Retaining Personal Expression for Social Search

Praphul Chandra
HP Labs,
Bangalore, India.
praphul.chandra@hp.com

Ajay Gupta
HP
Bangalore, India
ajay.gupta@hp.com

ABSTRACT
Web is being extensively used for personal expression, which includes ratings, reviews, recommendations, blogs. This user created content, e.g. book review on Amazon.com, becomes the property of the website, and the user often does not have easy access to it. In some cases, user's feedback may get averaged with feedback from other users e.g. ratings of a video. We argue that the creator of such content needs to be able to retain (a link to) her created content. We introduce the concept of MEB which is a user controlled store of such retained links. A MEB allows a user to access/share all the reviews she has given on different websites. With this capability users can allow their friends to search through their feedback. Searching through one's social network allows harnessing the power of social networks where known relationships provide the context & trust necessary to interpret feedback.

Categories and Subject Descriptors
H.3.4 [**Systems & Software**]: H.4.3 [**Communications Applications**];

General Terms
Algorithms, Management, Human Factors.

Keywords
Social search, user created content, web2.0, social networks.

1. INTRODUCTION
User created content, in the form of Star ratings, Tags, user Recommendations, user Attention and Text reviews/comments (henceforth, STRAT) is increasingly large & important part of the web [1,2]. Many sites now enable users to create such content and use this to rank content, do collaborative filtering. Such data is increasingly a critical success factor in the web2.0 business model. However, the user who creates this content often does not control or benefit from it. The lack of user's retaining & controlling their content means that if for e.g. my friends want to know which businesses I have positively reviewed on the web, there is no easy way for them to find this information. Relying on the experience of one's social network is critical human behavior since trustworthiness of the advisor is a critical factor. People often use the knowledge from their social networks to assess, interpret & trust comments from others.

2. PROBLEM STATEMENT
We identify the following challenges in capturing and utilizing STRAT content: (1) STRAT content is captured only on web-sites which are designed to accept such content e.g. Amazon.com - in other cases, such content is not captured even if the user is willing

to provide it. (2) When STRAT content is captured by the destination web site, it is owned by the web site and not the user. This means that the users do not control the content they create thus raising privacy issues. (3) When STRAT content is captured by the destination web site, it is often statistically aggregated by the destination site and not stored as a uniquely addressable entity. Furthermore, the destination site often does not maintain the 'link' between the individual STRAT content and the user who provided it. (4) STRAT content rarely contains hyperlinks - one consequence of this is that search engines which are strongly dependent on the hyperlink structure are not influenced by this form of user created content.

Our end-goal is to enable users to be control the content they create and support social queries like 'what did P say about E', or 'which books did my friends review favorably' [1,2].

3. EXISTING APPROACHES
Some web sites enable users to own the content they create for e.g. Amazon.com allows users to view all reviews she has given, YouTube allows users to create personal playlists & favorites lists. Social bookmarking sites like digg.com allow users to create & share collections of bookmarks. Review specific sites (Epinions, Zagat) allow users to give feedback on various products. Social networking sites enable users to share their reviews with their social circle. However none of these approaches enable the two important features of our proposed solution: (a) user's control access to their content and (b) opinionated queries over social network.

4. APPROACH
4.1 Overview
Our solution to this problem is to enable users to retain with them a "link" to the content they create. This user-retained link provides a foundation to new ways of navigating user created content - both by the creator and her social network.

In subsequent sections we describe the two key innovations of our solution (a) a new kind of link on the web which goes from a user to the content they have created thus allowing users to retain links to the content they create on different websites & (b) techniques for users to search their social networks using these links.

4.2 Retaining Personal Expression
We introduce a concept called MEB (ME-on-the-weB) that connects the creator with her created content and its subject (what is this content about). User created content will be one of STRAT type and the subject of the content will typically be a pointer to the website on which the user is giving her comment (typically a URL). Thus, a MEB conceptually is a tuple of the form *{user, subject_url, user_created_ STRAT} e.g. {John@gmail, Amazon.com/HarryPotter, "This is a good book"}.*
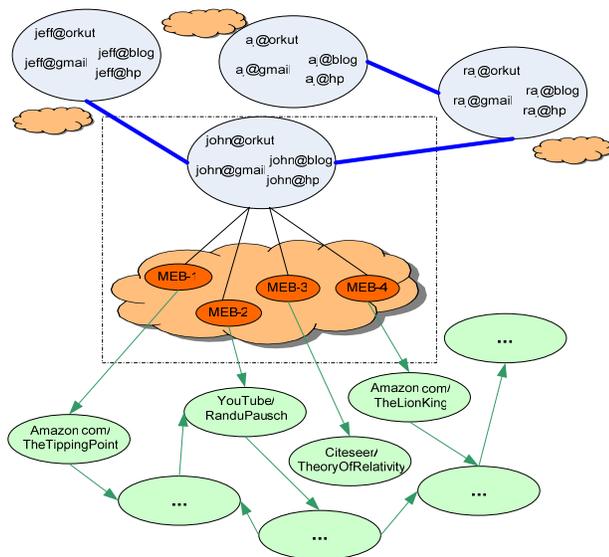
In figure 1, we show a collection of mebs for a user John. John has multiple web-identities (on different sites & communication mechanisms). John's social network will know him through such ids. Bold lines in the figure indicate these social relationships. The MEBs contain hyperlinks to the subject_URLs which they are about.

The information shown in the dotted box includes the multiple user identities of John and his MEB store. This information is owned by the user. Other users also have their MEB clouds. Retaining the link between the user and his content allows us to enable queries which take into account the opinion of my social network.

Conceptually, a MEB gets created as a by-product of a web interaction provided the user chooses to do so. One way of doing this is by using a browser agent to send notifications to a central repository server when posting new content as suggested in [2]. We have implemented MEB creation using browser plug-ins. A user may choose to store her MEBs at the entity-site (provided the site supports this) and/or publish it on one or more of her web-presences (blogs, facebook-page etc.). We believe that user-centricity for MEB is important for two reasons: (a) MEBs are created at different websites but must be stored in a user-centered fashion and (b) user must have absolute control over who can access their opinions captured in MEBs.

4.3 Searching Personal Expression on Social Networks

We enable opinion queries of the form, given a set of people V, and a user query σ, find V's STRAT content on σ. For e.g. assuming that John has access to Jeff's MEB cloud, John can query 'what has Jeff said about the book, Harry Potter and the Sorcerer Stone'.

The set of people specified in V should be communication-ids e.g. 'john@gmail' or id on a social networking site like Facebook etc. The querier may specify these ids explicitly (john@gmail) or

implicitly ('my social network'). The specified communication ids then define how the query is routed to the target meb-cloud. In a peer-to-peer implementation of this search, V will be an email-id. When the user types in a query (specifies V & σ), an email is sent to the email-ids specified in V. If the query is allowed as per access-policies, the receiving client triggers a search operation over the user's MEB clouds based on σ. Finally, the handler collects the results of the search operation and returns them via the same communication channel on which the request was received (email in this case). At the queried end, the results may be presented in an appropriate interface. We note the potential delay incurred as a function of the communication channel used. As another alternative implementation, V contains a social networking page URL. In this case, the query will be routed to an application installed on the user's social networking page (e.g. Facebook Apps). This application will then carry out the handler functions described above.

Further, we envision combining MEB search results with web search results. One approach may be to modify the link analysis algorithms on graphs for e.g. Hubs & Authorities [3] to include the authority conferred to a web page by a MEB-page. We can transfer the influence of a MEB page based on its creator - the notion here is that the meb pages may have a very low link-based score but the querier may value highly a friend's opinion. Formally, for each page p in (web+meb) search results, we calculate the following iteratively (a) authority score, $x^{<p>}$ = weighted sum of $y^{<p>}$ values of pages pointing to p:

$$x^{<p>} = \sum_{q:(q,p)\in E} w^{<q>} y^{<q>}$$ and (b) hub score, $y^{<p>}$ = weighted

sum of $x^{<p>}$ values of pages pointing to p:

$$y^{<p>} = \sum_{q:(p,q)\in E} z^{<q>} y^{<q>}$$. What we need now is to determine

the social weight matrix W (Z will be the transpose of W).

5. CURRENT STATUS
We have working prototypes to demonstrate MEB creation, sstorage & search over social networks. We are in the process of implementing using mebs to improve web search.

6. ACKNOWLEDGEMENTS

7. REFERENCES
[1] HA Kautz, B Selman, MA Shah. The Hidden Web. AI Magazine, 1997.

[2] Raghu Ramakrishnan, Andrew Tomkins, Towards a People Web. Computer, Vol. 40, No. 8. (2007)

[3] Kleinberg, J, Authoritative Sources in a Hyperlinked Environment. - Journal of the ACM, 1999