Mobile Search – Social Network Search Using Mobile Devices

Pedro Tiago
FCT, Universidade Nova de Lisboa
Caparica, PORTUGAL
ptiago@gmail.com

Niko Kotilainen
University of Jyväskylä
Jyväskylä, FINLAND
niko.kotilainen@jyu.fi

Mikko Vapa
University of Jyväskylä
Jyväskylä, FINLAND
mikko.vapa@jyu.fi

Heikki Kokkinen
Nokia Research Center
Helsinki, FINLAND
heikki.kokkinen@nokia.com

Jukka K. Nurminen
Nokia Research Center
Helsinki, FINLAND
jukka.k.nurminen@nokia.com

***Abstract*- During the last years progress in web search engines has been made to the point that relevant information can be reached easily most of the times. However very little empirical research has been carried to study web search in highly dynamic social network environments composed of mobile devices. The aim of this work was therefore to investigate novel approaches that took advantage of the social network environment inherent to mobile peer-to-peer paradigm. The work focused mainly on the development of a prototype for Mobile Search concept. The prototype was built on top of Drupal content site management system. This study suggests that the methods presented can be a complement to traditional web search engines.**

I. INTRODUCTION

Mobile phones' computational power has been improving approaching the capabilities of general purpose computers. Nowadays it is possible to host a web site on a mobile device. It is also expected that the number of mobile web sites will outnumber the static web servers [17].

Mobile phones possess an extra set of concerns that are not present in normal web servers (e.g. Personalization; Interactivity; Location and context dependence; Dynamicity) [17]. Those concerns can be further expanded by taking into consideration the social network formed by the contacts in the address book. This fact introduces paradigm shifts in relation to the Peer-to-Peer web search paradigm and the traditional centralized search approach.

Recently, there has been a growing interest in how to explore the mobile phone capabilities in the web search context and how to merge them with existing phone functionalities [15,17]. However the research has tended to focus on centralized approaches or Peer-to-Peer web search, rather than on the Peer-to-Peer web search in the social network context. The purpose of this article is to present different strategies that take advantage of the described type of an environment and extend the current web search mechanisms giving the end user new possibilities of exploring information.

In the future it will be common to have a web server running in mobiles devices. This represents a shift in normal web servers' webware. The biggest change is the possibility of users to freely manage their own content without being restricted by third parties. There is a need to categorize content in different ways in order to create new forms of navigation and search.

The content in mobile phones can be divided in two distinct logical groups: dynamic and static. Dynamic content usually is unique and generated by the mobile phone sensors. Static content on the other hand is not context dependent and is generated by the user. Both types of content can be easily replicated. Usually dynamic content can be easily characterized by tags, although static content can be categorized in a similar way. Content is distributed to overlapping data islands. Each user may belong to several data islands simultaneously because each user is connected to users who belong to different interest groups (even unknowingly) [5]. The connections are created based on the address book contacts forming presumably a power law graph [5]. It's assumed that the nearest neighbors of a node have higher probability to own relevant content to that node. In the information searching context it is important to have an ability to search through relevant data and take advantage of the overall network topology.

The article is structured as follows. The motivation behind the need for Mobile Search is presented in section II. Section III continues with the core concerns and major differences between this type of search and traditional centralized web search. Subsequently in section IV a brief description of the prototype is given and the related work within the topic is reviewed in section V. Finally, section VI describes the future work and section VII concludes the paper.

II. Mobile Search

This section describes a system for Mobile Search. The system is based on pure Peer-to-Peer architecture and it offers scalability, efficiency, resilience to failures and privacy at a higher degree than current centralized solutions. [4]

To take advantage of the portrayed scenario a new set of concepts were introduced. One is how to navigate through the data in a social network. Social network's connections are determined from an address book of a mobile device. Users search one graph level of their social network at a time usually starting from their neighbors. However, users may also start a query anywhere in the social network. Every time a user issues a search query the mobile device forwards it to all its neighbors. The neighbors answer back by returning a result set and a list of their neighbors. If the user who issued the query is not satisfied by the results he can always ask new results from the next level neighbors as long as there are non-visited nodes in the network. This concept was named *manual multi-hopping*. In manual multi-hopping the user needs to select which of the non-visited nodes will be used for querying the next level. Manual multi-hopping can be extended to *automatic multi-hopping* if an algorithm is used to sort which of the non-visited nodes to query further thus avoiding the need for user decision. One example of such algorithm is only to forward a query to neighbors of the nodes that previously returned results to that query. Automatically sorting the non-visited nodes leads to tradeoff between search accuracy and easiness of searching suggesting that both manual and automatic multi-hopping should be available for the user.

Another way of navigating is by searching neighbor content tags and getting the result set composed by the content links with the tags and the list of next level neighbors. Tags work as links between content categorized similarly. At each hop the user gets the list of contents tagged in a similar way by nodes in its neighborhood.

The Mobile Search system can be divided to two logical parts: local web search engine and meta crawling. Local web search engine is a search service, which manages the search index of a mobile device. Meta crawling term refers to a search service, which uses other local web search engines for getting the results and then combines different result sets into one. The part responsible for the meta crawler gets it's results from direct neighbors. The way the results are presented can always be changed thus the mobile device bears the load of processing the returned references. Any specific method to sort out the references in certain order can be employed. For example more relevance can be given to results from a certain source so they appear first in the result list. There is also the possibility to merge different types of mobile phone data with different type of content. For example user A may search for user B's meetings and after getting the results he may merge the results with his own agenda and display the meeting locations on a map.

The local web search engine gives a user the power to tailor the search results to his/her own needs. The search index can be updated every time the content changes. The user may allow certain information to be only searched by a specific group of users or to influence certain query results in a certain context. This feature allows users to create groups of trust. They can decide which information source is more relevant to them in different contexts. Also the level of privacy and who to trust is determined by each node following the motto: "I only display what I want to who I want".

III. Comparison

It may be pointed out that centralized solutions have a single point of failure, load balance and trust issues and may censor certain entities [11]. Although nowadays they have grown incredibly robust. One main advantage of Mobile Search is the total independence of the nodes. The system can operate without any central server and system load is fully distributed. Each node is responsible for processing the queries and search requests.

For example Google presents in its back end a highly scalable architecture [3] but it cannot address the premise that our friends are more likely to have interesting results to us and may not even be connected or linked to our content [8]. In this scenario the hyperlink concept is expanded by the network of connections formed by the mobile phone's address book. These types of links enable the blend of several groups of interest along the network. In several situations the link web structure of documents doesn't portray possible relations between people [10].

The search space indexed by centralized solutions is limited because central servers have limited crawling capacity. Index of a centralized solution can thus be characterized as one large result set. Also, crawling cannot easily find content without external references. In contrast, decentralized social network search consists of multiple small result sets, does not have indexing limitations and does not need external links to point out the content. Non-referenced content can be found by finding a neighbor of the owner of non-referenced content. Thus decentralized search potentially provides more results than centralized solutions when user continues navigating the social network further. However, queries executed in immediate surroundings of the querying node usually result in fewer and more accurate results than centralized solutions.

Web search engines do not allow tailoring results to individual needs. For example user A only wants to display a specific result list to a certain query from user B. Centralized solutions provide an efficient way of finding popular content but lack the ability to find more personal/social proximity content [8]. This situation is evident in a corporate setting where many documents are not available to the outside world. Other type of personal/social proximity content that is not indexed by web search engines is mobile phone data. One example is searching for a phone number or meeting information that is available in one of our neighbors. This capability avoids the use of third entities (e.g. number services, central servers) and enhances the information availability. In

the other hand Mobile Search due to the topic oriented network nature is not suited to find popular content. Conversely, it's a powerful mechanism in restricted topic set environment [8].

One major issue of Mobile Search in relation to the centralized approach is the quality of the results returned. Different sites may have different criteria to classify and rank information. This poses a problem how to merge the different results sets returned for a query [12]. In the other hand, this can highly increase the quality of the results in some scenarios. For example in a work context user A can give more weight to Document X in searches made by users from the workgroup because that document is more relevant to them.

Other issue is the high number of neighbors and free riding. Those factors are a risk to network traffic. They can be overcome first by limiting the search query to a pre-selected group of users and second by only returning back neighbors who have a higher probability of having meaningful content.

Centralized solutions update their index when content is crawled whereas in Mobile Search the owner can index the searchable content whenever he/she desires. This leads to up-to-date result sets without any increase in network traffic. And as long as the user sets the permissions for different content, other users authorized to view that content can find it without knowing the exact location. With centralized solutions everyone has to trust a single entity allowing possibilities for censorship or pressure from external entities.

Concern	Centralized solutions	Mobile Search
Load	*centralized/single point of failure*	*highly distributed*
Trust	*censorship/pressure from external entities*	*highly distributed*
Search space	*billions (single set)*	*hundreds to billions (multiple different sets)*
Index update	*days to months*	*every second*
Content type	*popular*	*personal/social proximity*

IV. Drupal Prototype

Drupal was used as a test platform for Mobile Search. Drupal is an open-source content management system. It allows managing and publishing several types of content. The meta crawler described in section II was built as a weakly coupled component on top of Drupal local web search engine. This component allows automatic multi-hopping and result interleaving.

The current implementation is single threaded because Mobile Apache doesn't support multiple threads [15,16]. Drupal tac_lite module and Drupal module were also used as fundamental elements in the prototype. These modules allow setting content access rules and to process user authentication in distributed fashion without any central servers.

An extra component that allows to do queries to local mobile phone content such as location, address book and meeting data was implemented. This feature was built as a proof of concept. However, the prototype is also able to gather search results from unmodified Drupal web sites.

One drawback during the elaboration of the prototype is related to the single-threaded nature of the meta crawler. This can have a negative impact on response time because site crawling is done in a serial way. A multi-thread implementation would speed up the system considerably.

V. Related Work

The concept of Peer-to-Peer web search has been harnessed before in the literature. Different approaches [2,8,13,14,20] have been tried before. Although these studies tended to focus on Peer-to-Peer web search, rather less attention has been paid to how to take advantage in this scenario of mobile sites' concerns and integration in the social network context.

Mislove et al. [8] studied how to integrate social network search with web search in order to complement search results. Also, how content publishing and locating influence the overall searching experience in the web perspective and in the social network context is discussed. Supported by the experiment made with PeerSpective prototype, [8] points out flaws in the traditional hyperlinked search like the difficulty of web search engines to index content not well linked to the general web or that is not publicly available. Similar to Mobile Search, [8] presents the idea that social networks, due to data islands formed by user communities, can lead to more timely and efficient searching experience.

Like in our work, [8] gives special importance to social network links but leaves as an open topic how the underlying social network links are formed. In Mobile Search social network structure is automatically defined by the mobile phone address book contacts and can be enhanced by linking content neighbor tags every time a search is performed. Ultimately, the Mobile Search presents the possibility of creating a virtual multi-level content social network. The mechanisms described in [8] could also be adapted and incorporated into Mobile Search.

Bawa et al. [2] introduce YouSearch, which allows searching dynamically changing content from personal web servers. YouSearch differs from Mobile Search approach by having a centralized server (registrar) for storing bloom filters of indexed keywords. This introduces a need to update bloom filters periodically to accommodate changes in content of the peers. Mobile Search is designed for mobile devices with a limited battery and therefore periodically occurring updates needs to be avoided. According to the calculations in [2] one registrar could serve approximately 10000 peers with a 1,5 Mbps network connection. In Mobile Search such an entity is not needed, because all functionalities are decentralized. YouSearch uses caching for storing search results on a querying peer to avoid re-executing a similar query later. This is a feature which could also be applied in Mobile Search.

Finally, YouSearch does not take into account social network connections and therefore searching needs to be explicitly directed to different groups or to specific registrar. This reduces the flexibility of searching.

Zhou et al. [20] states that the evaluation of resources by human users is more important for search quality than the traditional machine based approach. They present a novel page ranking algorithm - Peer-Rank. In this paper a simpler version to rank remote results is presented. First of all, in the problem context described in this study it's assumed that the content on the mobile phone can be divided in two sub-types: dynamic/unique (photos taken with mobile phone camera) and static/common (music files). It will be rare to have different sites returning the same content. Secondly, it's also supposed that the majority of the content will be dynamic/unique due to the nature of mobile phone. Furthermore, each mobile site can employ its own human/machine based methods to rank results. With these details in mind two ways of ranking the results are proposed: Explicit (Tagging content) and Implicit (Machine based methods).

Galanx [14] focuses on query forwarding in Peer-to-Peer web search context. Traditionally Peer-to-Peer web search studies try to "emulate" the behavior of centralized solutions. Those approaches are completely orthogonal to the one presented in this paper. One of the main concepts derived from the social network environment is the ability to navigate through neighbor sites and explore them like in a common social network site where users are able to follow friends' links and explore them. In this case links are created based on the search results. If users are not satisfied with the results they can always jump to the next set of nodes and continue searching. In the Galanx case, like in a centralized web search, only a set of results is provided and the users are unable to explore the network by themselves. The sites are presented as fully separated entities, although they can have hyperlinks between them allowing partial network navigation.

The query forwarding mechanism of Mobile Search can be described as a directed breadth first search with manual iterative deepening. The algorithm is similar to the one described in [19] and [6] with the exception of using manual iterative deepening. A search is only continued if the user is not satisfied with the results.

Other major source of inspiration was the social network tagging system. Similarly the same principle was applied to the system with minor modifications. Users are able to tag content freely. Some predefined tags related with mobile phone concerns will be always available (e.g., photo location). Generally user tags have only a local significance in the network [9]. The predefined tags try to create general tags present all over the network enhancing the navigation. Each time a user in a site can search for neighbor tags and navigate through them like in the normal web search presented in this report.

VI. Future Work

The concept of Mobile Search can be easily expanded and integrated as an extension to existing systems.

Query forwarding algorithms should be considered in order to minimize several problems like free riding [1] though in a different setting than previous studies. Algorithms like Ant search [18], K-Random walk, Expanding Ring and hybrid approaches should be considered.

Other way of extending the Mobile Search functionalities is by creating different ways of accessing the same content. Information could be accessed by a search result or by different entry point. An entry point is a link to a specific

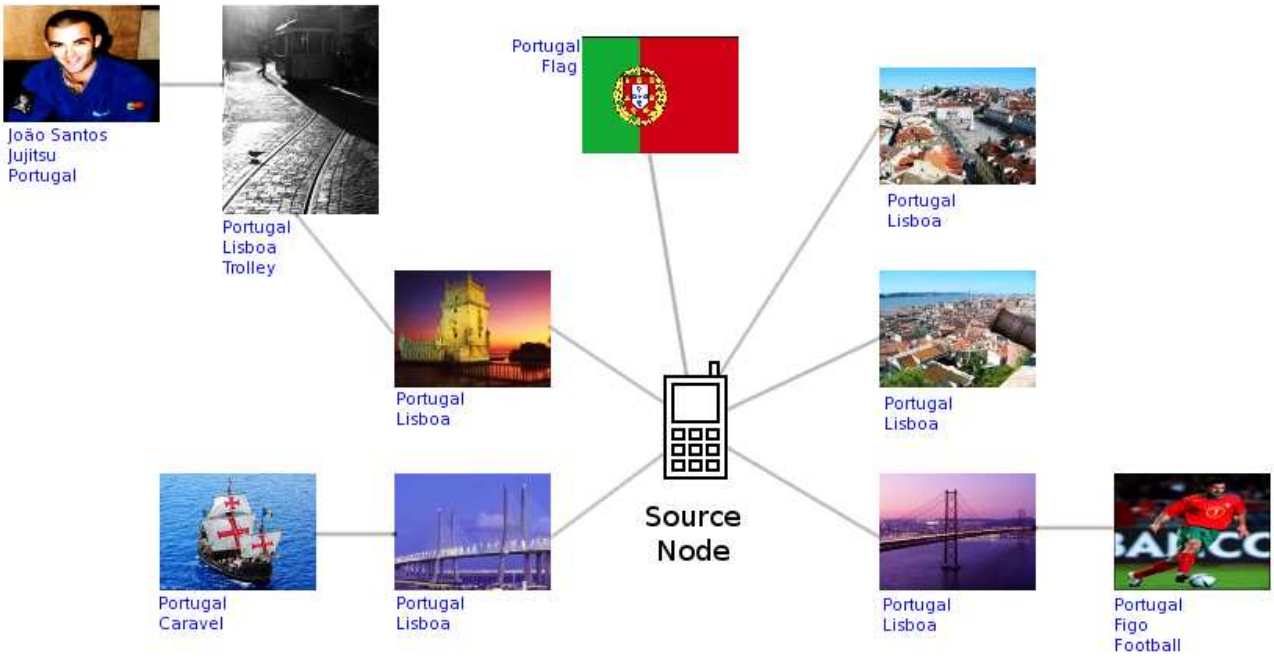

Figure 1. Tagging Concept

content. Tags are an example of creation of different entry points. A different way of creating an entry point is by merging different types of data.

Figure 1 shows an example of the tagging concept. The mobile phone represents the source node who issued a query searching for the tag Portugal. The figure represents the results returned by the neighbor nodes in different network levels (each image corresponds to a neighbor who returned a result).

For example if the source node issued the query Portugal it would obtain six results. If then the user chooses to navigate by the tag Lisboa he would get one result (the trolley image). If instead the user chooses the keyword Portugal he would get three results (the trolley, the caravel and Figo).

Mobile Search enables the creation of multi-social network fusion. With the Mobile Search the user doesn't need to know exactly where the different entry points are. The returned results will allow exploring vicinities following the links of the different tags or by asking for new results. The same user may present in its own site several data related to it's own interests. Certain data may only be available to a specific group of users. The data also may be presented in different ways for different groups. These features could be particularly valuable in an enterprise setting. One example would be a fully distributed enterprise portal [10] using the technology described in this paper.

Other feature worth exploring is adaptive ranking. Historical behavior of users who conducted similar searches or may have a similar role in an organization may be used to boost document rating. This concept may be expanded if more data is available by creating a profile to generate suggestions for documents based on user context and role in that particular social network [10].

All those features can be tweaked at different granularity for different group of users that access the system. For example a user may only generate profiles of work mates in order to make suggestions.

Other topic of interest is the usability of search results, and new paradigms of displaying different types of information and user interaction. Current Web2.0 may not be fully suitable for mobile device paradigm of interaction. This could also be an excellent opportunity to use a query language applied to this type of systems for example an adaptation of webSQL [7]. This would likely create a bigger interoperability and homogenization in this type of systems with easier deployment of new functionalities.

VII. Conclusions

Mobile Search complements traditional web search engines. It gives the user means to explore the neighbors' contents by traveling to the friends network topology. It covers a multitude of environments not covered by the centralized solutions.

One of the main advantages in relation to current centralized social network sites is the possibility to manage the site without interference from an external entity. Currently in a normal social network site a user can only display or use modules made available by a third entity. Due to this characteristic it is possible to merge different network sites that cover different topics and create a social network "melting pot". Each user can have what type of content he/she wishes in the site and display different content for different users.

This type of system is better suited for mobile devices due to the "always on" characteristic [18]. Content can be always updated on spot.

Mobile Search has an enormous potential to evolve and become a major tool in knowledge management technology. Adaptive Ranking, Role-based Recommendations, Locating Experts and Communities [10] can be taken to extreme. To sum up Mobile Search can be used to enhance the ability to search for critical information.

References

[1] E. Adar and B. Huberman. Free riding on gnutella. Technical report, Xerox PARC, 2000.

[2] M. Bawa, R. J. Bayardo Jr., S. Rajagopalan, and E. J. Shekita. Make it Fresh, Make it Quick – Searching a Network of Personal Webservers. Proceedings of the 12th International Conference on World Wide Web, ACM Press, pages 577-586, 2003.

[3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems, 30(1–7):107–117, 1998.

[4] D. Choon-Hoong, S. Nutanong, and R. Buyya, Peer-to-Peer Networks for Content Sharing, Peer-to-Peer Computing: Evolution of a Disruptive Technology, pages 28-65, Idea Group Inc., Hershey, PA, USA, 2005.

[5] E. F. Churchill and C. A. Halverson. Social networks and social networking. IEEE Internet Computing, 9(5):14–19, 2005.

[6] X. Li and J. Wu. Searching techniques in peer-to-peer networks. Handbook of Theoretical and Algorithmic Aspects of Ad Hoc, Sensor, and Peer-to-Peer Networks. Auerbach Publications, pages 613 – 642, 2006.

[7] A. O. Mendelzon, G. A. Mihaila, and T. Milo. Querying the world wide web. Int. J. on Digital Libraries, 1(1):54–67, 1997.

[8] A. Mislove, K. P. Gummadi, and P. Druschel. Exploiting social networks for internet search. In Proceedings of the 5th Workshop on Hot Topics in Networks, Irvine, CA, 2006.

[9] J. C. Paolillo and S. Penumarthy. The social structure of tagging internet video on del.icio.us. In Proceedings of the 40th Annual Hawaii International Conference on System Sciences, page 85, Washington, DC, USA, 2007. IEEE Computer Society.

[10] P. Raghavan. Social networks from the web to enterprise. IEEE Internet Computing, 6(1), 2002.

[11] P. Reynolds and A. Vahdat. Efficient peer-to-peer keyword searching. In Proceedings of International Middleware Conference, pages 21–40, June 2003.

[12] E. Selberg and O. Etzioni. The metacrawler architecture for resource aggregation on the web. IEEE Expert, 12(1):11-14, 1997.

[13] H. Tomiyasu, T. Maekawa, T. Hara, and S. Nishio. Social network applications using cellular phones with e-mail function. In Proceedings of 21st International Conference on Data Engineering. IEEE Computer Society, 2005.

[14] Y. Wang, L. Galanis, and D. J. DeWitt. Galanx: An efficient peer-to-peer search engine. Technical report, University of Wiscosin - Madison, 2003.

[15] J. Wikman. Mobile web server - eurooscon presentation, 2006.

[16] J. Wikman and Ferenc Dosa. Providing http access to web servers running on mobile phones. Technical report, Nokia Research Center, May 2006.

[17] J. Wikman, Ferenc Dosa, and Mikko Tarkiainen. Personal website on a mobile phone. Technical report, Nokia Research Center, 2006.

[18] C.-J. Wu, K.-H. Yang, and J.-M. Ho. Antsearch: An ant search algorithm in unstructured peer-to-peer networks. In Proceedings of the 11th IEEE Symposium on Computers and Communications, pages 429–434, Washington, DC, USA, 2006. IEEE Computer Society.

[19] B. Yang and H. Garcia-Molina. Improving search in peer-to-peer networks. In Proceedings of 22nd International Conference on Distributed Computing Systems, Vienna, Austria, 2002.

[20] J. Zhou, K. Li, and L. Tang. Towards a fully distributed p2p web search engine. In Distributed Computing Systems, 2004. FTDCS 2004. Proceedings. 10th IEEE International Workshop on Future Trends of, pages 332–338. IEEE Computer Society, 2004.