
Generalizing Email Messages Digests

Romain Vuillemot

Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205
F-69621 Villeurbanne, France
romain.vuillemot@insa-lyon.fr

Jean-Marc Petit

Université de Lyon, CNRS
INSA-Lyon, LIRIS, UMR5205
F-69621 Villeurbanne, France
jean-marc.petit@insa-lyon.fr

Mohand-Said Hacid

Université de Lyon, CNRS
Université Lyon 1, LIRIS, UMR5205
F-69622 Villeurbanne, France
mohand-said.hacid@univ-lyon1.fr

Abstract

An email digest is a message that results from the combination of other messages. Mailing list management systems implement digests to let subscribers reduce their email messages frequency. In this paper we address the issue of generalizing this digest technique for any message (i.e. not only issued from mailing lists). By generalizing we mean creating new message combinations while 1) keeping an email centric approach, and 2) generating a compact visualization to assist a user task. We implemented a preliminary prototype as a webmail and we will describe a series of digests providing users multiple visualizations in the context of a meeting planning by email.

Keywords

Email research, digest, templates, visualization.

ACM Classification Keywords

H.5.2 [Information Interfaces And Presentation]: User Interfaces, Graphical user interfaces

General Terms

Design

A context of (useful) email overload

The steady growth of exchanged emails makes Email Research still an active field. While a large branch of

Copyright is held by the author/owner(s).

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.

ACM 978-1-4503-0268-5/11/05.

research works separated SPAM (not solicited emails) and HAM (solicited emails), a recently introduced intermediate category known as BACN¹ deals with solicited emails, *but not right now*. BACN messages exist since email clients have become a hub to monitor, get updates and reminders from remote systems, via emails. BACN are difficult to automatically process because of their diversity and the variable attention users have towards them. To illustrate the diversity a brief list of BACN messages issued from one of the author's INBOX is listed below: call for papers (figure 1), newsletters (figure 2) and social updates (figure 3).

BACN is mainly generated by databases, but users themselves can also generate it. For example, organizing a meeting using email may lead to a large amount of emails and a potential overload [4]: an initial email proposes locations and dates to recipients; responses emails arrive progressively; some follow up messages could be sent and people can also change their mind over time, generating even more emails. External applications such as Microsoft Excel or Doodle [7] may be needed to collect and process answers, but induce context switches and distractions.

brdiczka@PARC.COM	Cfp: International Workshop on Hybrid...	07/01/2011 21:37
Willem-Paul Brinkma...	cfp - ECCE2011 workshop proposals - d...	08/01/2011 10:49
Mark Billinghurst	Computers and Graphics Special Issue...	10/01/2011 12:06
nemanja.memarovic...	CFP: EuroITV 2011 Workshop: **Publi...	10/01/2011 12:38
Henriette Cramer	CFP CHI 2011 Workshop: Ethics in Larg...	10/01/2011 13:29
David England	Final Reminder: Call for papers for the ...	10/01/2011 15:36

figure 1. A list of conferences call for papers and reminders from various mailing lists.

¹ Terms originally coined in August 2007 at PodCamp Pittsburgh



figure 2. A weekly newsletter from a company.

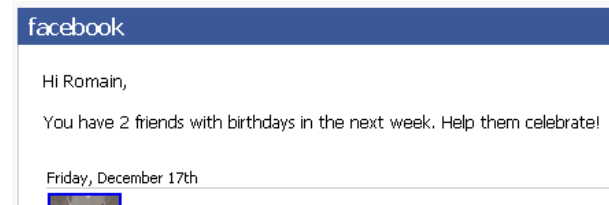


figure 3. A social update from a social network.

What is an email digest?

An email digest is a message that results from the combination of other messages (example figure 4). It is currently generated by mailing list management systems (MLMSs) such as Mailman [5] or Listserv [6]. It is composed of two parts: *wrapping* and *content*. The *wrapping* displays user's subscription that provides guidelines such as netiquettes and a summary of the compiled emails. The *content* is the digest itself, in other words the compilation of email messages that have been sent to the mailing list by subscribers, generally ordered by ascending order of date. Subscribers can set up their MLMS account to get digests on a time basis (daily, weekly) or on a number of messages basis (every 10 or 100 messages).

You can reach the person managing the list at
graphviz-interest-owner@research.att.com

When replying, please edit your Subject line so
than "Re: Contents of graphviz-interest digest..

Today's Topics:

1. Re: Large Dot Files and GraphViz (Emden R.
2. Re: Large Dot Files and GraphViz (Patrick

Message: 1

figure 4. Example of a current email digest generated by Mailman [5].

Email digests are an email-based approach compliant with the RFC 822 [3] standard. Thus it can be read with any email client, and classic email interactions on the digests such as reading, labeling and forwarding are possible. No add-ons, specific software or hardware is needed. Most digests implementation produce text with a minimal structure, but they can also have embed HTML.

Other systems than MLMS have a digests-like feature. Google Groups propose an "Abridged summary" of groups' activities, YouTube allows users to group their notification emails (such as new comments on videos). While some systems do not explicitly give users the possibility to create a digest, they reduce email numbers (e.g. for social updates such as birthdays reminders (figure 3)).

Motivation in visualizing digests templates

Generalizing digests means extending current digests to combine any set of email, and produce visualizations that are more compact than the classic linear digests (figure 4). Visualizations aim at assisting users to better perform a task by removing redundant email content, adding layout and graphical variables. In other words, making users focus on the task and preventing context switches. It can be assumed that users will then limit repetitive and error-prone interactions, and make better decisions by reducing their cognitive load.

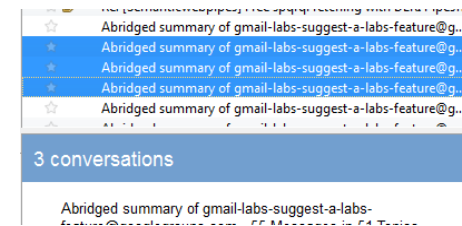


figure 5. Email messages selection with Mozilla Thunderbird.

Our approach differs from other systems such as Semanta [8] or any other external visualization [9] in the way that respectively no automatic processing is done on the data, and we remain within the current email client interface. We only optimize some currently underused screen real estate. Indeed when selecting multiple messages² (figure 5) no visualization is provided within the email content location, apart from information about the ongoing email selection.

² In the case where the email client interface is divided in multiple views: folders, messages list and content. This is mostly the case, with the notable exception of Gmail which is a thread-centric.

Digests creation steps

The digests creation steps follow a classic pipeline process (figure 6) that starts from the message combination (selected by the user in an email client), and ends with a digest that is viewed by the user (in his very same email client). Users have the possibility to change the digest on the fly while keeping the same selected emails set.

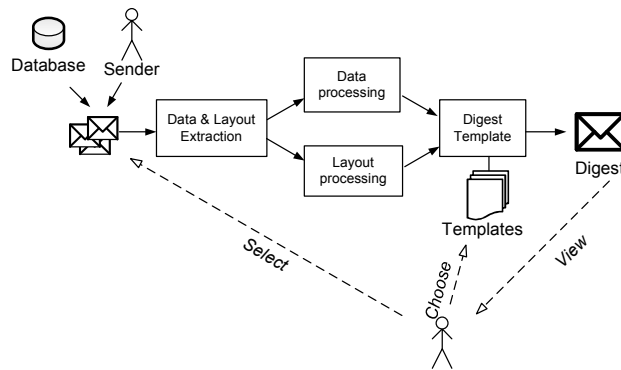


figure 6. Digest creation steps.

1. Data model and layout extraction. The first and most crucial step consists in separating data model from the view. Since emails are text, they contain both the model and the view with no separation, if not encoded in HTML. However the data model can be extracted using Natural Language Processing (NLP) techniques to identify locations, dates and acts of speech. Layout is parsed using line breaks and HTML, and can also aid to retrieve the model (e.g. via tables or lists) and to construct the digest if it requires an original layout.

2. Data processing. Data can be selected, aggregated, ordered, quantified or some arithmetic computation can be applied to them.

3. Layout processing. Operations on the layout such as deleting redundant graphical elements or adding graphical variables.

4. Digest templates. Both data model and layout are assembled using an email template that has been previously constructed by a designer and selected by the user regarding a specific task he wants to achieve.

Digests templates

One of the particular aspects of our approach is the use of digest templates. Templates are a common way to separately store the view from the data. Each template is dedicated to a particular task and results in a visualization embedded in an email. Visualization is mostly static due to the lack of minimal Javascript in email clients. A list of templates including well-known visualizations are described below (illustrated figure 7):

- *Classic.* Similar to existing digests where emails are combined linearly (similar as figure 1).
- *Geo-temporal.* Locations and dates are respectively displayed on a Google Map and a timeline.
- *Meeting.* Includes a view using the Doodle [7] website look and feel to perform a meeting organization.
- *Word-cloud.* A compact visualization where words are encoded according to their frequency (by size).
- *Graph.* A graph with people connection based on their communications, created using the GraphViz library.
- *Original template.* A reconstruction of the original emails layout, but removing redundant layout.

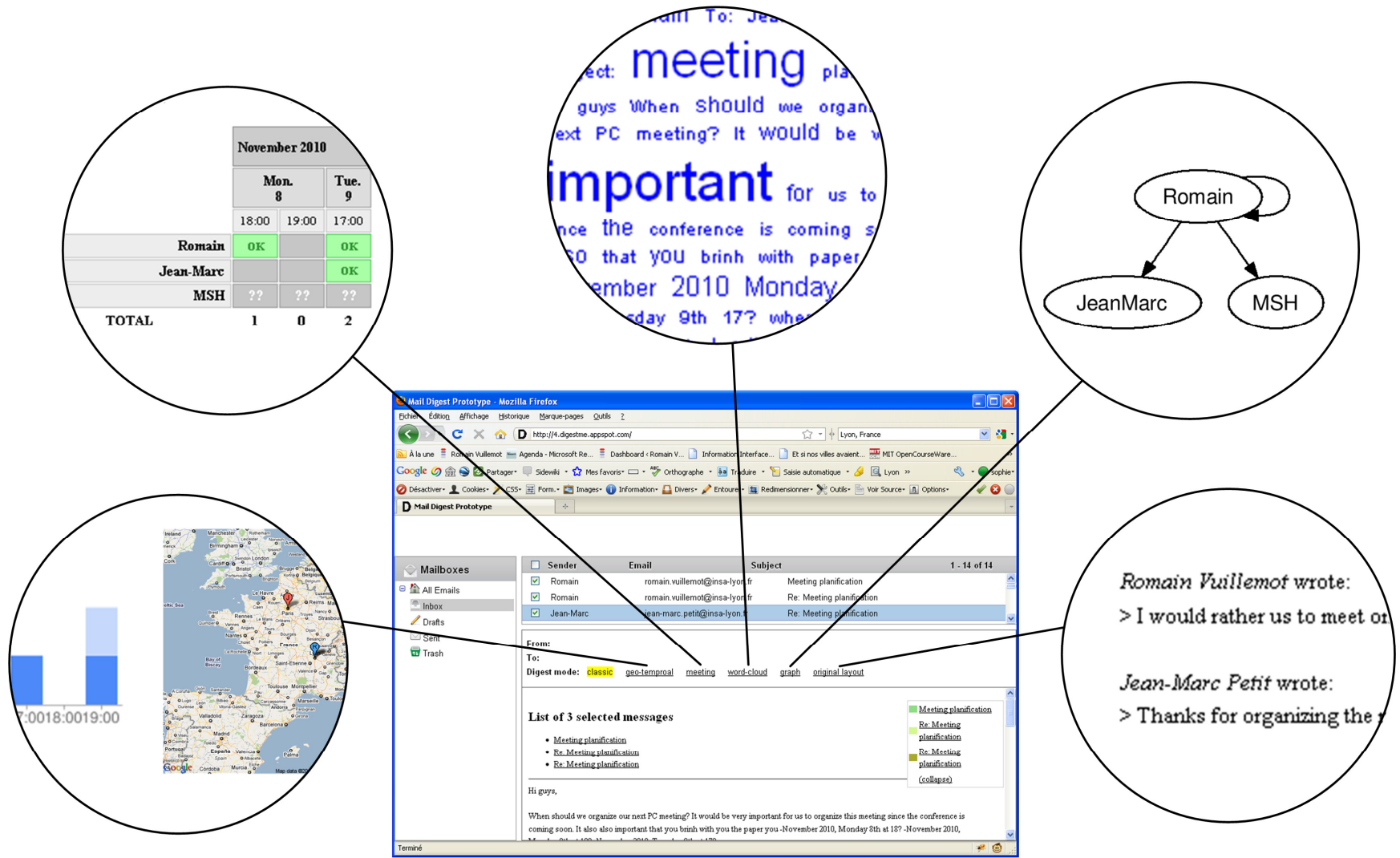


figure 7. A webmail (middle) implementing various digests. Available digest templates are listed between the email list and email content, and can be selected by users. Once a digest is selected, it appears in place of the email content. The digest templates we already implemented are the followings (clockwise starting from bottom left): geo-temporal with a timeline and a Google Map, meeting similar as Doodle [7], a word cloud where most frequent words appear bigger and where the order of words is the order they appear in emails, a communication graph connecting people based on the messages they exchanged and finally the original layout with redundant content removed.

Prototype

We implemented³ (figure 7) an early version of the digests as a webmail using the Google Web Toolkit's (GWT) [2] environment. Digests are remotely generated via web services, with the user's selected emails as parameters. This type of architecture enables extensibility and reusability, so digests can be used in any email client. Adding new templates (i.e. supporting new tasks or adding new visualizations) does not trigger any update on the client side. Finally, once a digest is generated and appears in the webmail, it can be archived, labeled or forwarded like any other email message since it is RFC 822 [3] compliant.

Next steps

Evaluating the system is the primary next step, to understand its acceptability and to compare performances with other external applications. Technical issues such as guarantying quick response time, privacy and security management must be taken care of due to the sensitivity of email communications.

NLP analysis may sometimes lead to error or just miss some features in the text. Digests must be improved to deal with such issues and communicate it to the user.

Emails headers and attachment have not been considered yet. They will also require some design strategies such to compile senders or subjects too.

Finally we think a collaborative way to design and share digests templates may be a pragmatic approach to provide users a wide variety of digests alternatives and

enable a wider variety of visualizations. Analyzing digests usage logs (from the digests web services) will provide a useful feedback loop to understand users' behavior and iterate on the digests design.

Acknowledgements

This work takes part in the ongoing research project DLM 3.0 (<http://www.dlm30.com/>) devoted to semantic email, and has been partially funded by the French Ministry of Economics and Labor.

References

- [1] BACN. <http://en.wikipedia.org/wiki/Bacn>. Retrieved January 2011.
- [2] Google Web Toolkit. <http://code.google.com/webtoolkit/>
- [3] Crocker, D. *Standard for the format of ARPA Internet text messages*. RFC 822 (August 1982).
- [4] Whittaker, S., Sidner, C. *Email overload: exploring personal information management of email*. In Proc. CHI 1996, ACM Press (1996), 276-283.
- [5] Mailman, the GNU Mailing List Manager. <http://www.gnu.org/software/mailman/>
- [6] LISTSERV - Email List Management Software. <http://www.lsoft.com/products/listserv.asp>
- [7] Doodle. <http://www.doodle.com/>
- [8] Scerri, S., Davis, B., Handschuh, S., Hauswirth, M. *Semanta - Semantic Email Made Easy*. In Proc. ESWC 2009. Springer-Verlag, Berlin (2009), 36-50.
- [9] Viégas, F., Golder, S., Donath, J. *Visualizing email content: portraying relationships from conversational histories*. In Proc. CHI 2006, ACM Press (1996), 979-988.

³ A live version is available at <http://4.digestme.appspot.com/>