

# A Cocktail Approach For Communication Between Different Social Networking Sites Through XML Schema

Meghana Kakde<sup>1</sup>, N.B.Kadu<sup>2</sup>

<sup>1</sup>Ms. Meghana V. Kakde  
Student, Department of Computer Engineering  
Pravara Rural Engineering College  
Loni, India  
kakdemeghana9096@gmail.com

<sup>2</sup>Prof. N. B. Kadu  
(Ass.Prof) Department of Computer Engineering  
Pravara Rural Engineering College  
Loni, India  
kamleshkadu@rediffmail.com

\*\*\*

**Abstract** - Today can see that so many people connect on social site. Communication is a strong media to communicate to each other by people. Today Social Networking Sites (SNS) have become a part of our day to day life. To share a lot of data on these sites. They helped us to make the world smaller and integrated with each other. There are many SNS available today and many more are piling each day. Thus a user uses many SNS each day and communicate and share data with friends and family. This communication medium gave rise to complex structure whether a user really like the SNS which he uses more or he needs another SNS other than he uses more. Also, Large Amount of data does not store properly so, To implement XML schema instead database As well as To Propose an approach that automatically diversifies XML keyword search based on its different contexts in the XML data. Given a short and vague keyword query and XML data to be searched, we first derive keyword search candidates of the query by a simple feature selection model.

**Key Words:** Keyword, Homogenous, Heterogeneous, offline

## 1.INTRODUCTION

Social Network Sites (SNSs) such as Facebook and Twitter have transformed today's society by providing easily accessible platforms for users to connect, communicate, and share vast amount of information. With SNSs, people keep in touch with their contacts, reconnect with old acquaintances, and establish new relationships with others based on shared features such as hobbies, interests, and overlapping friendship. The population of SNS users and the number of SNSs have been growing rapidly. For example, Facebook™ is estimated to have over one billion active users. Although it is hard to obtain an accurate estimate, there are thousands of

SNSs all over the world that provide different kinds of services. As a result, a user may register with multiple SNSs for different social network applications, carry multiple SNS accounts, interact with contacts from different SNSs, publish and access different web contents, and share contents within each SNS community. While SNSs offer different services, one key feature shared among SNSs is how they are built around users and users' existing social networks. Yet each SNS is isolated, so users manage their profiles and build relationship separately on different SNSs. The content for the same user in different SNSs may overlap, so it becomes a burden for users to manage contents across different SNSs. This is the landscape of heterogeneous SNSs. With growing influence and complexity of SNSs, researchers have been developing methods to connect users and aggregating data across SNSs so that each SNS no longer stands alone. For example, the study in [1] summarizes how social networks connect services allow users to leverage their information on multiple SNSs, from using single ID to access multiple SNS accounts to publish contents simultaneously on multiple SNSs. However, the aggregation of social relationship data has not been well studied, which motivates this article. Here, we propose a system architecture to integrate heterogeneous SNSs and investigate a model to characterize the social relationships among a large number of users across heterogeneous SNSs.

KEYWORD data has attracted much search on structured and semi-structured interest recently. While keyword query empowers ordinary users to search vast amount of data, the ambiguity of keyword query makes it difficult to effectively answer keyword queries, especially for short and vague keyword queries. To address this challenging problem, in this paper we propose an approach that automatically diversifies XML keyword search based on its different contexts in the XML data. Given a short and vague keyword query and XML data to be searched, we first derive keyword

search candidates of the query by a simple feature selection model. And then, we design an effective XML keyword search diversification model to measure the quality of each candidate. After that, two efficient algorithms are proposed to incrementally compute top-k qualified query candidates as the diversified search intentions. Two selection criteria are targeted: the k selected query candidates are most relevant to the given query while they have to cover maximal number of distinct results. At last, a comprehensive evaluation on real and synthetic data sets demonstrates the effectiveness of our proposed diversification model and the efficiency of our algorithms.

- 1) When the given keyword query only contains a small number of vague keywords, it would become a very challenging problem to derive the user's search intention due to the high ambiguity of this type of keyword queries.
- 2) Although sometimes user involvement is helpful to identify search intentions of keyword queries, a user's interactive process may be time-consuming when the size of relevant result set is large.

## 2. PROBLEM STATEMENT

The Problem is to determine As Follows,

- 1) how to communicate all SNS with the help of heterogenous technique By Using XML Schema.
- 2) Given a keyword query q and an XML data T, our target is to derive top-k expanded query candidates in terms of high relevance and maximal diversification for q in T. Here, each query candidate represents a context or a search intention of q in T.

## 3. LITERATURE SURVEY

### Privacy and Security for Online Social Networks: Challenges and Opportunities :

Chi Zhang and Jinyuan Sun , Xiaoyan Zhu , Yuguang Fang ,Online social networks such as Facebook, Myspace, and Twitter have experienced exponential growth in recent years. These OSNs offer attractive means of online social interactions and communications, but also raise privacy and security concerns. In this article we discuss the design issues for the security and privacy of OSNs. We find there are inherent design conflicts between these and the traditional design goals of OSNs such as usability and sociability. We present the unique security and privacy design challenges brought by the core functionalities of OSNs and highlight some opportunities of utilizing social network theory to mitigate these design conflicts.

### Measurement and Analysis of Online Social Networks :

Alan Mislove , Massimiliano Marcon , Krishna P. Gummadi , Online social networking sites like Orkut, YouTube, and Flickr are among the most popular sites on the Internet. Users of these sites form a social network, which provides a powerful means of sharing, organizing, and finding content and contacts. The popularity of these sites provides an

opportunity to study the characteristics of online social network graphs at large scale. Understanding these graphs is important, both to improve current systems and to design new applications of online social networks. This paper presents a large-scale measurement study and analysis of the structure of multiple online social networks. We examine data gathered from four popular online social networks: Flickr, YouTube, Live Journal, and Orkut. We crawled the publicly accessible user links on each site, obtaining a large portion of each social network's graph. Our data set contains over 11.3 million users and 328 million links. We believe that this is the first study to examine multiple online social networks at scale. Our results confirm the power-law, small-world, and scale free properties of online social networks. We observe that the in degree of user nodes tends to match the out degree; that the networks contain a densely connected core of high-degree nodes; and that this core links small groups of strongly clustered, low-degree nodes at the fringes of the network. Finally, we discuss the implications of these structural properties for the design of social network based systems.

### Social Network Sites: Definition, History, and Scholarship :

danah m. boyd , Nicole B. Ellison , Social network sites (SNSs) are increasingly attracting the attention of academic and industry researchers intrigued by their affordances and reach. This special theme section of the Journal of Computer-Mediated Communication brings together scholarship on these emergent phenomena. In this introductory article, we describe features of SNSs and propose a comprehensive definition. We then present one perspective on the history of such sites, discussing key changes and developments. After briefly summarizing existing scholarship concerning SNSs, we discuss the articles in this special section and conclude with considerations for future research.

### Searching Social Networks :

Bin Yu , Munindar P. Singh , A referral system is a multi-agent system whose member agents are capable of giving and following referrals. The specific cases of interest arise where each agent has a user. The agents cooperate by giving and taking referrals so each can better help its user locate relevant information. This use of referrals mimics human interactions and can potentially lead to greater effectiveness and efficiency than in single-agent systems. Existing approaches consider what referrals may be given and treat the referring process simply as path search in a static graph. By contrast, the present approach understands referrals as arising in and influencing dynamic social networks, where the agents act autonomously based on local knowledge. This paper studies strategies using which agents may search dynamic social networks. It evaluates the proposed approach empirically for a community of AI scientists (partially derived from

bibliographic data). Further, it presents a prototype system that assists users in finding other users in practical social networks.

#### **Keyword search on structured and semi-structured data**

:Yi Chen, Wei Wang, Ziyang Li, Xuemin Lin,

Empowering users to access databases using simple keywords can relieve the users from the steep learning curve of mastering a structured query language and understanding complex and possibly fast evolving data schemas. In this tutorial, we give an overview of the state-of-the-art techniques for supporting keyword search on structured and semi-structured data, including query result definition, ranking functions, result generation and top-k query processing, snippet generation, result clustering, query cleaning, performance optimization, and search quality evaluation. Various data models will be discussed, including relational data, XML data, graph-structured data, data streams, and workflows. We also discuss applications that are built upon keyword search, such as keyword based database selection, query generation, and analytical processing. Finally we identify the challenges and opportunities of future research to advance the field.

#### **XRANK: Ranked Keyword Search over XML Documents :**

Lin Guo ,Feng Shao ,Chavdar Botev , Jayavel Shanmugasundaram ,

We consider the problem of efficiently producing ranked results for keyword search queries over hyperlinked XML documents. Evaluating keyword search queries over hierarchical XML documents, as opposed to (conceptually) flat HTML documents, introduces many new challenges. First, XML keyword search queries do not always return entire documents, but can return deeply nested XML elements that contain the desired keywords. Second, the nested structure of XML implies that the notion of ranking is no longer at the granularity of a document, but at the granularity of an XML element. Finally, the notion of keyword proximity is more complex in the hierarchical XML data model. In this paper, we present the XRANK system that is designed to handle these novel features of XML keyword search. Our experimental results show that XRANK offers both space and performance benefits when compared with existing approaches. An interesting feature of XRANK is that it naturally generalizes a hyperlink based HTML search engine such as Google. XRANK can thus be used to query a mix of HTML and XML documents.

#### **Multi-way SLCA-based keyword search in xml data :**

Chong Sun , Chee-Yong Chan, Amit K. Goenka

Keyword search for smallest lowest common ancestors (SLCAs) in XML data has recently been proposed as a meaningful way to identify interesting data nodes in XML data where their subtrees contain an input set of keywords. In this paper, we generalize this useful search paradigm to support keyword search beyond the traditional AND semantics to include both AND, OR boolean operators as

well. We first analyze properties of the LCA computation and propose improved algorithms to solve the traditional keyword search problem (with only AND semantics). We then extend our approach to handle general keyword search involving combinations of AND, OR boolean operators. The effectiveness of our new algorithms is demonstrated with a comprehensive experimental performance study.

#### **Efficient keyword search for smallest lcas in xml databases :**

Yu Xu , Yannis Papakonstantinou

Keyword search is a proven, user-friendly way to query HTML documents in the World Wide Web. We propose keyword search in XML documents, modeled as labeled trees, and describe corresponding efficient algorithms. The proposed keyword search returns the set of smallest trees containing all keywords, where a tree is designated as "smallest" if it contains no tree that also contains all keywords. Our core contribution, the Indexed Lookup Eager algorithm, exploits key properties of smallest trees in order to outperform prior algorithms by orders of magnitude when the query contains keywords with significantly different frequencies. The Scan Eager variant is tuned for the case where the keywords have similar frequencies. We analytically and experimentally evaluate two variants of the Eager algorithm, along with the Stack algorithm. We also present the XK Search system, which utilizes the Indexed Lookup Eager, Scan Eager and Stack algorithms and a demo of which on DBLP data is available at <http://www.db.ucsd.edu/projects/xksearch>. Finally, we extend the Indexed Lookup Eager algorithm to answer Lowest Common Ancestor (LCA) queries.

#### **Top-k keyword search over probabilistic xml data :**

Jianxin Li, Chengfei Liu, Rui Zhou, Wei Wang,

Despite the proliferation of work on XML keyword query, it remains open to support keyword query over probabilistic XML data. Compared with traditional keyword search, it is far more expensive to answer a keyword query over probabilistic XML data due to the consideration of possible world semantics. In this paper, we firstly define the new problem of studying top-k keyword search over probabilistic XML data, which is to retrieve k SLCA results with the k highest probabilities of existence. And then we propose two efficient algorithms. The first algorithm Pr Stack can find k SLCA results with the k highest probabilities by scanning the relevant keyword nodes

only once. To further improve the efficiency, we propose a second algorithm Eager Top K based on a set of pruning properties which can quickly prune unsatisfied SLCA candidates. Finally, we implement the two algorithms and compare their performance with analysis of extensive experimental results.

#### **4. RELATED WORK**

In this paper most important task is image and video communication. How to transfer video as well image.

### 5. EXISTING SYSTEM

- Homogenous User only Communicate to each other eg. Facebook user only communicate to facebook user.
- No Heterogeneous user Communicate to each other.
- Message does not sent offline.
- The problem of diversifying keyword search is firstly studied in IR community. Most of them perform diversification as a post-processing or re-ranking step of document retrieval based on the analysis of result set and/or the query logs. In IR, keyword search diversification is designed at the topic or document level.
- When the given keyword query only contains a small number of vague keywords, it would become a very challenging problem to derive the user’s search intention due to the high ambiguity of this type of keyword queries.
- There is no guarantee that the structured queries to be evaluated can find matched results due to the structural constraints;
- The process of constructing structured queries has to rely on the metadata information in XML data.

### 6. PROPOSED SYSTEM

- Homogenous User as well as Heterogeneous User Communicate to each other eg. Facebook user only communicate to facebook user.
- Heterogeneous user Communicate to each other.
- Message is send offline.
- To address the existing issues, To develop a method of providing diverse keyword query suggestions to users based on the context of the given keywords in the data to be searched. By doing this, users may choose their preferred queries or modify their original queries based on the returned diverse query suggestions.
- To address the existing limitations and challenges, we initiate a formal study of the diversification problem in XML keyword search, which can directly compute the diversified results without retrieving all the relevant candidates.
- Towards this goal, given a keyword query, we first derive the co-related feature terms for each query keyword from XML data based on mutual information in the probability theory, which has been used as a criterion for feature selection. The selection of our feature terms is not limited to the labels of XML elements.

### 7. SYSTEM ARCHITECTURE

Homogenous:

Communication is possible between to each peer only.

Heterogeneous:

Communication is possible between to different peer only.

Admin :

- Admin set xml schema to access file respective request very fastly.
- Admin set category by using domain knowledge.

- With the help of category we can easily depart subcategory. To create xml schema to help access file and gives best result to client.

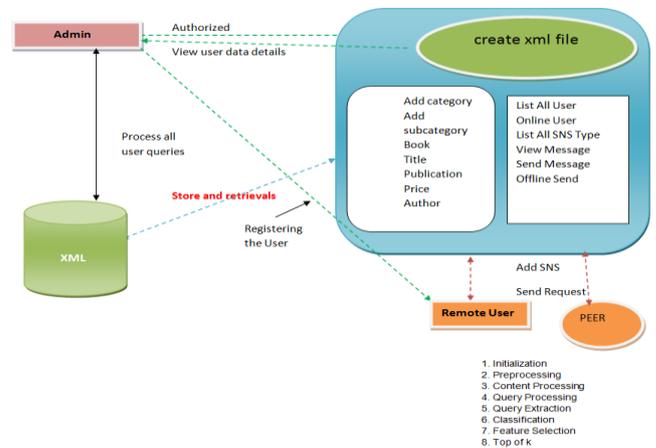


Fig 1 : SYSTEM DESIGN

User:

- 1) It is client.
- 2) User is access query keyword base so xml schema to help best data from server.
- 3) XML having following stage,
  - Initialization
  - Preprocessing
  - Content Processing
  - Query Processing
  - Query Extraction
  - Classification
  - Feature Selection

With help of xml to give follows result,

- Search file
- Search top K
- View details
- download
- ❖ A method of providing diverse keyword query suggestions to users based on the context of the given keywords in the data to be searched. By doing this, users may choose their chosen queries or modify their original queries based on the returned diverse query suggestions.
- ❖ To address the accessible limitations and challenges, we initiate a formal study of the diversification problem in XML keyword search, which can directly figure out the diversified results without retrieving all the relevant candidates.

### 8.ALGORITHMS

1) i-Search Mechanism:

Input: s, r, P, Z(P)

Output: Pnew, Z(Pnew)

```

1 foreach v : v OE G - P do
2 if v = r then
3 Pnew ← P ← {s ← v};
4 Z(Pnew) ← Z(P)F(s, v);
5 return;
6 else if v is online, and Z(P)F(s, v) > D then
7 v.i Search(v, r, P ← {s ← v}, Z(P)F(s, v));
8 else if v is off-line, or Z(P)F(s, v) ← D then
9 quit;
10 end
11 end
    
```

2) Smallest Lowest Common Ancestor :

Input:

```

Q(Z) = { n1,n2, n3, n4..... }
XML-> T(dataset)
m->left subtree
n-> right subtree
s(x,y) -> Score
l(x,y)->length
temp= null
    
```

Output:

Top -K of Q(Z)

```

1. START
2. M(m,n) ->getFeatureTerms(g,G)
3. while( qnew = M !=null ) do
    l(x,y) -> getNodeList(s,T)
    .....x<m<1<y<n
4. Pscore = Count (Mod (x,y) /getNodeSize(x,y,T))
5. temp -> computeSLCA(x,y)
6. if (temp is empty then )
    score(qnew)->Pscore
else
for all result candidates rx -> temp do
for all result candidates ry -> temp do
if( rx == ry ) then
return temp.remove(rx) .....ancestor
else if( rx == ry ) then
return temp.remove(ry) .....descendant
score(qnew)->temp
if( Mod( Q ) < K ) then
score( qnew )-> Q
else if(score(qnew) > score(q'new) ) then
Q.replace(qnew,q'new) return Q
    
```

9.RESULT

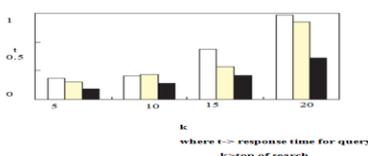


Fig 2: RESULT ANALYSIS

```

Q(Z)->{q1,q2,q3....qn}
    
```

```

T(Z)->{t1,t2,t3,...tn}
K(Z)->{k1,k2,k3...kn}
k->(qn,tn)
    
```

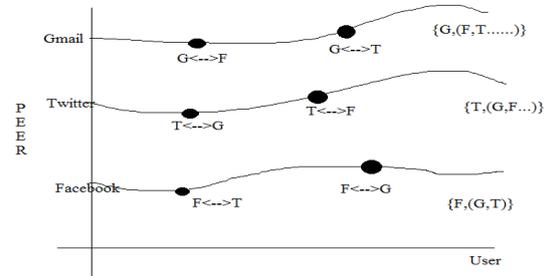


Fig 3: RESULT GRAPH

Our proposed SNS system provides a tremendous advantage that Image and video communication is very important. In this the transfer of video and transfer of Image has taken into consideration.

CONCLUSION

The proposed system shows that how to communicate Homogenous as well as Heterogeneous Schema through XML as well as how to provide diverse keyword query suggestions to users based on the context of the given keywords in the data to be searched. By this, users may choose their ideal queries or modify their original queries based on the returned diverse query suggestions.

The proposed system is designed to be cost effective and very easy to use. Users can easily work with the system almost immediately and the messages sent to one another becomes fast and very rapidly.

REFERENCES

[1] C. Zhang et al., "Privacy and Security for Online Social Networks: Challenges and Opportunities," IEEE Network, vol. 24, no. 4, July/Aug. 2010, pp. 13-18.

[2] <http://newsroom.fb.com/content/default.aspx?NewsAreaId=22>.

[3] A. Mislove et al., "Measurement and Analysis of Online Social Networks," Proc. 7th ACM SIGCOMM Conf. Internet Measurement, 2007, pp. 29-42.

[4] N. Ellison and D. Boyd, "Social Network Sites: Definition, History, and Scholarship," J. Computer-Mediated Communication, vol. 13, no. 1, Oct. 2007, pp. 210-30.

[5] M. N. Ko et al., "Social-Networks Connect Services," Computer, vol. 43, no. 8, Aug. 2010, pp. 37-43.

[6] J. Bae, "A Global Social Graph as a Hybrid Hypergraph," Proc. 5th Int'l. Joint Conf. INC, IMS and IDC, Aug. 2009, pp. 1025-31.

[7] <http://openid.net>.

[8] <https://developers.facebook.com/docs/reference/api/>.

[9] <http://developer.android.com/reference/packages.html>.

[10] L. Katz, "A New Status Index Derived From Sociometric Analysis," *Psychometrika*, vol. 18, no. 1, Mar. 1953, pp. 39–43.

[11] S. B. Ronald, *STRUCTURE: A General Purpose Network Analysis Program Providing Sociometric Indices, Cliques, Structural and Role Equivalence, Density Tables, Contagion, Autonomy, Power and Qquilibria in*

[12] *Multiple Network Systems (Version 4.2)*. New York: Columbia University Press, 1991.

[13] B. Yu and M. P. Singh, "Searching Social Networks," *Proc. 2nd Int'l. Joint Conf. Autonomous Agents and Multiagent Systems (AAMAS '03)*, 2003, pp. 65–72.

[14] N.B. Chang and M. Liu, "Controlled Flooding Search in a Large Network," *IEEE/ACM Trans. Net.*, vol. 15, no. 2, Apr. 2007, pp. 436–49

[15] S. Jiang et al., "LightFlood: Minimizing Redundant Messages and Maximizing Scope of Peer-to-Peer Search," *IEEE Trans. Parallel and Distributed Systems*, vol. 19, no. 5, May 2008, pp. 601–14.

[16] Y. Chen, W. Wang, Z. Liu, and X. Lin, "Keyword search on structured and semi-structured data," in *Proc. SIGMOD Conf.*, 2009, pp. 1005–1010.

[17] L. Guo, F. Shao, C. Botev, and J. Shanmugasundaram, "Xrank: Ranked keyword search over xml documents," in *Proc. SIGMOD Conf.*, 2003, pp. 16–27.

[18] C. Sun, C. Y. Chan, and A. K. Goenka, "Multiway SLCA-based keyword search in xml data," in *Proc. 16th Int. Conf. World Wide Web*, 2007, pp. 1043–1052.

[19] Y. Xu and Y. Papakonstantinou, "Efficient keyword search for smallest lcas in xml databases," in *Proc. SIGMOD Conf.*, 2005, pp. 537–538.

[20] J. Li, C. Liu, R. Zhou, and W. Wang, "Top-k keyword search over probabilistic xml data," in *Proc. IEEE 27th Int. Conf. Data Eng.*, 2011, pp. 673–684.

[21] J. G. Carbonell and J. Goldstein, "The use of MMR, diversity based reranking for reordering documents and producing summaries," in *Proc. SIGIR*, 1998, pp. 335–336.

[22] R. Agrawal, S. Gollapudi, A. Halverson, and S. Jeong, "Diversifying search results," in *Proc. 2nd ACM Int. Conf. Web Search Data Mining*, 2009, pp. 5–14.