A Review on Social Network Aggregator using Crawler and Document Clustering

Neetu, Deepti Sharma

M.Tech(pursuing), Dept. of CSE, Advanced Institute of Technology & Management, Palwal, Haryana under the

Affiliation of Maharshi Dayanand University at Rohtak, Haryana, India

Head, Dept. of CSE, Advanced Institute of Technology & Management, Palwal, Haryana under the Affiliation of

Maharshi Dayanand University at Rohtak, Haryana, India

ABSTRACT: Organizing contacts, links friends; sharing thought, emotions and content; looking content and knowledge arE the essential suggests that a social networking web site provides. By sharing, and managing content the users type a social network as an example, Facebook, Linkedin, twitter, Orkut, etc. A user might exist on many completely different social networking sites and therefore the drawback to take care of the account on these networks prevails. This brings US to outline a Social network aggregation i.e. aggregation the social content from completely different social networks and integrate it at one location/site. it's a trial to arrange a user's social networking expertise as whole. This paper provides a review on completely different Social network Aggregators and problems in group action social network.

KEYWORDS: Social Web, Web Aggregators, Document Clustering, Focused Crawlers.

I. INTRODUCTION

The social internet may be a network of relationships that link individuals through the Social networking sites. on-line social internet sites like Instagram, Facebook, and Twitter area unit among the foremost fashionable sites on the web. A user will have accounts on multiple social networking sites/services; the plain question are the overhead to manage and track their content/contact or alternative social activities that is formed by their friends' everyday and primarily scattered across many networks. This 'social data' (which may be posts, status, photos, links, tweets, scraps etc that area unit a part of updates in social network service) might cause important info overload to users. this type of knowledge overload referred to as "network overload" [1]. Alex Patriquin of Compete.com reported on the overlapping of users having accounts at numerous on-line social network services.

Table 1: Survey of user having account at multiple sites

Site	LinkedIn	Facebook	Friendster	Bebo	Orkut	Plaxo	Ning	MySpace	HiS
LinkedIn	100	42	8	4	3	3	8	32	2
Facebook	2	100	2	4	1	9	1	64	2
Friendster	6	23	100	5	1	0	2	40	4
Bebo	1	25	2	100	0	0	1	65	3
Orkut	8	26	4	3	100	1	2	29	7
Plaxo	54	48	8	5	4	100	14	34	2
Ning	19	35	6	6	2	2	100	44	1
MySpace	0	20	1	3	0	0	0	100	1
HiS	1	24	4	7	2	0	0	69	100

A 2009 study of 11,000 users reported that the mainstream of MySpace, LinkedIn, and Twitter users also have Facebook accounts [7]. Social Network Aggregator may be a wise resolution to the on top of drawback and problems in social world. It's the method of collating/aggregating/organizing information unfold across multiple social network services. Plan is to arrange and ease the data retrieval method for a user maintaining multiple social networks actively. Social network aggregators consolidate the varied social activities/data in such the simplest way that user isn't needed to login to every web site and perform same group action. A user performs the group action at one web site and also the info is synchronized to any or all of the social networks that the user specifies. Aggregation tools area unit in situ that has users to consolidate messages, track social information across networks.

ISSUES IN AGGREGATING MULTIPLE SOCIAL NETWORKS

There are many issues related to integration of Social Networks   [6][9][10]. Some of them are:

1. Disparity in schema of different databases is a major issue in integrating data from multiple networks
2. In spite of data aggregation, single sign in for same user is still an open issue.
3. There are no means for monitoring any unauthorized access to data in social networks and real-time application updates may result in unwanted access to private information.
4. Representation of social data is another important challenge as each Social Network Services have their own syntaxes and terms for representing social data specific to their network. The academic and open web communities have put great effort to develop standard ontology for the representation of social data. There are several major standards, including FOAF (foaf-project.org), XFN (gmpg.org/xfn/), GUMO and Activity Stream.
5. Another important challenge to overcome is the transformation among the different meanings attached to same standards.
6. Authentication of a web services to a user or another social network cannot be guaranteed as well.
7. The correctness of information and accuracy provided by aggregator can also be not guaranteed.

TECHNIQUES TO AGGREGATE THE SOCIAL NETWORKS

A lot of research has been done in the area of integrating social networks from the mid 1990s.
There are four main ways by which we can integrate the social network [11] [13]:
1 Content Aggregators: Content aggregators collect social data from multiple sources of specific topic and provide analytics based on relationships across multiple data sources or networks. Content aggregators analyzes features set content and correlate them to user data (e.g. preferences, interests), based on a user model derived by analyzing the previous actions on data by the user.
2 Comparison Analytics: A comparative result is evaluated on the basis of user specified criteria
3 Relationship Aggregation: It Provides a delta of relationships between a user and company services/ information sources with which the user has a business relationship.
4 Process Aggregation: Business Process which require coordination across a variety of services/ information and managed and a common point of contact is provided.

EXISTING SOCIAL NETWORK AGGREGATORS

There are number of social network aggregators available now days to manage the profiles at a single point. These aggregators provide specialized features to maintain the profile wisely. Some of them are listed below:
SocConnect:

SocConnect is a web browser that aggregates social data from different Social Networks and allows users to create personalized social and semantic contexts for their social data. Users can combine and cluster friends and rate Network. SocConnect also features personalized suggestion of friends' activities that may be fascinating to the user, using machine learning techniques. SocConnect allows users to define their individual perspective of social data aggregated from different social networks that may indicate their presence for each environment. SocConnect provides content-based recommendations for social updates in social network services. Based on extensive estimation, it provides a set of user preferences that can provide the best performance on personalized recommendation.

Flock:

Flock is a web tool provides management tools for social networking and other Web 2.0 services. It integrates multimedia services, including Myspace, Facebook, Twitter etc. When signing into any of the supporting service, Flock can get status updates and friend's updates. In addition, Flock can also search in Twitter to update multiple services at once, and use Facebook chat service from the browser. Other features are:

- Customary sharing of scraps/posts, links, photos and videos.
- A media bar showing pictures and videos.
- News reader with RSS feeds
- A reader and editor of blogs
- An email client

Flock doesn't require you to provide authentication to any other site that maintains online security. However aggregation of multiple social networks is one aspect of Flock, it's a complete Web browser.

XeeMe:

XeeMe lets users/brands manage their entire social identity, identify new networks and people and nurture their presence. It has a long number of supported networks and it offers useful analytics and with XeeGraph the user has a point of reference about his presence value and network relevance. It offers the user the possibility of organizing all social networks at a single point and shares their social presence with one URL with friends, customers, partners and people. Through the application the user can discover new networks or people who are in other networks and offer the possibility of connecting with them. By sharing the URL on each post, the number of visits to the social site of the user will increase.

Hootsuite:
Hootsuite is an aggregator for businesses and organizations to collaboratively provide a web dashboard that executes campaign across multiple social networks. The user must sign in the social networks he wants to use and give some permission. After that the user can see the updates in his network, publish in one or more networks at the same time, launch marketing promotions, identify new users so that network grows, and distribute targeted messages.

- Scheduling: Option for live updates or pre-schedule posts and shares in advance.
- Files: Integrate images, video and files along with your messages.
- RSS feed; Connect to your RSS and send your blog to your social media streams.
- Bookmark functionality: Hootsuite Hootlet can be useful to share and collaborate the social data easily
- Integration with Handheld: flexible integration with Hootsuite from your iPhone.
- Popularity of HootSuite is due to fact that it is run via a web browser and there is no need to additional patch/software.

People Aggregator:

People Aggregator is basically a social identity hub. It runs only on Linux. The site is also a Digital Identity Hub, meaning that users will be able to utilize a single sign on, using, for example, their Flickr or Blogger ID; profile data can also be imported/ exported without losing any content. People Aggregator will allow users to create blogs, media galleries, podcasts, blogcasts, and forums. As an open social network, users can freely move between similar networks without having to reestablish profile info, blog posts, etc.

Comparison of Different Social Network Aggregators

There are various solutions available to integrate the social network but no one has tried to integrate the information available within multiple social networks. Flock can get updates from friends, status updates and photos submitted at Multiple Social networks whereas SocConnect users can creates a personalized social and semantic contexts for their social data. Users can combine and cluster friends and rate Network. HootSuite aggregates organizations and

businesses to collaboratively execute promotions across multiple social networks and XeeMe Organizes Social presence, discovers new network and people. It organizes the entire social presence of the user, determine new networks and people and develop their presence and influence. But none of the aggregator has mined the multiple social networks and extracted some useful information after collecting data from different Social Networks.

Table2: Comparison of different social networks

Networks → Features ↓	SocConnect	Flock	Xeeme	HootSuite	People Aggregator
Analytics	No	No	No	Yes	No
Scheduling	Yes	No	yes	Yes	Yes
Team Colllaboration	Yes	Yes	yes	Yes	No
Group friends	Yes	No	No	No	No
Rating of activities & Friends	Yes	No	No	No	No
Social Networks	Facebook LinkedIn Twitter	Facebook LinkedIn Twitter	Facebook LinkedIn Twitter	Facebook LinkedIn Twitter	Facebook LinkedIn Twitter
Analysis & extraction of social data	No	No	No	No	No

II.   LITERATURE REVIEW

More and more web users keep up with newest information through information streams such as the popular micro-blogging website Twitter. In this paper we studied content recommendation on Twitter to better direct user attention. In a modular approach, we explored three separate dimensions in designing such a recommender: content sources, topic interest models for users, and social voting. We implemented 12 recommendation engines in the design space we formulated, and deployed them to a recommender service on the web to gather feedback from real Twitter users. The best performing algorithm improved the percentage of interesting content to 72% from a baseline of 33%. We conclude this work by discussing the implications of our recommender design and how our design can generalize to other information streams.[2]

Large numbers of college students have become avid Facebook users in a short period of time. In this paper, we explore whether these students are using Facebook to find new people in their offline communities or to learn more about people they initially meet offline. Our data suggest that users are largely employing Facebook to learn more about people they meet offline, and are less likely to use the site to initiate new connections.[1]

Social network and data portability has recently gained a lot of interest as one of the issues for social media sites on the Web. In this paper, we will show how Semantic Web technologies and especially the FOAF and SIOC vocabularies can be used to model user information and user-generated content in a machine-readable way. Thus, we will see how data and network information can be reused among various services and applications, at almost zero-cost for developers of such tools.[3]

A literature search was carried out in July 2012. The multidisciplinary search engine ISI Web of Science was used that stores articles from more than 250 disciplines and covers over 12,000 of the highest impact journals worldwide,

including Open Access journals and over 150,000 conference proceedings. Only peer-reviewed papers were considered, conference abstracts were not included in the literature review. The following search terms as topics were used: "social networking" (B), "online communities" (C), "social media" (D), "computer-mediated communication"(E), "discussion boards" (F), "networking sites" "social network sites" (G), website (H), "social computing" (I), "Virtual community" (J), "Web 2.0" (K), and "user-generated content", each search term combined with the terms "older adults", elder*, senior*, and "age differences". Search result numbers ranged from 0 or 1 results (user-generated content AND older adults) to 89 results (social networking AND older adults). Title and abstract were screened to assess suitability to include the article into the review. Additional literature was identified by a literature search of Xie et al in press, where articles were used which sufficed our search criteria (see below) (A). Further literature was searched with Google scholar, and in references and citing articles of the literature found in Web of Science (marked with add. search). Articles that focused on ICT use of older adults in general (e.g., Nasi et al 2012) establishing new communication technologies without the use of publicly available sites or discussion boards were not included in the review (e.g., Garattini, Wherton & Prendergast 2011; Rothenpieler et al., 2011; Sun et al. 2007; 2010). Articles that investigated age differences of people in young adulthood (e.g. Baams et al 2011) were not included in the review. Articles identified by the literature search were screened if topic was indeed linked to social media use and included conceptual or empirical links to older adults. Since the results of this search were quite scarce, an additional search in Google scholar was carried out. Additionally, the articles found via ISI were searched for their references and citing articles to identify more articles. Only articles were included which reflect some kind of innovative form of communication, such as commenting online, interacting with instant messaging or in a chat room, or participating in virtual game or social worlds as defined by Kaplan and Haenlein (2011). Articles which merely included the use of email as online communication channel were not included in the review, since this form of internet-based communication reflects great similarities to more traditional forms of communication like writing a letter and may thus not involve specifics of social media use.[4]

Social networks are playing an important role in personal as well as corporate environments. However, perceived issues and evolving challenges may hinder further expansion of social networks to meet new opportunities. In this paper, we review inherent concepts and properties of social networks and highlight major analytical evaluation criteria, which are used to identify key findings that reveal degrees of benefits and shortcomings of social networks. We also discuss some proposed solutions related to decentralized social networks in the context of business implications as well as their effects on privacy, identity and trust issues.[5]

III.   PROPOSED SOLUTION

We propose to develop an aggregator that will integrate several social web sites together and responds to a user's query; extracting the relevant data as specified in Query written in a natural language from multiple social networks and presenting data appropriately as result; thereby helping users who belong to multiple networks manage diverse profiles. The proposed network will mine the profiles of a user existing at multiple social sites and provide the information after combining the different profiles of their friends and their activities. For example: "Friends who studied from Advanced Institute of Technology and Management passed in 2016". The resultant will be user's all friends who studied from Advanced Institute of Technology and Management passed in 2016 and are there in the profiles of his/her existing multiple networks. The Proposed aggregator will maintain several accounts at one place and extracts the relevant publicly available data. It will also offer a better improvement over keyword Searching. It will perform a combine search about their friend's likes, places, pictures, events, applications, things they love or care about etc.. The Search will be based on both the content of the user and their friends' profiles.

For this approach, we propose a social web where people can come and interact with each other. They can connect with Friends, share thoughts, photos, media, events and knowledge. They can also search for the information available at their several accounts of social networking sites in a natural language. We will develop a crawler which will extract the information of the user from its several profiles existing at different networking sites. It will thus create a graph of his/her network and cluster the events accordingly and store it in a database.

User's query will be divided into the tokens and these keywords will be given to query processor to search from the database and thus the result is submitted to the user. The resultant will be user's all information that exists in the profile of his/her multiple networks.
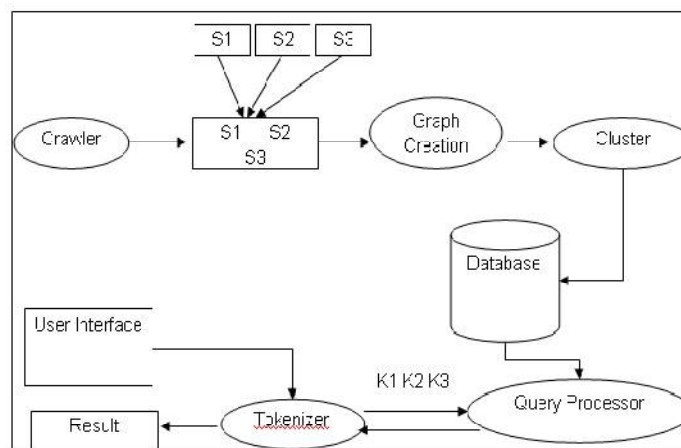

Figure 1 Proposed Model of Social Network Aggregator

IV. CONCLUSION

Social Networking sites contain a tremendous amount of data such as text and multimedia which can be leveraged for mining of a variety of applications. The success of Online Social Networks emerges the interest in social network integration amongst the researchers. Social Network Aggregators maintains one single point of social data where user can keep track of the various social activities from all of the social networking sites where he/she is having account. There are issues pertaining to the aggregation of data across social networks. This has been sighted with examples that exist to integrate multiple social network services. A number of social media aggregators have revealed up in recent years, but social media services still require to research and implement more effective and in efficient ways to provide aggregation. No such example exist which collects and excavate the data. Thus there is a need of a Social network Aggregator which can extract the profiles of a user existing at multiple social sites and provides information after combining the different profiles

REFERENCES

1. Cli Lampe, Nicole Ellison, and Charles Steineld. A face (book) in the crowd: Social searching vs. social browsing. In Proceedings of the ACM Special Interest Group on Computer-Supported Cooperative Work, 2006.
2. Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams. In CHI '10: Proceedings of the 28th international conference on Human factors in computing systems, pages 1185{1194, New York, NY, USA, 2010. ACM
3. Uldis Bojars, Alexandre Passant, John Breslin, and Stefan Decker. Social network and data portability using semantic web technologies. In Proceedings of the Workshop on Social Aspects of the Web, 2008.
4. Petter Brandtzaeg and Jan Heim. Why People Use Social Networking Sites. In Ant A. Ozok and Panayiotis Zaphiris, editors, Online Communities and Social Computing, volume 5621, chapter 16, pages 143-152. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
5. Francesca Carmagnola, Fabiana Vernero, and Pierluigi Grillo. Sonars: A social networks-based algorithm for social recommender systems. In Procedddings of the 17th International Conference on User Modeling, Adaptation, and Personalization, 2009
6. Michael Chisari, The future of social networking. In Proceedings of the W3C Workshop on the Future of Social Networking, 2009
7 . Rohani, Vala Ali; Ow Siew Hock (2010). "On Social Network Web Sites: Definition, Features, Architectures and Analysis Tools".Journal of Advancesin Computer Research (2): 41-53. Retrieved 2011-06-11.
8. Heckmann, D., Schwarzkopf, E., Mori, J., Dengler, D., & Kroner, A. (2007). The user model and context ontology gumo revisited for future web 2.0 extensions. In Proceedings of the international workshop on contexts and ontologies: Representation and reasoning
9. Erétéo, G., Buffa, M., Gandon, F., Leitzelman, M., & Limpens, F. (2009). Leveraging social data with semantics. In Proceedings of the W3C workshop on the future of social networking.
10 Chen, J., Nairn, R., Nelson, L., Bernstein, M., & Chi, E. (2010). Short and tweet: Experiments on recommending content from information streams. In CHI '10: Proceedings of the 28th international conference on human factors in computing systems (pp. 1185–1194).
11. S. Catanese, P. De Meo, E. Ferrara, and G. Fiumara. "Analyzing the Facebook Friendship Graph." In Proceedings of the 1st Workshop on Mining the Future Internet, pages 14 - 19, 2010.
12 Resnick, P., Lacovou, N., Suchak, M., Bergstrom, P., Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In Proceedings of the ACM conference on computer supported cooperative work