Using Tag Clouds to Promote Community Awareness in Research Environments

Alexandre de Spindler, Stefania Leone, Matthias Geel, and Moira C. Norrie

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{despindler,leone,geel,norrie}@inf.ethz.ch

Abstract. Tag clouds have become a popular visualisation scheme for presenting an overview of the content of document collections. We describe how we have adapted tag clouds to provide visual summaries of researchers' activities and use these to promote awareness within a research group. Each user is associated with a tag cloud that is generated automatically based on the documents that they read and write and is integrated into an ambient information system that we have implemented.

Keywords: Tag Clouds, Ambient Information, Community Awareness.

1 Introduction

Community awareness within working environments promotes knowledge sharing among members of an organisation and extends cooperation beyond formal project boundaries. Keeping a community informed about its members and their activities is becoming increasingly challenging as the sheer amount of information available to us, and directed at us, rapidly grows and our capabilities to filter and absorb relevant, but non-critical, information are reaching their limits. At the same time, there is a tendency for users to become locked into virtual communities, not only for reasons of geographical distribution, but also the time spent communicating and working within a digital world.

To address these challenges, we developed an ambient news service, AwareNews [1], that integrates information about activities within an organisation as local news items within a more general news service. By providing information published by various media services such as breaking news stories, sports results and technology news alongside local news, we aim to attract users and sustain interest. News items are automatically integrated through newsfeeds and we developed additional local newsfeeds for a range of applications including calendars and software repositories.

Recently, we decided to investigate ways in which we could extend AwareNews to improve awareness of the past and present research activities of members of a research group. Tag clouds are widely used in Web 2.0 sites to provide visual summaries of various forms of data collections[1] and have now been adopted by

[1] http://www.flickr.com, http://delicious.com, http://wordpress.com

Y. Luo (Ed.): CDVE 2010, LNCS 6240, pp. 3–10, 2010.

some research groups to summarise the topics of their research based on titles of publications and/or projects[2]. We therefore decided to experiment with the use of tag clouds to visualise the research activities of individual members of a group based on, not only the publications that they had written, but also those recently read. The visualisation schemes were carefully selected so that a single tag cloud could be used to represent the overall expertise of a researcher as well as recent areas of interest.

In Sect. 2, we discuss the background to this work before going on to present the details of our tag cloud representations in Sect. 3 and 4. We then provide an overview of the system that we have implemented in Sect. 5. Open issues and directions of future research are discussed in Sect. 6 before giving concluding remarks in Sect. 7.

2 Background

Awareness has been recognised as a key factor for increasing efficient and effective cooperation within shared workspaces [2]. Community awareness is not related to any specific task, but rather aims to make users more aware of other members of the community and their activities in order to promote interactions that may lead to ad-hoc collaboration and knowledge sharing. This must be done in a manner that places few or no demands on the users and does not distract them or disrupt their normal work practices. For this reason, secondary displays [3] or ambient displays [4] are often used to display awareness information.

In [5], a general infrastructure for communication walls is presented that enables users to post any information items (text, images, videos) of interest. Apart from the disadvantage of relying on active user participation, another problem of such a system is the fact that the public displays are generally not aesthetically pleasing and hence not attractive to potential information consumers. Some projects have therefore looked at designing aesthetically pleasing information collages [6] or adapting recognised art forms to display information [7].

The kinds of information used to promote awareness and also the sources of that information are important in determining not only what users will become more aware of, but also the level of interest and involvement of the users. One informal study related to a personal awareness system found that 80% of participants check news and financial web sites on a regular basis [8]. This suggested that integrating various forms of news information into a community awareness system could be beneficial in terms of attracting users to the system and this is what we did in our AwareNews system [1]. In AwareNews, RSS feeds developed for local applications such as calendars and software repositories are integrated with feeds from various global news services. This allows users to be informed about the activies of other users related to events that they attend and also software projects that they are working on. Further, by using Apache SubVersion to control shared access to text documents as well as software, information about activities related to the writing of documents could also be integrated into the system. AwareNews provides access to this information by means of peripheral

[2] For example using http://www.citeulike.org/groups/browse

displays installed in offices and public areas. News items are selected at random and displayed one at the time using a very simple layout and slow rate of change to minimise distraction. Details of a qualitative user study are given in [1].

Although AwareNews provides some information about the current activities of users and was shown to have a positive effect on community awareness, it is very limited in terms of raising awareness about the knowledge and interests of users and therefore stimulating knowledge exchange and ad-hoc collaborations as proposed in [9]. In the remainder of this paper, we describe how we extended the system for a university setting to provide more information about the past and present work of researchers in terms of their reading and writing activities.

3 Profiling Researchers through Tag Clouds

Tag clouds have become extremely popular in Web 2.0 applications as a means of providing visual summaries of collections of documents. Although very simple, they can be used to support search, browsing and recognition as well as forming and presenting impressions [10]. We therefore decided to investigate their use as a means of profiling researchers.

We conducted a simple user study to assess if, and to what extent, members of a research organisation would be able to identify persons using a tag cloud generated from the titles of the papers that they had authored. The tag clouds were generated using the Web application Wordle[3] with a merged list of publication titles given as the input string. We chose to generate monochrome tag clouds and the only significant variable is the size of the tags which correlates with the relative term frequency in the input string.

Since the tag cloud of an individual represents their accumulated expertise, we decided to create a second augmented version which highlights the recent work of the user in terms of the two most recent publications. This was done by copying the tag cloud and colouring the tags appearing in the last two publication titles in red. An example of an augmented tag cloud is given in Fig. 1 with the monochrome tags shown in grey and the red coloured tags shown in black.

For the study, we chose three research groups[4] within our department. From each of these groups, we selected six PhD students with at least five publications, three of them to be *test persons* and three of them to be *represented persons*. The nine test persons were asked to identify nine represented persons, each represented by a basic and an augmented version of their tag cloud. We first showed the basic tag cloud and asked the test person to identify the research group. Second, we asked the test person to identify the represented person. Third, we showed the augmented tag cloud. If the test person named a person before, we asked whether they had changed their mind. If they were unable to name someone before, we asked if they could do so now. Finally, we presented the tested person with their own tag cloud and asked whether they liked it.

[3] www.wordle.net/create

[4] A group is either a research group associated with a single Chair or an Institute composed of multiple Chairs all working in the same area of research.

Fig. 1. Tag cloud of an individual researcher

The main results of the study are summarised in Table 1. In the first row, the number of correct identifications are given as a percentage of all attempts with the standard deviations written in brackets. In the second row, we provide the percentage of incorrect identifications. The remaining cases are those where the user was not able to name a group or individual.

The first two columns give the percentages for the cases where represented persons were members of the same group as the test persons or different groups, respectively. In the former case, they were able to identify the research group in 92.6% of all attempts and the represented person in 74.1%. In the latter case, the ability to identify the group decreased to 46.3% and the identification of the represented person to 7.4%. In both cases, the identification of groups was more successful than that of individual persons and this difference was even more prominent when the test person and represented person were members of different groups. The fault rates were higher when represented persons were members of different groups. We note that test persons were more cautious when identifying persons as opposed to identifying groups.

The last two columns indicate the effect of augmented tag clouds overall. With basic tag clouds, research groups were identified in 56.8% of cases while represented persons were identified in 25.9% of cases. The presentation of the augmented tag cloud increased these numbers to 61.7% and 29.6%, respectively. The fault rates decreased when identifying persons rather than groups, however, in both cases the fault rates increased when presenting the augmented tag clouds.

Table 1. Identification rates, fault rates and standard deviations in brackets

Same Group		**Different Group**		**Basic Cloud**		**Augmtd. Cloud**	
Identify		**Identify**		**Identify**		**Identify**	
Group	**Person**	**Group**	**Person**	**Group**	**Person**	**Group**	**Person**
+ 92.6 (0.4)	74.1 (1.1)	46.3 (1.6)	7.4 (0.5)	56.8 (1.4)	25.9 (1.0)	61.7 (1.7)	29.6 (0.9)
− 11.1 (0.7)	11.1 (0.7)	40.7 (1.6)	25.9 (1.6)	24.7 (1.2)	14.8 (1.4)	30.9 (1.6)	21.0 (1.5)

In summary, the study indicates that test persons had a good recognition rate for members of their own research group based on tag clouds. Generally, they were better at identifying groups than individual persons with the ability to represent individuals outside their own group being relatively low. It was interesting to note that many cases of false identification of both groups and individuals resulted from overlaps in research areas that reflected general hot topics in research such as mobile computing. Finally, seven out of nine test persons revealed a positive attitude toward their own tag cloud.

4 Using Tag Clouds for Awareness

We decided to integrate researchers' tag clouds into the AwareNews system as a means of raising the awareness of members of a research group about the past and current interests of other members of the group. In addition to providing a visualisation of topics of papers authored by a researcher, we also wanted to integrate information about the topics that they were currently reading about. Further, in line with the general goals of AwareNews, we wanted to be able to capture this information and generate the tag clouds automatically without requiring a change in work practices.

Since it is common nowadays for researchers to use some kind of tool for managing bibliographies of papers read as well as papers written[5], we decided to use this as our information source. When a user creates a new bibliography item, a simple check is made to determine whether or not the user is an author of that paper. If so, then the title of the paper is added to profile data of that user under *authored-papers*. If the user is not an author, then the title is added to the profile data under *read-papers*. The only change to work practices that we are considering is to have researchers create a bibliography entry for papers as soon as they start working on them as a technical report in order to ensure that the tag clouds reflect current work and not just work that is already published. We note that some research groups already enforce such a practice.

The tag clouds used in our study were based on only authored publications and we created an augmented version with the tags in red to highlight topics of recent papers. Our study has shown that this simple mechanism is also effective in helping to identify researchers within a group. To represent recently read publications, we incorporate the tags appearing in the titles of the last two papers read into the tag cloud and then highlight these terms in blue. If a tag appears in both *authored-papers* and *read-papers*, then it will be coloured red since writing about a topic is considered to subsume reading about it. Note that there are issues about how many papers should be taken into account in representing recent activities and also whether tags should be weighted in some way so that the size of tags of recent topics would be larger relatively and hence more easily noticed. A disadvantage of giving more weight to recent activities is that this may have a negative effect on raising awareness of past experience if these tags would become very small. For the moment, we have therefore chosen to experiment

[5] http://www.endnote.com, http://www.mendeley.com, http://www.connotea.org

with two versions of the tag clouds—one where no weighting is applied and one where tags are weighted according to recency. In the case of publications authored, recency is determined by the year entry of the bibliography item, whereas in the case of publications read, it is determined by the date that it was entered in the bibliography.

An issue of importance in all awareness systems that use shared displays and especially displays in public areas is that of privacy. Here the results from our study also indicate some of the benefits of tag clouds. On the one hand, they are an abstract representation of researchers' activities and for people outside of the group they would tend to simply see the tag clouds as visually pleasing representations of the activities of a group. On the other hand, within a research group, users can easily identify individual researchers and their current activities. This enables them to recognise if other members of the group have worked, or are working on, topics that are of interest to them.

5 Architecture and Implementation

The AwareNews system consists of several secondary and wall-mounted public screens driven by small computers which are connected to a context-aware data aggregation engine responsible for acquiring local and global news sources. News sources are RSS feeds where feed items are pushed to the displays on a regular basis. Therefore, additional news sources can be integrated by implementing an RSS feed generator. To support the work presented in this paper, we have implemented such a generator that produces feeds containing tag clouds.

Figure 2 gives an overview of the system architecture and highlights the processes involved in generating and displaying tag clouds. To the left, we show the users of AwareNews who are exposed to the ambient displays shown in the middle, while the software system is outlined on the right. The screens display tag clouds which *aggregate* the titles of those documents which the *represented* users have either *written* or *read* as exemplified by the grey circles and diamond-shaped relationships.

While Wordle was used to manually generate tag clouds for our study, we had to develop our own *tag cloud generator* for the AwareNews system which is shown in the top right. The tag cloud generator accesses the *bibliography*

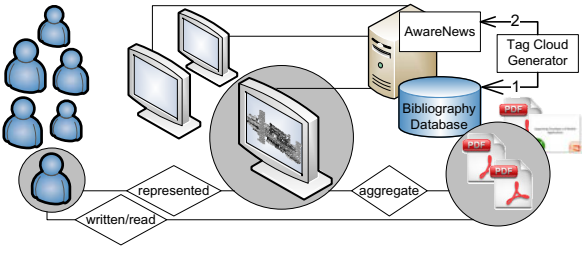

Fig. 2. System architecture

database (1) which is in charge of managing the bibliographies of all users and includes items for the documents authored by the users as well as for those that they have read. Whenever a document authored by a user is entered into the system, a deferred task is spawned that refreshes the tag clouds affected by the current addition. For each author of the new document, the tag cloud generator queries the database in order to retrieve the titles of all documents written or read by the author. Similarly, a user's tag cloud will be updated if they add a new document to the bibliography database that was not authored by any members of the group, but in this case it will be added to the set of read documents.

In order to obtain more useful clouds, some basic natural language processing is performed on the accumulated titles, such as filtering insignificant words and removing certain inflections. In a last step, a tag cloud algorithm determines the sizes of the tags according to their frequencies and places them spatially according to a layout strategy. Based on this layout and additional aesthetic options such as font type, font size and colouring, the tag cloud is directly rendered to an image and stored in the file system. The whole process is concluded by updating a single RSS feed with a URL pointing to the newly created image (2). Consequently, the next time the AwareNews system polls this feed, the images are retrieved from the file system and displayed.

6 Discussion

Following our study, we have a first implementation of the system and are experimenting with different visual representations. As mentioned previously, one of the issues is to investigate possible weighting schemes to reflect the recency of tags and whether this could be beneficial. One of the factors that has to be taken into account is the size of the tag cloud in terms of the number of tags. If one considers the tag cloud of a professor who has been publishing over many years, then clearly the characteristics of their tag cloud is quite different from that of a new PhD student and it is easier for individual tags to get lost within the cloud if they are recent but infrequent. Therefore, it may be that we require a weighting scheme that is even dependent on the size of the tag cloud. However, we note that the ability to recognise more senior researchers from junior researchers through the size of the tag cloud can be seen as another benefit of using tag clouds as profiles of individual researchers. We also want to experiment with the number of publications considered as defining recent activities. At the moment, we are using the two most recently written and read publications, but this is mainly because our study focussed on PhD students who tend to have a relatively small publication record.

One of the features of the existing AwareNews system is that it also displays information about the presence of members of a group based on their MSN status. We are experimenting with different ways of capturing and visualising presence and status information and plan to include the tag cloud profiles as one possible visualisation. This means that a researcher's tag cloud would only be displayed when they are actually present and it could be visually augmented to indicate the status.

AwareNews is currently an ambient information service that involves no user interaction. However, it could be useful for users to be able to interact with the tag clouds to find out more information about the publications or even the projects related to them. We have also considered the possibility of allowing users to send a message to the researcher associated with a tag cloud as a first means of opening up discussions. As future work, we therefore plan to extend the system to support such forms of interaction.

7 Conclusions

We have proposed the use of tag clouds as a means of visualising researchers' past and present activities and described how we integrated this concept into an ambient information service designed to promote community awareness. We note that tag clouds can play a double role in such a setting since they provide an attractive display for visitors that provides them with an overview of the group, while, at the same time, being used by members of the group to identify researchers and keep them informed about their activities.

References

1. Decurtins, C., Norrie, M.C., Reuss, E., Weibel, N.: AwareNews - a Context-Aware, Ambient News Service. In: Proc. 4th Intl. Conf. on Intelligent Environments, Seattle, USA (July 2008)
2. Dourish, P., Bellotti, V.: Awareness and Coordination in Shared Workspaces. In: Proc. CSCW 1992, Toronto, Canada (November 1992)
3. Grudin, J.: Partitioning Digital Worlds: Focal and Peripheral Awareness in Multiple Monitor Use. In: Proc. CHI 2001, Seattle, USA (March 2001)
4. Wineski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., Yarin, P.: Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In: Streitz, N.A., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370, p. 22. Springer, Heidelberg (1998)
5. Ferscha, A., Kathan, G., Vogl, S.: WebWall - An Architecture for Public Display WWW Services. In: Proc. WWW 2002, Hawaii, USA (May 2002)
6. Fogarty, J., Forlizzi, J., Hudson, S.: Aesthetic Information Collages: Generating Decorative Displays that Contain Information. In: Proc. UIST 2001, Orlando, USA (November 2001)
7. Redström, J., Skog, T., Hallnäs, L.: Informative Art: Using Amplified Artworks as Information Displays. In: Proc. DARE 2000, Elsinore, Demark (April 2000)
8. Miller, T., Stasko, J.: The InfoCanvas: Information Conveyance through Personalized, Expressive Art. In: Proc. Extended Abstracts CHI 2001, Seattle, USA (March 2001)
9. Farrell, S., Lau, T., Nusser, S., Wilcox, E., Muller, M.: Socially Augmenting Employee Profiles with People-Tagging. In: Proc. 20th ACM Symp. on User Interface Software and Technology, New York, NY, USA (October 2007)
10. Bateman, S., Gutwin, C., Nacenta, M.: Seeing Things in the Clouds: The Effect of Visual Features on Tag Cloud Selections. In: Proc. 19th ACM Conf. on Hypertext and Hypermedia, Pittsburgh, USA (June 2008)