

Organizing Email

BarCampZurich, 28 Oct 2006

Gabor Cselle, mail@gaborcselle.com

<http://www.gaborcselle.com/>



Overview

We focus on the problem of
sensibly structuring emails in the inbox.

1. Problem: Email Overload
2. Existing Approaches
3. Topic-Based Email: BuzzTrack
4. BuzzTrack Demo
5. BuzzTrack Architecture
6. Future Work



Email Overload

- Email clients were not designed to handle volume and variety of messages users are dealing with today:
 - Large volumes of email
 - Task Management
 - Personal Archiving or Filing
 - Keeping Context



Existing Approaches

- Today's Email Clients
- Visualizations
 - Conversations
 - Exploring the Past
 - Analyzing Contacts
- Organization Schemes
 - Automatic Foldering
 - People-Centered Email Clients
 - Task-based Email



Today's Email Clients: The Three-Pane View

The screenshot displays the Thunderbird email client interface in a three-pane view. The top bar includes a menu (File, Edit, View, Go, Message, Tools, Help) and a toolbar with icons for Get Mail, Write, Address Book, Reply, Reply All, Forward, Delete, Junk, Print, and Stop. Below the toolbar is a search bar with the text "Subject or Sender".

Pane 1: Folder list (left): Shows a tree view of folders for "mail@yourmap.com" and "Local Folders". The "Inbox (5)" folder is selected, showing sub-folders like Fabian, Keno, BarCamp, Blog, Crap, Cron Daemon, Google Alerts, Google Groups, Logcheck, VIS, Unsent, Drafts, Sent, Junk (2), Trash, TestSet, Autosort, AWork, GaborTestSet, LastMails, LastMailsSet, LSATest, and May2006.

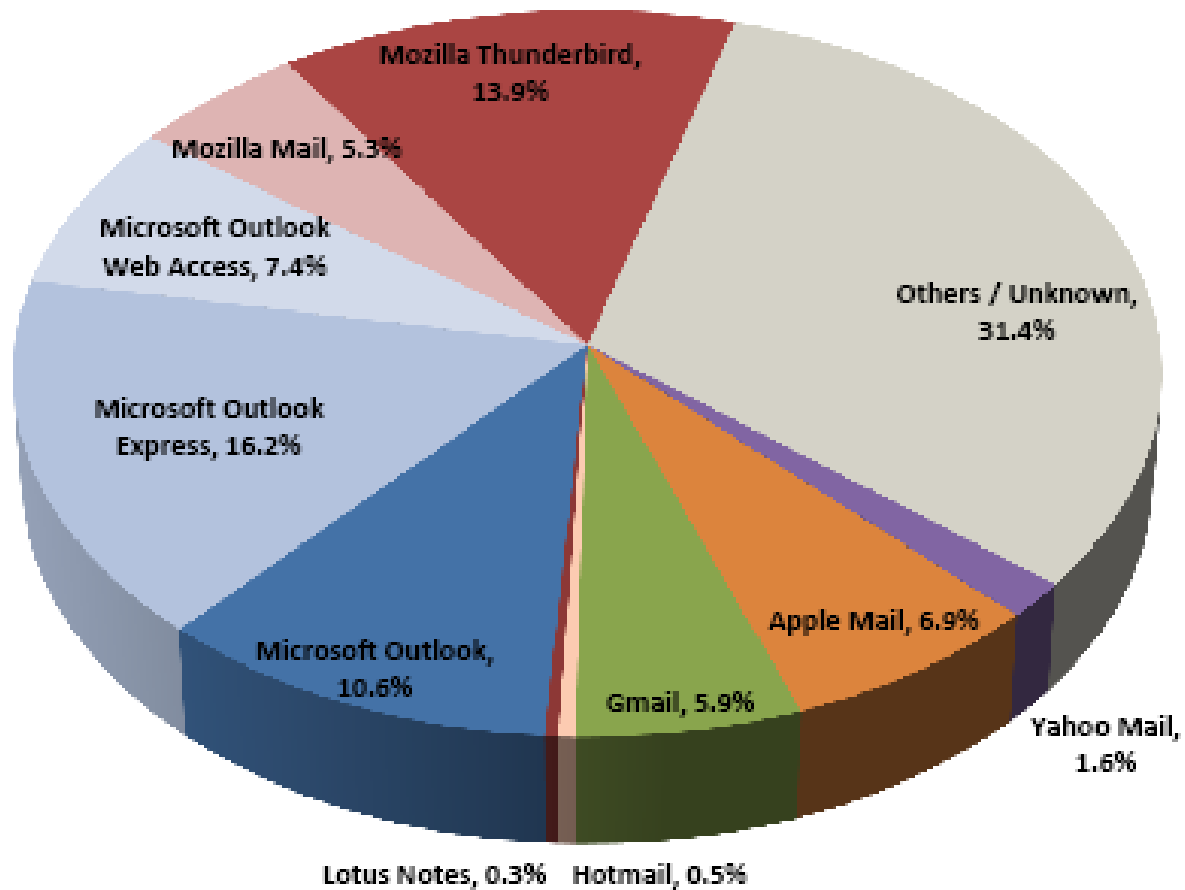
Pane 2: Email list (middle): A table of email messages with columns for Subject, Sender, Date, and Size. The selected message is "Re: Social Networking Reference" by Keno Albrecht, dated 7/12/2006 10:24 AM, with a size of 3KB.

Subject	Sender	Date	Size
Re: Big G	fabian.bannwart@ubs.com	7/12/2006...	2KB
Re: Weekend?	Fischli, Nicolai (CH - Zurich)	7/12/2006...	3KB
Big G	fabian.bannwart@ubs.com	7/12/2006 1...	5KB
Gabor /.ed	Bernhard Ager	7/12/2006 1...	1KB
einer der letzten beach partie...	Yvonne Anne Oswald	7/12/2006 1...	3KB
AW: Lost contact	TCselle	7/12/2006 1...	4KB
WG: Please update platit.com ...	TCselle	7/12/2006 1...	14KB
Re: Social Networking Reference	Keno Albrecht	7/12/2006 1...	3KB
Re: Ich liebe Klischees...	Kuhn Moritz (KSCC)	7/12/2006 9...	2KB
Re: Semesterferien...	Gabriel Dondi	7/12/2006 9...	4KB

Pane 3: Email content (right): Shows the content of the selected email. The subject is "Re: Social Networking Reference". The sender is Keno Albrecht <kenoa@tik.ee.ethz.ch>. The date is 7/12/2006 10:24 AM. The recipient is Gabor Cselle. The body text reads: "Hm, eigentlich hatte ich an eine standard Social-Networking Referenz gedacht, aber jetzt bin ich eigentlich sehr gluecklich ueber das erste Paper (hatte ich vergessen). Gruesse, Keno". Below this, it says "Gabor Cselle wrote:" followed by "Hallo Keno, bin heute wahrscheinlich nicht da. Meinst du Spam + Social Networks? Dann dieses Paper: @article{leveraging_spam,".

At the bottom right, there are status indicators: "Unread: 5" and "Total: 9994".

Today's Email Clients



Source: User agent strings / domain names in author's email.

Visualizations



Visualizations: Conversations

Gmail (Google)

The screenshot shows a Gmail conversation thread titled "BTNode ACL Pkt Acknowledgement" in the "Inbox". The thread consists of six messages from different contacts: Suguna Subramanian, Matthias Ringwald, and Kevin Martin. Annotations highlight specific UI features: "common conversation title" points to the thread title; "preview of email content" points to the first line of the first message; "Individual emails in conversation: fold out on click" points to the star icon on the left of each message; and "one color per contact" points to the colored name of the sender in each message.

common conversation title

BTNode ACL Pkt Acknowledgement

preview of email content

[Inbox](#)

Individual emails in conversation: fold out on click

one color per contact

★ Suguna Subramanian Hello, Fine..So the function deals with the ATTI Aug 18

★ Matthias Ringwald hi That's what I guess: In order to get only the mill Aug 18

★ Suguna Subramanian Hi, Many Thanks for your reply... I have few qt Aug 21

★ Matthias Ringwald Hi On 21.08.2006, at 10:18, Suguna Subramaniar Aug 21

★ Suguna Subramanian Hi, In this function sync_generate_timestamp| Aug 22

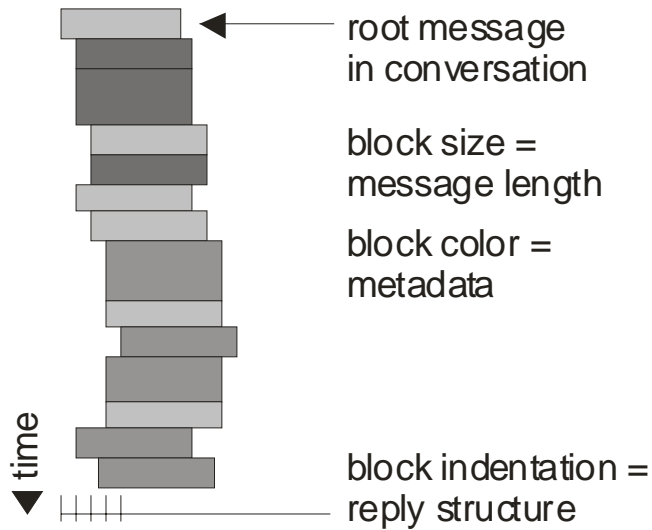
★ Kevin Martin to Suguna, btnode-develop. [More options](#) Aug 22

Hi Suguna

first of all, we are just about rewriting the time synchronization protocol. The new version of it will be available on CVS at the end of this week...

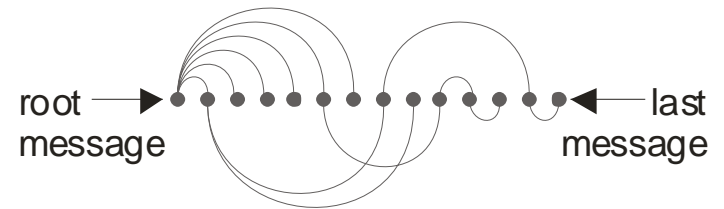
Visualizations: Conversations

Conversation Thumbnails

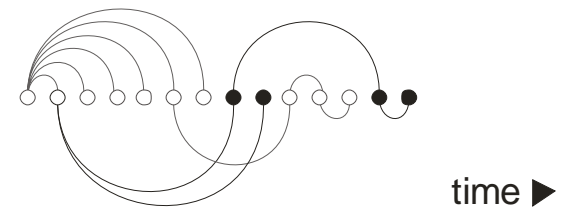


Thread Arcs

Normal State



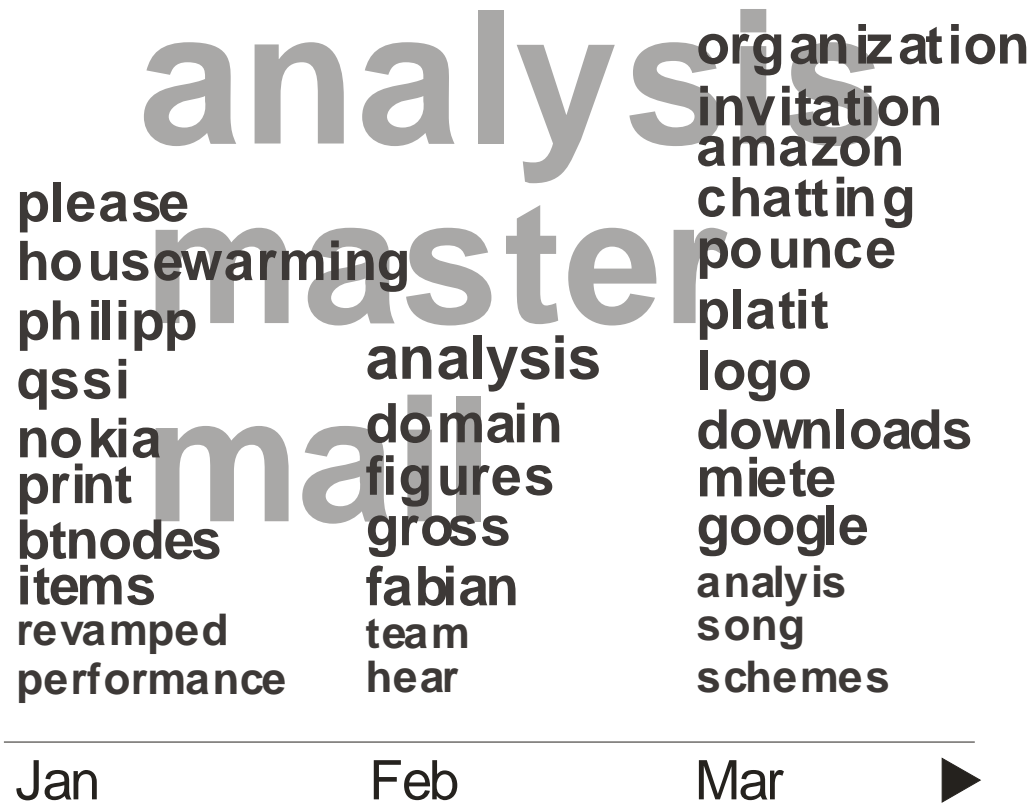
Message 2 Selected



Visualizations: Exploring the Past

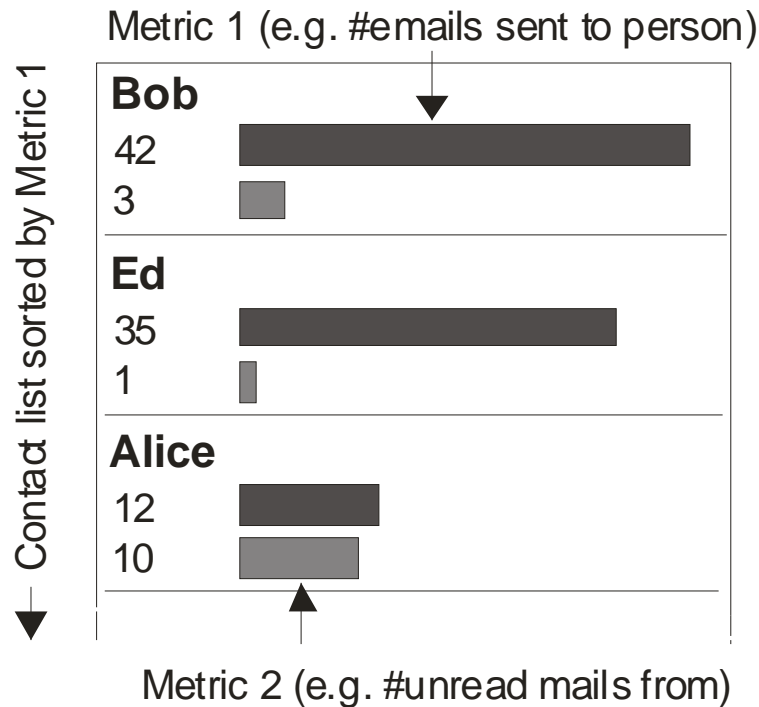
Themail

large, background: yearly words
small, foreground: monthly words

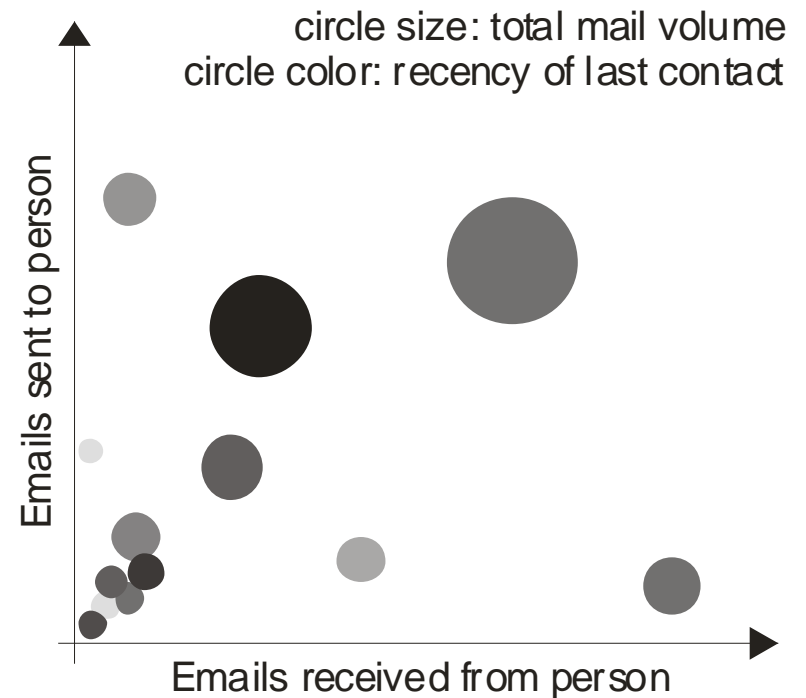


Visualizations: Analyzing Contacts

SNARF



Correspondent Crowds

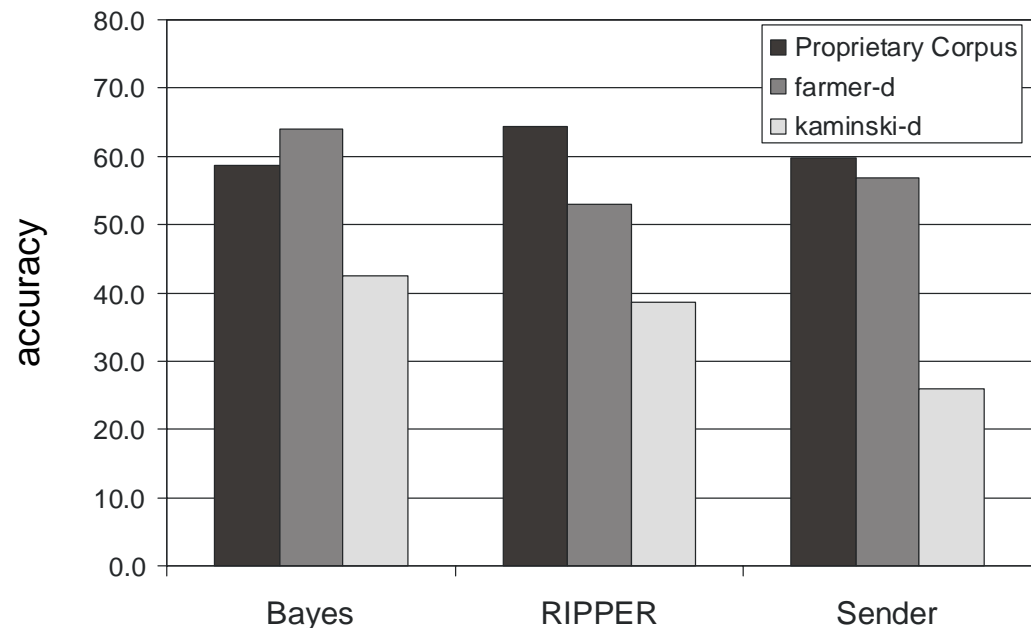
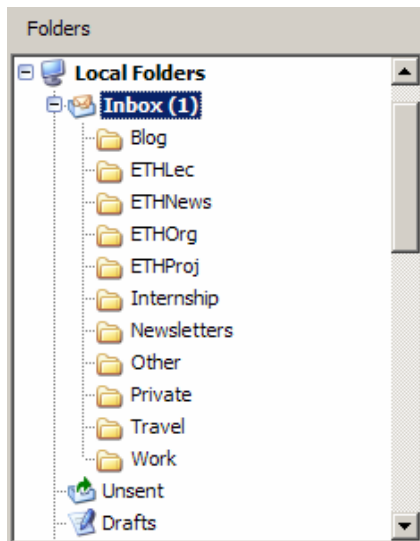




Organization Schemes

Automatic Foldering

- Using machine learning techniques to automatically move emails into folders upon arrival
- Low accuracy rates, conceptual problems:
 - Users need to manually create folders and seed them with data.

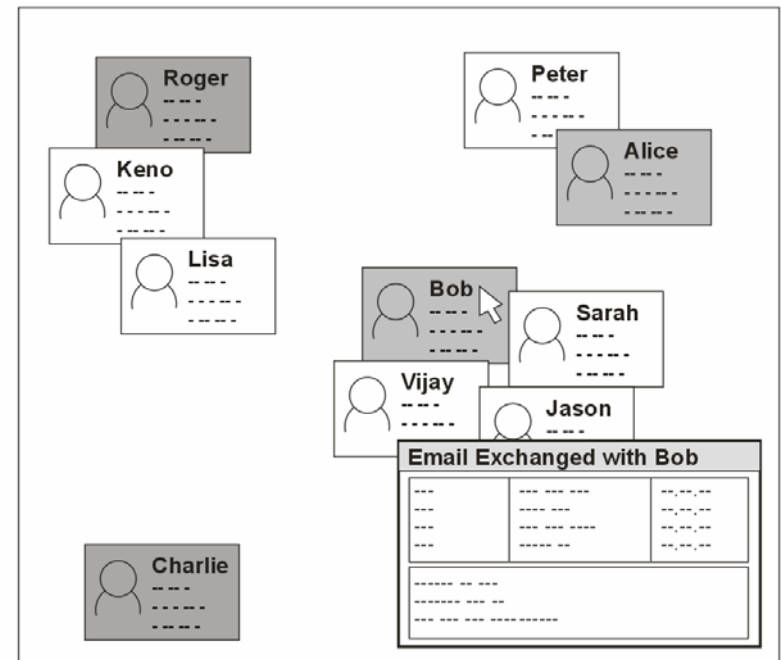


People-Centered Email Clients

Bifrost

	Who	Date	Subject
0 Timely			
	Sarah Miller	12/5/06	InfSec Colloqum TODAY
1 VIP Platinum			
•	Jack Boss Dan Chief, VP	13/5/06 12/5/06	You should work more Performance evaluations
2 VIP Gold			
	Julie Spouse Andrew Brother	12/5/06 10/5/06	You should work less Family video
3 Personal			
•	Stanley Slegel AAA Mechanics	12/5/06 12/5/06	Please review report Car repair bill
	<i>Replies</i> Fablan Frank Chris Bolliger Jan Drexler	13/5/06 12/5/06 11/5/06	Re: Screentaker Re: PIN for Nokia phone Re: Lunch
4 Small Distrib.			
	Roger Schneider Moritz Sam Jonas Burri	13/5/06 12/5/06 12/5/06	Monthly Report Please Barbecue? Someone should clean the ki...
5 Large Distrib.			
	Mathias Beutel Moritz Sam	13/5/06 12/5/06	[bt-node-devel] Re: set baudr... [bt-node-devel] set baudrate ...

ContactMap



Task-based Email

The screenshot shows the Taskmaster email client interface. At the top, there's a menu bar with 'File', 'Message', and 'Thask'. Below it are buttons for 'New', 'Reply', 'Reply to All', and 'Forward'. The main area is divided into two sections: a list of tasks and a detailed view of a selected email.

Task List:

- GOOGLE
- UPCOMING TALKS BY ME
- UPCOMING AND PAST TALKS BY OTHERS
- FOOTBALL
- CONTACTS
- First Meeting of Web-Based Collaboration Reading Group Tomorrow
- CHI 2003 Paper
- Fw: CFP - CSCW 2002 Workshop: Redesigning Email for the 21st Century
- MAC STUFF
- Renew your ACM Membership today!
- CS paper - submission (ask Peter first)
- File Server Outage: parcweb, thelma, louise - saturday 8AM-3PM
- G4 Tips

Email List Table:

Sender	Subject	Due Date	Received
Victoria Bellotti	It's mine ALL MINE		07/15 22:58
Victoria Bellotti	Is anyone around today?		07/17 11:54
[Saved Message]	Fw: Latest Draft		07/17 16:18
Ducheneaut, Nicolas <nicolas@parc.com>	Fw: Latest Draft		07/17 16:18
Victoria Bellotti	Latest Draft of CHI 2003 Paper... Your tu		07/17 16:30
[Attachment]	CHI2003-Taking_email_#3A402.doc	07/18/2002	07/17 16:30
Smith, Ian <iansmith@parc.com>	call Chris about paper		07/17 16:38
Victoria Bellotti	CHI Papers Guidelines		07/17 16:47
[Link]	http://www.chi2003.org/papers.html#pref		07/17 16:47

Email Detail View:

To: [Redacted]
 Subject: CHI2003-Taking_email_#3A402.doc
 Due Date: 07/18/2002

ABSTRACT

There is widespread agreement that email plays a central role in task and project management for many users, yet email tools' features have remained surprisingly static over the past decade and do not reflect their users' evolving practices. The Taskmaster system seeks to bridge this gap by recasting email as task management and embedding the necessary resources directly in the email client. In this paper, we describe an innovative interface breaking the barrier between the traditionally compartmentalized elements of most email applications (e.g. tools, calendar, day [e.g. Hiltz and Turoff, Whitake

Acknowledging this problem, a small number of users have recently tried to overhaul email's traditional interface. Among them, many have tried to diminish the cognitive burden associated with filing messages by automating this process (papers [see for instance XX]). This, however, only addresses the problem of the volume of email users receive. research projects have explored different approaches to these, CAFE [Takkinen 1998] is proposed.

Example:
TaskMaster
 (PARC [2])

thasks

thask
 contents

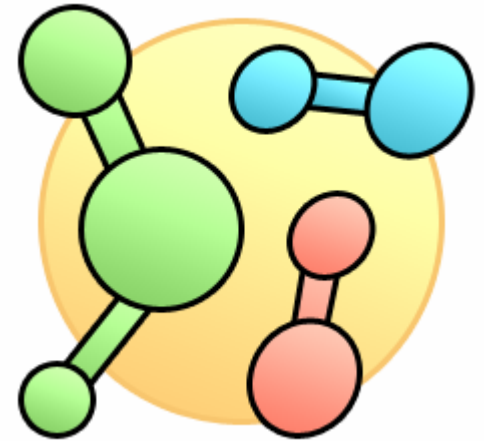
item
 contents
 (emails,
 documents, etc.)

BuzzTrack



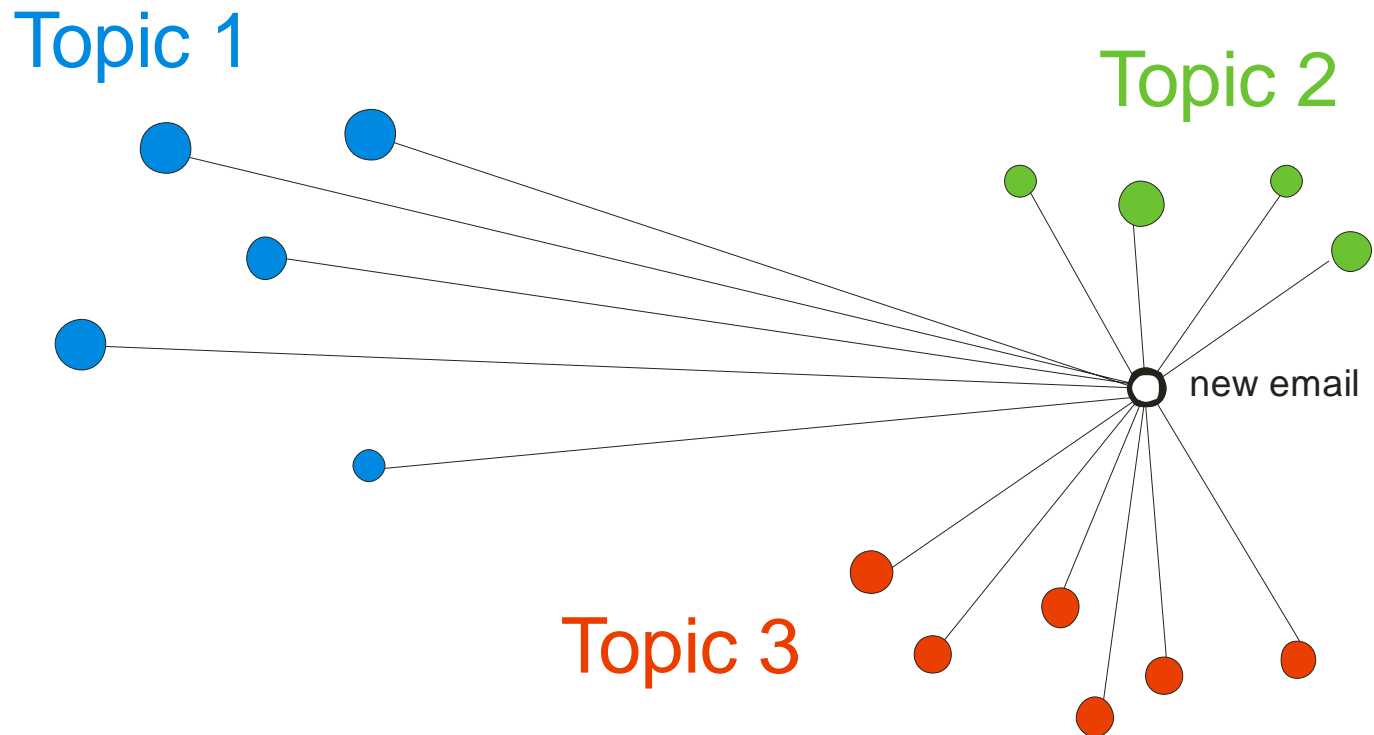
BuzzTrack

- **Mozilla Thunderbird extension** to automatically group related emails into topics.
- Will be distributed through website: **www.buzztrack.net**
- **Topics** are groups of emails that relate to the same idea, action, event, task, or question.
- We focus on organizing the user's inbox only.
- **Paradigm shift:** Fast full-text search is today's solution to finding emails in archives



BuzzTrack: Clustering Methods

- Single-link clustering: Newly incoming emails are compared to every email in existing topics:
 - Similarity value $>$ threshold: assigned to topic
 - Similarity value \leq threshold: email starts new topic



BuzzTrack: Clustering Features

- How do we generate similarity values between emails?
- Via a linear combination of several similarity features.
- **Examples:**
 - Text similarity (TFIDF Value, cosine similarity metric)
 - People similarities (comparing sets of people in the From / To / Cc lines of email headers)
 - Thread membership

BuzzTrack: Clustering Features

Other features for deriving similarities:

- Subject similarity
- Sender domain overlaps
- Sender rank and percentage
- % of email from sender that is answered
- Time passed since last email in topic
- People and reference count for email
- Known people and reference %
- Cluster size
- Has attachment

BuzzTrack: Decision Score

Similarities are combined into a decision score for each email / cluster pair through a linear combination of feature values:

$$\text{dec}_{i,j} = w_a * \text{sim}_a(m_i, C_j) + w_b * \text{sim}_b(m_i, C_j) + \dots$$

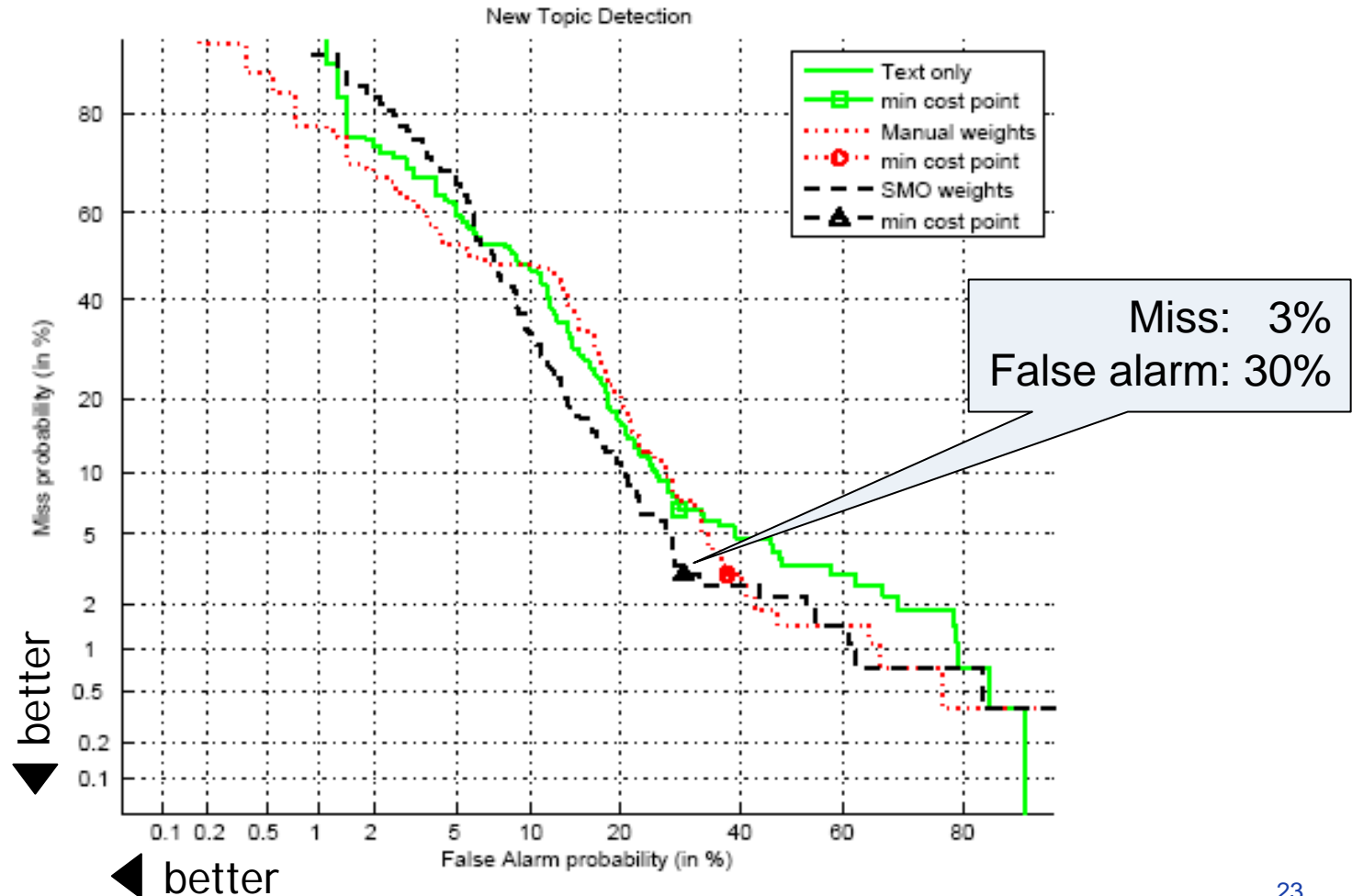
Weights derived by manual trial-and-error, linear SVM on a development set of emails.

BuzzTrack: Clustering Results

- How do we evaluate clustering quality?
- Topic Detection and Tracking competitions by National Institute of Standards and Technology (NIST) [3,4]. Aimed at clustering news articles.
- We created a corpus from 2403 own messages.
- Clustering Task is split into subtasks:
 - **New Topic Detection (NTD):**
Given stream of emails, which ones start new topics?
 - **Topic Tracking (TT):**
Given a fixed topic, which newly incoming emails belong to it?
- DET Curves plot miss rate vs. false alarm rate for possible threshold for decision scores $\text{det}(m_i, C_j)$

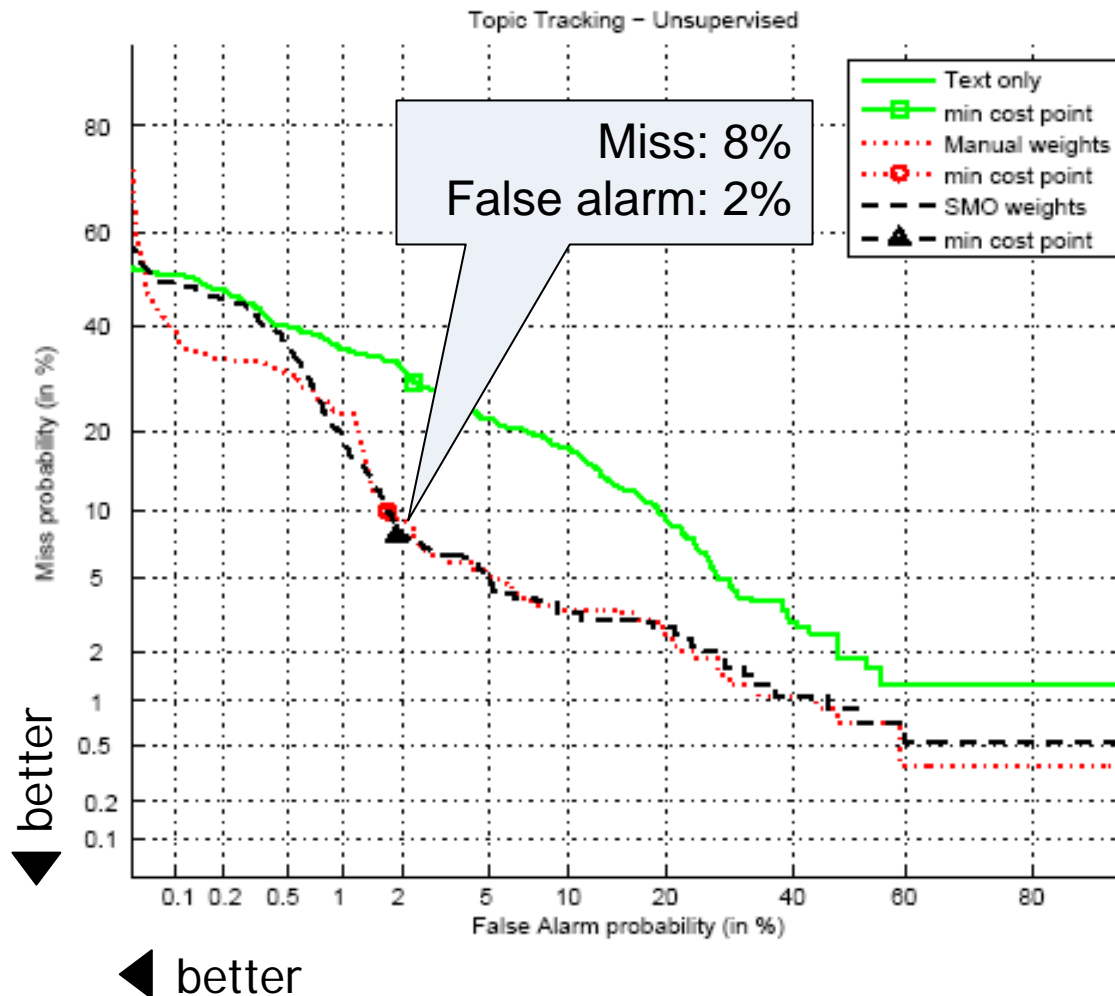
BuzzTrack: Clustering Results

- TDT New Topic Detection Task



BuzzTrack: Clustering Results

- TDT Topic Tracking Task



BuzzTrack Alpha Demo

Inbox - BuzzTrack View - Thunderbird

File Edit View Go Message Tools BuzzTrack Help

Get Mail Write Address Book Reply Reply All Forward Delete Junk Print Stop BuzzTrack

BuzzTrack

BuzzTrack Inbox Topics

Subject	Sender	Date
in ZRH soccer Apu		
Re: [Invitation] Spiel Schweiz-France @ Tue, Jun 1...	Apu Shah	3:42 PM
Re: In ZRH: 6/12-6/14	Apu Shah	6/13/200...
Re: In ZRH: 6/12-6/14	Apu Shah	6/13/200...
In ZRH: 6/12-6/14	Apu Shah	6/12/200...
vote Reddit submission spam		
Re: Reddit Kommentar	Keno Albrecht	1:15 PM
Re: Reddit Kommentar	Keno Albrecht	9:55 AM
Vote for submission reddit	Keno Albrecht	6/12/200...
test our website		
Re: test our website	Ed Freyfogle	6/12/200...
Re: test our website	Javier Extebeste	6/12/200...
test our website	Ed Freyfogle	6/9/2006 ...
Xobni Beta Outlook relationship		
Re: New Beta	Adam Smith	10:02 AM
Re: Relationship view	Matt Brezina	6/12/200...
New Beta	Adam Smith	6/10/200...
Xobni update	Matt Brezina	6/9/2006 ...

Subject: In ZRH: 6/12-6/14
From: Apu Shah <...>
Date: 6/12/2006 2:24 PM
To: Gabor Cselle

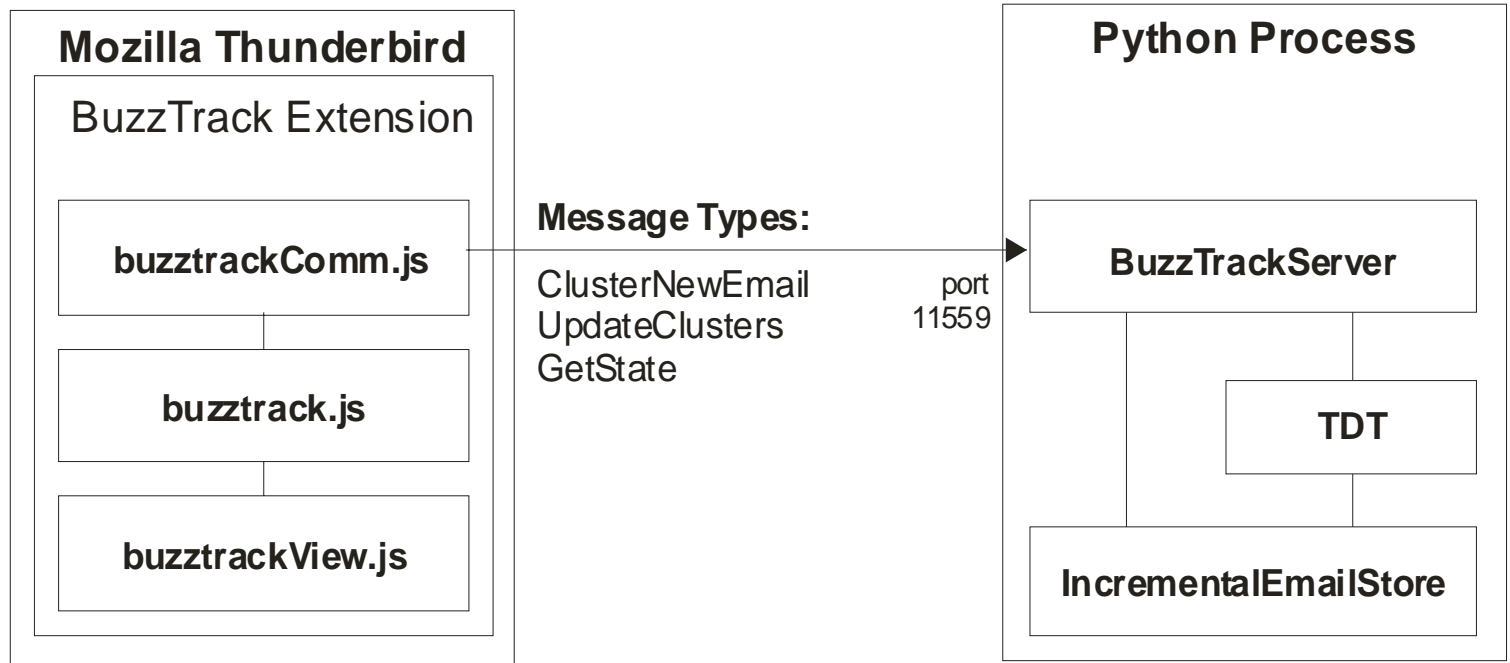
Hey Gabor,

I am in Zurich for the next couple of days. Sorry for the short notice. Let me know if you have time to hang out (dinner? worldcup? something else). I am pretty much free tonight and tomorrow night and am leaving during the evening of the 14th. I am staying

Unread: 2 Total: 9906

BuzzTrack Architecture

- BuzzTrack Plugin Architecture.



Future Work

- Distribute software, manual evaluations.
- Hierarchical clustering?
- More advanced visualizations of similarity
- Interactive reclustering, different clustering styles
- Hope BuzzTrack will be valuable tool for real-world users

Questions?

References:

<http://www.gaborcselle.com/msthesis/>

Discussion:

- What do you need in an email clients?
- What functionality would you like to see?
- How could you use email to work better?
- Can we solve email overload via software, or only through social customs?
- For grouped views such as BuzzTracks, what are the main grouping criteria?
- What are the features of important emails?

Contact: Gabor Cselle, mail@gaborcselle.com

Thanks to: Keno Albrecht, Roger Wattenhofer!