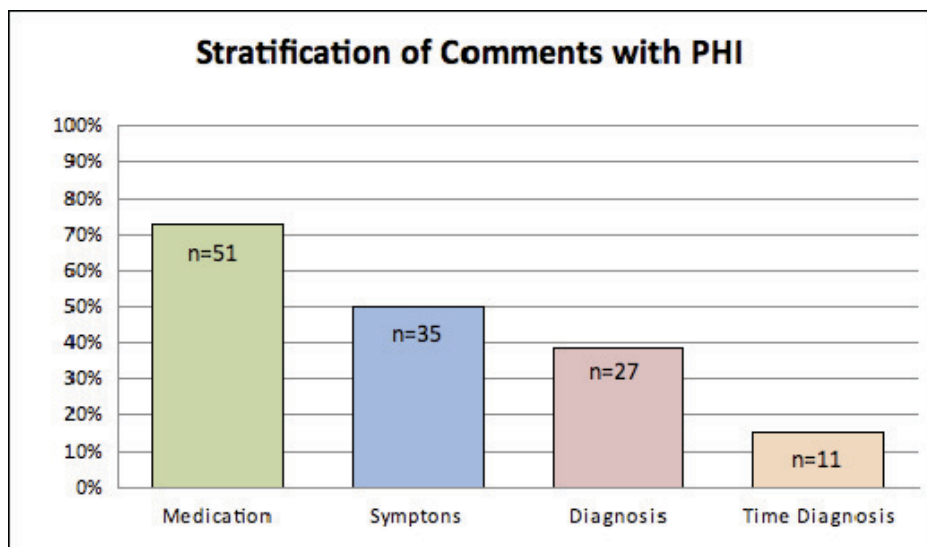
Figure 2. Total number of comments with personal health information by category

Comments also encouraged other patients to take or change their medications; for example: *"im in a wheelchair is it worthwhile taking tysabri?"* The response by the video author was: *"In a word: ABSOLUTELY! It is amazing how much freedom I recovered when I started walking!"*

4. Discussion

Our results show that although the comments with personal health information had a low prevalence, the importance of the information disclosed within is extremely complex. Comments contained detailed user information. The public availability of these comments could constitute a risk to the user's privacy even in if their comments were written under an "anonymous" alias. For example, a YouTube's nickname is often found in other Web 2.0 sites like Flickr or Blogger. This could lead a "curious" investigator to find other profiles in social networking sites, like Facebook and consequently access family member information (e.g., the Facebook profile of the daughter diagnosed with MS). This not as unlikely as it appears when, for example, 25% of the Norwegian population have a profile on Facebook [8]. We also found that some of the commenter's YouTube profiles had been deleted, but their comments remained in the video database. This would make it nearly impossible to delete this public information. The public availability of this information could even have further implications that reach beyond our current ethical paradigms. For example, if the comment's author is identified, could it be used to deny health insurance coverage? Or, could a patient legally sue another for providing medical advice through YouTube?

The majority of comments with personal health information were related to medications. Several users were seeking advice on whether to start or change medication. We were surprised by the number of comments and videos about medications, especially about Tysabri, one of the newest drugs for Multiple Sclerosis. Tysabri is prescribed when a patient is un-responsive to other treatments or when the patient is too affected by side effects. However, many video authors reported recovery since they started Tysabri treatment. The majority of comments in these videos came from patients who were considering this new treatment. For these reasons, PGC and corresponding patient risks need to be studied in further detail. Further research is necessary to understand the implications of PGC in relation to privacy, security, and ethical paradigms.

At present, we are developing an automated semantic-enhanced modeling system of educational resources (e.g., YouTube videos) as part of a project aimed at delivering personalized recommendations for health education. Semantic-based modeling and text analysis techniques could be used for analyzing the comments, e.g., to identify the topics of the videos. Berendt and Navigli [9] have proposed two methods for semantic-enhanced text analysis that allow analysts to integrate domain-specific and general background knowledge. A similar approach has also been suggested by Sheth and Nagarajan [6] where SWRL [10] rules can be defined on top of ontology and serve for classification and information decision-making.

5. Conclusion

To our knowledge, this study is the first of its kind to analyze PGC for the disclosure of personal health information. Our study, although limited to a unique dataset and condition, has shown some of the risks users may be subject to, when disclosing medical information in PGC. Although our results cannot be generalized to all conditions, they highlight the important need for further research in this subject. It is particularly important to understand patient's intents and awareness on risk when disclosing personal health information. This research may also contribute to the formation of health consumer guidelines for safety in Web 2.0 service utilization such as not reusing web aliases in platforms where personal health information is publicly disclosed.

Videos analyzed for this study represent a small sample within YouTube's database of approximately 121,200 total video views. It is essential to standardize and promote the adherence to health information quality standards (i.e., HONcode or Patient Blogger Code of Ethics) by the health consumers who are creating content on the Internet. In addition, we believe that studying patients as providers of health information could improve the access and quality of health information on the Internet.

References

[1] Andreassen, H.K., Bujnowska-Fedak, M.M., Chronaki, C.E., Dumitru, R.C., Pudule, I., Santana, S., Voss, H., Wynn, R. (2007) European citizens' use of e-health services: A study of seven countries. *BMC Public Health* 7:53.

[2] Luque, L.F., Basagoiti, I., Johnsen, E. et al. (2008) Study of the ePatient as a generator of health content in the Internet. In Eysenbach, G. (Ed.) *Medicine 2.0 Proceedings*, Toronto, Canada, 28–29.

[3] Keelan, J., Pavri-Garcia, V., Tomlinson, G., Wilson, K. (2007) YouTube as a source of information on immunization: A content analysis. *Journal of the American Medical Association* 298(21):2482–2484.

[4] Burton, A. (2008) YouTube-ing your way to neurological knowledge. *Lancet Neurology* 7(12):1086–1087.

[5] Bonander, J. (2008) Personally tailored health information: A health 2.0 approach. In Eysenbach, G. (Ed.) *Medicine 2.0 Proceedings*, Toronto, Canada, 12.

[6] Sheth, A., Nagarajan, M. (2009) Semantics-empowered social computing. *IEEE Internet Computing* 13, 1:76–80.

[7] Burkow, T.M., Vognild, L.K., Krogstad, T., Borch, N., Ostengen, G., Bratvold, A., Risberg, M.J. (2008) An easy to use and affordable home-based personal eHealth system for chronic disease management based on free open source software. In *Proceedings of MIE 2008, Studies in Health Technology and Informatics* 136:83–88.

[8] insidefacebook.com. (2008) Tracking Facebook's 2008 international growth by country July 29th, 2008. http://www.insidefacebook.com/2008/07/29/tracking-facebooks-2008-international-growth-by-country/. (Archived by WebCite® at http://www.webcitation.org/5eILehtpM).

[9] Berendt, B., Navigli, R. (2006) Finding your way through blogspace: Using semantics for cross-domain blog analysis. In Nicolov, N., Salvetti, F., Martin, J.H., Liberman, M. (Eds.) *Proceedings of the AAAI 2006 Symposium on Computational Approaches to Analysing Weblogs*, Stanford, Technical Report SS-06-03, AAAI Press, Menlo Park, 1–8.

[10] W3C Member Submission (2004) *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*, http://www.w3.org/Submission/SWRL/.

Paper VIII

Paper IX

Personalized Recommendation of Socially Relevant Images

Randi Karlsen
Department of Computer Science,
UiT The Arctic University of Norway
randi.karlsen@uit.no

Najeeb Elahi
Department of Computer Science,
UiT The Arctic University of Norway
najeeb.elahi@uit.no

Anders Andersen
Department of Computer Science,
UiT The Arctic University of Norway
anders.andersen@uit.no

ABSTRACT

We present a social image recommender system that offers a hybrid filtering approach, combining content- and knowledge-based filtering with a novel social-based filtering, that selects images of social interest to the user, by e.g. being posted by close friends or family. User activity on social media is used when generating a user profile reflecting user interests and social context, and images are recommended according to a combination of social relevance and topic of interest. Our system handles both cross-source user profiling and image recommendation across social media, currently focusing on images from Facebook and Flickr.

CCS CONCEPTS

• **Information systems → Recommender systems**;

KEYWORDS

Social Media, Image Recommendation, Social-based Filtering

1 INTRODUCTION

Social media, such as Facebook[1], Flickr[2], Instagram[3] and Snapchat[4], give people access to an enormous amount of images[5,6,7], where they can follow their friends' activities through images, or seek images based on topic of interest. A consequence of rapidly growing digital image collections, is information overload and the increased problem of finding images of personal interest to the user. This is a problem even if we restrict retrieval to only include images from friends on online social media.

---

[1] https://www.facebook.com/
[2] https://www.flickr.com/
[3] https://www.instagram.com/
[4] https://www.snapchat.com/
[5] https://expandedramblings.com/index.php/flickr-stats/
[6] https://zephoria.com/top-15-valuable-facebook-statistics/
[7] https://socialpilot.co/blog/151-amazing-social-media-statistics-know-2017/

Relevance of an image can be determined with respect to different aspects, including *user interests* and *social context*. On social media we have friends of different types, e.g. close friends, colleagues, class mates, acquaintances and even people the user has never met. An image from a close friend can be socially relevant, by depicting situations or activities attended by the user's best friends. An image from a distant friend may be relevant to the user's interests but is not necessarily socially relevant.

Our goal is to provide users with recommendations of images that are both socially relevant and relevant to user interests. A contribution of our work is a novel hybrid image recommendation approach, combining social-based filtering with content- and knowledge-based filtering. Images are selected based on the social closeness to the image owner (i.e. closeness of friends) and closeness between user interests and image content.

Recommendations are supported by a novel user profiling method, that uses social media information to describe both user interests and social context. User interests are inferred based on user activity and metadata from the user's images, while social context describes the user's relation to other members of the network through classification of friends.

Social network information has been used in recommender systems for various purposes, e.g. to improve quality of recommendations, and provide new recommender systems. Most commonly, social information has been used in collaborative filtering techniques, trust-based techniques and to tackle cold-start problems [6, 7, 11, 15]. The amount of information participants voluntarily reveal, is also attractive for user profiling [1, 2, 13].

Our work contrast previous work, in our focus on the social relevance aspect, and how social information is used to generate a social context profile and the use of social-based filtering as part of a hybrid filtering approach.

2 ARCHITECTURE FOR USER MODELING AND IMAGE RECOMMENDATION

Recommender systems are software tools and techniques, which suggest items that are most likely of interest to a particular user. Based on a user profile, relevant items are selected using filtering techniques such as content-based, collaborative, demographic and hybrid filtering [4, 6, 9, 10]

Our image recommendation system is depicted in Figure 1. The lower part (below the dotted line) shows user profile construction, where the *extraction manager* collects user information from Facebook and Flickr, selects terms representing user interests and categorizes friends to generate a social context profile. Term are semantically enriched [1] using WordNet[8] and DBpedia[9] before a

---

[8] https://wordnet.princeton.edu
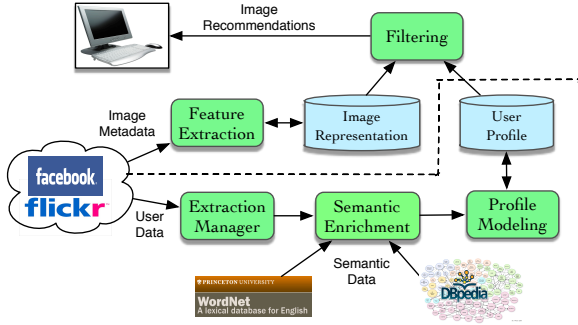[9] http://wiki.dbpedia.org

Figure 1: Framework of the system

user profile, containing both user interests and social context, is generated.

To recommend images through a *filtering* process (depicted in the upper part of Figure 1), the user profile is compared with textual representations of images. *Feature extraction* collects and process image metadata to identify representative terms describing the image.

3 USER MODELING BASED ON SOCIAL MEDIA INFORMATION

User activity information from Facebook and Flickr is analyzed to automatically create a user profile consisting of two components; user interests and social context. *User interests* are represented by a vector of terms, reflecting topics of interest, while *social context* is represented as a classification of friends, distinguishing between e.g. family, close friends, colleagues and distant friends.

We use client-side user profiling, where knowledge of, and access to, the user's Facebook and Flickr accounts are easily available, and where privacy of the information can be supported. As user information from both Facebook and Flick are used, cross-source user profiling [3] is handled.

3.1 User interest profiling

Assuming image collections reflect user interests (by e.g. depicting historical buildings, hiking or football), we use *image collection descriptions* as input to the user profile. Such descriptions are automatically created based on image metadata, using a technique described in [12]. A user may have several collections on Flickr, while we consider all Facebook images owned by the user, as one collection.

User interests are also identified based on *Facebook activity information*. Analysis of title, description and category for the user's *Facebook groups*, can obviously reveal user interests. The same can frequently recurring query topics, which are identified through analyzing the *Facebook search log*. The search log reflects progress in user interests, in that recent queries identify the more recent interests. Short-term, long-term and current interests influence the term weight in the interest vector.

Processing information from Facebook and Flickr, results in a number of interest vectors; for each image collection and for Facebook searches and groups. An aggregated user interest profile

is constructed by combining the vectors and using the most frequent terms for the profile [3].

Assume vectors $\{(V_1(u), k_1), \ldots, (V_n(u), k_n)\}$ for user u, where k_n represents the weight of vector $V_n(u)$. The weight (or importance) of term t is a function of the weight of t in each vector, multiplied with the vector's weight. For example, if term t_i occurs in vector V_2 and V_5, the accumulated weight of t_i is $w(u, t_i) = w_2(u, t_i) * k_2 + w_5(u, t_i) * k_5$.

Interest vector aggregation can result in either a single or multiple vectors. It is argued that multiple keyword profiles, one per interest area, provides a more accurate picture of the user compared to a single profile[9]. Investigating a multi-vector approach is part of our future work.

3.2 Social context profiling

A social context profile describes a user's relation to other members of the social network, by classifying friends based on type and frequency of interactions. We assume that users having a close social relationship will have a higher likelihood to comment and like each other's posts. A relationship score between two users, u and p, is thus calculated based on the comment and like activities between them.

The interaction activity from user p to user u is described in Formula (1), while the relationship score [8], $F_{(u,p)}$, reflecting the two way interaction between u and p, is described in Formula (2).

$$Act_{(u,p)} = \frac{|L_{(u,p)}| + |C_{(u,p)}|}{|Post^u(p.d)|} \qquad (1)$$

In Formula (1), $|L_{(u,p)}|$ and $|C_{(u,p)}|$ represents the number of likes and comments (respectively) from user p to user u, and $|Post^u(p.d))|$ the number of posts made by u since the date when p and u became friends. This normalizes the activities from p over the number of available posts from u.

$$F_{(u,p)} = Act_{(u,p)} + Act_{(p,u)} \qquad (2)$$

The relationship score, F, and interaction activity score, Act, classifies friends into *close friends, contacts, distant friends, followers* and *following*, as shown in Table 1. A high relationship score indicates a close relationship between users. By comparing the two Act scores between u and p, we can detect a predominant one-way interaction and classify the friend as either *follower* or *following*.

The threshold value, s_1 in Table 1 distinguishes close friends from contacts, while s_2 distinguishes contacts from distant friends. The class weight, f_w, indicates the closeness of friends to user u, and is used for calculating relevance of images (see Section 4.3). The f_w values in Table 1, are tentative and can be adjusted.

We also use Facebook Friend Lists[10] to categorize friends. Users can manually add friends to the Close Friends list, or set up smart lists (such as Family, Work, School and Your Area) that automatically stay up-to-date based on profile information users have in common. In Table 1, the sets Family, CloseFriends, Work, School, YourArea and Following refer Facebook's classification of friends.

4 IDENTIFYING RELEVANT IMAGES

We use a novel, hybrid approach for image recommendations, combining social-based filtering with content- and knowledge-based

---

[10]https://www.facebook.com/help/204604196335128/

Table 1: Classification of friend p with respect to user u

Class	Description	Condition	f_w
Family	Included in the Family list	$p \in Family(u)$	1
Close friends	Included in the Close Friends list or frequent, two-way interaction	$p \in CloseFriends(u)$ or $F_{(u,p)} \geq s_1$	1
Contacts	Occasional, two-way interaction	$s_2 \leq F_{(u,p)} < s_1$	0.5
Distant friends	Sparse or non-existing interaction	$F_{(u,p)} < s_2$	0.2
Colleagues	Included in the Work list	$p \in Work(u)$	0.6
Classmate	Included in the School list	$p \in School(u)$	0.4
Local friend	Included in the Your Area list	$p \in YourArea(u)$	0.3
Following	Included in the following list or one way interaction, from user u to friend p	$(F_{(u,p)} \geq s_2 \wedge Act_{(u,p)} \approx 0)$ or $p \in Following(u)$	0.6
Follower	One way interaction, from friend p to user u	$F_{(u,p)} \geq s_2 \wedge Act_{(p,u)} \approx 0$	0.3

filtering. *Content-based filtering* tries to match users to items that are similar to what they have liked in the past, while *knowledge-based filtering* base recommendations on similarities between customer requirements and item descriptions [4, 6].

In our notion of *Social-based filtering* we try to match users to items that are of social importance, by being owned by friends that are considered socially important to the user in target. Social-based filtering focuses the social context of the user, where classification of friends is a central component.

4.1 Finding Facebook users on Flickr

To seamlessly recommend images from both Facebook and Flickr, we need to link Facebook and Flickr accounts. This is challenging since user identity information, for the same person, can be rather diverse across different social media sites [14]. Useful, but not generally applicable, are approaches that already links between accounts, by allowing users to explicitly link profiles, or supporting single sign-on services that implicitly match accounts.

Our automatic account matching compares features from Facebook and Flickr user profiles, following the approach of [16]. Personal data (PD) in a user profile, can be viewed as a set of (field, value) pairs, i.e. $PD = \{(f_1, v_1), \ldots, (f_n, v_n)\}, n \geq 1$, where the fields can be e.g. *nickname, name* and *email*.

Given two sets of user profiles, B from Facebook Q from Flickr, we calculate personal data similarity scores $PD_{sim}(b, q) \forall b \in B, q \in Q$. If $PD_{sim}(b, q) > r$ for some threshold r, we deem the two user profiles to belong to the same person. The similarity function (3) compares the profiles field-by-field from the available information, considering some fields more important than others.

$$PD_{sim}(b, q) = \sum_{f \in Fd} w[cmp(v_f(b), v_f(q))] \quad (3)$$

Let b, q be two different profiles, the set of possible fields be denoted Fd, and the values for some $f \in Fd$ denoted $v_f(b)$ and $v_f(q)$ for profile b and q respectively. The comparison function cmp depends on f, and w is the weight assigned to field f. In our system, the possible fields are, in order of importance, email, full name, nickname, and location, and we support partial matching functions and fuzzy-matches [16].

Since the number of user profiles is huge, scalability of profile matching is an issue. To reduce computation costs, blocking functions can be defined, constructing e.g. blocking keys for profile, content or network features, so that user identities not matching on the key are not compared [14].

4.2 Extracting image features

Content-based filtering is about measuring similarity between user interests and image features. Feature extraction is based on textual image metadata (such as tags, title, description and other surrounding text), and results in a vector containing weighted terms, representing a description of the image.

For images with little metadata, collection(s) where the image is included, can be used to infer image content and contribute to determine possible relevance or not. Image collections often have a clear topic (or objective), and each image within a collection is somehow related to that topic.

Collection descriptions can also be used as a first step for filtering images, by first identify image collections that are likely to contain relevant images, and subsequently identify relevant images within these collections [12].

4.3 Model for recommendations

Relevance of an image I to user u is determined by combining interest relevance of I to u and the social relevance of the friend owning I. A friend's classification determines her social relevance (i.e. weight). Interest relevance is determined by comparing two vectors, representing user interests and image features. This is done using a traditional vector similarity measure, $sim(a, b)$, shown in Formula (4), where $|\vec{a}|$ and $|\vec{b}|$ are the norms of the vectors a and b and $\vec{a} \bullet \vec{b}$ is the internal product of the two vectors [5].

$$sim(a, b) = \frac{\vec{a} \bullet \vec{b}}{|\vec{a}||\vec{b}|} \quad (4)$$

The combined relevance measure for image I to user u, is expressed in Formula (5). Here $P_{int}(u)$ represents the interest vector from user u's profile, $v(I)$ a vector representing image I, and $f_w(I.owner)$ the weight of the friend owning I. An image I is considered relevant, and thus presented to the user, if the relevance score is higher than a threshold value, r, i.e. $Rel(u, I) > r$.

$$Rel(u, I) = x \bullet sim(P_{int}(u), v(I)) + y \bullet f_w(I.owner) \quad (5)$$

The variables x and y determines the importance of interest relevance and social relevance, respectively. Normally we require that $x + y = 1$. Equal importance of the components gives the value 0.5 for both x and y, but the values can be set to favor either interests or social context.

We recognize that a user may want to customize image recommendations through some coarse-grained filtering. For example, the user may want interest relevant images from all her friends, or all images (regardless of topic) from the closest friends. She may

Table 2: Specific filtering alternatives

	Values	Relevance	Source	Images recommended
Default case	$x > 0 \wedge y > 0$	$Rel(u, I)$	F_u	Combined interest and social relevant images from all friends.
Special cases	$x = 1 \wedge y = 0$	$sim(P_u(int), v(I))$	F_u	Interest relevant images from all friends. Social relevance is not considered.
	$x = 0 \wedge y = 1$	$f_w(I.owner)$	F_u	Any type of images from friends. Social relevance determines image relevance. User interests are not considered.
	$f_w(C_i{}^u) = -1$	$Rel(u, I)$	$F_u - C_i{}^u$	User u's friends in class C_i, i.e. $C_i{}^u$, are disregarded, and no images from these friends are recommended.
	$y = -1$	$sim(P_u(int), v(I))$	all users	Social relevance is disregarded. Interest relevant images are searched among any user.

also want interest relevant images from strangers (i.e. persons not in her friends list). Such customization is reflected in Formula (5), by allow special values for x, y, and f_w.

Table 2 presents our filtering alternatives, by showing values for x, y and f_w, relevance function used in each case, source (i.e. users) from which images are retrieved, and a description of which images are recommended. The F_u in source, represents user u's friends, and $C_i{}^u$ represents user u's friends in class C_i.

We first show the default case, where interest relevance and social relevance is combined, and the user receives relevant images, based on relevance function $Rel(u, I)$, from all friends. The following 4 alternatives, are special cases caused by user customization.

Using values $x = 1$ and $y = 0$, the user receives friends' images related to interests only. There is no distinction between friends, i.e. images from close and distant friends are considered equally interesting. Relevance is determined using only the interest relevance component of $Rel(u, I)$.

With values $x = 0$ and $y = 1$, the user receives any type of image from her friends, regardless of topic. Image relevance is determined by the social relevance component of $Rel(u, I)$, and threshold value, r, determines which classes of friends that are relevant. Thus, all friends having a f_w value $f_w > r$, are relevant.

The user may want to avoid recommendations of images from one or more classes of friends, for example the most distant friends. This is reflected by setting the value of f_w to -1 for the unwanted class, i.e. $f_w(C_i{}^u) = -1$.

By setting $y = -1$, social relevance is disregarded all together. Interest relevant images are searched among any user, both friends and strangers. Relevance is determined using the interest relevance component of $Rel(u, I)$.

5 CONCLUSION

We have described a social image recommender system that provides users with images that are both socially relevant and relevant with respect to user interests. The specific focus is recommendation of images that are of personal interest to the user, by e.g. depicting close friends or family, situations or events that are familiar or specifically interesting because it happens to people known by the user. To facilitate filtering based on social relevance, we use information from social media to both automatically generate a user profile, and act as a source for image selection. Our contribution is a novel social-based filtering approach, combined with content-based and knowledge-based filtering. The relevance measure combines interest relevance and social relevance, and allow users to customize the relative importance for these components.

REFERENCES

[1] Ahmad Abdel-Hafez and Yue Xu. 2013. A Survey of User Modelling in Social Media Websites. *Computer and Information Science* 6, 4 (2013), 59–71. https://doi.org/10.5539/cis.v6n4p59

[2] Fabian Abel, Qi Gao, Geert Jan Houben, and Ke Tao. 2011. Semantic enrichment of twitter posts for user profile construction on the social web. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6643 LNCS, PART 2 (2011), 375–389. https://doi.org/10.1007/978-3-642-21064-8_26

[3] F Abel, E Herder, G Houben, N Henze, and D Krause. 2013. Cross-system user modeling and personalization on the SocialWeb. *User Model User-Adap Inter* 23 (2013), 169–209.

[4] C. Aggarwal. 2016. *Recommender Systems*. Springer. https://doi.org/10.1007/978-3-319-29659-3

[5] R.A. Baeza-Yates and B. Ribeiro-Neto. 2011. *Modern Information Retrieval*. Addison-Wesley Publishing Company, Boston, MA, USA.

[6] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. 2013. Recommender systems survey. *Knowledge-Based Systems* 46 (2013), 109–132. https://doi.org/10.1016/j.knosys.2013.03.012 arXiv:arXiv:1011.1669v3

[7] Mohamed Reda Bouadjenek, Hakim Hacid, and Mokrane Bouzeghoub. 2016. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. *Information Systems* 56 (2016), 1–18. https://doi.org/10.1016/j.is.2015.07.008

[8] Najeeb Elahi and Randi Karlsen. 2014. Relation based image retrival in online social network. *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication - ICUIMC '14* (2014), 1–7. https://doi.org/10.1145/2557977.2558019

[9] Susan Gauch, Mirco Speretta, Aravind Chandramouli, Micarelli Alessandro, and Alessandro Micarelli. 2007. User Profiles for Personalized Information Access. *The Adaptive Web* 4321 (2007), 54–89. https://doi.org/10.1007/978-3-540-72079-9_2

[10] M. Rami Ghorab, Dong Zhou, Alexander O'Connor, and Vincent Wade. 2013. Personalised Information Retrieval: Survey and classification. *User Modelling and User-Adapted Interaction* 23, 4 (2013), 381–443. https://doi.org/10.1007/s11257-012-9124-1

[11] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. 2010. Social media recommendation based on people and tags. *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval - SIGIR '10* (2010), 194. https://doi.org/10.1145/1835449.1835484

[12] R. Karlsen, D. Sundby, and J. Nordbotten. 2016. Development of image collection representations for intelligent distributed systems. *Concurrency Computation Practice and Experience* 28, 4 (2016), 1336–1355. https://doi.org/10.1002/cpe.3649

[13] Manel Mezghani, Corinne Amel Zayani, Ikram Amous, and Faiez Gargouri. 2012. A user profile modelling using social annotations. *Proceedings of the 21st international conference companion on World Wide Web - WWW '12 Companion* October 2015 (2012), 969. https://doi.org/10.1145/2187980.2188230

[14] Kai Shu, Suhang Wang, Jiliang Tang, Reza Zafarani, and Huan Liu. 2017. User Identity Linkage across Online Social Networks. *ACM SIGKDD Explorations Newsletter* 18, 2 (2017), 5–17. https://doi.org/10.1145/3068777.3068781

[15] Zhoubao Sun, Lixin Han, Wenliang Huang, Xueting Wang, Xiaoqin Zeng, Min Wang, and Hong Yan. 2015. Recommender systems based on social networks. *Journal of Systems and Software* 99 (2015), 109–119. https://doi.org/10.1016/j.jss.2014.09.019

[16] Jan Vosecky, Dan Hong, and Vincent Y Shen. 2009. User identification across multiple social networks. *Networked Digital Technologies, 2009. NDT '09. First International Conference on* (2009), 360–365. https://doi.org/10.1109/NDT.2009.5272173

References

[1] AHMAD ABDEL-HAFEZ AND YUE XU. A Survey of User Modelling in Social Media Websites. *Computer and Information Science*, 2013. 87

[2] FABIAN ABEL, QI GAO, GEERT JAN HOUBEN, AND KE TAO. Semantic enrichment of twitter posts for user profile construction on the social web. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2011. 87

[3] CHARU C AGGARWAL ET AL. *Recommender systems*. Springer, 2016. 82

[4] SYED ZAMBERI AHMAD. Evidence of the characteristics of women entrepreneurs in the Kingdom of Saudi Arabia: an empirical investigation. *International Journal of Gender and Entrepreneurship*, **3**[2]:123–143, 2011. 34

[5] BOANERGES ALEMAN-MEZA, MEENAKSHI NAGARAJAN, CARTIC RAMAKRISHNAN, LI DING, PRANAM KOLARI, AMIT P SHETH, I BUDAK ARPINAR, ANUPAM JOSHI, AND TIM FININ. Semantic analytics on social networks: experiences in addressing the problem of conflict of interest detection, 2006. 56

[6] ZAFAR ALI. SOCIAL MEDIA IMPLICATION ON POLITICS OF PAKISTAN; MEASURING THE IMPACT OF FACEBOOK. *The International Asian Research Journal*, **1**[01]:13–21, 2013. 33

[7] HEGE K ANDREASSEN, MARIA M BUJNOWSKA-FEDAK, CATHERINE E CHRONAKI, ROXANA C DUMITRU, IVETA PUDULE, SILVINA SANTANA,

Henning Voss, and Rolf Wynn. European citizens' use of E-health services: A study of seven countries. *BMC Public Health*, **7**[1]:53, 2007. 36

[8] Miriyam Aouragh. Social media, mediation and the Arab revolutions. *tripleC: Communication, Capitalism & Critique. Open Access Journal for a Global Sustainable Information Society*, **10**[2]:518–536, 2012. 30

[9] Shlomo Avineri. *The Social and Political Thought of Karl Marx*. Cambridge University Press, 1968. 17

[10] Francis P Barclay, C Pichandy, Anusha Venkat, and Sreedevi Sudhakaran. India 2014: Facebook 'Like'as a Predictor of Election Outcomes. *Asian Journal of Political Science*, [ahead-of-print]:1–27, 2015. 33

[11] J. A. Barnes. Class and Committees in a NORWEGIAN ISLAND PARISH. *Human Relation 39-58*, 1954. 17

[12] M Baur. visone-Software for the Analysis and Visualization of Social Networks. 2008. 55

[13] Peter Beaumont. The truth about Twitter , Facebook and the uprisings in the Arab world, 2011. 31

[14] Nicholas J. Belkin. Intelligent Information Retrieval: Whose Intelligence? 16

[15] M Bender, T Crecelius, M Kacimi, S Michel, T Neumann, J X Parreira, R Schenkel, and G Weikum. Exploiting social relations for query expansion and result ranking. In *Data Engineering Workshop, 2008. ICDEW 2008. IEEE 24th International Conference on*, pages 501–506, 2008. 90

[16] Tim Berners-Lee, James Hendler, Ora Lassila, and Others. The semantic web. *Scientific american*, **284**[5]:28–37, 2001. 23, 24, 25

[17] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 2013. 82, 87

[18] Jason Bonander. Personally tailored health information: a health 2.0 approach. In *Medicine 2.0 Conference*. JMIR Publications Inc., Toronto, Canada, 2008. 87

[19] Pia Borlund and Jesper W. Schneider. Reconsideration of the simulated work task situation. In *Proceeding of the third symposium on Information interaction in context - IIiX '10*, page 155, New York, New York, USA, 2010. ACM Press. 16

[20] Mohamed Reda Bouadjenek, Hakim Hacid, and Mokrane Bouzeghoub. Social networks and information retrieval, how are they converging? A survey, a taxonomy and an analysis of social information retrieval approaches and platforms. 2015. 87

[21] Danah Boyd. Friendster and publicly articulated social networking. In *CHI '04: CHI '04 extended abstracts on Human factors in computing systems*, pages 1279–1282. ACM Press, 2004. 42

[22] Ulrik Brandes. A faster algorithm for betweenness centrality*. *The Journal of Mathematical Sociology*, **25**[2]:163–177, jun 2001. 55

[23] John G Breslin, Stefan Decker, Andreas Harth, and Uldis Bojars. SIOC: an approach to connect web-based communities. *International Journal of Web Based Communities*, **2**[2]:133–142, 2006. 27

[24] Dan Brickley and Libby Miller. The Friend of a Friend (FOAF) project, 2000. 26

[25] Dan Brickley and Libby Miller. FOAF Vocabulary Specification 0.98, 2010. 63

[26] P. J. Brown and G. J. F. Jones. Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering. *Personal and Ubiquitous Computing*, **5**[4]:253–263, dec 2001. 15

[27] V Bush and AWM Think. The atlantic monthly. *As we may think*, 1945. 11

[28] Edward Caropreso and Sue-Jen Chen. Effects of Personality on Small Group Communication and Task Engagement in an. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, **2005**, pages 1921–1930, 2005. 75

[29] PJ Carrington, J Scott, and S Wasserman. *Models and methods in social network analysis*. 2005. 20

[30] Guanling Chen and David Kotz. A Survey of Context-Aware Mobile Computing Research. nov 2000. 15

[31] D E Comer, David Gries, Michael C Mulder, Allen Tucker, A Joe Turner, and Paul R Young. Computing As a Discipline. *Commun. ACM*, **32**[1]:9–23, jan 1989. 6, 7

[32] Colleen Cool and Amanda Spink. Issues of context in information retrieval (IR): an introduction to the special issue. *Information Processing & Management*, **38**[5]:605–611, sep 2002. 16

[33] Berners-Lee; T.; Bray; Connolly; D.; Cotton. Architecture of the World Wide Web, Volume One. 25

[34] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z. Wang. Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, **40**[2]:1–60, apr 2008. 1

[35] Marc Davis, Simon King, Nathan Good, and Risto Sarvas. From context to content. In *Proceedings of the 12th annual ACM international conference on Multimedia - MULTIMEDIA '04*, page 188, New York, New York, USA, oct 2004. ACM Press. 16

[36] B Debatin, J P Lovejoy, A.-K. Horn, and B N Hughes. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Journal of Computer-Mediated Communication*, **15**:83–108, 2009. 46

[37] Peter J Denning. Is computer science science? *Communications of the ACM*, **48**[4]:27–31, 2005. 6

[38] J DiMicco, DR Millen, and W Geyer. Motivations for social networking at work. In *ACM conference on Computer supported cooperative work*, pages 711–720, 2008. 21

[39] Robin Dunbar. *Grooming, Gossip, and the Evolution of Language.* Harvard University Press, Cambridge, MA, 1998. 44

[40] Najeeb Elahi and Randi Karlsen. User behavior in online social networks and its implications. In *Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics - WIMS '12*, page 1, New York, New York, USA, jun 2012. ACM Press. 28, 74

[41] Najeeb Elahi, Randi Karlsen, and Sigmund Akselsen. A context centric approach for semantic image annotation and retrieval. In *2009 Computation World: Future Computing, Service Computation, Cognitive, Adaptive, Content, Patterns*, pages 665–668. IEEE, 2009. 48

[42] Najeeb Elahi, Randi Karlsen, and Einar J Holsbø. Personalized Photo Recommendation By Leveraging User Modeling On Social Network. In *Proceedings of International Conference on Information Integration and Web-based Applications & Services*, IIWAS '13, pages 68:68—-68:71, New York, NY, USA, 2013. ACM. 70

[43] Najeeb Elahi, Randi Karlsen, and Waqas Younas. Ontology-Based Image Annotation by Leveraging Social Context. *International Journal of Handheld Computing Research (IJHCR)*, **3**[3]:53–66, 2012. 48

[44] D Elgesem and Nordbotten.J. The Role of Context in Image Interpretation. *Context-based Information Retrieval CIR'07*, aug 2007. 15

[45] Ramez Elmasri and Shamkant Navathe. *Fundamentals of database systems.* Addison-Wesley Publishing Company, 2010. 14

[46] Mona Eltahawy. Tunisia's Jasmine Revolution. *The Washington Post. www. washingtonpost. com. Available at: http://www. washingtonpost. com/wp-dyn/content/article/2011/01/14/AR2011011405084. html*, 2011. 30

[47] Nahed Eltantawy and Julie B. Wiest. The Arab Spring— Social Media in the Egyptian Revolution: Reconsidering Resource Mobilization Theory, sep 2011. 29, 34, 87

[48] G Erétéo, M Buffa, F Gandon, P Grohan, M Leitzelman, and P Sander. A state of the art on social network analysis and its applications on a semantic web, 2008. 22, 25, 56

[49] Guillaume Erétéo, Michel Buffa, Fabien Gandon, and Olivier Corby. Analysis of a Real Online Social Network Using Semantic Web Frameworks. In Abraham Bernstein, David Karger, Tom Heath, Lee Feigenbaum, Diana Maynard, Enrico Motta, and Krishnaprasad Thirunarayan, editors, *The Semantic Web - ISWC 2009*, **5823**, pages 180–195. Springer-Verlag, 2009. 22

[50] O Erling and I Mikhailov. {RDF Support in the Virtuoso DBMS}. **221**, 2007. 62

[51] Luis Fernandez-Luque, Najeed Elahi, and Francisco J Grajales. An analysis of personal medical information disclosed in YouTube videos created by patients with multiple sclerosis. *Studies in health technology and informatics*, **150**:292–296, 2009. 28, 36

[52] Claudiu S. Firan, Wolfgang Nejdl, and Raluca Paiu. The Benefit of Using Tag-Based Profiles. In *2007 Latin American Web Conference (LA-WEB 2007)*, pages 154–154. IEEE, oct 2007. 90

[53] LC Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 1977. 57

[54] Linton C Freeman. Centrality in social networks conceptual clarification. *Social Networks*, **1**[3]:215–239, 1978. 21, 56

[55] Shenghua Gao, Zhengxiang Wang, Liang-Tien Chia, and Ivor Wai-Hung Tsang. Automatic image tagging via category label and web data. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1115–1118. ACM, 2010. 89, 90

[56] Scott Golder, Dennis Wilkinson, and Bernardo Huberman. Rhythms of social interaction: messaging within a massive online network. 2006. 46, 74

[57] Asuncion Gomez-Perez, Mariano Fernández-López, and Oscar Corcho-Garcia. Ontological engineering. *Computing Reviews*, **45**[8]:478–479, 2004. 23, 24

[58] Mark Granovetter. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, **1**:201, 1983. 71

[59] Ralph Gross and Alessandro Acquisti. Information revelation and privacy in online social networks, 2005. 29, 43, 87

[60] Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, **5**[2]:199–220, 1993. 24

[61] Arif Hasan. *The Unplanned Revolution: Observations on the Processes of Socio-economic Change in Pakistan* . Oxford University Press, 2009. 45

[62] Patrick Hille, Gianfranco Walsh, Simon Brach, and D Dose. Why online identity theft poses a major threat to e-business. In *Proceedings of the ACM WebSci'11*, pages 1–2, 2011. 42

[63] Matthew Horridge, Holger Knublauch, Alan Rector, Robert Stevens, and Chris Wroe. A Practical Guide To Building OWL Ontologies Using The Protege-OWL Plugin and CO-ODE Tools Edition 1.0. 2004. 24

[64] Andreas Hotho, Robert Jschke, Christoph Schmitz, and Gerd Stumme. FolkRank: A Ranking Algorithm for Folksonomies. In *Proc. FGIR 2006*, pages 111–114, 2006. 90

[65] Victoria Hougham. Sociological Skills Used in the Capture of Saddam Hussein, 2005. 21

[66] Chia-Chuan Hung, Yi-Ching Huang, Jane Yung-jen Hsu, and David Kuan-Chun Wu. Tag-Based User Profiling for Social Media Recommendation. Technical report, 2008. 87

[67] Peter Ingwersen. Cognitive perspectives of information retrieval interaction: Elements of a cognitive IR theory, 1996. 16

[68] Yongick Jeong and Reaz Mahmood. Reading the world's mind: Political, socioeconomic and cultural approaches to understanding worldwide Internet search queries. *International Communication Gazette*, **73**[3]:233–251, 2011. 45

[69] J.Kahan and Marja-Ritta Koivunen. Annotea: an open RDF infrastructure for shared Web annotations, 2001. 49

[70] Gareth JF Jones, Daragh Byrne, Mark Hughes, Noel E O'connor, and Andrew Salway. Automated annotation of landmark images using community contributed datasets and web resources. In *International Conference on Semantic and Digital Media Technologies*, pages 111–126. Springer, 2010. 89, 90

[71] Dhiraj Joshi, Jiebo Luo, Jie Yu, Phoury Lei, and Andrew Gallagher. Using geotags to derive rich tag-clouds for image annotation. In *Social media modeling and computing*, pages 239–256. Springer, 2011. 89, 90

[72] Jason J Jung, Krzysztof Juszczyszyn, and Ngoc Thanh Nguyen. Centrality measurement on semantically multiplex social networks: divide-and-conquer approach. *International Journal of Intelligent Information and Database Systems*, **1**[3]:277–292, jan 2007. 18

[73] Andreas M Kaplan and Michael Haenlein. Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, **53**[1]:59–68, 2010. 1

[74] Randi Karlsen, Najeeb Elahi, and Anders Andersen. Personalized recommendation of socially relevant images. In *Proceedings of the*

8th International Conference on Web Intelligence, Mining and Semantics, page 41. ACM, 2018. 70, 82

[75] Randi Karlsen, Martin Hætta Evertsen, and Najeeb Elahi. Metadata-based automatic image tagging. *International Journal of Metadata, Semantics and Ontologies*, **8**[4]:298–308, 2013. 48

[76] Inga Carboni Kate Ehrlich and I B M Technical Report. Inside Social Network Analysis, 2005. 56, 71

[77] Jennifer Keelan, Vera Pavri-Garcia, George Tomlinson, and Kumanan Wilson. Youtube as a source of information on immunization: a content analysis. *jama*, **298**[21]:2482–2484, 2007. 87

[78] Jihad B Khazen. Censorship and State Control of the Press in the Arab World. *The Harvard International Journal of Press/Politics*, **4**[3]:87–92, jul 1999. 30

[79] Hak Lae Kim, Sung-Kwon Yang, Seung-Jae Song, John G Breslin, and Hong-Gee Kim. Tag Mediated Society with SCOT Ontology. In *Semantic Web Challenge*, 2007. 27

[80] Alan Kirman. Ants, Rationality, and Recruitment. *The Quarterly Journal of Economics*, **108**[1]:137–156, feb 1993. 21

[81] Ralf Klamma and Christian Haasler. Wikis as Social Networks: Evolution and Dynamics. 2008. 18

[82] Graham Klyne and Jeremy J Carroll. Resource description framework (RDF): Concepts and abstract syntax. 2006. 25

[83] Jerzy Kozyra. Why your friends have more friends than you do? 23

[84] Grzegorz Kukła, Przemysław Kazienko, Piotr Bródka, and Tomasz Filipowski. Recommendation boosted query propagation in the social network. pages 113–124, oct 2010. 18

[85] CLIFF LAMPE, NICOLE ELLISON, AND CHARLES STEINFIELD. A face(book) in the crowd: social Searching vs. social browsing, 2006. 43

[86] HOLLINK LAURA, SCHREIBER GUUS, WIELEMAKER JAN, AND WIELINGA BOB. Semantic Annotation of Image Collections, 2003. 88

[87] SMITH A ANDERSON M DUGGAN M PERRIN A LENHART A. Teens, Technology and Friendships. *Pew Research Center Pew Research Center*, aug 2015. 23

[88] ZHENG LIU, HUA YAN, AND HUIJIAN HAN. Mining large-scale social images with rich metadata and its application. *JSW*, **7**[4]:749–756, 2012. 89, 90

[89] HP LUHN. A new method of recording and searching information. *American Documentation*, 1953. 11

[90] HP LUHN. A statistical approach to mechanized encoding and searching of literary information. *IBM Journal of research and development*, 1957. 11

[91] WENDY E MACKAY. Triggers and barriers to customizing software, 1991. 44, 46

[92] AMEESH MAKADIA, VLADIMIR PAVLOVIC, AND SANJIV KUMAR. Baselines for image annotation. *Int. J. Comput. Vision*, **90**[1]:88–105, October 2010. 89

[93] CD MANNING, P RAGHAVAN, AND H SCHÜTZE. *Introduction to information retrieval*. 2008. 12, 13

[94] EE MARIA. Is Centrality the Key to High Performance? *Journal of Competitiveness*, 2010. 20

[95] J MCCARTHY. Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence*, **13**[1-2]:27–39, apr 1980. 24

[96] DEBORAH L MCGUINNESS, RICHARD FIKES, JAMES HENDLER, AND LYNN ANDREA STEIN. DAML+ OIL: an ontology language for the Semantic Web. *IEEE Intelligent Systems*, **17**[5]:72–80, 2002. 25

[97] Deborah L McGuinness, Frank Van Harmelen, and Others. OWL web ontology language overview. *W3C recommendation*, **10**[10]:2004, 2004. 25

[98] Pablo Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. DBpedia Spotlight: Shedding Light on the Web of Documents. 2011. 62

[99] Manel Mezghani, Corinne Amel Zayani, Ikram Amous, and Faiez Gargouri. A user profile modelling using social annotations. 2012. 87

[100] JC Mitchell. *Social networks in urban situations: analyses of personal relationships in Central African towns.* 1969. 20

[101] Stefano Mizzaro and Luca Vassena. A social approach to context-aware retrieval. *World Wide Web*, **14**[4]:377–405, feb 2011. 90

[102] JL Moreno. Emotions mapped by new geography. *New York Times*, 1933. 22

[103] Emily Moxley, Jim Kleban, and B S Manjunath. Spirittagger: a geo-aware tag suggestion tool mined from flickr, 2008. 88, 89, 90

[104] Emily Moxley, Jim Kleban, Jiejun Xu, and B S Manjunath. Not all tags are created equal: learning Flickr tag semantics for global annotation, 2009. 88

[105] Dennis M Murphy and James F White. Propaganda: Can a word decide a war? Technical report, DTIC Document, 2007. 30

[106] Mark A Musen and the Protégé Team. The Protégé Project: A Look Back and a Look Forward. *AI matters*, **1**[4]:4–12, jun 2015. 62

[107] Mor Naaman, Ron B Yeh, Hector Garcia-Molina, and Andreas Paepcke. Leveraging context to resolve identity in photo albums, 2005. 88

[108] Elahi Najeeb and Karlsen Randi. Relation Based Image Retrival In Online Social Network, paper presented ACM (COSN 2013). *paper presented ACM Conference on Online Social Networks (COSN 2013)*, 2013. 70

[109] M. E. J. Newman. The structure and function of complex networks. page 58, mar 2003. 18

[110] MEJ Newman. The structure of scientific collaboration networks. *Proceedings of the National Academy of . . .*, 2001. 21

[111] Mark Nixon. *Feature extraction & image processing.* Academic Press, 2008. 50

[112] Amanda Nosko, Eileen Wood, and Seija Molema. All about me: Disclosure in online social networking profiles: The case of FACEBOOK. *Computers in Human Behavior*, **26**:406–418, 2010. 44

[113] Benjamin Nowack. CONFOTO: Browsing and annotating conference photos on the Semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, **4**[4]:263–266, 2006. 88

[114] Martin O'connor, Holger Knublauch, S W Tu, and M A Musen. Writing rules for the semantic web using SWRL and Jess. *Prot{é}g{é} With Rules WS, Madrid*, 2005. 25

[115] Paul Ormerod. *Why most things fail: Evolution, extinction and economics.* John Wiley & Sons, 2007. 21

[116] Brian L Ott. The age of Twitter: Donald J. Trump and the politics of debasement. *Critical Studies in Media Communication*, **34**[1]:59–68, 2017. 47

[117] Alexander P Pentland, Rosalind W Picard, and Stan Scarloff. Photobook: Tools for content-based manipulation of image databases. In *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*, pages 34–47. International Society for Optics and Photonics, 1994. 50

[118] A GLOOR PETER. Web Science 2.0: Identifying Trends through Semantic Social Network Analysis. **4**, pages 215–222, 2009. 53

[119] ADRIAN POPESCU AND PIERRE-ALAIN MOËLLIC. Monuanno: automatic annotation of georeferenced landmarks images. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, page 11. ACM, 2009. 89, 90

[120] J J POTTERAT, L PHILLIPS-PLUMMER, S Q MUTH, R B ROTHEN-BERG, D E WOODHOUSE, T S MALDONADO-LONG, H P ZIMMERMAN, AND J B MUTH. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sexually transmitted infections*, **78 Suppl 1**:i159–63, apr 2002. 21

[121] JASON POTTS, STUART CUNNINGHAM, JOHN HARTLEY, AND PAUL ORMEROD. Social network markets: a new definition of the creative industries. *Journal of Cultural Economics*, **32**[3]:167–185, jun 2008. 17, 18, 21

[122] CH PROCTOR AND CP LOOMIS. Analysis of sociometric data. *Research methods in social*, 1951. 57

[123] MASSIMILIANO RUOCCO. *Geo-Temporal Mining and Searching of Events from Web-based Image Collections.* PhD thesis, Citeseer. 12

[124] YASMINE RYAN. Tunisia's bitter cyberwar, 2011. 31

[125] G SABIDUSSI. The centrality index of a graph. *Psychometrika*, 1966. 57

[126] DANIEL SALBER, ANIND K. DEY, AND GREGORY D. ABOWD. The context toolkit. In *Proceedings of the SIGCHI conference on Human factors in computing systems the CHI is the limit - CHI '99*, pages 434–441, New York, New York, USA, may 1999. ACM Press. 15

[127] G SALTON. The SMART retrieval system—experiments in automatic document processing. 1971. 11

[128] GERARD SALTON AND MICHAEL J MCGILL. *Introduction to Modern Information Retrieval.* McGraw-Hill, Inc., New York, NY, USA, 1986. 60

[129] B SCHILIT, N ADAMS, AND R WANT. *Context-aware computing applications.* Xerox Corp., Palo Alto Research Center, 1994. 15

[130] B.N. SCHILIT AND M.M. THEIMER. Disseminating active map information to mobile hosts. *IEEE Network*, **8**[5]:22–32, sep 1994. 15

[131] ALBRECHT SCHMIDT, KOFI ASANTE AIDOO, ANTTI TAKALUOMA, URPO TUOMELA, KRISTOF VAN LAERHOVEN, AND WALTER VAN DE VELDE. Advanced Interaction in Context. pages 89–101, sep 1999. 15

[132] JOHN SCOTT. *Social network analysis.* Sage, 2nd ed edition, 2000. 18

[133] HATEM MOUSSELLY SERGIEH, GABRIELE GIANINI, MARIO DÖLLER, HARALD KOSCH, ELÖD EGYED-ZSIGMOND, AND JEAN-MARIE PINON. Geo-based automatic image annotation. In *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*, page 46. ACM, 2012. 89, 90

[134] AMIT SHETH AND MEENAKSHI NAGARAJAN. Semantics-empowered social computing. *IEEE Internet Computing*, **13**[1]:76–80, 2009. 87

[135] BÖRKUR SIGURBJÖRNSSON AND ROELOF VAN ZWOL. Flickr tag recommendation based on collective knowledge. In *Proceeding of the 17th international conference on World Wide Web - WWW '08*, page 327, New York, New York, USA, apr 2008. ACM Press. 90

[136] ABRAHAM SILBERSCHATZ, HENRY F KORTH, SHASHANK SUDARSHAN, ET AL. *Database system concepts*, **6**. McGraw-Hill New York, 2010. 13

[137] ANA SILVA AND BRUNO MARTINS. Tag recommendation for georeferenced photos. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, pages 57–64. ACM, 2011. 89, 90

[138] PINAKI SINHA AND RAMESH JAIN. Classification and annotation of digital photos using optical context data, 2008. 50

[139] ARNOLD W M SMEULDERS, MARCEL WORRING, SIMONE SANTINI, AMARNATH GUPTA, AND RAMESH JAIN. Content-Based Image Retrieval at the End of the Early Years. *IEEE Trans. Pattern Anal. Mach. Intell.*, **22**[12]:1349–1380, 2000. 50, 88

[140] RAYMOND M SMULLYAN. *First-order logic*, **21968**. Springer, 1968. 25

[141] SUSAN SONTAG. On Photography. *New York Times*, 1977. 1

[142] ZHOUBAO SUN, LIXIN HAN, WENLIANG HUANG, XUETING WANG, XI-AOQIN ZENG, MIN WANG, AND HONG YAN. Recommender systems based on social networks. *Journal of Systems and Software*, 2015. 87

[143] J TRAVERS AND S MILGRAM. An experimental study of the small world problem. *Sociometry*, **32.4**:425–443, 1969. 17

[144] CHIH-FONG TSAI AND CHIHLI HUNG. Automatically annotating images with keywords: A review of image annotation systems. *Recent Patents on Computer Science*, **1**[1]:55–68, 2008. 89

[145] MISCHA TUFFIELD, STEPHEN HARRIS, DAVID P DUPPLAW, AJAY CHAKRAVARTHY, CHRISTOPHER BREWSTER, NICHOLAS GIBBINS, KIERON O'HARA, FABIO CIRAVEGNA, DEREK SLEEMAN, YORICK WILKS, AND NIGEL R SHADBOLT. Image annotation with Photocopain, 2006. 88

[146] WAN-SHIOU YANG, JIA-BEN DIA, HUNG-CHI CHENG, AND HSING-TZU LIN. Mining Social Networks for Targeted Advertising. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, **6**, pages 137a–137a. IEEE, 2006. 17, 18

[147] MENG WANG, BINGBING NI, XIAN-SHENG HUA, AND TAT-SENG CHUA. Assistive tagging: A survey of multimedia tagging with human-computer joint exploration. *ACM Computing Surveys (CSUR)*, **44**[4]:25, 2012. 89

[148] Xin-Jing Wang, Lei Zhang, Feng Jing, and Wei-Ying Ma. Annosearch: Image auto-annotation by search. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, **2**, pages 1483–1490. IEEE, 2006. 89, 90

[149] S Wasserman. *Social network analysis: methods and applications.* Cambridge University Press, 1994. 17, 18, 20, 21, 56

[150] Dengsheng Zhang, Md Monirul Islam, and Guojun Lu. A review on automatic image annotation techniques. *Pattern Recognition*, **45**[1]:346–362, 2012. 89