

T@gz : Intuitive and Effortless Categorization and Sharing of Email Conversations

Parag Joshi
HP Laboratories
1501 Page Mill Rd
Palo Alto, CA 94304
parag.joshi@hp.com

Claudio Bartolini
HP Laboratories
1501 Page Mill Rd
Palo Alto, CA 94304
claudio.bartolini@hp.com

Sven Graupner
HP Laboratories
1501 Page Mill Rd
Palo Alto, CA 94304
sven.graupner@hp.com

ABSTRACT

In this paper, we describe T@gz, a system designed for effortless and instantaneous sharing of enterprise knowledge through routine email communications and powerful harvesting of such enterprise information using text analytics techniques.

T@gz is a system that enables dynamic, non-intrusive and effortless sharing of information within an enterprise and automatically harvests knowledge from such daily interactions. It also allows enterprise knowledge workers to easily subscribe to new information. It enables self organization of information in conversations while it carefully avoids requiring users to substantially change their usual work-flow of exchanging emails.

Incorporating this system in an enterprise improves productivity by:

- discovery of connections among employees with converging interests and expertise, similar to social networks naturally leading to formation of interest groups,
- avoiding the problem of information lost in mountains of emails,
- creating expert profiles by mapping areas of expertise or interests to organizational map.

Harvested information includes folksonomy appropriate to an organization, tagged and organized conversations and expertise map.

Categories and Subject Descriptors

H.4.3 [Information Systems Applications]: Communications Applications; H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces

General Terms

Design, Experimentation, Human Factors

Keywords

Email analysis, social network, folksonomy, collaboration, information extraction, groupware, social tagging

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

1. INTRODUCTION

Huge amount of enterprise knowledge remains locked in email communications. Not all of this needs to be private to the individual although it can remain confidential within the organization. A great deal of information in emails concerns the team or the organization as opposed to the individual only. Due to private nature of all email communication regardless of actual level of confidentiality, the rest of the organization cannot benefit from the information captured in previous emails. Often, individuals also find it hard to locate the information which they know exists in their own inbox because it is not properly categorized. Moreover, it is not possible for them to access information captured in the email box of their colleagues. Problems are compounded when people have to go through re-organizations or leave the company and some knowledge is lost with them. There is a tremendous business opportunity in allowing organizations to tap into this vast and extremely valuable information resource.

T@gz enables a novel and dynamic publish and subscribe model for enterprise knowledge primarily through emails as opposed to other collaborative means such as, wiki or similar document sharing systems. We have observed that despite prevalence of wikis and sharepoint sites, these tools are not as effortless as using emails and so most knowledge is exchanged by email interactions. Users generally have email client open all day and it is part of their workflow. This also contrasts with traditional mailing lists which require special effort to set up, configure and manage subscriptions and generally they are hard to discover and it is harder for users to subscribe and unsubscribe. By contrast, T@gz works like a dynamically generated mailing list that is very easy to use.

Important contributions of T@gz are: 1) easy sharing of enterprise conversations using web and email, 2) effortless web 2.0 style tagging and categorization of emails, 3) leveraging web technologies such as, RSS and web folksonomies for enhancing email experience.

2. RELATED WORK

Knowledge workers in a modern enterprise have a myriad of social networking or knowledge management tools for sharing project information, documents, etc. Many companies spend great deal of time and money on installing tools such as, MS Sharepoint, various flavors of Wiki, etc. This mind boggling array of new tools and their unique peculiarities in configuration and usage puts great deal of strain on workers. Learning curve of these tools adds to resistance compared to simply using email. Many times workers can-

not agree on a specific tool and so settle on email as lowest common denominator. These have been cited [1] [2] as main causes for low adoption of such Web 2.0 tools in enterprise. Regardless of standardization of any tool, these systems have their own logins and they require users to initiate a browser session and in general these tools are not as intuitive and effortless as using email with an email client that generally remains open and where they are already authenticated. Researchers have discussed [3] many other existing approaches and their shortcomings over email.

Some of these Web 2.0 style tools also involve asking users to create their profile and ask them to list their areas of expertise, and the projects they are working on, etc. However, unless such practices are mandated and policed, busy workers are reluctant to find time to create and keep the profile current. By contrast, T@gz automatically tags people and conversations from daily email interactions.

Although there have been problems in adoption of Web 2.0 tools in enterprise, it has been seen that Web 2.0 style tagging is valuable. And there are great benefits from tagging documents, tagging people, etc. in an enterprise context if it can be done in an effortless manner.

T@gz system is much more intuitive and effortless compared to existing enterprise knowledge sharing systems such as, wikis and MS Sharepoint. Main advantages over wikis and document sharing systems are that with T@gz a user does not have to authenticate to a separate system and can effortlessly share information using emails as part of daily workflow. Also, wikis and document sharing systems do not support effortless tagging as supported by T@gz system.

In the literature, mailing lists and other email based approaches have been described to overcome these shortcomings. For example, Nelson, et al. describe an email based tagging system called Mail2Tag [3]. Here the approach is to use an email alias specific to each tag.

Following lists advantages of our system over Mail2Tag and other email distribution or mailing list systems:

- Mailing lists can allow formation of interest groups similar to T@gz system. However, it takes extra effort to have an administrator create a new mailing list and manage subscriptions. It is hard for employees to discover existing mailing lists and subscribe to them. T@gz system makes it completely effortless where putting # in front of a word anywhere in subject or body of email would create a mailing list.
- By cleverly leveraging web standard RSS and creating topic specific RSS feeds, we have created a functionality similar to mailing lists that is much more intuitive to use than mailing lists.
- Major advantage of our system is in-line indication of tags in email by using #, as opposed to having to type a long email address in mail2tag. Similar approach was initially pioneered in IRC (Internet Relay Chat) chat systems and later has been popularized by Twitter for micro messages on the web. In mail2Tag system, the users have to send an email to tag-specific emails (e.g. healthcare@share.company.com). Users might think of appropriate tags as they are typing words in the body or subject of email and it takes extra cognitive effort and context switching to click the list of recipients and add a long tag-specific email address

(e.g. healthcare@share.company.com) or similar mailing list alias whereas, in the system we have invented it is much more effortless to just type several such tags like #healthcare as part of email body or subject (e.g., Our technology solutions for #healthcare and #pharmaceutical industry). So, with this T@gz system, it only takes typing one extra letter as opposed to much more tedious process of switching context while composing email and typing a long email address.

- Intuitive options for allowing intuitively hierarchical and multi-word tags (as described in the Scenario section)
- Since with our system, multiple tags can appear in the same sentence or the same paragraph or the same email, semantic similarity and contextual information among tags is not lost for further data mining. This is impossible in mail2tag system as all the tags have to appear as recipients and the tag co-occurrence or contextual information in sentences or paragraphs is completely lost.

3. APPROACH

As shown in Figure 1, the system architecture comprises an Email Server (it can use any existing email server used by the enterprise), an email address automatically monitored by our system (e.g. an enterprise or group wide generic sharing email like tagz@company.com, in our case tagz@hpl.hp.com), Email Analysis System and Web UI.

Core Email Analysis system has been built using Ruby and PostgreSQL for meta-data management. Web interface has been developed using REST and RSS interfaces exposed from RubyOnRails and Java.

3.1 Scenario

- User composes an email to other recipients. To allow the email to be automatically categorized and indexed, the user can involve T@gz in conversation by including tagz@company.com in 'To' or 'CC' fields along with other recipients. This would also mark the email for sharing among users within the enterprise or group. Optionally, an email client plug-in can retrieve tag suggestions from Tag Suggester to help the user in tagging.
- The T@gz web UI presents an organized view of enterprise information automatically harvested from public conversations. T@gz also harvests information about people by recognizing names of people in email conversations using Natural Language Processing (NLP) techniques (as shown in Figure 1) and connects them to their activities. This way, in addition to tagging conversations, people are also tagged with topics such as, areas of expertise or project activities.
- The user can assign tags to the email by prefixing words by #. For example, if a user writes #SMBMarketing anywhere in the subject or body, the conversation (the original email as well as all replies/forwards on the same thread) will be marked as belonging to the subject "SMBMarketing" (Small and Medium Business Marketing).

System Architecture

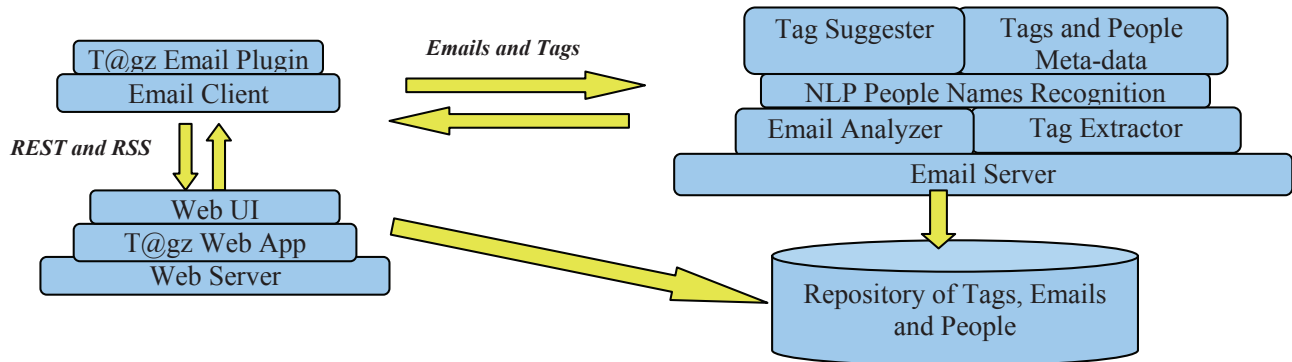


Figure 1: T@gz System Diagram

- This approach allows labeling the same email with multiple topics e.g., #laptops, #printers, #marketing, etc.
- To use multi-word tags users can use one of these forms: Using underscore instead of space, such as, #social_network or mixed case such as, # SocialNetwork. T@gz allows freedom to users instead of enforcing a style. Other options that we have designed are: a) Quoting with # as in # “United States of America” or, b) ‘Quoting’ with # as in #United States of America#
- The system is extensible to allow hierarchical tag categories using notation such as, #marketing.smallbusiness, #marketing.enterprise, or #marketing.asiapacific etc.
- T@gz generates RSS feeds for all emails as well as feeds for each tag. By automatically generating RSS feeds for each tag, we have enabled functionality similar to mailing lists for various topics/tags that users can subscribe or unsubscribe using their RSS reader or Outlook RSS client as shown in Figure 3. Any time there is a new email conversation matching topic of interest for the user, the message would come up in the RSS client of the user similar to new emails from mailing list. Since users can use MS Outlook RSS client or any other RSS client with similar integration to an email client the experience is just like typical mailing lists but they do not face the hassle of discovering, subscribing or unsubscribing from mailing lists.
- The T@gz web-based UI allows users to browse and search all tagged conversations as shown in Figure 2. It is possible to search with partial hash-tag words or part of people names.
- Our system also enables scenario where users can subscribe by sending email to tagz@company.com with a subject line such as, “subscribe #computers, #marketing, #asiapacific”. Such simple subscription for multiple topics is not supported by mailing lists.
- Using the T@gz UI people can create folksonomies and converge on popular concepts specific to their working domain similar to how people converge on categories

in Wikipedia. Here tags like #cloud or #service would have a meaning very different from the meaning in general English language.

- The system can enable users to define relations among tags such as, “USA”, “United States” or “America” and merge tags so that email conversations tagged with any of these words would belong to the same topic.

4. CONCLUSIONS

We have built a working system that can be demonstrated and our team has been using it for information sharing. We have also been using it to easily log feature requests or bug reports in our project.

By allowing effortless tagging from emails, the system facilitates answers to these key questions : What (hash-tags in email conversations representing topics), Who (Enterprise professionals connected to the topics) and When. Thus it can create snapshots of a project at various points of time.

We have found this to be a tool for enhancing enterprise productivity as opposed to social networking tools in public Internet which might focus on personal entertainment.

5. REFERENCES

- [1] M. Brzozowski, T. Sandholm, and T. Hogg. Effects of feedback and peer pressure on contributions to enterprise social media. In *Proceedings of the ACM SIGCHI International Conference on Supporting Group Work*, 2009.
- [2] C. Dugan, W. Geyer, and D. Millen. Lessons learned from blog muse: audience-based inspiration for bloggers. In *ACM CHI*, 2010.
- [3] L. Nelson, R. Nairn, E. Chi, and G. Convertino. Mail2tag: Augmenting email for sharing with implicit tag-based categorization. In *Proceedings of the International Conference on Collaboration Technologies and Systems (CTS)*, 2011.

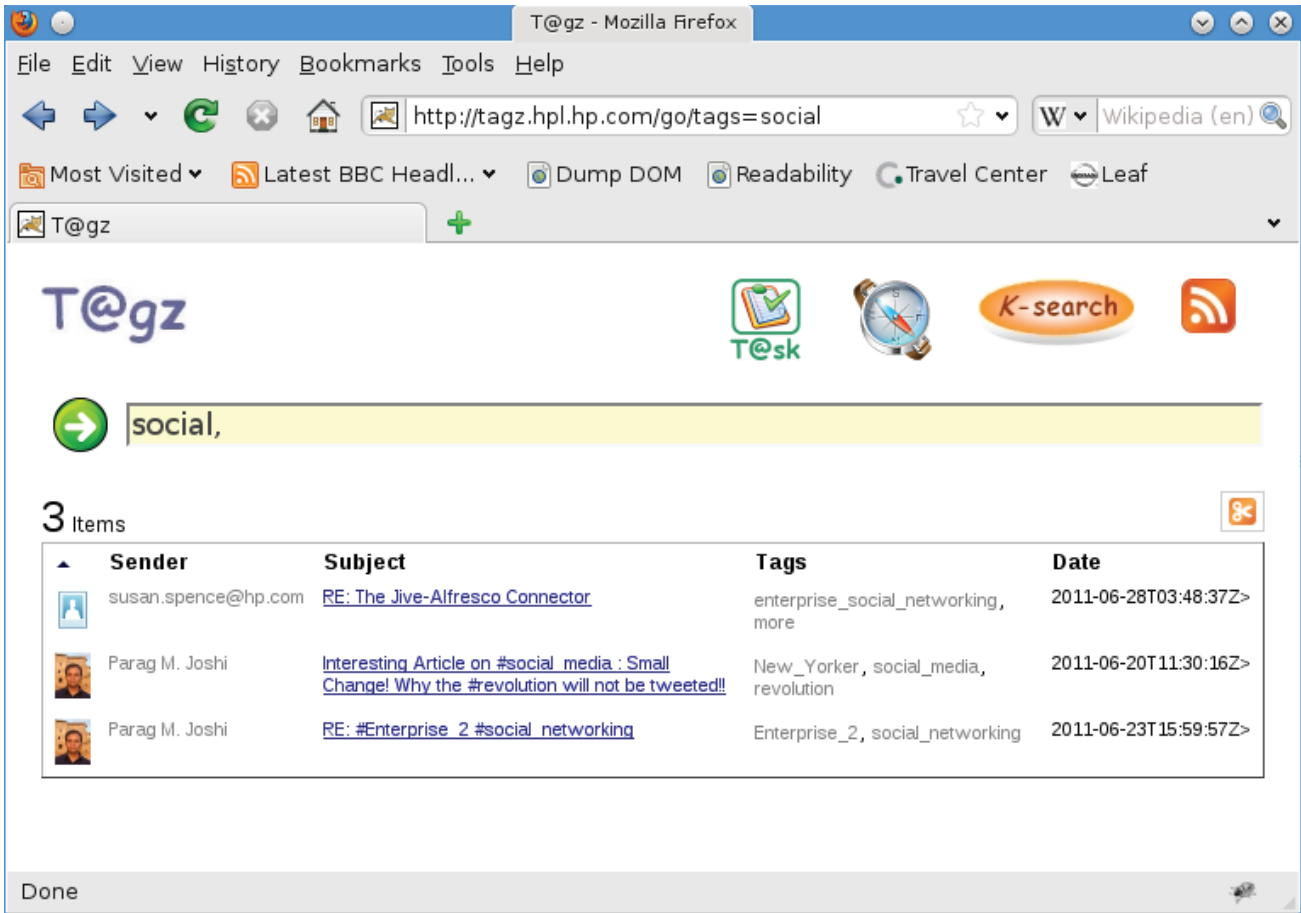


Figure 2: T@gz UI screenshot

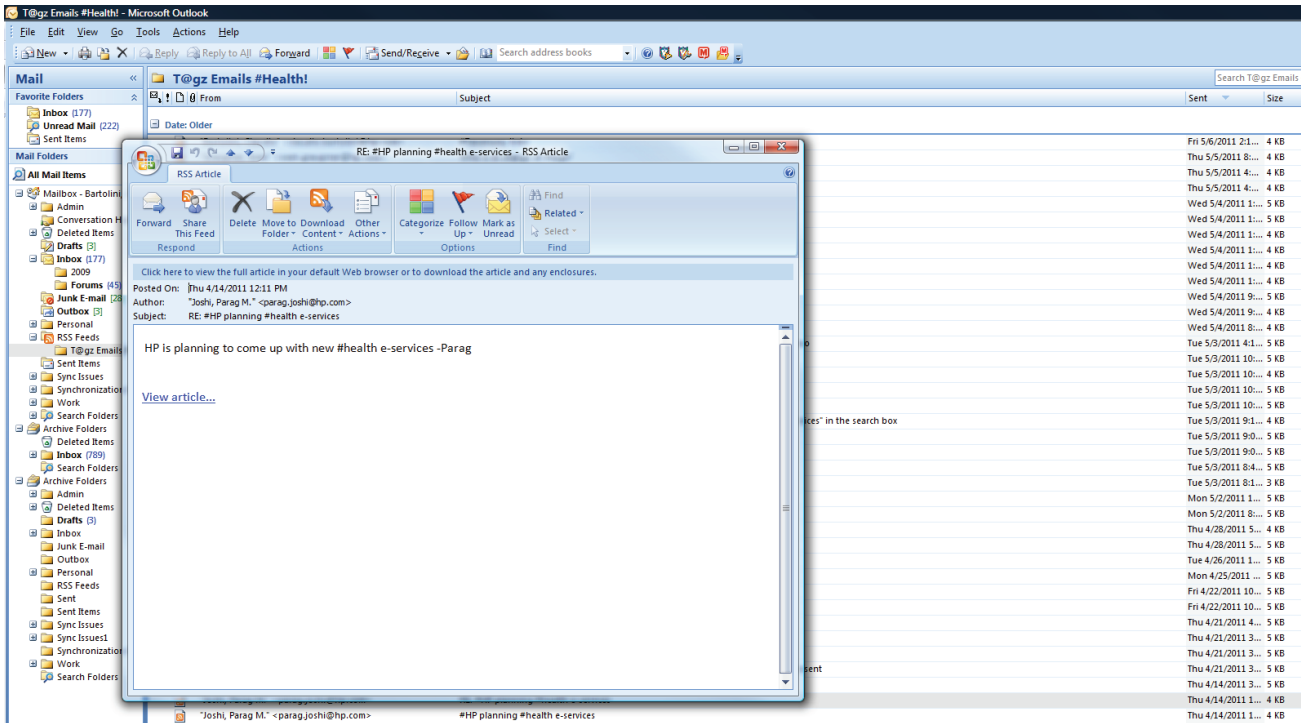


Figure 3: Screenshot of T@gz RSS UI in MS Outlook