Information Retrieval from Social Network

Muthuselvi.G Guru Prasath.B

ABSTRACT
Social networks are very popular these days, and it facilitates search and retrieval of information through internet. Hence, our understanding and usage of the WWW significantly differs from early days. The Internet, particularly the WWW is not only being used and approached for gathering information anymore. It is used routinely to reach other people through various applications ranging from email, chats, community platforms and social network sites. The current development of the WWW increasingly allows more people to publish information in social networks. Whereas users have changed from passive recipients of information to active content consumers and creators, and the nature of information has also changed from static text to dynamic multimedia. With the widespread use of social networks, live User-Defined Information (UDI) has begun to overcome the Internet. Such UDI covers a range of social network. More research works have focused on analyzing, organizing, indexing, and retrieving information, search for dynamic and live UDI's from the web. However, research efforts to make sense of the huge and unstructured UDI have begun. Social networks and multimedia content sharing Web sites have become increasingly popular in recent years. Information Retrieval is one of the most focused research area which motivate us to concentrate on this field.

KEYWORDS:
User Defined Information (UDI), Information Retrieval(IR), Internet, Social networks, multimedia, World Wide Web(WWW), traditional IR and web IR systems.

INTRODUCTION
Internet and the World Wide Web (WWW) have dramatically and dynamically changed since the Internet became popular. Social network service typically focuses on building online communities of people who share interests and activities, or are interested in exploring the interests and activities of others. This trend has resulted in a continuously growing volume of publicly available multimedia content on content sharing Web sites from text (tweets, forums, and Face book messages [4]) to images (Instagram and Flickr [1]), photos Picasa [2], videos (YouTube [3]), and community questioned-answer forums (Yahoo! Answers and Wiki Answers), which have created new challenges for access, search, and retrieval of the shared content. For instance, Flicker has hosted more than 6 billion photos and Face book has approximately 100 billion photos stored on its servers [5]. Every minute, 48 h of video are uploaded to YouTube [6], and 20 million videos are uploaded.

Retrieval of User Defined Information from the following areas:

- Image and location-based information, including shared photos and check-in venues;
- Topic-based information, such as tweets, discussion forums, and community question and-answer forums;
- Application-based information covering local mobile apps and associated information and discussions;
- Structured information, including cultural and historical information.

All these social sites generally contain publicly available data, information, interactions and their collective preferences and interests in the particular area. At the same time, they have become a popular way to share and disseminate information. For example, users upload their personal photos and share them through online communities, letting other people comment or rate them. UDI permits us to better understand the social networks and society. This paper describes our initial work on dynamic monitoring of raw UDI and events as they unfold. Specifically, this paper highlights six of the research projects carried out by real time to generate higher-level analytics to understand topics of interests to society, their sentiments, evolving live events and social communities, as well as fashion habits and trends. User-Defined Information offers a wealth of research, business, and societal opportunities. Our long term aim is to generate social representations of data, information, topics and people, documenting the activities and concerns of people within the society.

Monitored User-Defined Information sources (as of 24 August 2012).

Figure 1 describes familiar social networks and their usages are described below.

- Amazon has 889,867 Images, 1,813,545 Products and 5,814,731 Reviews
- Flickr has 4,622,134 Images, 11,022,514 Reviews, and 2,460,357 Users.
- Foursquare 2,543,267 Check-ins, 1,492,096 Venues, 60,667,476 Tweets, 25,101,730 Users, 26,101,730 User relationships.
- Twitter 706,815 Images, 1,908 Tweets, Users 357,483,444.
- Yahoo! Answers Question &Answer pairs 4,158,912
- Wiki Answers Question &Answer pairs 51,018,100.

Dynamic Monitoring

Our project crawls and monitors UDI related to a city. Along this aim, we have continually crawled the image and location-based UDI, topic oriented UDI, UDI related to local mobile applications, and structured information. A dynamic monitor is designed to view and visualize the live UDIs as they are being created and propagated through the Internet.

Figure 2 shows an example of Independence Day event map. This example visualizes user-generated check-ins, Flickr image uploads, and tweets that reference Independence Day. We found 25,822 records in all sources, and this map displays only 1,300 of them in 1.011 seconds. On visualizing the check-ins and tweets related to the topic ''Independence Day.''

We analyze these UDI to generate first order analytics. For example, by analyzing the tweets related to any topic at different time stamps, we can generate keyword clouds that depict the differences in context related to this entity.

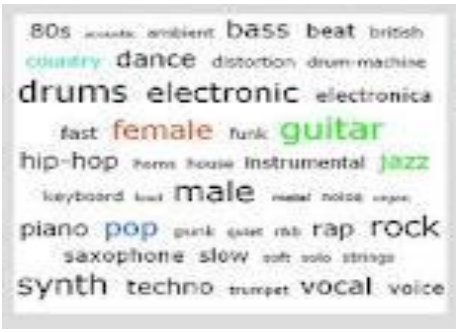

Figure 3. Keyword clouds

It shows the keyword clouds for Instagram before and after it was acquired by social network.

Information Retrieval Frequency

Obviously, a user community is an important factor in relating social networks.

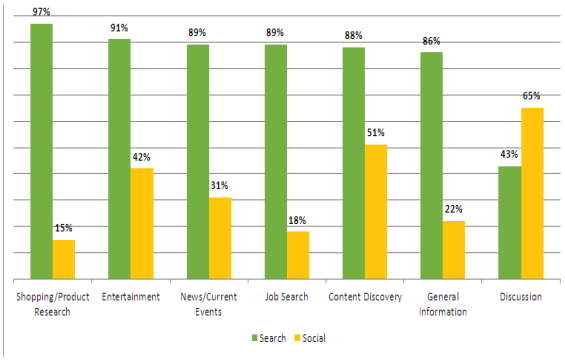

Figure: 4 show the analysis of discovering user communities is difficult because users might have completely dissimilar interests, friendship groups, discussion topics, and so forth. Thus, another research issue is to automatically identify key users, hot and emerging discussions, and events related to an organization. Discovering such information requires the design of sophisticated multimedia analysis methods.

We have cumulated this research into a system for discovering the frequency of information retrieval and monitoring emerging discussions and topics.

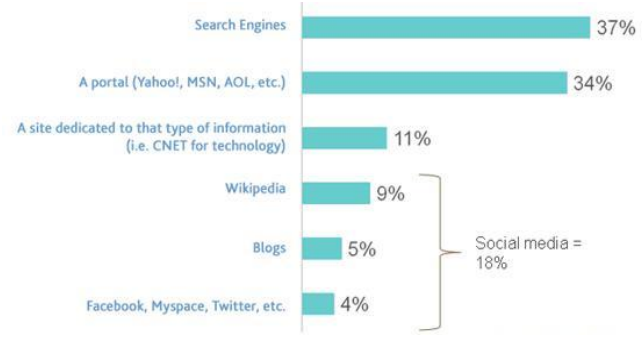

Figure.5 Frequency of Search Engine

When we look for new information online, these websites we prefer to search. 18% of social networks are really involved in information retrieval.

Observation of Social Network Evolution

Information about personal relations and social interaction in the WWW is not only freely accessible, but it is also archived to some degree. Web pages might not be available forever, but it can be assumed that information is kept for a certain amount of time.

Therefore, information about past social interactions or personal relations is also available as current ones. This allows the extraction and analysis of social networks that may have existed in the past in combination with current social networks. Structural changes within social networks might become visible and useful for improvement of web applications.

Integration of social network information into the information retrieval process:

Essential steps we have to follow while information integration and retrieval

Information Source Selection

It is necessary to identify possible information sources for a meaningful extraction of social networks.

Disambiguation of Entities

The identification and differentiation of person names is challenging as the quality of web pages is low and misspellings, errors and invalid content are common. Names can also be written in different styles, but actually refer to the same person or vice versa a single name may refer to several different persons.

Information Retrieval in the Web and Social Network

There are significant differences between traditional IR and web IR systems. Although web IR systems use similar information models to analyze and index textual documents, these systems differ significantly from traditional IR systems as the WWW has different requirements.

Weighting of Social Relations

The social context and the type of social interactions within these information sources need to be carefully analyzed in order to obtain a meaningful understanding of the social network structure and the selection of appropriate SNA methods.

Location oriented Social network

In location-oriented social networks, users implicitly interact with each other by visiting places, posting comments, and uploading photos. These heterogeneous interactions convey latent information we can use to identify meaningful user groups (namely, social communities) that exhibit unique location-based characteristics. In particular, social network users can perform check-ins, upload photos and post comments at various locations.

Organizational Analysis

Social media portals such as Twitter contain the everyday thoughts, opinions, and experiences of their online users. Parts of these UDI reflect and reveal their sentiments about organizations or events. Motivated by real-world industrial needs, this project focuses on discovering the public perception of organizations from live social media. We can ascertain the public perception of an organization by identifying users, locations and their opinions our work focuses on tackling the key challenges in identifying this information. First, different organizations might share the same acronym. Hence, we need to differentiate the discussions and user communities of such organizations. However, because tweets are no more than 140 characters long, the disambiguation task is challenging, especially for organizations that share the same context (for example, two universities). To tackle this issue, we utilize the posts' content and geolocation as well as the community of the authors (users) to accurately relate the posts to target organizations. However, because users often turn off the GPS tracking services of their devices, the posts' geolocations are usually unavailable [7]. Therefore, the second challenge is to automatically predict a post's geolocation.

Applications of social network.

Social networks are very popular in all fields.

Friend-of-a-Friend (FOAF)

It is a machine-readable ontological representation of Persons and their relationships to other persons. Thus, it uses and builds on ontology descriptions and languages that are widely used in semantic web initiatives. FOAF is another language that builds upon RDF and describes in a machine-readable format personal attributes (email, homepage, projects, interests, etc). More importantly, it describes links to individual friends. Semantically-Interlinked Online Communities (SIOC) SIOC is a framework that provides a Semantic Web ontology based on RDF/OWL for the representation and integration of online community information. It standardizes a semantic description of social media content like blog posts, online forums, mailing lists or Usenet groups. SIOC focuses on the relation between content, creator and the context in which the content was created. While this focus does not explicitly provide information about direct social relations between people, it provides information about the interests of people and groups to which people are related. SIOC provides a vocabulary to describe the relation between a usergroup and its members.

XHTML Friends Network (XFN)

The XHTML Friends Network (XFN)26 is a simple extension of the hyperlinks tag to express social relationships with hyperlinks. Hence, XFN makes use of the HTML hyperlink tag attribute rel, which is intended to describe the relationship between the current document and the document referred to by the hyperlink. It further defines a very basic vocabulary for different kinds of social relationships that can be applied to this attribute in order to express the social relation that is related to the hyperlink.

This vocabulary is grouped in the following categories: friendship (contact, acquaintance, friend), professional (co-worker, colleague), physical (met), geographical (co-resident, neighbor), family (child, parent, sibling, spouse, kin), romantic (muse, crush, date, sweetheart), identity (me). Through the attribution of hyperlinks which are directed relations, and with data about social relation a social network with directed connection can be extracted from XFN.

OpenSocial

OpenSocial provides an API specification for social networking sites to open up their databases for external applications. It defines a REST interface which provides information about people. It also allows adding persons and relations to the database of the OpenSocial provider.

Twitter

Twitter is a service that allows people to publish short messages on the internet from diverse communications devices, such as from a computer or a mobile phone. It is commonly referred to as microblogging services because messages are short (140 characters) and in most cases inform readers about what a user is doing in a certain moment.

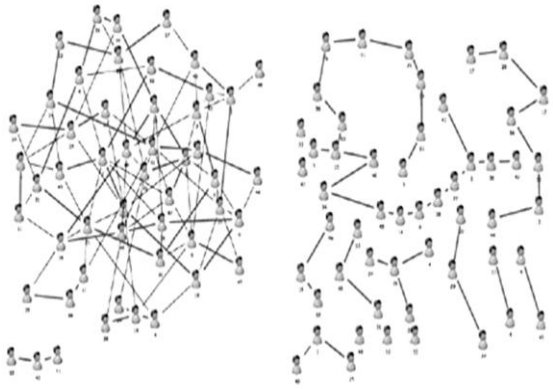

Figure 6. a) Social network-followers
b) Social network–friendship

Conclusion

UDI permits to learn about the growth of various concepts and topics within the society over time. Such information is valuable to normal users going about their daily lives, can help organizations better serve their customers, and can help governmental organizations better plan for social services and provisions as a result of social trends and concerns. This work is just beginning.

We aim to share our data, analytics, and resources in an effort to increase collaboration and make our world a better place to live in.

REFERENCES

[1] Flickr Web site. (Online). Available: http://www.flickr.com

[2] Picasa Web site. (Online). Available: http://picasa.google.com

[3] YouTube Web site. (Online). Available: http://www.youtube.com

[4] Facebook Web site. (Online). Available: http://www.facebook.com

[5] E. Barnett. (2011, Oct.). 3.4 billion photographs on Google+ in 100 days.(Online).Available: http://www.telegraph.co.uk/technology/google/8838196/3.4-billion-photographs-on-Google-in-100-days.html

[6] Google Inc. (2011, Dec.) YouTube statistics. (Online). Available: http://www.youtube.com/t/press_statistics.

[7] A. Sadilek et al., ''Finding Your Friends and Following them to Where You Are,'' Proc. 5th ACM Int'l Conf. Web Search and Data Mining (WSDM), ACM Press, 2012

[8] A. Saha et al., ''Learning Evolving and Emerging Topics in Social Media: A Dynamic NMF Approach with Temporal Regularization,'' Proc. 5th ACM Int'l Conf. Web Search and Data Mining (WSDM), 2012, pp. 693_702.

[9] Z. Ming et al., ''Prototype Hierarchy Based Clustering for the Categorization and Navigation of Web Collections,'' Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, ACM Press, 2010, pp. 2_9.

[10] L. Tang et al., Community Detection and Mining in Social Media, Morgan & Claypool Publishers, 2010.

[11] H. Liu et al., ''Robust Clustering as Ensemble of Affinity Relations,'' Proc. 24th Ann. Conf. Neural Information Processing Systems (NIPS), Curran Associates Publisher, 2010, pp. 1414_1422.

[12] S. Liu et al., ''Street-to-Shop: Cross-Scenario Clothing Retrieval via Parts Alignment and Auxiliary Set,'' Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition (CVPR), IEEE CS Press, 2012.

[13] G. Li et al., ''Desks: Direction-Aware Spatial Keyword Search,'' Proc. Int'l Conf. Data Eng. (ICDE), IEEE CS Press, 2012, pp. 459_470.

[14] L. Shi et al., ''S3: An Efficient Shared Scan Scheduler on MapReduce Framework,'' Proc. 2011 Int'l Conf. Parallel Processing (ICPP), IEEE CS Press, 2011, pp. 325_334.

[15] D. Carmel, N. Zwerdling, I. Guy, S. Ofek-Koifman, N. Har'El, I.Ronen, E. Uziel, S. Yogev, and S. Chernov, "Personalized social search based on the user's social network," in Proc. CIKM, 2009, pp.1227–1236.

[16] Y. Cai and Q. Li, "Personalized search by tag-based user profile and resourceprofile in collaborative tagging systems," in Proc. CIKM, 2010, pp. 969–978.