The Future of Information Retrieval

Steve Whittaker
University of Sheffield

http://dis.shef.ac.uk/stevewhittaker

Three Main Trends

- Document Retrieval is passé, age of Multimedia is upon us
- Information Glut (Overload) not Information Famine
- Growth of Personal Information

What I will do

- Discuss the new set of IR problems
  - Multimedia
  - Information Glut
- Present potential solutions to those problems
- With reference to work from my own lab

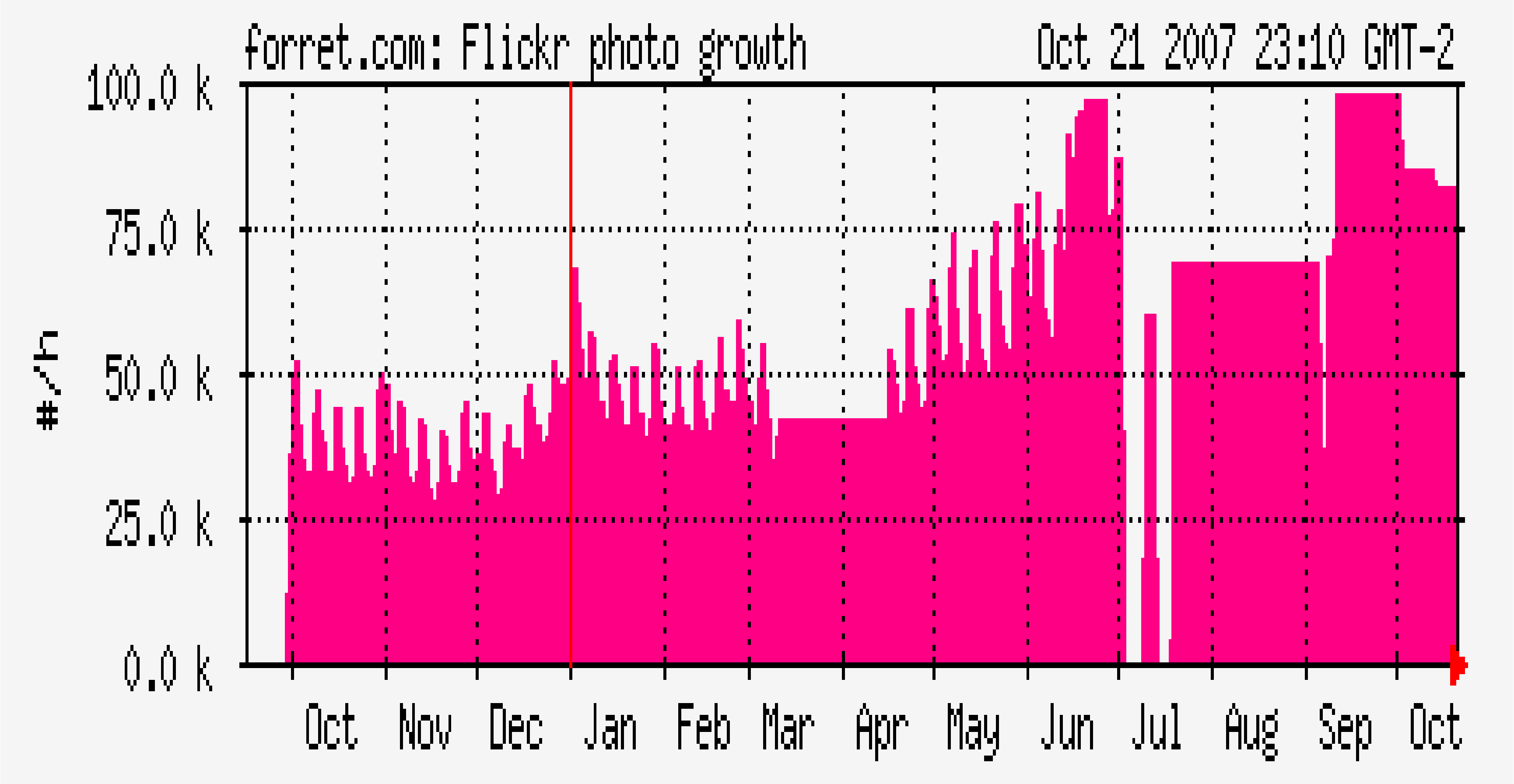
From text to multimedia IR

- Inception of the web
  - Text with a few images
  - Now text IR is a *solved problem* on the web
  - (but not for personal information or enterprise search)
- Now:
  - Video (Youtube), Images (Flickr)
  - Blogs?? – sociologically rather than technically interesting

YouTube

- In one month (Aug.2006), the number of videos grew 20% to 6.1 million
- YouTube has 45 terabytes of videos
- Video views reached 1.73 billion
- Total time people spent watching YouTube since it started last year is 9,305 years
- *Wall Street Journal 2006*

Flickr



http://blog.forret.com/projects/statistics/photo-hosting/flickr-statistics/

The technical problem with multimedia IR

- ‘Semantic gap’ – low level automatically extracted features do not generate useful information for users
- E.g. even face recognition in a picture collection is still not reliable

Closing the semantic gap

- *Content-oriented (CBIR):* Use automatic speech recognition (ASR) to generate text and then use text techniques
    - (only works when there is speech in the video)
- *Metadata*
    - Explicit content-oriented tags: *terminator, schwarzenegger, hastalavista*
    - Implicit tags: user behaviour with respect to the source

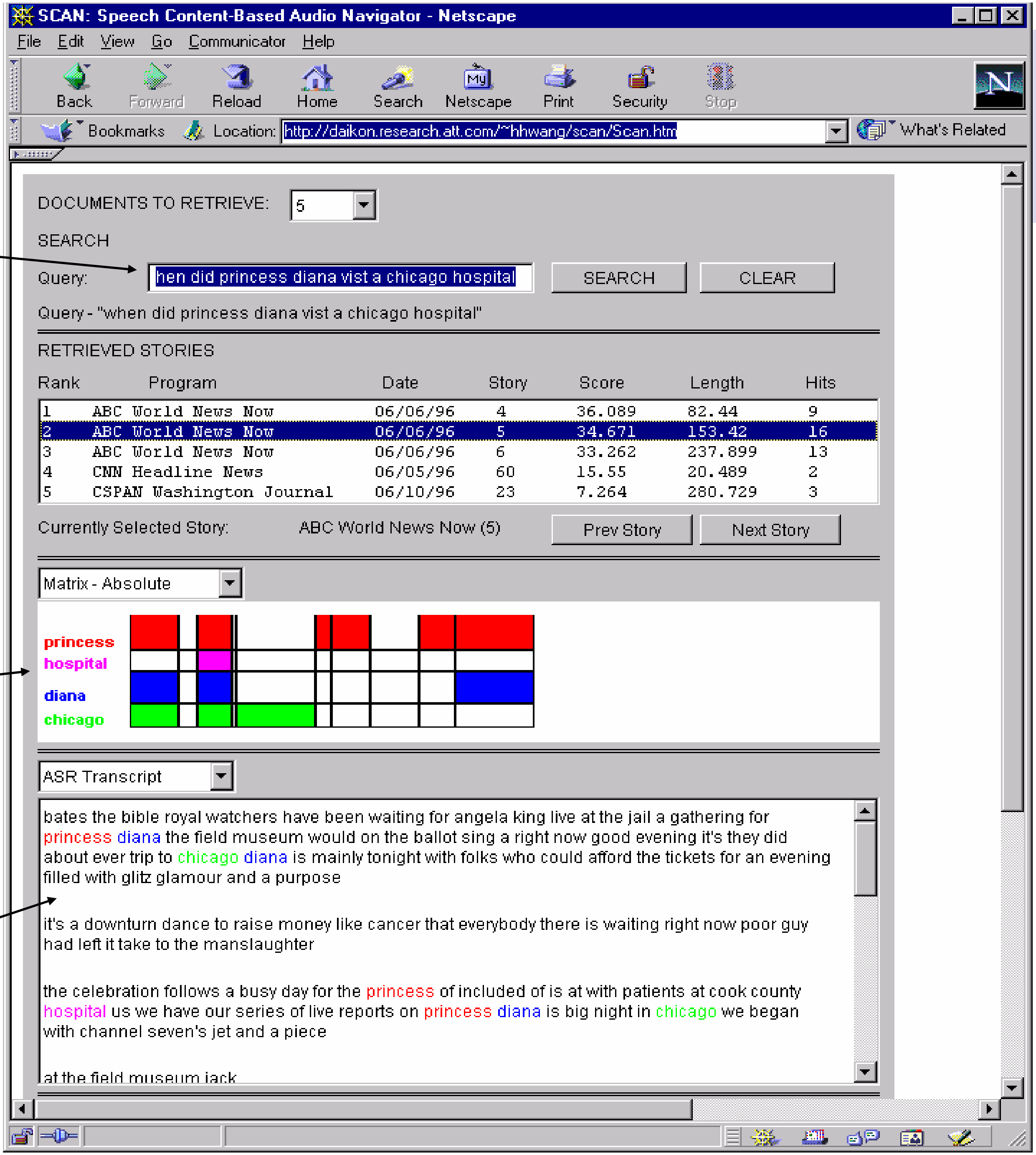
Content-based multimedia retrieval

- Speech->Text via ASR
- Then use text search techniques
- It doesn’t have to be perfect
- SCAN – retrieve broadcast news
- SCANMail – retrieve your voicemail
- Whittaker et al., 1999, Whittaker et al., 2004.

Search

Overview

Transcript


SCAN: Speech Content-Based Audio Navigator - Netscape
File Edit View Go Communicator Help
Back Forward Reload Home Search Netscape Print Security Stop
Bookmarks Location: http://daikon.research.att.com/~hhwang/scan/Scan.htm
What's Related
SEARCH
CLEAR
Prev Story
Next Story
princess
hospital
diana
chicago

DOCUMENTS TO RETRIEVE: 5

SEARCH

Query: hen did princess diana vist a chicago hospital

Query - "when did princess diana vist a chicago hospital"

RETRIEVED STORIES

Rank	Program	Date	Story	Score	Length	Hits
1	ABC World News Now	06/06/96	4	36.089	82.44	9
2	ABC World News Now	06/06/96	5	34.671	153.42	16
3	ABC World News Now	06/06/96	6	33.262	237.899	13
4	CNN Headline News	06/05/96	60	15.55	20.489	2
5	CSPAN Washington Journal	06/10/96	23	7.264	280.729	3

Currently Selected Story: ABC World News Now (5)

Matrix - Absolute

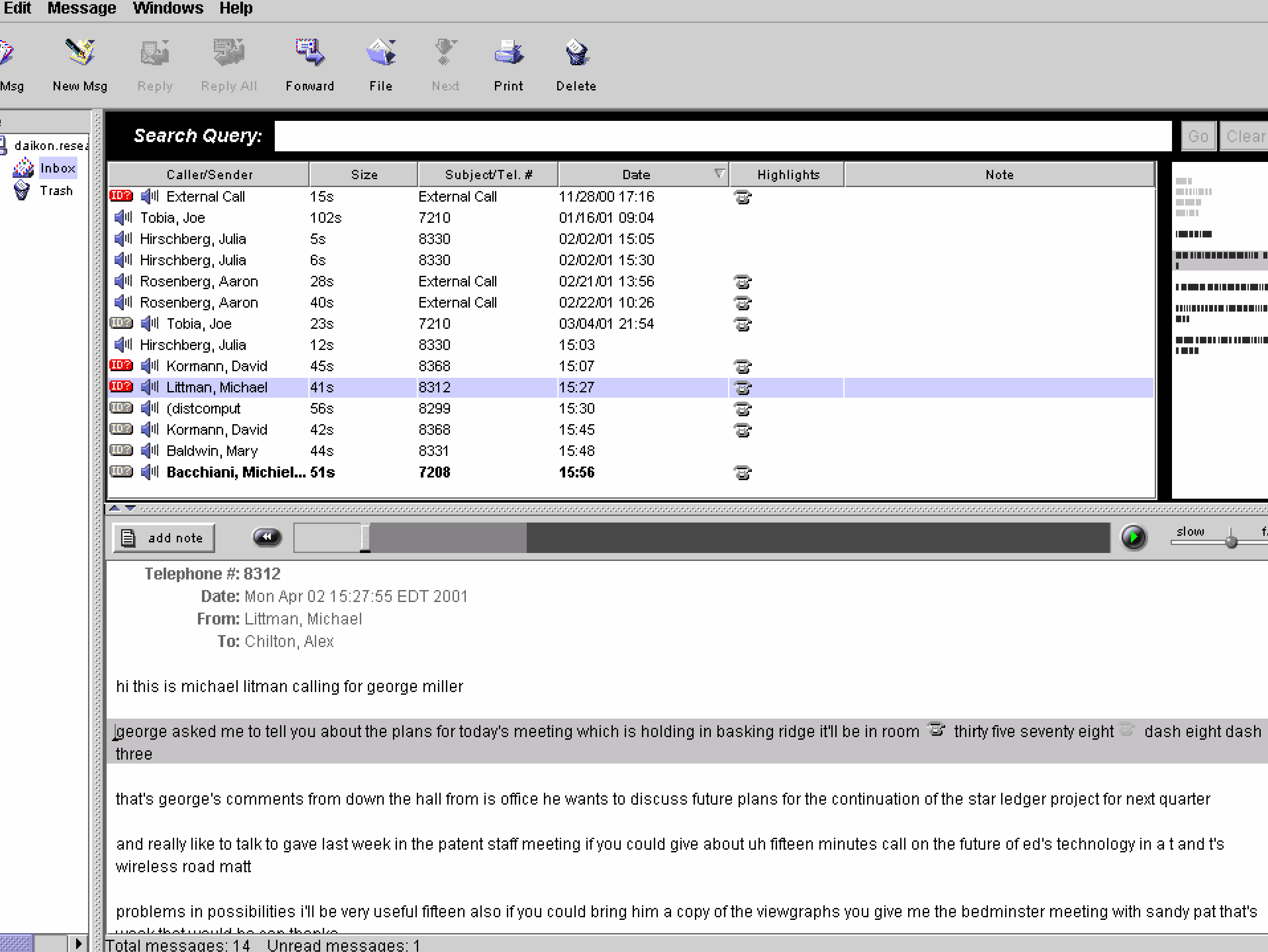
ASR Transcript

bates the bible royal watchers have been waiting for angela king live at the jail a gathering for princess diana the field museum would on the ballot sing a right now good evening it's they did about ever trip to chicago diana is mainly tonight with folks who could afford the tickets for an evening filled with glitz glamour and a purpose

it's a downturn dance to raise money like cancer that everybody there is waiting right now poor guy had left it take to the manslaughter

the celebration follows a busy day for the princess of included of is at with patients at cook county hospital us we have our series of live reports on princess diana is big night in chicago we began with channel seven's jet and a piece

at the field museum jack



Caller/Sender	Size	Subject/Tel. #	Date	Highlights	Note
External Call	15s	External Call	11/28/00 17:16	☎	
Tobia, Joe	102s	7210	01/16/01 09:04		
Hirschberg, Julia	5s	8330	02/02/01 15:05		
Hirschberg, Julia	6s	8330	02/02/01 15:30		
Rosenberg, Aaron	28s	External Call	02/21/01 13:56	☎	
Rosenberg, Aaron	40s	External Call	02/22/01 10:26	☎	
Tobia, Joe	23s	7210	03/04/01 21:54	☎	
Hirschberg, Julia	12s	8330	15:03		
Kormann, David	45s	8368	15:07	☎	
Littman, Michael	41s	8312	15:27	☎	
(distcomput	56s	8299	15:30	☎	
Kormann, David	42s	8368	15:45	☎	
Baldwin, Mary	44s	8331	15:48		
Bacchiani, Michiel...	**51s**	**7208**	**15:56**	☎	

Telephone #: 8312
Date: Mon Apr 02 15:27:55 EDT 2001
From: Littman, Michael
To: Chilton, Alex

hi this is michael litman calling for george miller

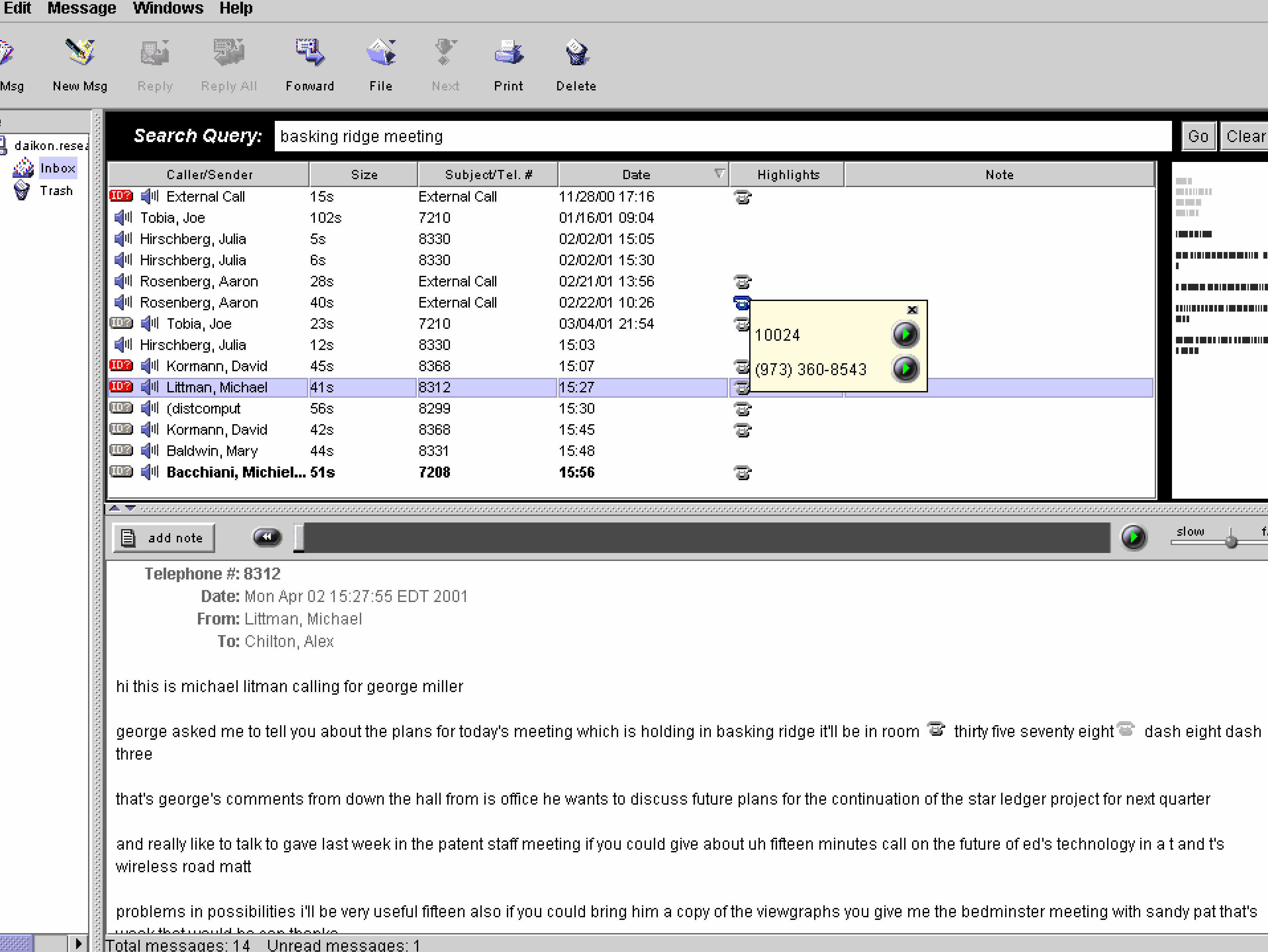
george asked me to tell you about the plans for today's meeting which is holding in basking ridge it'll be in room ☎ thirty five seventy eight ☎ dash eight dash three

that's george's comments from down the hall from is office he wants to discuss future plans for the continuation of the star ledger project for next quarter

and really like to talk to gave last week in the patent staff meeting if you could give about uh fifteen minutes call on the future of ed's technology in a t and t's wireless road matt

problems in possibilities i'll be very useful fifteen also if you could bring him a copy of the viewgraphs you give me the bedminster meeting with sandy pat that's week that would be con thanks

Total messages: 14 Unread messages: 1



Search Query: basking ridge meeting

Caller/Sender	Size	Subject/Tel. #	Date	Highlights	Note
External Call	15s	External Call	11/28/00 17:16	☎	
Tobia, Joe	102s	7210	01/16/01 09:04		
Hirschberg, Julia	5s	8330	02/02/01 15:05		
Hirschberg, Julia	6s	8330	02/02/01 15:30		
Rosenberg, Aaron	28s	External Call	02/21/01 13:56	☎	
Rosenberg, Aaron	40s	External Call	02/22/01 10:26	☎	
Tobia, Joe	23s	7210	03/04/01 21:54	☎	
Hirschberg, Julia	12s	8330	15:03		
Kormann, David	45s	8368	15:07	☎	
Littman, Michael	41s	8312	15:27	☎	
(distcomput	56s	8299	15:30	☎	
Kormann, David	42s	8368	15:45	☎	
Baldwin, Mary	44s	8331	15:48		
Bacchiani, Michiel...	**51s**	**7208**	**15:56**	☎	

10024

(973) 360-8543

Telephone #: 8312
Date: Mon Apr 02 15:27:55 EDT 2001
From: Littman, Michael
To: Chilton, Alex

hi this is michael litman calling for george miller

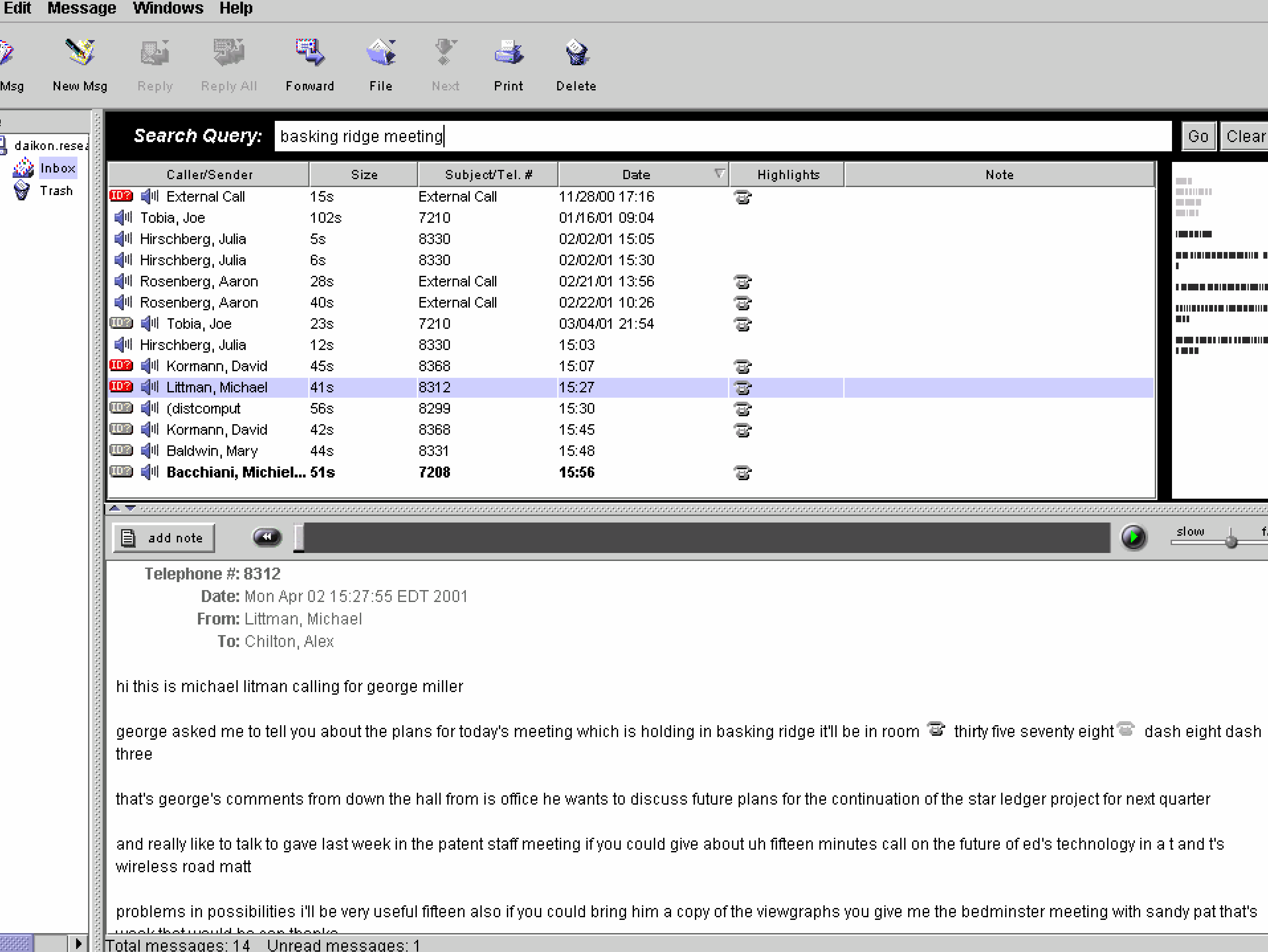
george asked me to tell you about the plans for today's meeting which is holding in basking ridge it'll be in room ☎ thirty five seventy eight ☎ dash eight dash three

that's george's comments from down the hall from is office he wants to discuss future plans for the continuation of the star ledger project for next quarter

and really like to talk to gave last week in the patent staff meeting if you could give about uh fifteen minutes call on the future of ed's technology in a t and t's wireless road matt

problems in possibilities i'll be very useful fifteen also if you could bring him a copy of the viewgraphs you give me the bedminster meeting with sandy pat that's week that would be con thanks

Total messages: 14 Unread messages: 1



Search Query: basking ridge meeting

Caller/Sender	Size	Subject/Tel. #	Date	Highlights	Note
ID? External Call	15s	External Call	11/28/00 17:16	☎	
Tobia, Joe	102s	7210	01/16/01 09:04		
Hirschberg, Julia	5s	8330	02/02/01 15:05		
Hirschberg, Julia	6s	8330	02/02/01 15:30		
Rosenberg, Aaron	28s	External Call	02/21/01 13:56	☎	
Rosenberg, Aaron	40s	External Call	02/22/01 10:26	☎	
ID? Tobia, Joe	23s	7210	03/04/01 21:54	☎	
Hirschberg, Julia	12s	8330	15:03		
ID? Kormann, David	45s	8368	15:07	☎	
ID? Littman, Michael	41s	8312	15:27	☎	
ID? (distcomput	56s	8299	15:30	☎	
ID? Kormann, David	42s	8368	15:45	☎	
ID? Baldwin, Mary	44s	8331	15:48		
ID? **Bacchiani, Michiel...**	**51s**	**7208**	**15:56**	☎	

Telephone #: 8312
Date: Mon Apr 02 15:27:55 EDT 2001
From: Littman, Michael
To: Chilton, Alex

hi this is michael litman calling for george miller

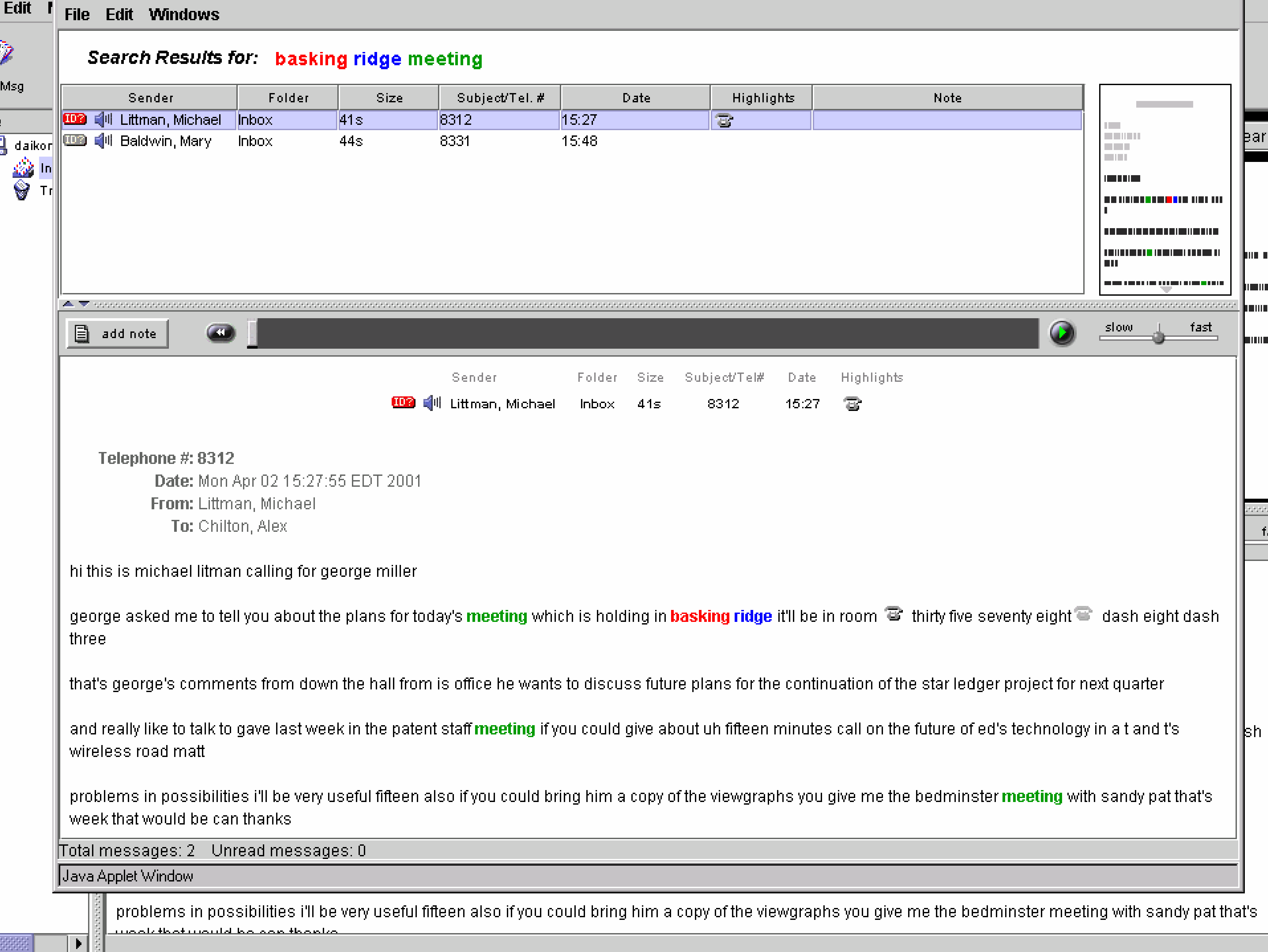
george asked me to tell you about the plans for today's meeting which is holding in basking ridge it'll be in room ☎ thirty five seventy eight ☎ dash eight dash three

that's george's comments from down the hall from is office he wants to discuss future plans for the continuation of the star ledger project for next quarter

and really like to talk to gave last week in the patent staff meeting if you could give about uh fifteen minutes call on the future of ed's technology in a t and t's wireless road matt

problems in possibilities i'll be very useful fifteen also if you could bring him a copy of the viewgraphs you give me the bedminster meeting with sandy pat that's week that would be con thanks

Total messages: 14 Unread messages: 1



Search Results for: basking ridge meeting

Sender	Folder	Size	Subject/Tel. #	Date	Highlights	Note
ID? Littman, Michael	Inbox	41s	8312	15:27	☎	
ID? Baldwin, Mary	Inbox	44s	8331	15:48		

Sender	Folder	Size	Subject/Tel#	Date	Highlights
ID? Littman, Michael	Inbox	41s	8312	15:27	☎

Telephone #: 8312
Date: Mon Apr 02 15:27:55 EDT 2001
From: Littman, Michael
To: Chilton, Alex

hi this is michael litman calling for george miller

george asked me to tell you about the plans for today's **meeting** which is holding in **basking ridge** it'll be in room ☎ thirty five seventy eight ☎ dash eight dash three

that's george's comments from down the hall from is office he wants to discuss future plans for the continuation of the star ledger project for next quarter

and really like to talk to gave last week in the patent staff **meeting** if you could give about uh fifteen minutes call on the future of ed's technology in a t and t's wireless road matt

problems in possibilities i'll be very useful fifteen also if you could bring him a copy of the viewgraphs you give me the bedminster **meeting** with sandy pat that's week that would be can thanks

Total messages: 2   Unread messages: 0

Java Applet Window

problems in possibilities i'll be very useful fifteen also if you could bring him a copy of the viewgraphs you give me the bedminster meeting with sandy pat that's

Content-based approach

- ASR doesn’t have to be perfect
- We can do very well with 60% correct
- ASR fails on words we don’t care about
- Problem: there isn’t always speech, e.g. with images
- -> user tags

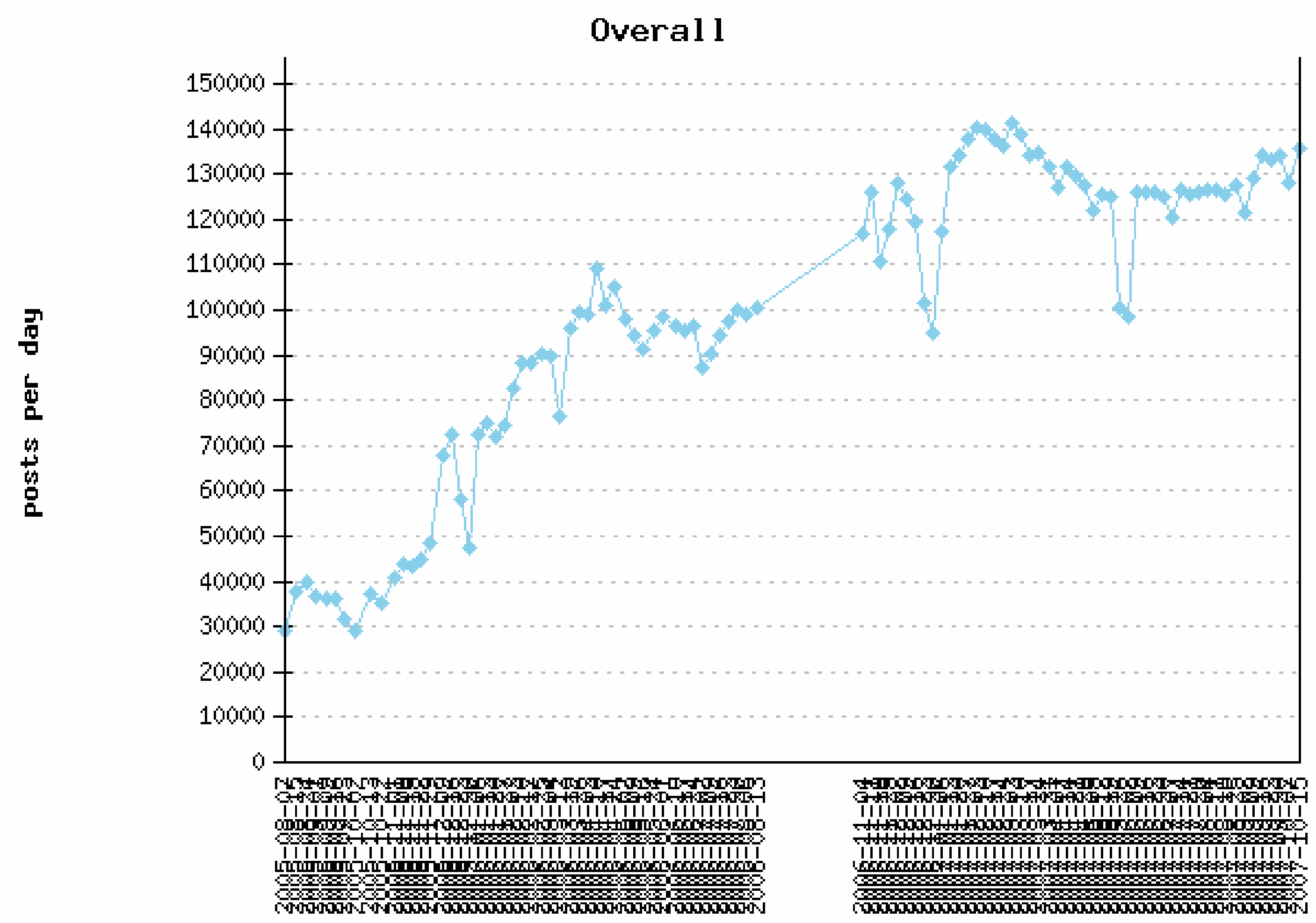
Tags

- Two types
  - Explicit
  - Implicit

Explicit tags

- Youtube, Flickr, Digg, Citeulike….
- Users generate content-oriented tags
- (Social tagging)

Del.icio.us tags

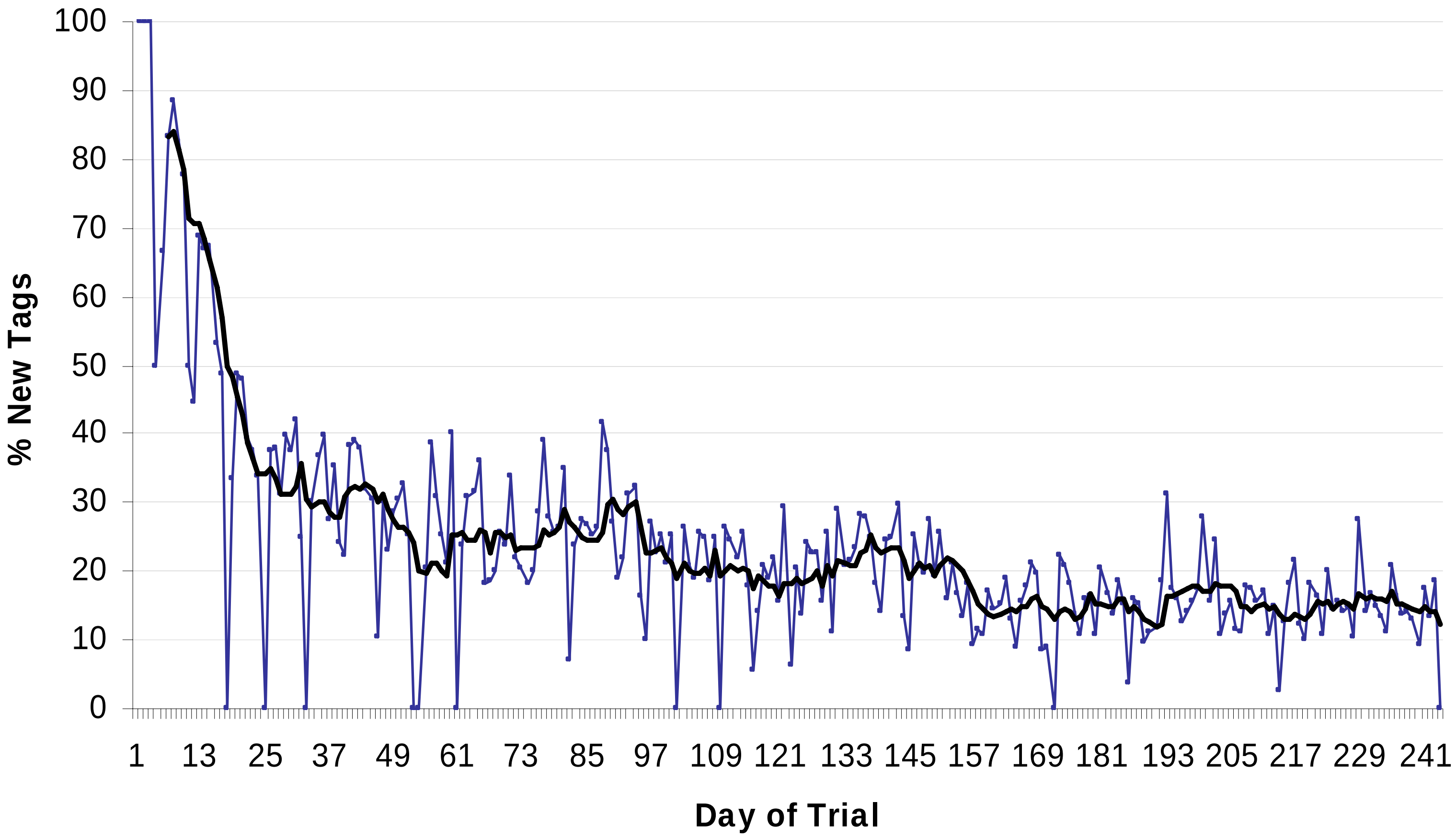


http://deli.ckoma.net/stats

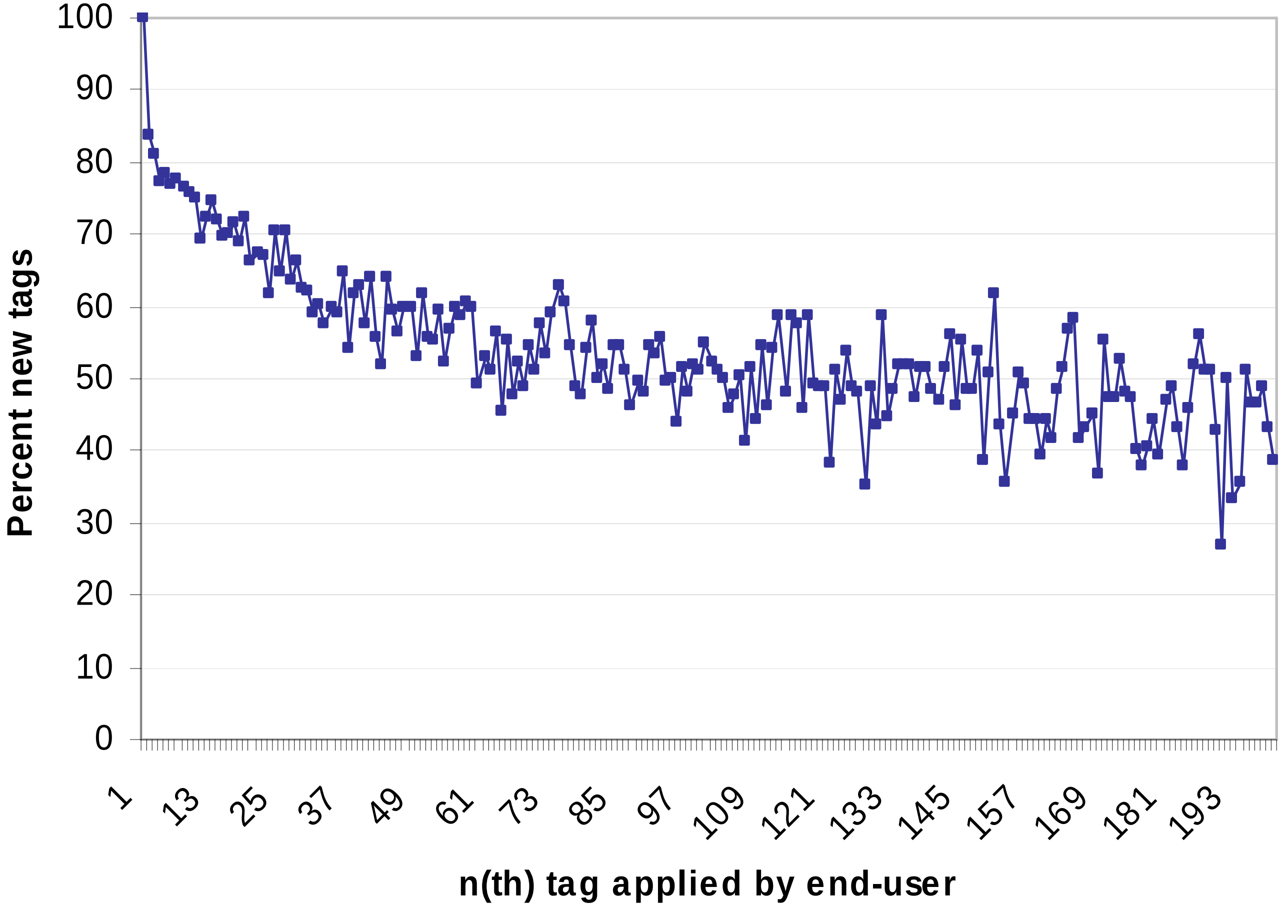
Problems with social tagging?

- Concerns of the librarians (Hammond et al. 2005)
- Ordinary people can’t create reliable taxonomies
- Inconsistent terms and ontologies

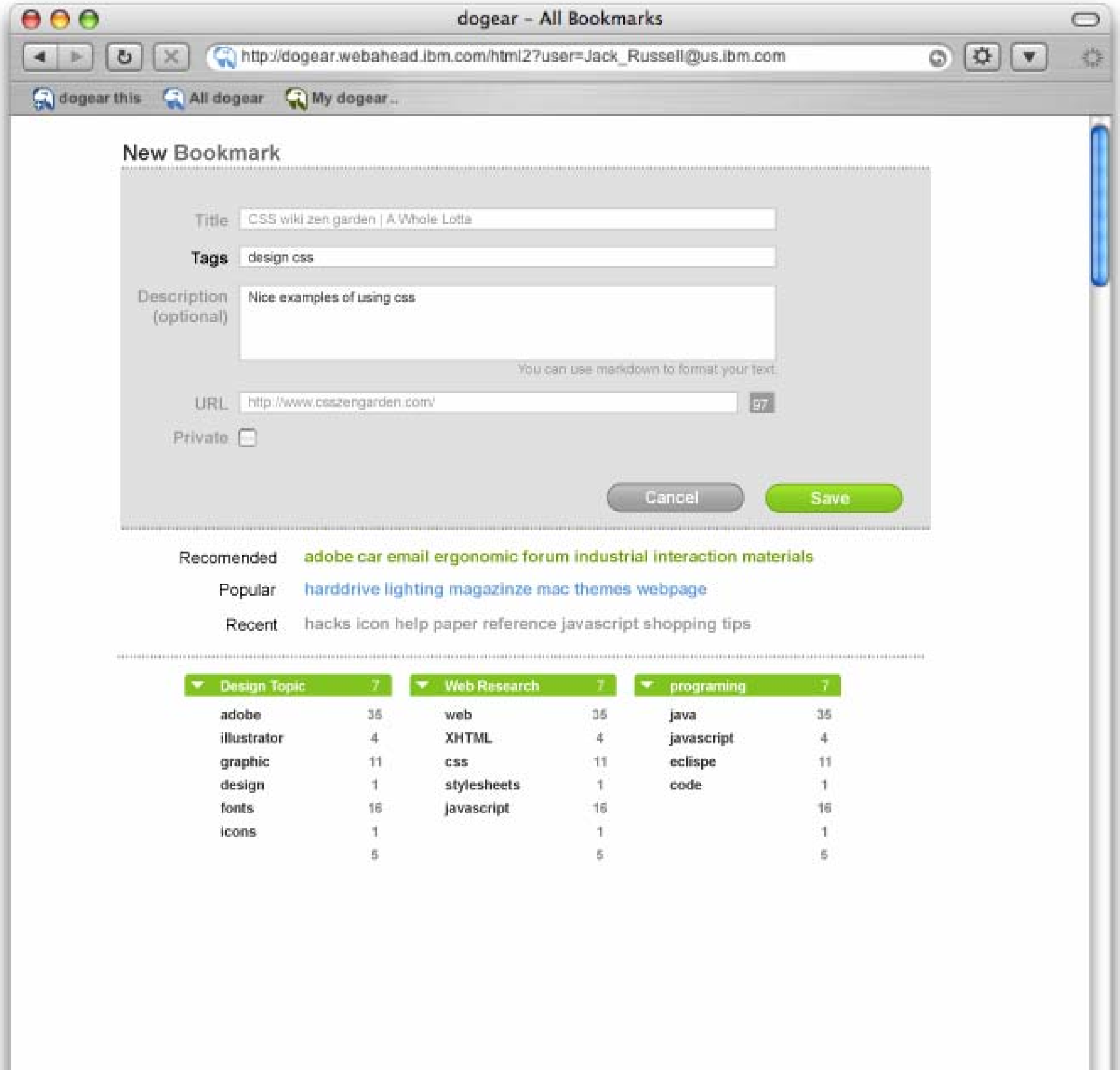
Convergence: Creation of a corporate folksonomy (Millen et al.,2006)



Convergence: Individual end-user tag histories (Millen et al., 2006)



Dogear post form



http://dogear.webahead.ibm.com/html2?user=Jack_Russell@us.ibm.com

New Bookmark

Title: CSS wiki zen garden | A Whole Lotta

Tags: design css

Description (optional): Nice examples of using css

You can use markdown to format your text.

URL: http://www.csszengarden.com/ 97

Private □

Recomended adobe car email ergonomic forum industrial interaction materials

Popular harddrive lighting magazinze mac themes webpage

Recent hacks icon help paper reference javascript shopping tips

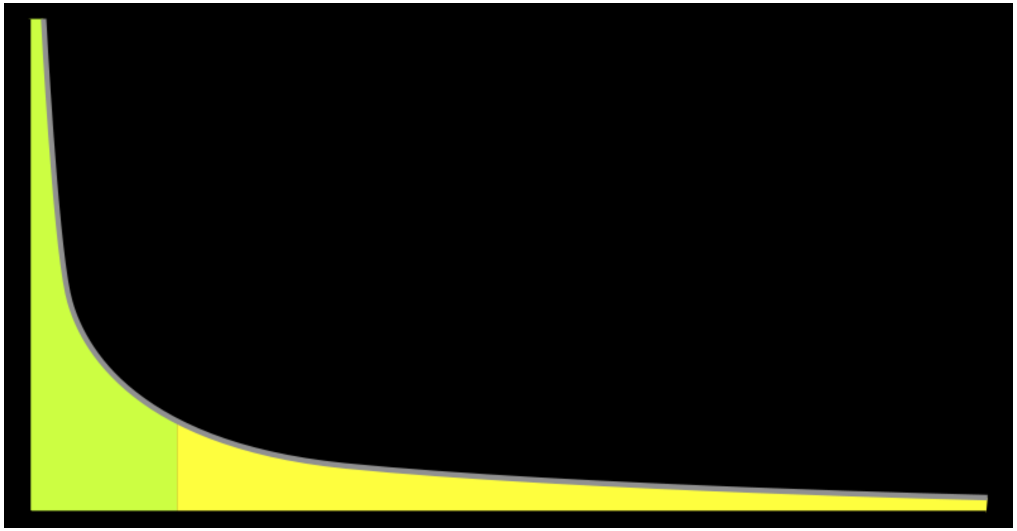
Design Topic	7	Web Research	7	programing	7
adobe	35	web	35	java	35
illustrator	4	XHTML	4	javascript	4
graphic	11	css	11	eclispe	11
design	1	stylesheets	1	code	1
fonts	16	javascript	16		16
icons	1		1		1
	5		5		5

User Interface helps tag convergence

- Completion
  - Shows others' tags – increase tag similarity
  - Reduces spelling mistakes
  - Makes it efficient to tag, so more tags
- How others tagged this page
- Frequent tags
- All increase likelihood of convergence

Outstanding problems with tagging

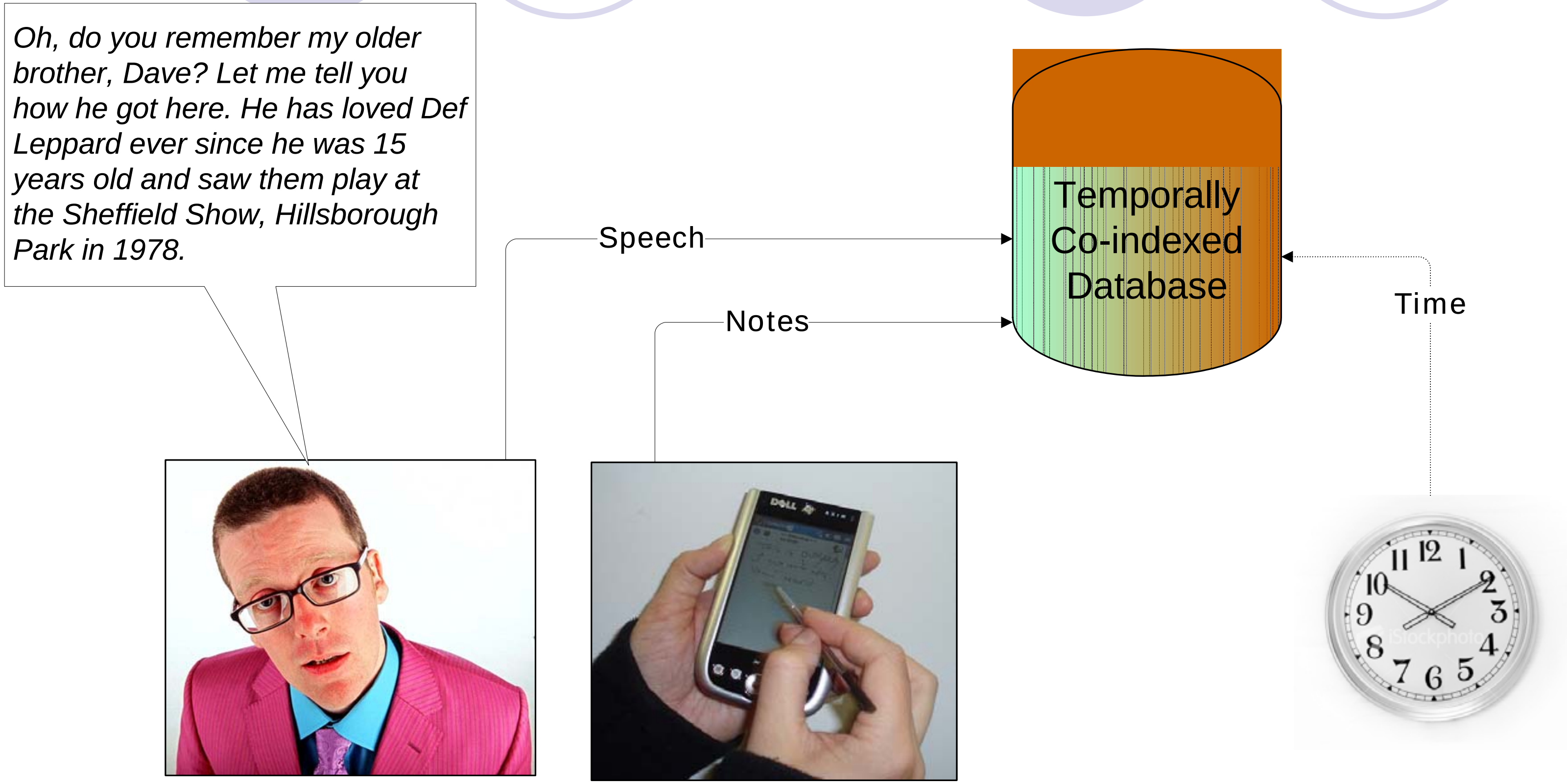
- Annotation costs
- ‘Long tail’ – power law
  - Most tags are generated by a small subgroup
  - Not everything gets tagged



So… Implicit Tags

- Reduce annotation costs, democratise, by using *implicit behaviours*
- Temporal tagging, ‘swarm’ effects
- E.g.
  - Edit/print a photo, registers interest
  - Take a note about a specific slide, registers interest

Temporal Tags – handwritten notes



Oh, do you remember my older brother, Dave? Let me tell you how he got here. He has loved Def Leppard ever since he was 15 years old and saw them play at the Sheffield Show, Hillsborough Park in 1978.

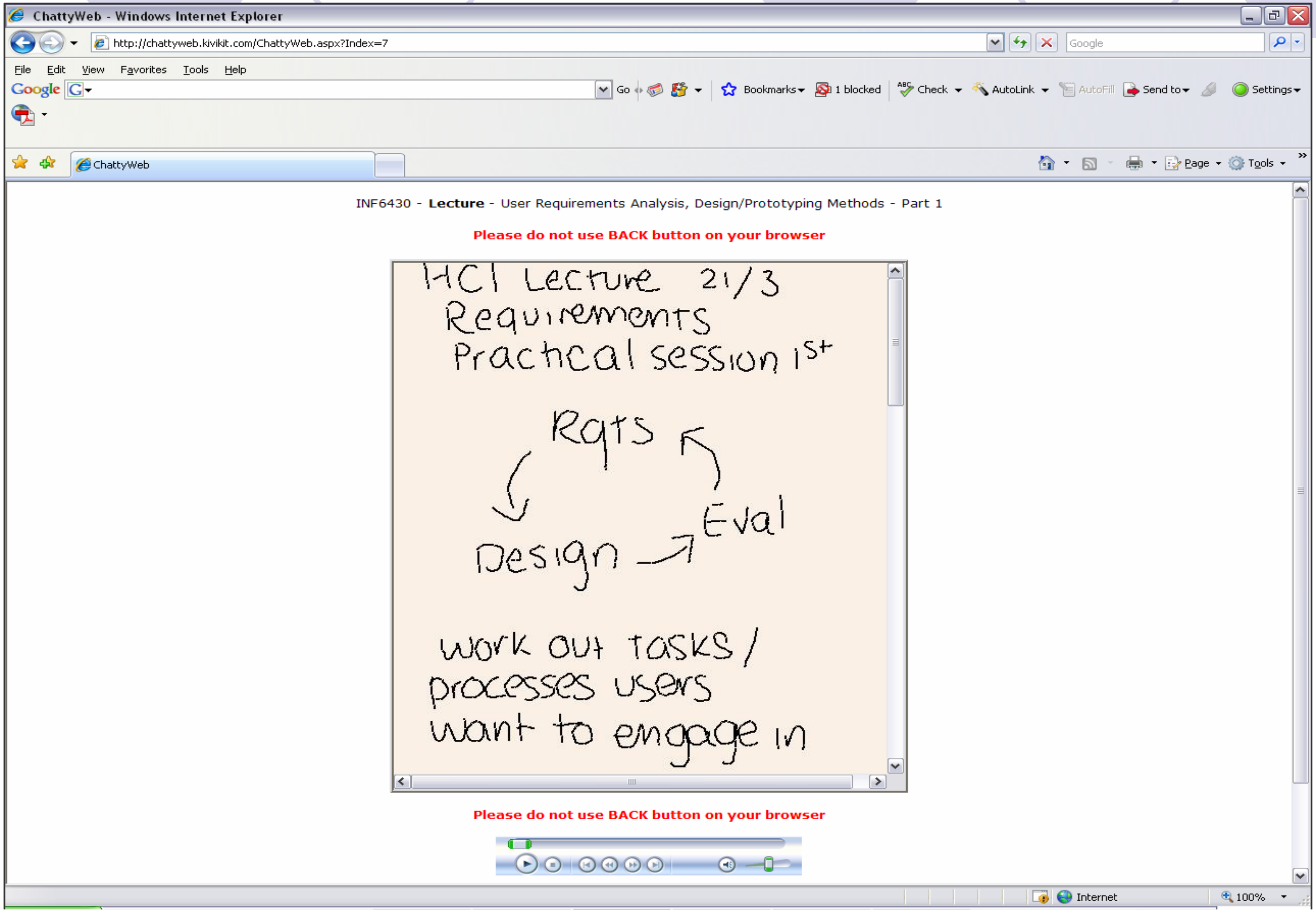
Speech

Notes

Temporally Co-indexed Database

Time

Retrieval Interface



ChattyWeb - Windows Internet Explorer
http://chattyweb.kivikit.com/ChattyWeb.aspx?Index=7
File Edit View Favorites Tools Help
Go
Bookmarks
1 blocked
Check
AutoLink
AutoFill
Send to
Settings
ChattyWeb
Page
Tools
Internet
100%

INF6430 - **Lecture** - User Requirements Analysis, Design/Prototyping Methods - Part 1

Please do not use BACK button on your browser

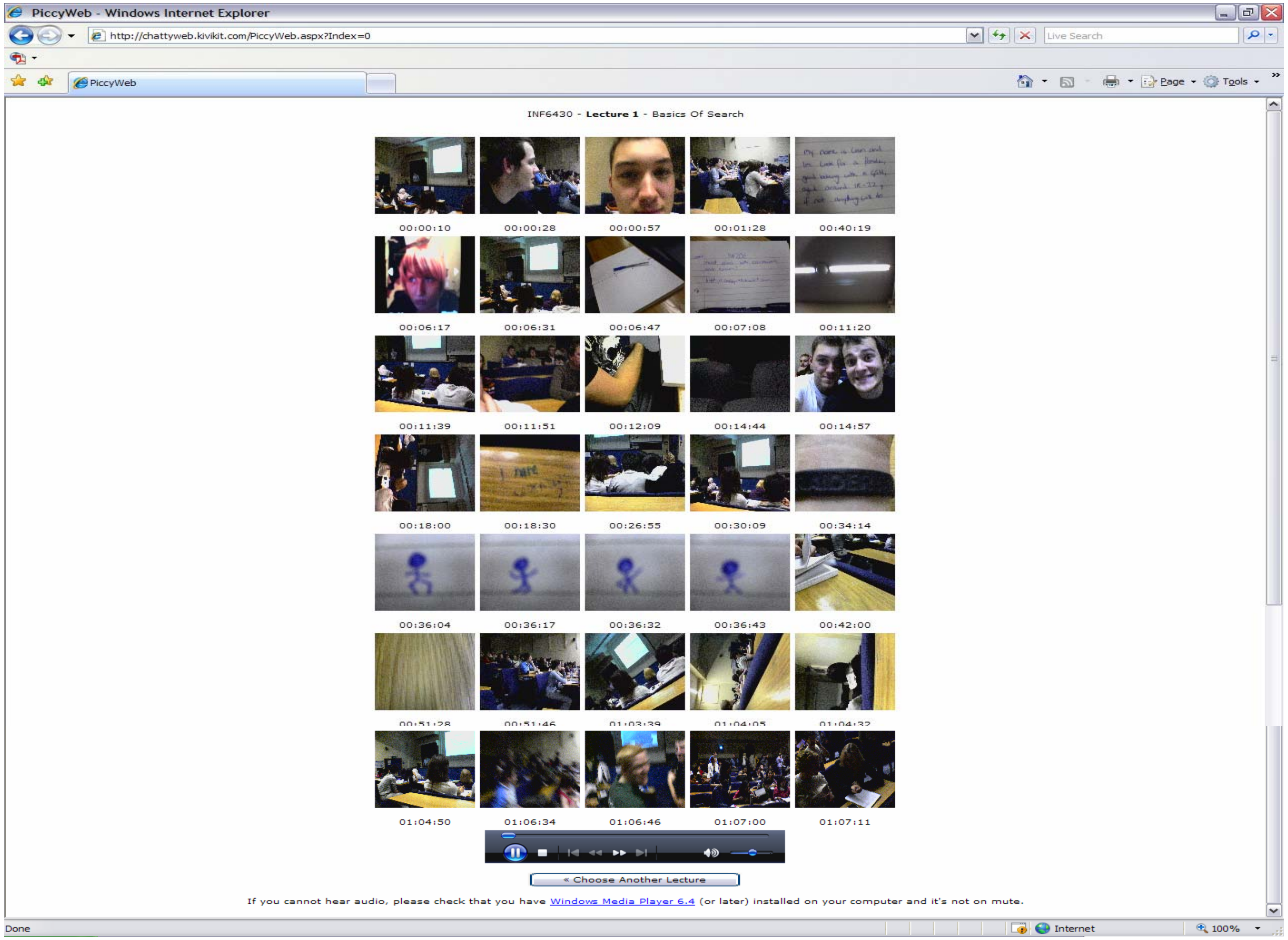
HCI Lecture 21/3
Requirements
Practical session 1st

Rqts
Design
Eval

work out tasks / processes users want to engage in

Please do not use BACK button on your browser

Pictorial Tags



PiccyWeb - Windows Internet Explorer
http://chattyweb.kivikit.com/PiccyWeb.aspx?Index=0
Live Search
PiccyWeb
Page
Tools

INF6430 - **Lecture 1** - Basics Of Search

If you cannot hear audio, please check that you have Windows Media Player 6.4 (or later) installed on your computer and it's not on mute.

Chitty Chatty demo

- Chattyweb.kivikit.com

Naturalistic Evaluation of ChattyWeb

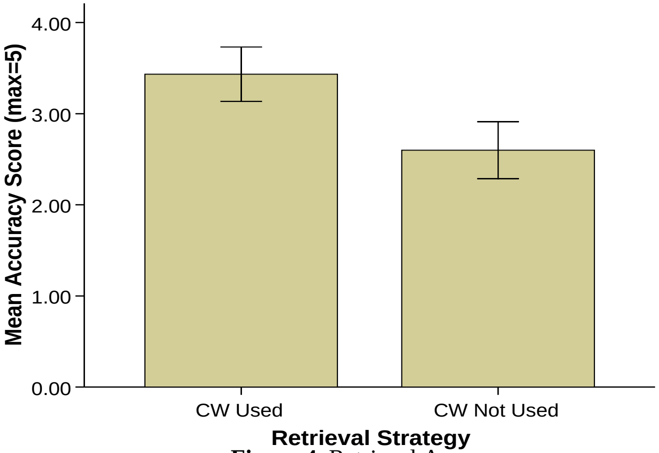


Figure 4. Retrieval Accuracy

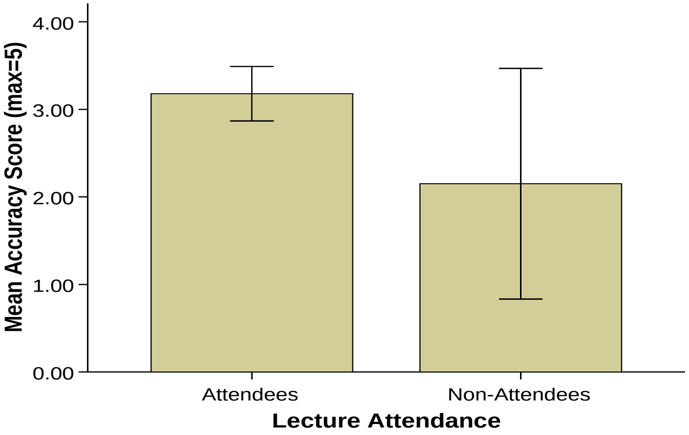


Figure 5. Attendance & Accuracy

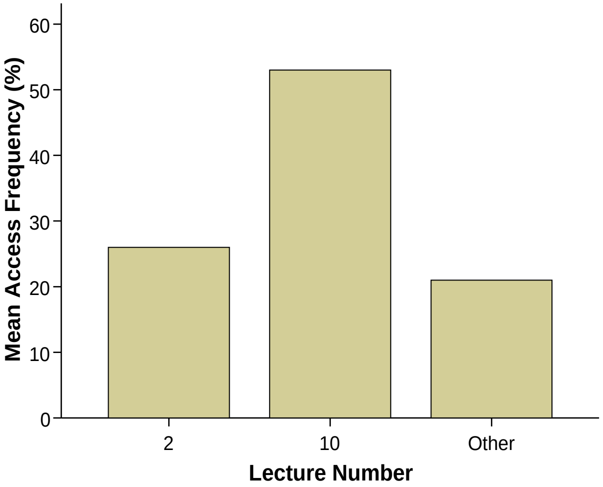
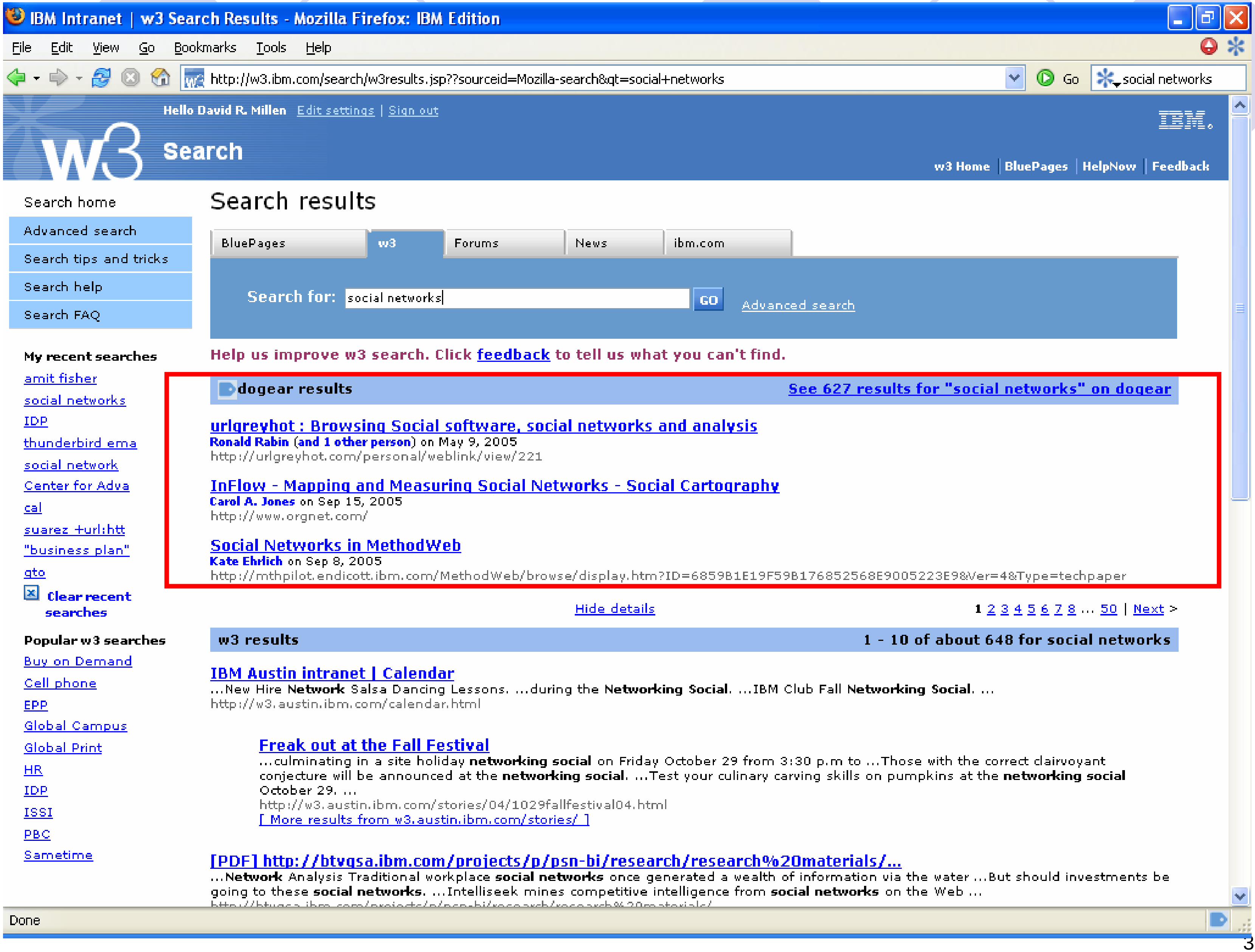


Figure 6. Frequency of Lecture Access

Tags can also improve text search

- Google search vs. Dogear (results reordered by tag frequency)
- Preference for Dogear search over Google (Millen et al., 2006)


IBM Intranet | w3 Search Results - Mozilla Firefox: IBM Edition
File Edit View Go Bookmarks Tools Help
http://w3.ibm.com/search/w3results.jsp??sourceid=Mozilla-search&qt=social+networks
Go
social networks
Hello David R. Millen Edit settings | Sign out
w3
Search
IBM.
w3 Home | BluePages | HelpNow | Feedback
Search home
Advanced search
Search tips and tricks
Search help
Search FAQ
BluePages
w3
Forums
News
ibm.com
GO
Advanced search
Done

Search results

Search for: social networks

Help us improve w3 search. Click feedback to tell us what you can't find.

My recent searches

- amit fisher
- social networks
- IDP
- thunderbird ema
- social network
- Center for Adva
- cal
- suarez +url:htt
- "business plan"
- qto
- ☒ **Clear recent searches**

Popular w3 searches

- Buy on Demand
- Cell phone
- EPP
- Global Campus
- Global Print
- HR
- IDP
- ISSI
- PBC
- Sametime

dogear results — See 627 results for "social networks" on dogear

urlgreyhot : Browsing Social software, social networks and analysis
Ronald Rabin (and 1 other person) on May 9, 2005
http://urlgreyhot.com/personal/weblink/view/221

InFlow - Mapping and Measuring Social Networks - Social Cartography
Carol A. Jones on Sep 15, 2005
http://www.orgnet.com/

Social Networks in MethodWeb
Kate Ehrlich on Sep 8, 2005
http://mthpilot.endicott.ibm.com/MethodWeb/browse/display.htm?ID=6859B1E19F59B176852568E9005223E9&Ver=4&Type=techpaper

Hide details — 1 2 3 4 5 6 7 8 ... 50 | Next >

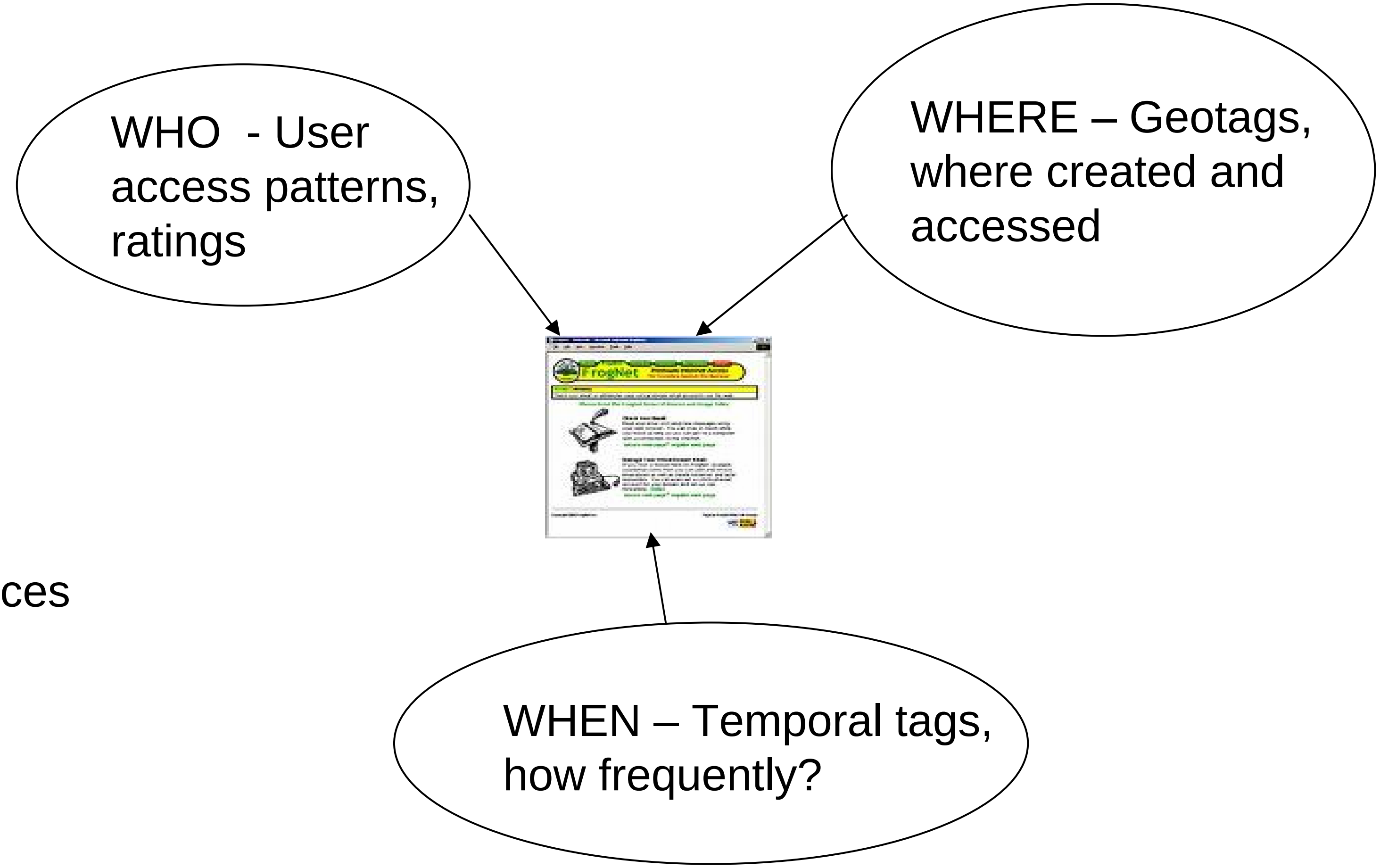
w3 results — 1 - 10 of about 648 for social networks

IBM Austin intranet | Calendar
...New Hire **Network** Salsa Dancing Lessons. ...during the **Networking Social**. ...IBM Club Fall **Networking Social**. ...
http://w3.austin.ibm.com/calendar.html

Freak out at the Fall Festival
...culminating in a site holiday **networking social** on Friday October 29 from 3:30 p.m to ...Those with the correct clairvoyant conjecture will be announced at the **networking social**. ...Test your culinary carving skills on pumpkins at the **networking social** October 29. ...
http://w3.austin.ibm.com/stories/04/1029fallfestival04.html
[More results from w3.austin.ibm.com/stories/]

[PDF] http://btvqsa.ibm.com/projects/p/psn-bi/research/research%20materials/...
...**Network** Analysis Traditional workplace **social networks** once generated a wealth of information via the water ...But should investments be going to these **social networks**. ...Intelliseek mines competitive intelligence from **social networks** on the Web ...

Future: Making sense of tags



WHO - User access patterns, ratings

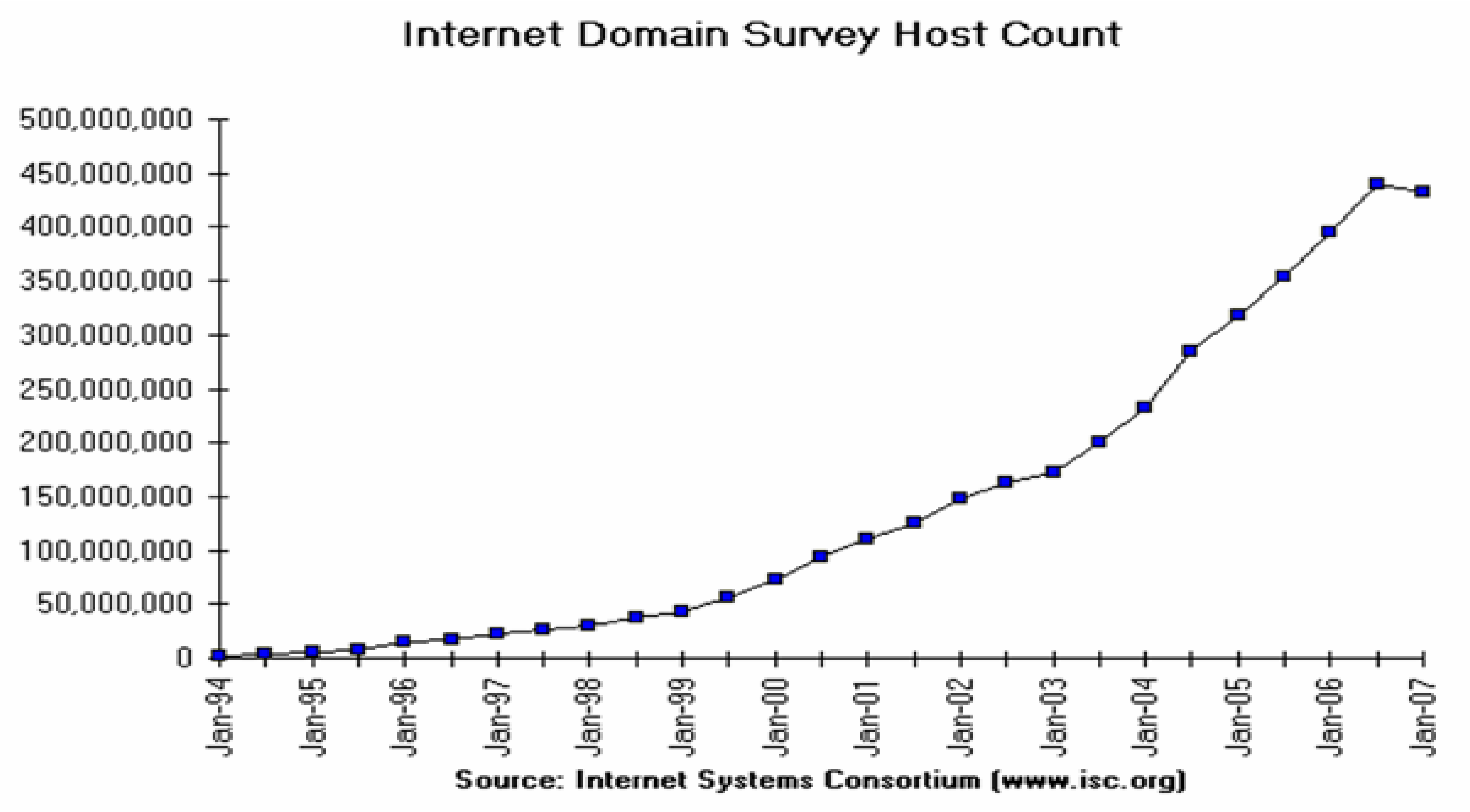
WHERE – Geotags, where created and accessed

WHEN – Temporal tags, how frequently?

Algorithms

User Interfaces

Information Glut: Domain hosts



Source: Internet Systems Consortium (www.isc.org)

From Information Famine to Glut

- Past: high cost of information access
  - Information held in libraries, interlibrary loan, xerox
  - Commend students for finding relevant resources
- Now: Cost of information access is much reduced
  - Education: Plagiarism is the problem
  - Allocating attention to *what’s important*
  - Can people make sense of found information?

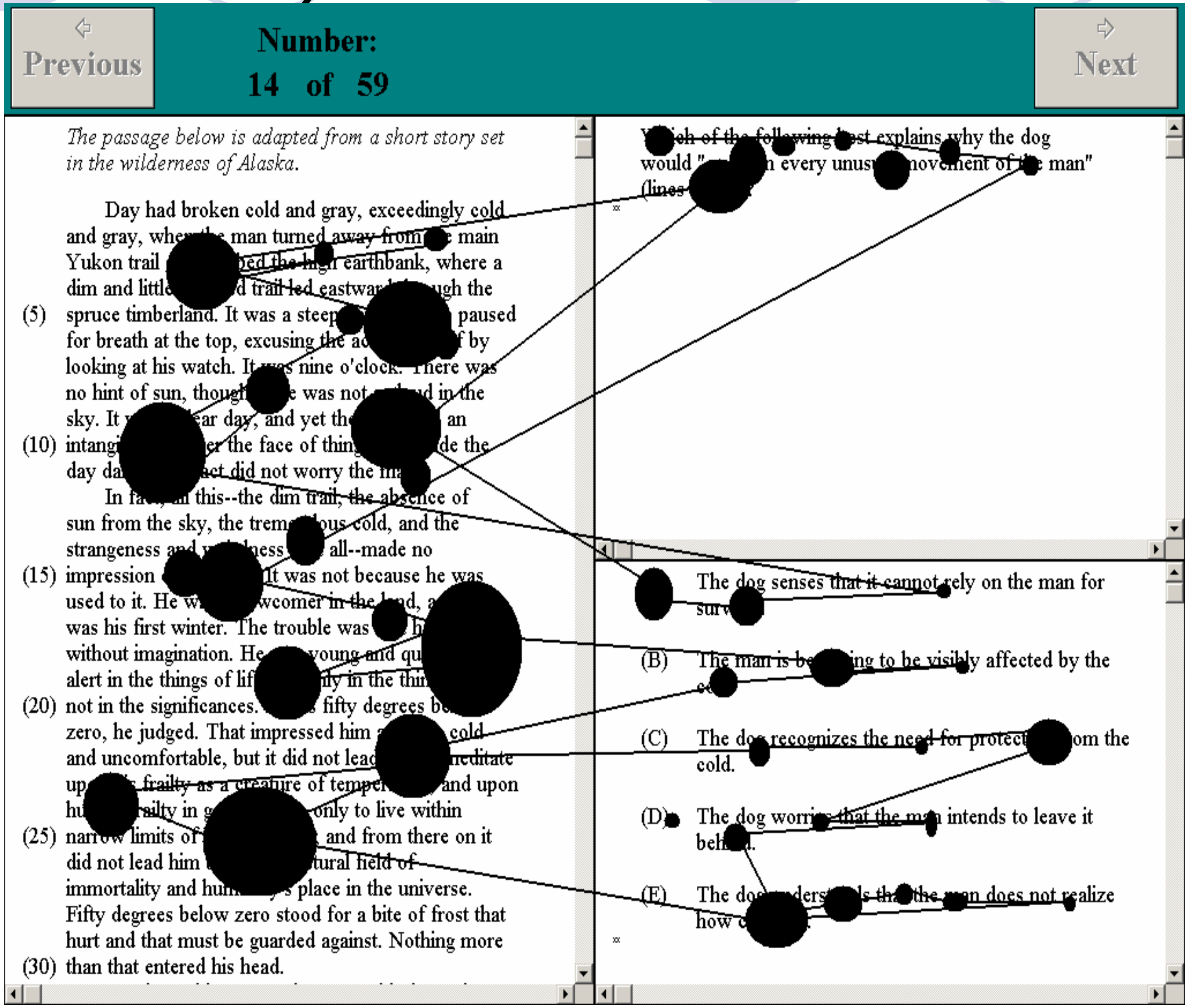
New tools needed to:

- *Skim* – identify which parts of retrieved documents are important
  - n.b. also skim to find whether document is relevant at all
- *Combine sources* (sense-making)

Skimming

- Allow users to focus on the relevant parts of a document
- Is this document relevant?
- If so, what parts of it are relevant?

Replicate and improve on the human eye (Gorin, 2005)



The passage below is adapted from a short story set in the wilderness of Alaska.

Day had broken cold and gray, exceedingly cold and gray, whe[illegible] the man turned away from [illegible]e main Yukon trail [illegible]bed the high earthbank, where a dim and littl[illegible]d trail led eastwar[illegible]ugh the
(5) spruce timberland. It was a steep [illegible] paused for breath at the top, excusing the a[illegible]f by looking at his watch. It was nine o'clock. There was no hint of sun, though [illegible]e was not [illegible]d in the sky. It [illegible]ear day, and yet the [illegible] an
(10) intangi[illegible]er the face of thin[illegible]de the day da[illegible]act did not worry the ma[illegible]
In fa[illegible] this--the dim trail, the absence of sun from the sky, the tremendous cold, and the strangeness and [illegible] all--made no
(15) impression [illegible] It was not because he was used to it. He w[illegible]wcomer in the land, a[illegible] was his first winter. The trouble was [illegible]h without imagination. He [illegible] young and qu[illegible] alert in the things of lif[illegible]ly in the thin[illegible]
(20) not in the significances. [illegible] fifty degrees b[illegible] zero, he judged. That impressed him [illegible] cold and uncomfortable, but it did not lea[illegible]editate up[illegible] frailty as a creature of temper[illegible] and upon hu[illegible] frailty in g[illegible] only to live within
(25) narrow limits of [illegible] and from there on it did not lead him [illegible]tural field of immortality and hu[illegible]s place in the universe. Fifty degrees below zero stood for a bite of frost that hurt and that must be guarded against. Nothing more
(30) than that entered his head.

W[illegible]ich of the following best explains why the dog would "[illegible]n every unus[illegible]movement of the man" (lines [illegible]

The dog senses that it cannot rely on the man for surv[illegible]

(B) The man is be[illegible]ing to be visibly affected by the co[illegible]

(C) The do[illegible] recognizes the need for protect[illegible]om the cold.

(D) The dog worri[illegible] that the man intends to leave it beh[illegible].

(E) The do[illegible]ders[illegible]ls that the man does not realize how c[illegible]

Interactive Compression

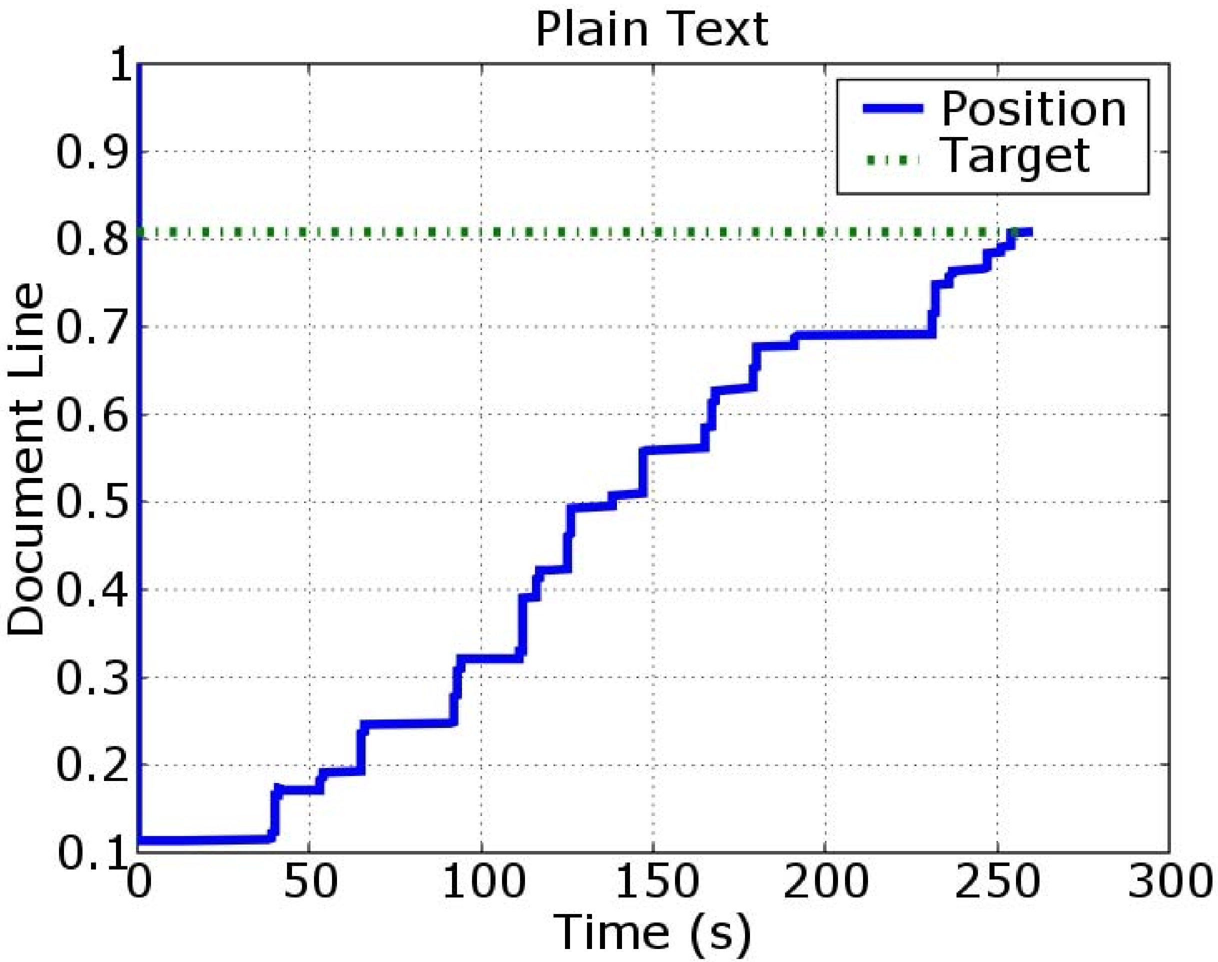
- Interactive tools that allow users
  - focus on relevant information
  - ignore irrelevant information
- Increase efficiency of skimming

Interactive Compression

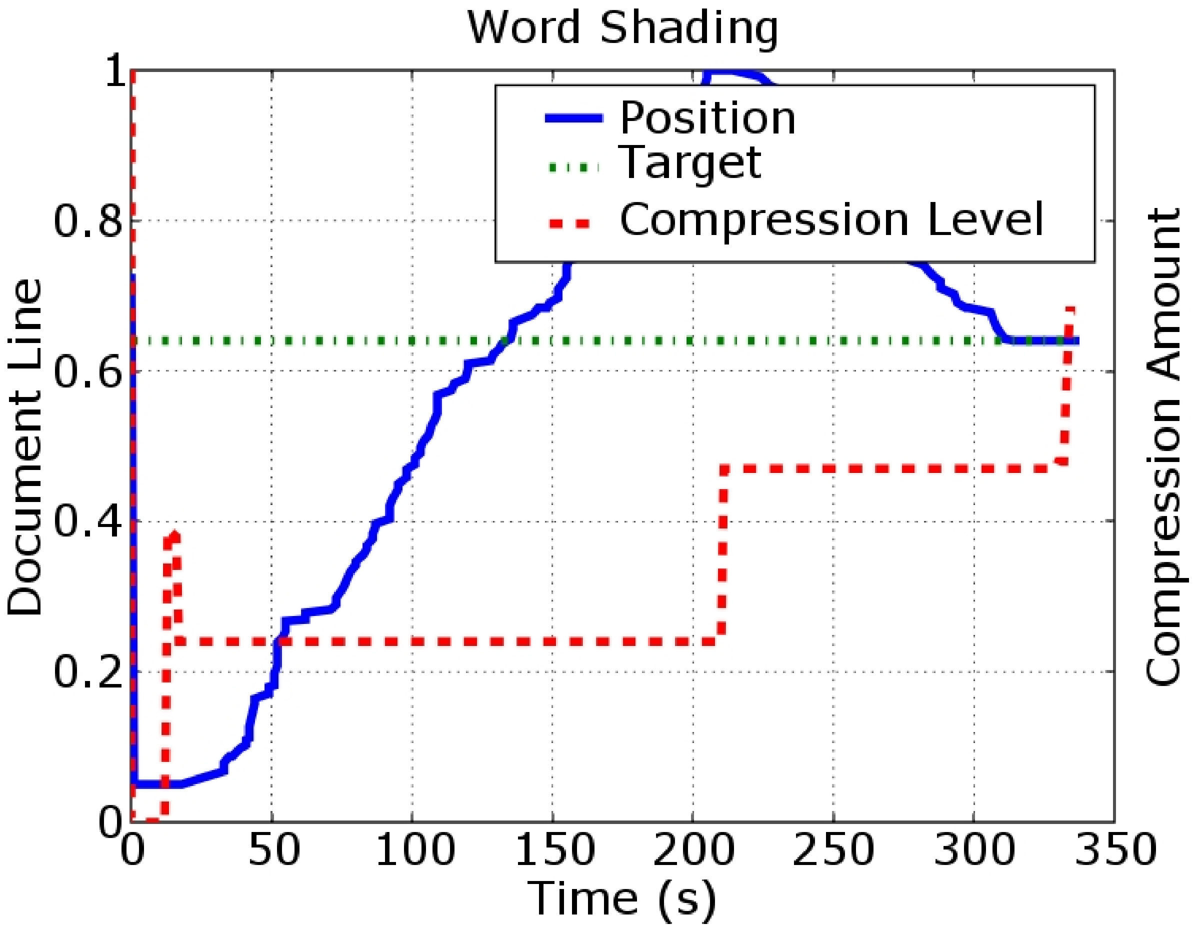
- Allow users to control
  - *Amount* of information that they see
  - *Presentation* of that information
- Two main methods
  - *Remove* irrelevant information (destructive)
  - *Highlight* relevant information (non-destructive)
- Automatic Methods

Demos

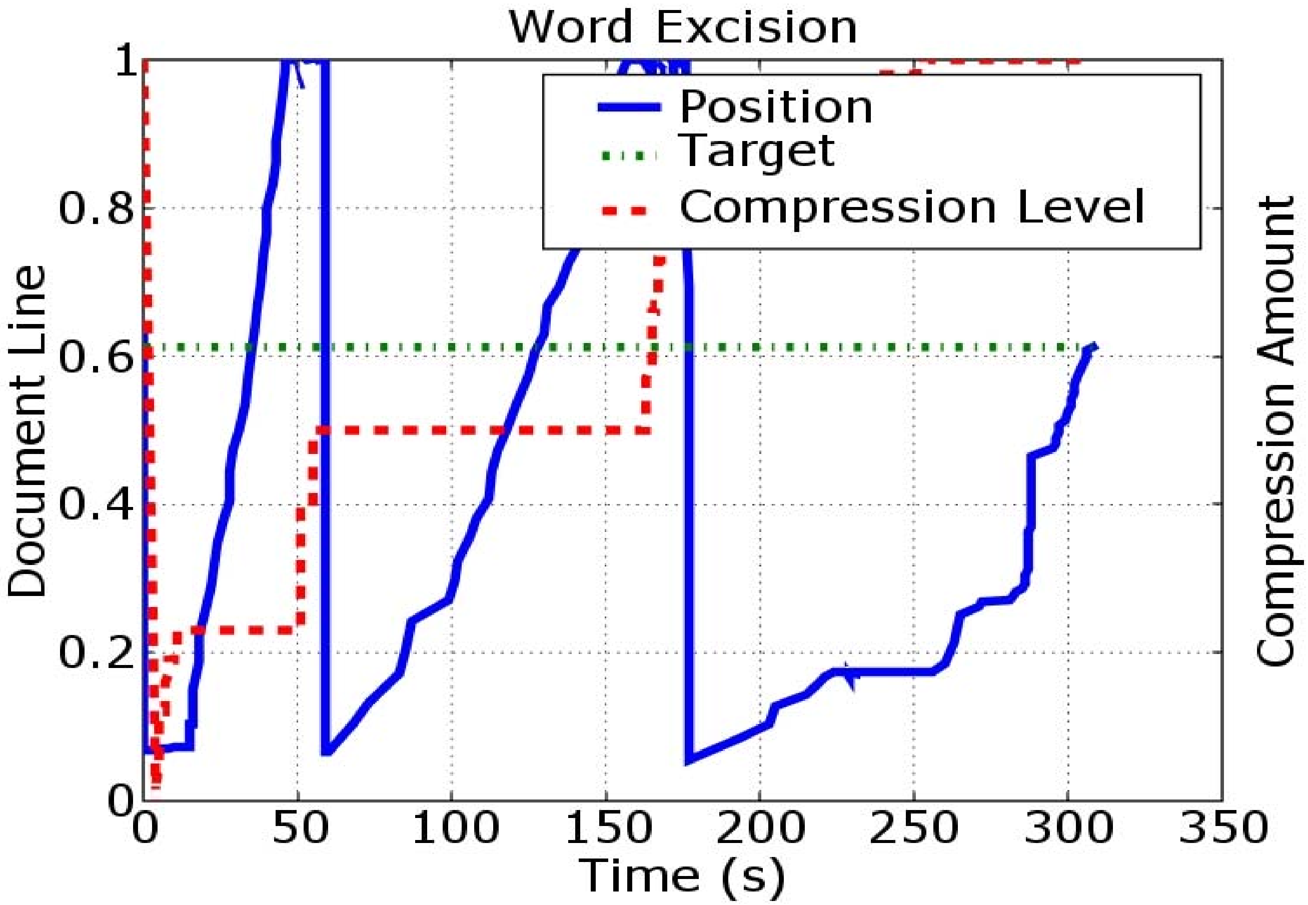
Typical behaviours: plain text



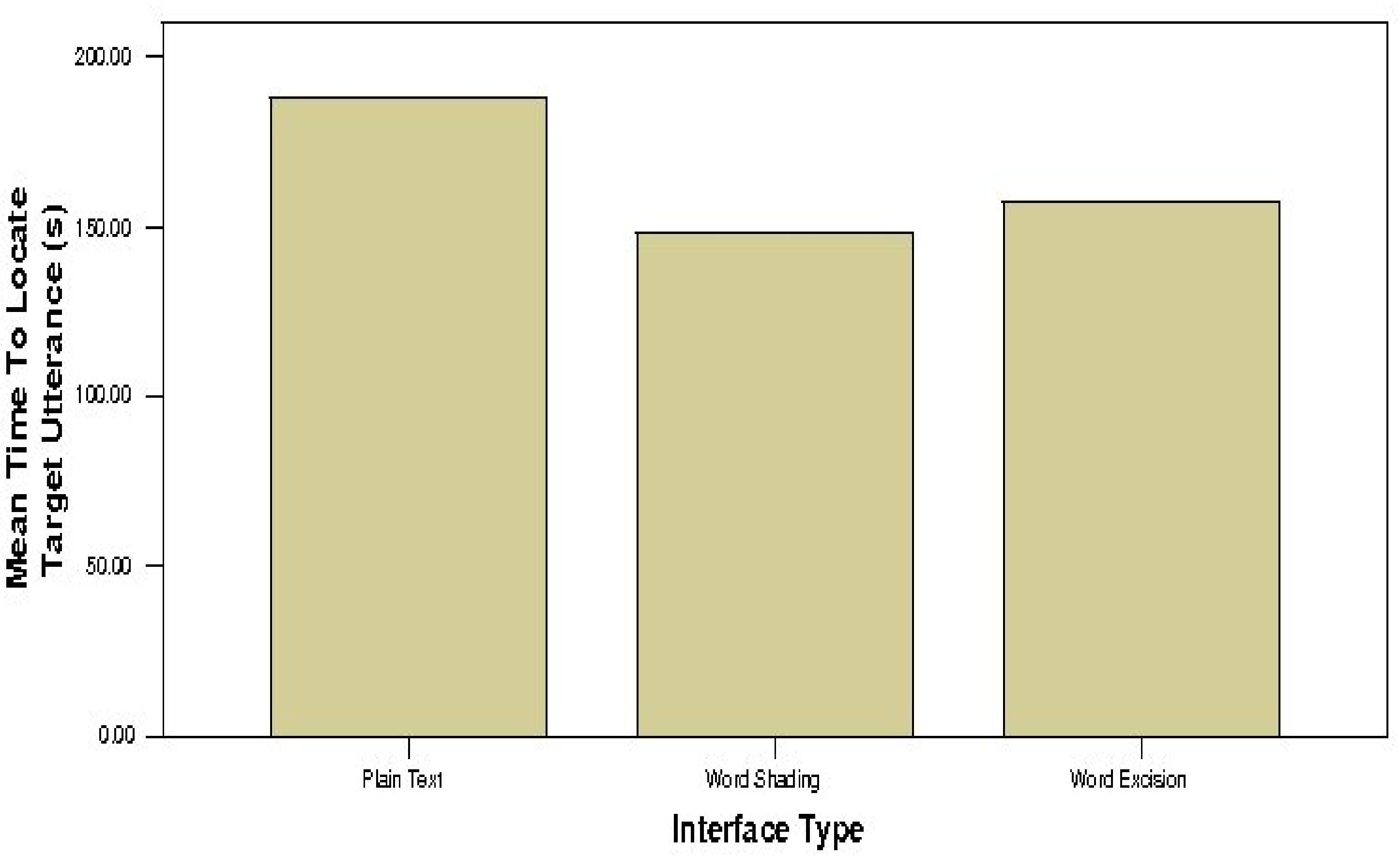
Word Shading



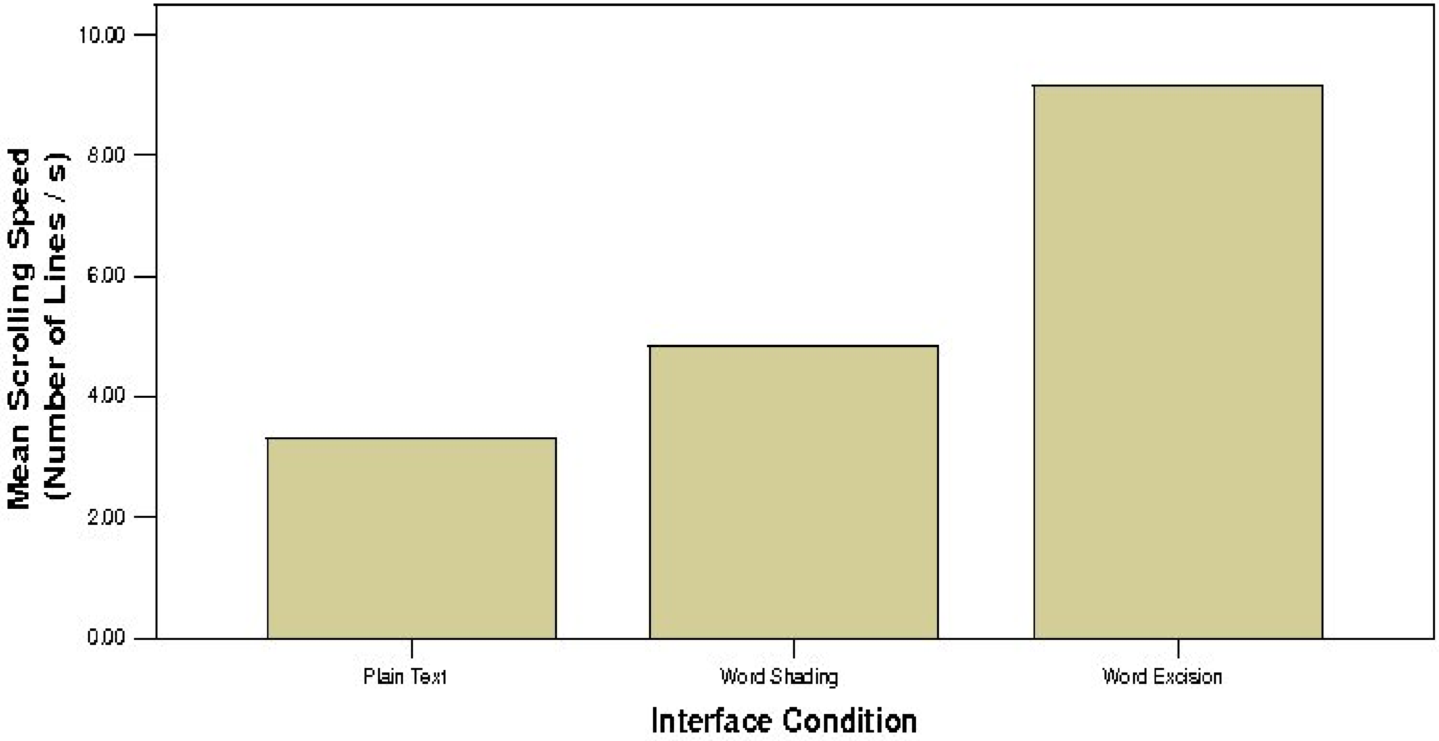
Word excision



Interactive compression reduces retrieval time



Improved scanning speed



Comparing Methods

- Overall efficiency
  - Excision = Shading > Plain
- Speed through document
  - Excision > Shading > Plain
- Is there a cost?
  - Excision > Shading > Plain - number of passes through document
  - But - Error rates, ‘overshoots’ were equivalent

Conclusions

- Beyond document retrieval
- Multimedia IR
  - Content-based: speech oriented methods
  - Tagging, especially implicit tags
  - Future – tag interpretation, combination
- Skimming
  - Interactive compression methods
  - Future – 'sensemaking'
- Implications?
  - Education, science, society