Originally published in:
Proceedings of the 10th International Conference on Information Visualisation (IV'06), London, England, 05-07 July 2006 / E. Banissi et al. (eds.), pp. 124-129

Visualising Collaboration via Email: Finding the Key Players

Onn Azraai Puade, Theodor G Wyeld
ITEE, University of Queensland, Queensland, Australia
{puade@itee.uq.edu.au, twyeld@itee.uq.edu.au}

Abstract

Email is an important form of asynchronous communication. Visualizing analyses of email communication patterns during a collaborative activity help us better understand the nature of collaboration, and identify the key players. By analysing the contents of email communication and adding reflective comments on its perceived importance from the participants of a collaboration new information can be gleaned not immediately obvious in its original flat form. This paper outlines a proof-of-concept prototype collaborative email visualisation schema. Data from a collaboration case study is analysed and subsequently employed to construct a display of the relative impact of both key players and the types of email used.

1. Introduction

Email as a form of asynchronous communication allows large communities to exchange information in a low cost environment. Email often forms the backbone to research, industry, educational and other collaborations. In this paper, we report on a new and novel method for identifying the key players in a collaboration exercise based on their impact on the group as a whole. It forms its conclusions based on how the individual players rate the importance of each other's emails to the collaboration. 176 emails were collected over a period of 6 months (197 days). The data from a peak period, over one week when the main workshop was run, was used in the proof-of-concept visualization method described here.

The analysis of this subset of emails reveals the patterns of interaction that existed within the collaboration project email archives. From this we can gain a better understanding of the nature and characteristic of cooperation and collaboration in general. More specifically, it is used to identify the key collaborators.

This paper is organized into four sections. The following section provides an overview other research in this field. Then we present our case study including three specific analyses of the results and their visualisation. This is followed by a discussion on interpretation of the results and their visualisation. This paper concludes with a brief discussion on how email visualisation can assist future collaborations.

1.2 Collaboration Visualization

Roschelle and Teasley [14] define collaboration as "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem.'' Brna [6] suggests collaboration can be "seen as a set of possible interrelationships between participants." Collaboration in a virtual team context is the act of working together on a common task or process [4]. All these definitions rely on communication enabling technologies. Central to these is email. Card et al [7] reminds us that computer-supported, visual representations of abstract data amplifies our ability to make sense of large collections of data. Visual representations aid and enable the user to understand the different kinds and forms of data. The structural modeling and graphical representation of collaboration assists in its visualization. Creating an overview plays a crucial role in getting the user to see the big picture; only to later focus on what they consider to be an important part of it. Such an overview [15] of the structural and the behavioral aspects of collaboration helps define how all the elements in particular environments interrelate. Within this, the behavioral actions and interactions can be seen to evolve over time [5]. We can visualize the various user roles and their grand purpose to the collaboration as a whole [12].

1.3 Collaboration Email Visualization

A key factor in any collaboration exercise is the exchange of email [9, 18]. There are several advantages to studying email as a measure of collaboration. These include the social networks formed, its ubiquitous usage, and high volume. Moreover, within email, its structural elements (sender and receiver), cumulative structural characteristics (frequency, reciprocity) and temporal dimensions (timestamps) are automatically recorded along with its contents [1]. These can be used to construct visualisations for displaying their interrelationships graphically. For Divitini and Farshchian [8], email is seen as a key collaboration medium. They organise the roles of email as those used to access experts, resolve issues and decisions, provide awareness to project-related issues and support for irregular synchronous collaboration. Applying visualizations and its various techniques to email archives helps aid information retrieval processes and make analyses of the trends embedded [10]. Such salient

Proceedings of the Information Visualization (IV'06)
0-7695-2602-0/06 $20.00 © 2006 IEEE

structures cannot be readily comprehended in their original text-based flat file forms.

Visualization of communication patterns can thus help identify collaborative innovation networks – groups of self-motivated individuals with a common vision working together on a new idea, via the internet [11]. Analysis of interaction logs (including email) can enable identification of, among others, key contributors and important collaborators in organizations. It is within this frame that the focus of our study is set. Of the many email visualization studies conducted to-date, Perer et al's [13] emerges as pertinent to our study. Perer et al extract contextual information by visualizing the temporal rhythms of social relationship in email archives. Their classification of the different types of interaction with email collections can be seen in Table 1.

Table 1 Types of interaction with email collections

	Individual	Organizational	Social
Current	Managing an individual user's current inbox (A)	Managing current email within an organization (B)	Managing current conversations within a public space (C)
Archived	Exploring an archive of an individual's message (D)	Exploring an archive of an organization's messages (E)	Exploring an archive of a public space (F)

Our study addresses the category in Cell E. Our collection of archived messages provided the raw data for a socio-centric study. With this, we have identified also the roles individuals adopt and types of interactions between their social networks.

Table 2 Categorization of email via its contents functionality

Category	Description
Awareness	Email with contents that make users aware of issues, by announcement
Decision making	Email with elements of decision making on key issues in the collaboration which determine the outcome of project related issue, local or global
Accessing expert	Enquiry directed to an expert on technical and administrational issues, local
Feedback	Feedback to project related announcement or enquiry, local or global
Resolving issues	Mass enquiry soliciting comment on a specific issue, global

The socio-centric approach was chosen over the alternative, egocentric analysis, because, rather then focusing on a single individual's emails, our study uses many message archives to extract patterns representing an overall social structure. It is hoped our visualization of the email archives will serve as a tool for self reflection for its participants to enable and motivate better collaboration outcomes. Other socially organized visualizations include Viegas et al [16, 17], and Begole et al [3]. Divitini and Farshchian's [8] key roles of email in collaboration was extended to create an expert system to organize and classify our email collection, within Perer et al's [13] typology (see Table 2). It was used as a control study for later comparison.

2. Case Study

2.1 Setting

The participants in this study came from diverse backgrounds: facilities manager, project coordinator, technical assistant, chief executive officer, secretary, research assistant, external consultant, project leader, artist, programmer and project manager. They represent also a wide range of ages 21-51. Their acculturation to email as a communication tool was assumed. This study focuses on email visualization of a specific, time-constrained, event-driven collaboration. The collaboration involved the organization and running of a workshop to develop resources for a massive multi-user game. The workshop ran for three days. The CEO from a participating organisation and their chief computer programmer traveled from Canada for the workshop as part of other business in Australia. 20 individuals from 6 organizations were involved in the activity over 197 days (i-mmersion Canada, ACID Australia, University of Queensland, Queensland University of Technology, Silicon Graphics Inc, Institute of Modern Art Brisbane). The period chosen for analysis in this study represents the period just before and after the workshop was run. There were 24 emails sent by 10 participants over this period. 11 participants from 3 organizations are identified in the email collection. The collaboration involved various activities in different location and continents. This included the workshop studio, the research organization, the private company (Canada), the host university, and private residences.

2.2 Process

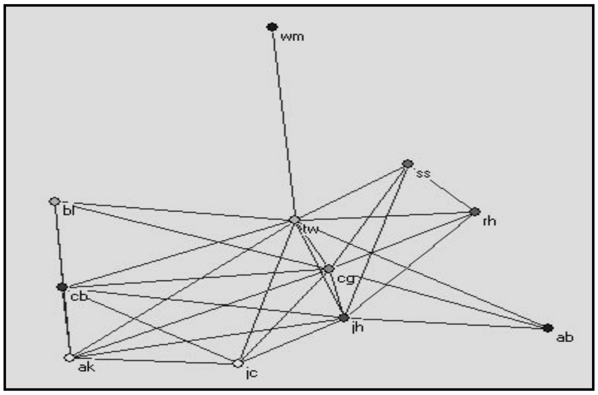
The contents of the 24 emails were analyzed. Each email included embedded prior emails, subject descriptions, sender, receiver(s), date and message. From this data, we were able to plot the connection between the various collaborating participants and the types of topics discussed, and the temporal sequence. Initial analysis conducted on this information yielded structural results. However, what many email analyses and subsequent information visualizations fail to do is seek reflective information from their participants to include in their datasets. To this end, we conducted a survey with all those participants recorded in the 24 email subset.

This information was then used to construct visualization schemas that better reflect users' reasonings for how the importance of an email is perceived within such a collaboration exercise. This yielded surprising results and is finally compared with our initial 'expert categorization system'.

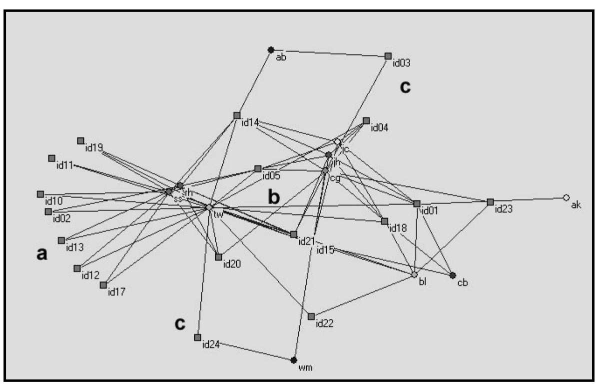
2.3 Analyses

2.3.1 Network Diagram Analysis

In the first round of analyses, network graphs were constructed from the collection of emails. Node and link graphs were generated using Pajek [2], a social network analysis visualization tool. Undirected graphs were used, with the typical vertices representing emails and nodes representing participants. Figure 2a shows how the 11 participants were connected by the 24 emails; and, in figure 2b, by adding email nodes, we can see which emails connect participants as individuals or as part of a group (Sender nodes are connected to email nodes which are connected to receivers).



(a)



(b)

Figure 2 (a) People-to-people and (b) People-to-email diagram

From Figure 2a the people-to-people diagram [19], we can see the three participants (tw, cg and jh) at the centre of the graph are the most connected. This suggests they had a more active participation in the overall collaboration. In Figure 2b, we can see that, by adding the emails as nodes, a sender-email-receiver pattern emerges that now includes more participants in the highly connected category. This diagram suggests there are more active participants than what is first suggested by the people-to-people schema. However, as we will see later, these visualization methods can be misleading in terms of a particular participants 'perceived' importance and contribution when rated by all participants.

2.3.2 Email Content Analysis

Automatic classification by data mining and information retrieval techniques can be seen in many research studies which focus on document management and organization [1, 20, 21]. To demonstrate this we recast Divitini and Farshchian's [8] email roles as a classification system. We used this to categorized our emails according to their content where: A = Awareness; D = Decision making; E = Accessing expert; F = Feedback; and, R = Resolving issues. We found some emails had multiple categories. This categorization formed the basis of our expert control system used later to compare with other systems (see Table 2).

Table 3 Tabulated category and ranking for 24 emails

Email	Category	Aggregated Rating
1	A	20
2	E, A	20
3	A	16
4	A	13
5	A, F	20
6	A, D, R	20
7	A, R	19
8	F, A, D	21
9	A, D	18
10	A	14
11	E, R	14
12	F, A	12
13	F, R	18
14	A	21
15	A	17
16	A, F	17
17	F	18
18	F	23
19	A	20
20	F	17
21	A, F	20
22	A	15
23	A	18
24	A	24

After categorizing the emails based on this system, we conducted a survey with the participants identified in the 24 email collection. Each participant was given a printed copy of the 24 emails in sequential order. At the end of each email was a check box survey field. They were asked to rate each individual email in term of its importance on a scale of 0 - Not applicable, 1 - Not important, 2 - Important, and 3 - Very important. A comment field was included for participants to provide reasons why they gave the email a particular rating. When no check box was marked for an email, this was recorded as '0' or 'Not applicable'. Some chose not to give reasons for their rating of an email. The overall expert category and aggregated participants' ratings by email are provided in Table 3.

Table 4 Data by participants, sorted by average rating

Participants	No of emails	Average rating	No of emails x Average rating
WM	1	2.4	2.4
CB	1	2.3	2.3
AK	1	2	2
CG	5	1.9	9.5
JH	1	1.9	1.9
TW	5	1.84	9.2
RH	3	1.83	5.5
BL	2	1.65	3.3
AB	1	1.6	1.6
SS	4	1.45	5.8

From Table 3, we see that the highest rating (24) was for an email of type A (Awareness) and the lowest (12) was for an email of type FA (Feedback - Awareness). To gain a better understanding of these extremes in the context of the overall collaboration, we need to re-organize the table by number of emails per participant, type, ratings and average ratings (see Table 4).

From this organisation we can make some preliminary interpretations of it: if we say that the average rating represents the 'loudness' (L) of a participant's message within the collaboration, and multiply their loudness by the number of emails (N) sent, we can say this is a measure of their overall 'impact' (I) on the collaboration.

$$L \times N = I$$

Table 5 Data by email category, sorted by average rating

Category	No of emails	Loudness	Impact
A	10	1.78	17.8
AF	4	1.725	6.9
F	3	1.93	5.79
FAD	1	2.1	2.1
EA	1	2	2
ADR	1	2	2
AR	1	1.9	1.9
AD	1	1.8	1.8
FR	1	1.8	1.8
ER	1	1.4	1.4

Hence, from table 4 we find the highest (average) rated email (2.4), or 'loudest' emailer, was sent by participant WM, and the lowest (average) rated email (1.45), or the least 'loud' emailer, was sent by participant SS. What is interesting here is that the loudest emailer sent only a single email, whereas the least loud emailer sent many emails. If we now look at these results in terms of impact we find the loudest emailer has low impact whereas the least loud emailer has higher impact. When we compared these interpreted results with table 3 it demonstrates how different representations of information can tell very different stories. Such as, many less loud emails can have a greater overall impact than a single louder email.

We can organize the categorization of emails in a similar manner (see Table 5). From Table 5 we see that, while there is little 'loudness' variation across the different types of emails, there are clearly more emails of type 'awareness' which, in turn, generate the greatest impact.

However, in Table 6, which compares participant impact with type of email, we notice that the participants with the lowest and highest impact both include the same type of email (awareness) which is also the loudest. Hence, it appears it may not be useful to correlate the type of email with participant impact. Clearly, other factors are contributing to participants impact beyond simply the type and number of emails sent. As we will see in the next section, it is the rating applied to individual emails by all participants that, when averaged and multiplied by the number emails they sent, determines a participant's overall impact.

Table 6 Impact value and type of emails sent for each participant

Participants	Impact	Type of emails
CG	9.5	A, FAD, AF, ADR
TW	9.2	EA, AF, AD, A
SS	5.8	ER, AF, A
RH	5.5	FR, A, F
BL	3.3	A
WM	2.4	A
CB	2.3	F
AK	2	A
JH	1.9	AR
AB	1.6	A

2.3.3 Visualizing Collaboration Impact

We can visualize the results of these tabulations. We chose a system which should allow one to gain an easy understanding of the various forms of information contained. For the impact by participant tabulation in table 4, this includes participants' roles (bracketed terms), the number of emails they sent (orange dots), their average rating or loudness (blue dots) and their combined rating by number of emails or impact (dashed circle). The concentric rings this information is arranged within indicate the four rating scales: 0 - Not applicable, 1 - Not important, 2 - Important, and 3 - Very important (see Figure 3).

We can similarly visualize the impact by type tabulation in table 5. It includes, type of email (capital letters), the number of emails (orange dots), their average rating or loudness (blue dots) and their combined rating by number of emails or impact (dashed circle). The concentric rings this information is arranged within indicate the four rating scales: 0 - Not applicable, 1 - Not important, 2 - Important, and 3 - Very important (see Figure 4).

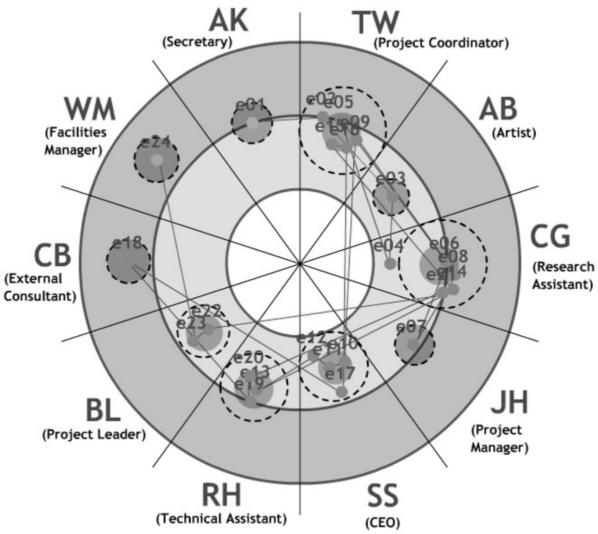


Figure 3 The visualization widget showing individual-oriented data

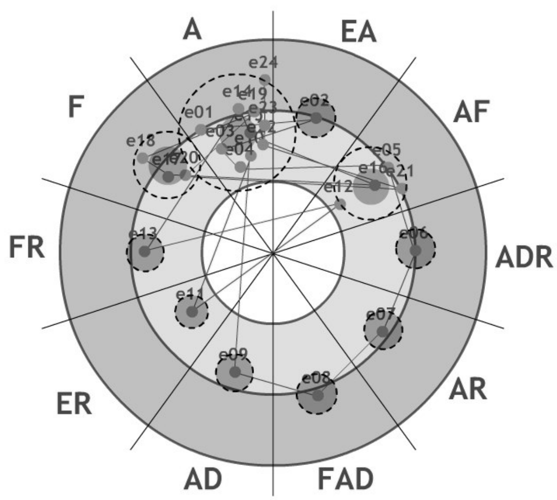


Figure 4 The visualisation widget with the expert system categorization data

Both these visualisations help us to gain 'at a glance' a better understanding of the information contained in the tables whence the data came. For example, we notice in figure 3 (impact by participant) the loudest emailer is not necessarily the one with the greatest impact. Similarly, in figure 4 (impact by type), we notice that there is little variation in loudness across all types of emails yet one particular email type (awareness) has a much greater impact than all others. In this sense, the graphical visualisation of collaboration via email assists quick identification of key players and types of email with the greatest impact in an easily assimilable form.

3. Discussion

From the three types of analyses discussed in this paper, we can draw some preliminary conclusions. In the first analysis, we were able to isolate the connectivity of participants – to each other and through the emails they sent and received. This was based on the objective information available from the emails themselves.

In the second analysis, after conducting a survey with the participants to add their reflective input to the dataset in the form of an importance rating for each email, we were able to expand our connectivity visualisation to include how important emails, type of emails, and individual participants, were perceived by all participants. We could also aggregate and average these ratings to make comparisons from the various features captured.

To make a reasonable comparison between these features, we used Divitini and Farshchian's [8] roles for email within Perer's [13] typology as a control. Of the 5 noted by Divitini and Farshchian [8], 10 subtypes were identified. We used these to group emails by type and compare the ratings applied by participants collected in the survey. We found there was little variation across types as an average rating. However, when we applied the average rating to individual participants, we found more variety. Indeed, the highest rated email was from a participant who only sent a single email. This raised the question: "how to define importance?"

As the survey asked participants to rate emails in terms of their importance, we needed to find a method for correlating how important a participant's email was with how many emails were sent by that participant. We considered the average rating of importance applied by all participants to a particular email as a measure of how 'loudly' this participant's message was received. Hence, how important a participant's contribution was to the collaboration was a factor of both how loudly their emails where received and how many emails they sent. If we simply multiply their loudness (or average rating per email) by the number of emails sent, we find a measure of their 'impact' on the overall collaboration. We applied this concept also to our control system to find the impact of types of emails.

What we notice from the visualisation of these two analyses is that there was more variation between participants in both loudness and impact. While, in the case of the control system, there was little variation in loudness between the different types of email, there was greater variation between impact than that displayed in the participant impact.

From this, as a system for tracking collaboration, we can say: predefined roles, such as project leader, manager, coordinator, and so on, do not necessarily generate the greatest impact on a collaborative project over time. However, the traditional purpose of these roles – to make announcements on progress, meetings, and queries – is supported by the 'by-type' visualisation.

4. Conclusion

We started out by trying to find the key players in a collaboration by analysing a subset of emails from a larger collection. What we found was that there is more variation between how the importance of participant's emails is perceived compared to the importance of the types of email sent. However, emails of type 'awareness' clearly have a greater impact on the collaboration because of their dominant use. On the other hand, the traditional purveyors of announcements – managers, leaders, coordinators, and so on – were not identified as having the greatest participant impact. Hence, we can conclude that despite being a team effort, this collaboration relied to a large degree on the activities of the research assistant – an often overlooked role. Furthermore the importance of providing both an expert system analysis of email roles and feedback from the participants in a collaboration exercise is crucial to a comprehensive understanding of the importance and impact of contributors. This is demonstrated by our two visualisation schemas.

As a visualisation tool for tracking collaboration participation, it is envisaged future implementations of the proof-of-concept prototype described here could be used to foster greater awareness of individual roles within a dynamic collaboration over time. This would involve incremental updates on email importance as they are received, entered into the visualisation tool for others to see. The nature of the visualisation schema developed here is also scalable – more participants can be added by simply including more sectors. In time, this could be useful also in gauging other forms of collaboration research impact.

References

[1] M. V. Alstyne and J. Zhang, "EmailNet - A System for Mining Social Influence and Network Topology in Communication," 2003.

[2] V. Batagelj and A. Mrvar, "Pajek - Program for Large Network Analysis," Connections, vol. 21, pp. 47-57, 1998.

[3] J. B. Begole, J. C. Tang, R. B. Smith, and N. Yankelovich, "Work rhythms: analyzing visualizations of awareness histories of distributed groups " in Proceedings of the 2002 ACM conference on Computer supported cooperative work New Orleans, Louisiana, USA ACM Press, 2002 pp. 334-343

[4] R. P. Biuk-Aghai, "Visualization of Web-Based Workspace Structures " in Proceedings of the First International Conference on Web Information Systems Engineering (WISE'00)-Volume 1 - Volume 1 IEEE Computer Society, 2000 pp. 302

[5] R. P. Biuk-Aghai, "Visualizing Structural and Behavioural Aspects of Virtual Collaboration " in Proceedings of the 10th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises IEEE Computer Society, 2001 pp. 279-284

[6] P. Brna, "Models of Collaboration," 2002.

[7] S. K. Card, J. D. Mackinlay, and B. Shneiderman, Readings in information visualization: using vision to think. San Francisco: Morgan Kaufmann Publishers, 1999.

[8] M. Divitini and B. A. Farshchian, "Collaboration and Coordination through Basic Internet Tools: A Case Study," presented at World MultiConference SCI/ISA, Orlando, Florida, 1999.

[9] N. Ducheneaut and V. Belotti, "Email as a habitat: An exploration of embedded personal information management," ACM Interactions, vol. 8, pp. 30-38, 2001.

[10] S. Frau, J. C. Roberts, and N. Boukhelifa, "Dynamic Coordinated Email Visualization," presented at WSCG 2005, Plzen, Czech Republic, 2005.

[11] P. A. Gloor, R. Laubacher, S. B. C. Dynes, and Y. Zhao, "Visualization of Communication Patterns in Collaborative Innovation Networks - Analysis of Some W3C Working Groups " in Proceedings of the twelfth international conference on Information and knowledge management New Orleans, LA, USA ACM Press, 2003 pp. 56-60

[12] B. Otjacques, M. Noirhomme, and F. Feltz, "Visualization of Interactions Within a Project: The IVF Framework," Lecture Notes in Computer Science, vol. Volume 3675, pp. Pages 61 - 70, 2005.

[13] A. Perer, B. Shneiderman, and D. W. Oard, "Using Rhythms of Relationships to Understand Email Archives," Journal of the American Society for Information Science and Technology, 2005.

[14] J. Roschelle and S. D. Teasley, "The construction of shared knowledge in collaborative problem solving," in Computer supported collaborative learning Berlin, Springer, 1995 pp. 69-97

[15] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," presented at Visual Languages, 1996. Proceedings., IEEE Symposium on, 1996.

[16] F. B. Viégas and J. Donath, "Social Network Visualization: Can We Go Beyond the Graph?," presented at CSCW'04 Workshop on Social Networks, Chicago, IL, USA, 2004.

[17] F. B. Viégas, D. Boyd, D. H. Ngunyen, J. Potter, and J. Donath, "Digital Artifacts for Remembering and Storytelling: PostHistory and Social Network Fragments," presented at Proceedings of the 37th Annual Hawaii International Conference on System Sciences, 2004.

[18] S. Whittaker and C. Sidner, "Email overload: exploring personal information management of email," presented at Proceedings of the SIGCHI conference on Human factors in computing system, Vancouver, British Columbia, Canada, 1996.

[19] R. Xiong, M. A. Smith, and S. M. Drucker, "Visualizations of Collaborative Information for End-Users," Microsoft Research 1998.

[20] E. Crawford, J. Kay and E. McCreath, "IEMS - The Intelligent Email Sorter" in Proceedings of the Nineteenth International Conference on Machine Learning Sydney, Australia, Morgan Kaufmann Publishers Inc., 2002 pp. 83-90

[21] E. Giacoletto and K. Aberer, "Automatic Expansion of Manual Email Classifications Based on Text Analysis" in 2nd International Conference on Ontologies, Databases, and Applications (ODBASE) Sicily, 2003 pp. 785-802.