

iCollaborate : Harvesting Value from Enterprise Web Usage

Ajinkya Kale[†], Thomas Burris[¶], Bhavesh Shah[†], T L Prasanna Venkatesan[†],
Lakshmanan Velusamy[†], Manish Gupta[§], and Melania Degerattu[‡]

[†]India Software Lab, IBM India Pvt. Ltd., [¶]CIO Innovation Initiatives, IBM India Pvt. Ltd.,

[§]India Research Lab, IBM India Pvt. Ltd., [‡]GWWE, CIO Office, IBM Corp.

{ajinkyakale, thomburris, bhaveshshah, prasannav, lakshmanan.v,
gmanish}@in.ibm.com, mdegera@us.ibm.com

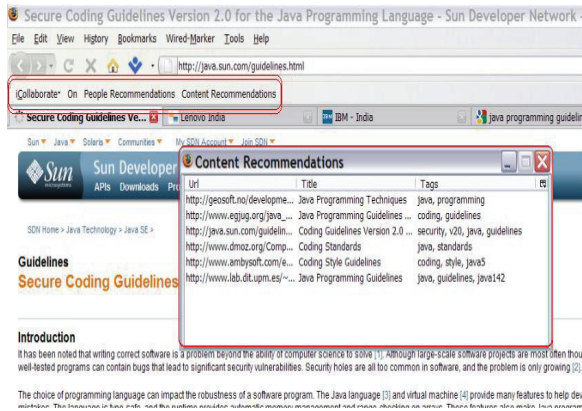


Figure 1: A screenshot of iCollaborate in action

Categories and Subject Descriptors : H.3.3 [Information storage and retrieval]Information filtering

General Terms : Algorithms

Keywords : Enterprise Social Data, Social browsing

1. INTRODUCTION

We are in a phase of ‘Participatory Web’ in which users ‘add value’ to the information on the web by publishing, tagging and sharing. The Participatory Web has enormous potential for an enterprise because unlike the users of the internet an enterprise is a community that shares common goals, assumptions, vocabulary and interest and has reliable user identification and mutual trust along with a central governance and incentives to collaborate. Everyday, the employees of an organization locate content relevant to their work on the web. Finding this information takes time, expertise and creativity, which costs an organization money. That is, the web pages employees find are knowledge assets owned by the enterprise. This investment in web-based knowledge assets is lost every time the enterprise fails to capture and reuse them. iCollaborate is tool to capture user’s web interaction, persist and analyze it, and feed that interaction back into the community - the enterprise.

Pain points : Current web applications that explicitly monitor web browsing activity, such as eyebrowse[2], are limited by a number of factors. First, they exist out on the

public web, where user identity, trust, and governance are issues. Second, the extent to which they are aimed at providing recommended content and views into browsing trends are challenged by the lack of tightly organized, closed communities. Other utilities like Delicious, StumbleUpon usually need an explicit user actions to mark the page on the user profile and to rate the page.

Crowd Intelligence in an Enterprise : The ideal community for web data mining and services is an enterprise. In an enterprise, user identity and trust can be assured. The sub-net of web pages that employees use is, by definition, focused on the mission and business of the enterprise. These user groups can be utilized to enhance the web browsing knowledge of the enterprise as a whole.

Solution outline : iCollaborate specifically targets a community of enterprise employees. We track user’s web interaction via a browser plugin using a page usage score algorithm to determine the extent to which an employee actually used the page. This data, along with data derived from the content of the pages themselves, is used to create clusters of ‘similar’ users and content. We map users into two different cluster groups based on their long term interest ie browsing history and their short term interests ie their recent web browsing sessions. These clusters are used to provide real-time people and content recommendations. Using a combination of Locality Sensitive Hashing[1], Min-hash and Co-Visitation techniques we bind content-based and collaborative filtering for common interest clustering of users and content. These services derived from this analysis is a direct function of the coherence of the community from which it is derived. To an enterprise, seeking competitive advantage by maximizing the use of the information at its disposal, a system like iCollaborate can provide much value.

2. REFERENCES

- [1] P. Indyk, R. Motwani. Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality. In *Proc. of the 30th Annual ACM Symposium on Theory of Computing*, 1998, pp. 604-613.
- [2] B. Moore, M. V. Kleek, D. Karger, *eyebrowse* : <http://eyebrowse.csail.mit.edu/>