



Computation Institute

SIDGrid

The Social Informatics Data Grid

Mark Hereld

**Computation Institute
Argonne National Laboratory &
University of Chicago**



THE UNIVERSITY OF
CHICAGO



Collaborators

- **Bennett Bertenthal (UC, psychology)**
- **Robert Grossman (UIC, cs)**
- **David Hanley (UIC, cs)**
- **Mark Hereld (UC/ANL, cs)**
- **Sarah Kenny (UC, cs)**
- **Susanne Lefvert (UC, cs)**
- **Gina Levow (UC, cs)**
- **David McNeill (UC, psychology)**
- **Michael E. Papka (UC/ANL, cs)**
- **Steve Porges (UIC, psychology)**
- **Kavithaa Rajavenkateshwaran (UC, cs)**
- **Rick Stevens (UC/ANL, cs)**

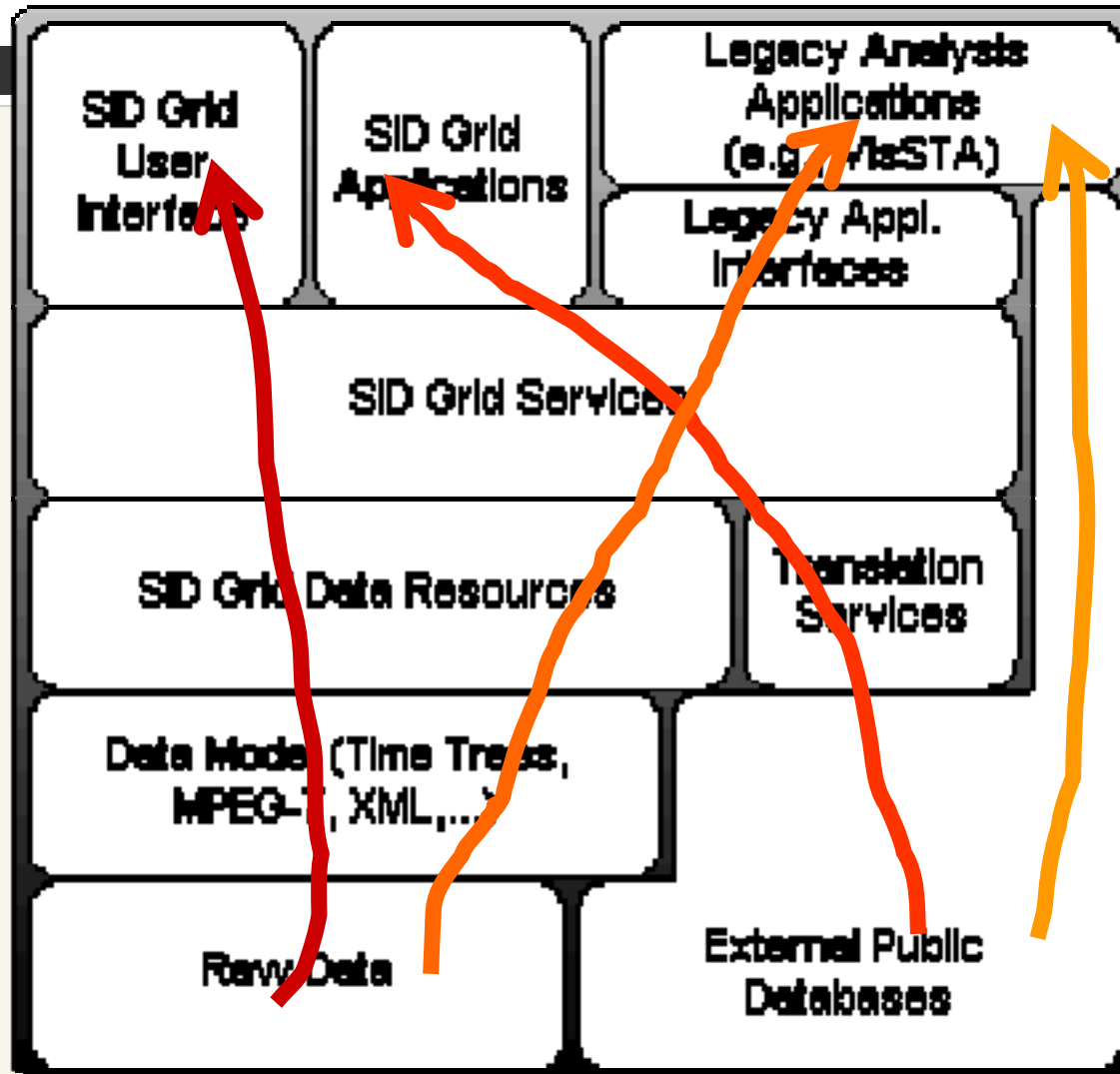
Goals

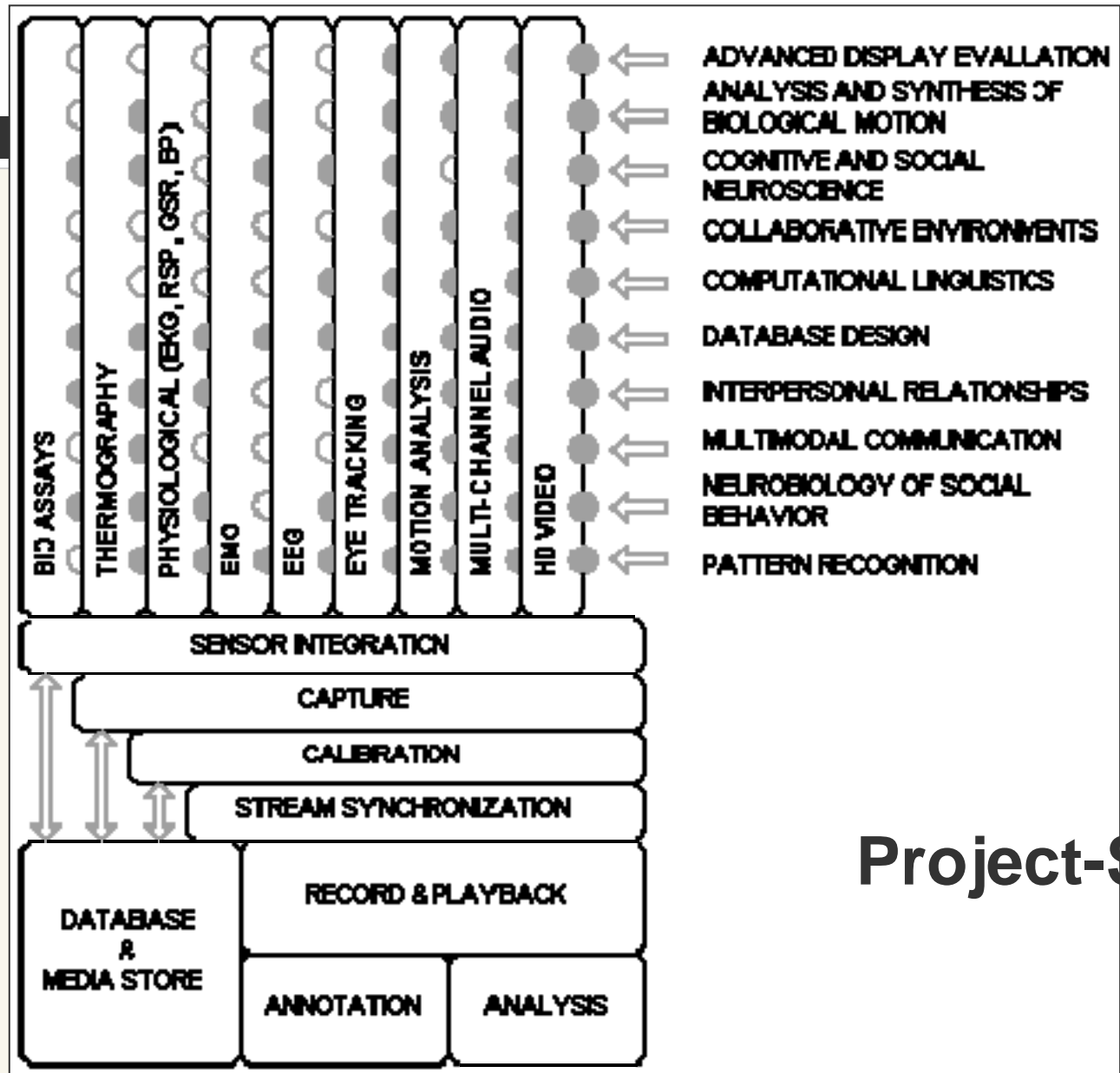
- **Fundamentally change the kind of experiments that can be attempted**
- **Enable new class of virtual experiments**
- **Redefine Social Informatics**
- **Promote collaboration in creation and analysis of complex multimodal data archives**

Research Project Areas

- **Analysis & Synthesis of Biological Motion**
 - **Cognitive and Social Neuroscience**
 - **Collaborative Environments**
 - **Interpersonal Relationships**
 - **Multimodal Communication**
 - **Neurobiology of Social Behavior**
- **Advanced Display Evaluation**
 - **Collaborative Environments**
 - **Computational Linguistics**
 - **Database Design**
 - **Pattern Recognition**

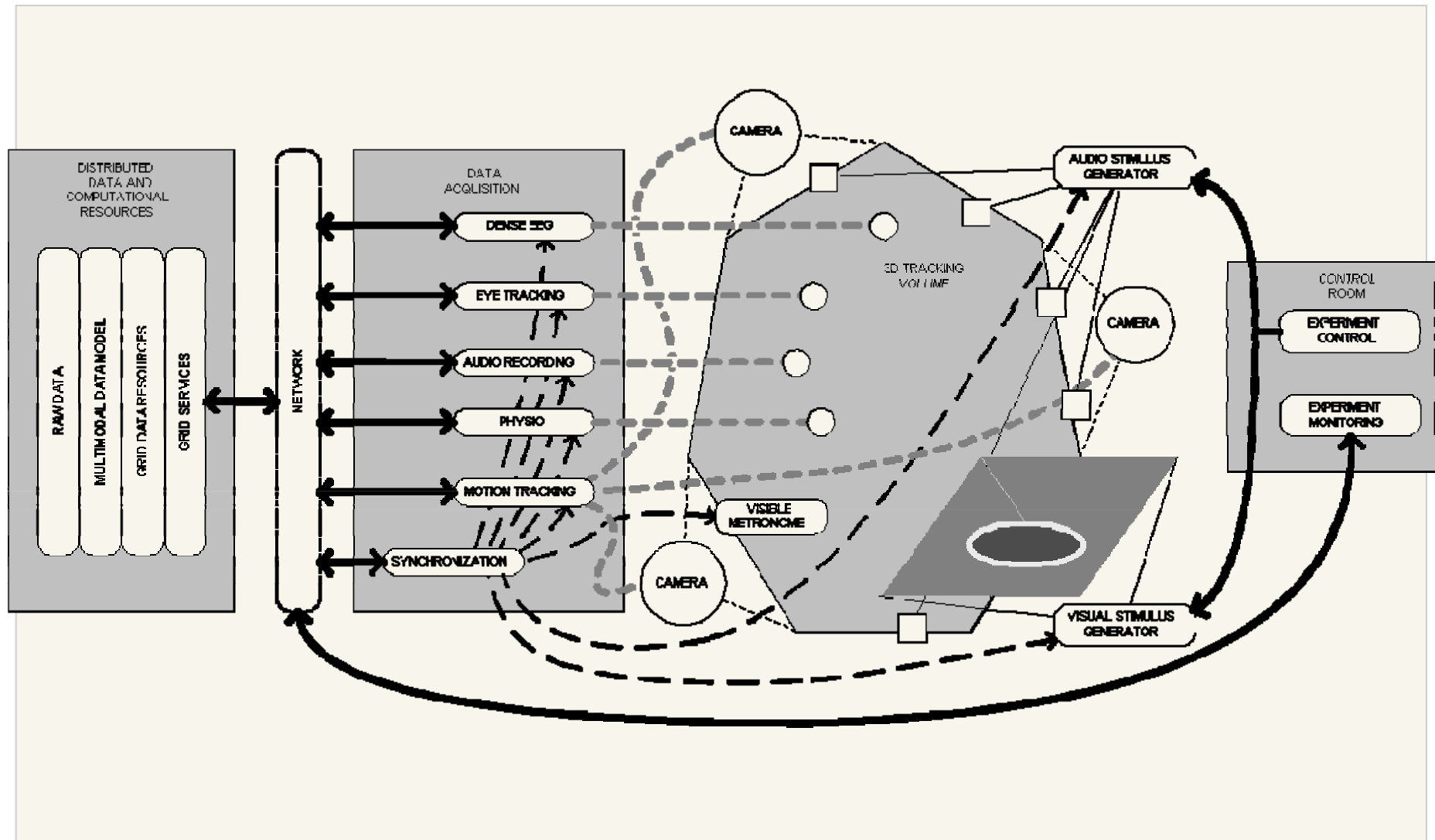
SIDGrid





Project-Sensor Matrix

SuperLab



Sensor and Stimulus Systems

- Stimulus Display
- Audio Stimuli
- Video Recording
- Audio Recording
- Eye Tracking
- Motion Analysis
- IR Thermography
- Physiological Recording
- EEG

- Scene
- Face
- Full body
- Ambient array
- Personal boom mic

- Multi-projector wall
- Transparent display
- High density display
- Audio feedback
- Spatialization

- electrocardiogram
- electromyogram
- galvanic skin response
- respiration
- blood pressure
- High density electroencephalogram

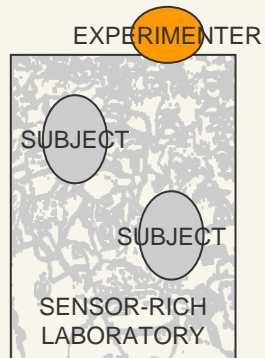


- Head mounted
- “Remote optics”
- Visualeyez
 - 512 active markers
- Mid-infrared 320x240 NTSC



Enabling New Experimental Modes

HERE - NOW

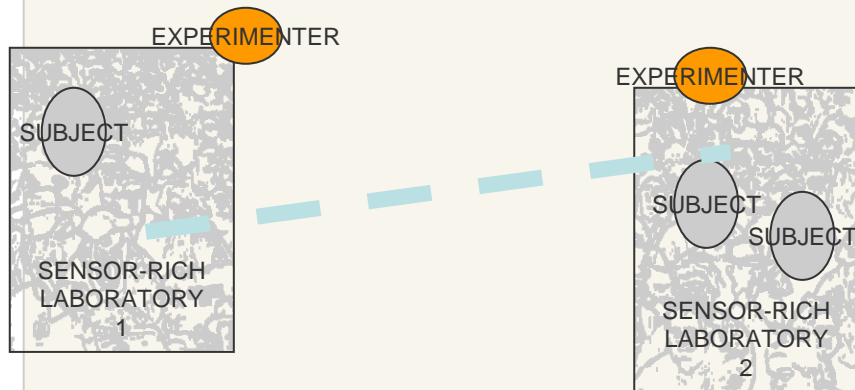


- **Control**
- **Monitoring**
- **Annotation**
- **Analysis**
- **Archiving**

REMOTE - NOW



DISTRIBUTED - NOW



ANYWHERE - LATER



Client Side

- **Leveraging efforts of ELAN development team**
 - Annotation and analysis of multi-modal meeting streams
 - Familiarity in the user community
 - XML, Java
 - Cross platform
- **Adding SIDGrid functionality**
 - Minimally intrusive
 - avoid complicated co-development w/ELAN team
 - Browsing SIDGrid data
 - Additional data types
 - Upload / Download to SID Grid server
- **Web services interfaces**
 - “roll your own” client

Elan - DemoProject.eaf

File Edit Search View Options Help

Grid Text Subtitles Controls

Volume: 100

Rate: 100

00:00:24.587 Selection: 00:00:00.000 - 00:00:00.000 0

Selection Mode Loop Mode

00:00:23.000	00:00:24.000	00:00:25.000	00:00:26.000	00:00:27.000	00:00:28.000	00:00:29.000	00:00:30.000
K-Spch	there is another			rotunda		ja	
W-Spch	n t	rhine eh valley	yeah that's another eh		kind of rotunde	and then you fo	
W-Words	he	Rhine eh valley	yeah th 'another eh		kin o rotunde	an th you foll	
W-POS	irt	n po n	int de adj part		n p n	co ad pro v r	
W-IPA	tʊ ð	raɪn ə vælɪ	je: ðætʰ ənəðə ə		kaɪnd əv rɒtʊnd	ænd ðen ju: fɒl	
W-RGU							

Elan - DemoProject.eaf

File Edit Search View Options Help

- New... ⌘N
- Open... ⌘O
- Open from SidGrid
- Save to SidGrid**
- Save ⌘S
- Save As...
- Save As Template...
- Save Selection As .eaf...
- Merge Transcriptions...
- Automatic Backup
- Page Setup
- Print Preview
- Print ⌘P
- Export As
- SidGrid Transformation
- New from SidGrid
- Import
- Exit

Grid Text Subtitles **Controls**

Volume: 100

Rate: 100

Selection: 00:00

587

00:00:24.000 00:00:25.000 00:00:30.000

Open

20/files.eaf	66 GB	5 mov	2 wav	...
21/Laursen.eaf	368 GB	23 mov	6 wav	...
22/Cassell.eaf	5 GB	1 mov	0 wav	...
23/Pres2.eaf	21 GB	3 mov	12 wav	...
24/1995.eaf	4 GB	9 mov	1 wav	...
25/ISL.eaf	4 GB	4 mov	1 wav	...
26/1985.eaf	1 GB	0 mov	2 wav	...
27/1991.eaf	945 GB	1 mov	66 wav	...
	8 GB	3 mov	0 wav	...
	20 GB	13 mov	2 wav	...

OK Cancel

K-Spch there is another

W-Spch rhine eh valley yeah that's another eh kind of rotunde and then you fo

W-Words he Rhine eh valley yeah th another eh kin o rotunde an th you foll

W-POS n po n int de adj part n p n co ad pro v i

W-IPA raɪn ə vælɪ jɛ: ðætʰ ənaðə ə kaɪnd əv rɒtʊnd ənd ðen ju: fɒlə

W-RGII

Server Side

- **Web services**
 - Query
 - Data download / upload
- **Portal interface**
 - Security
 - Data and metadata browsing
 - Preview
 - Tags, attributes
 - Projects
 - Groups
 - Search
 - Data transformation using grid resources

Social Informatics Data Grid

SIDGrid

[home](#)
[about](#)
[people](#)
[news & events](#)
[partners](#)
[portal](#)
[log out](#)

Welcome here!

[projects](#)
[videos](#)
[administrate](#)

Search for like

my projects

all projects

GROUPS:

talkbank

ALL PROJECTS

	.mov	.wav	.eaf	GB	NEXT ▶
▶ files	10	0	0	45	2006-11-30 17:18:35
▶ Cassell	4	30	0	20	2006-11-30 17:18:35
▶ Pres2	2	2	1	3	2006-11-30 17:18:35
▶ 1995	12	100	9	200	2006-11-30 17:18:35
▶ ISL	1	1	1	1	2006-11-30 17:18:35
▶ 1985	6	2	0	12	2006-11-30 17:18:35
▶ 1991	400	0	1	1001	2006-11-30 17:18:35
▶ 1989	0	666	1	312	2006 11 30 17:18:35
▶ 1992	0	0	13	0.1	2006-11-30 17:18:35
▶ 1986	0	0	0	0.0	2006-11-30 17:18:35
▶ CLIPS	18	4	0	66	2006-11-30 17:18:35

Social Informatics Data Grid SIDGrid

[home](#) [about](#) [people](#) [news & events](#) [partners](#) [portal](#) [log out](#)

Welcome here!

[projects](#) [videos](#) [administrate](#)

Search for like

CLICK ON A THUMBNAIL TO DOWNLOAD THE CORRESPONDING VIDEO

NEXT ▶



group5.mov



2010CZ304.mov



dec19f.mov



7004JP405.mov



8-5.mov



clip18.mov



dec21a.mov



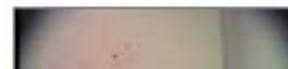
ste_administration_issues.mov



2011CZ406.mov



dec19e.mov



Search for like

projects > DemoProject

DemoProject created by: susanne || updated:

KEYWORDS:

demo yoto testkey funfun

PROJECT FILES:

 elan-example1.mpg elan-example1.wav

ANNOTATION TIERS:

K-Spch W-Spch W-Words
 W-POS W-IPA W-RGU
 W-RGph W-RGMe K-RGU
 K-RGph K-RGMe

Export range from to (Ranges from -1 to 36180 milliseconds)

Process Data (select script and new experiment name)

New Experiment Name:

Script:

SIDGrid Preview



Color	Annotation Tier
	K-Spch
	W-POS

Social Informatics Data Grid SIDGrid

home about people news & events partners portal log out

job id	status
ID000001	in queue
ID000002	in queue
ID000003	in queue
ID000004	in queue
ID000005	in queue

Welcome here!

projects videos administrate

Search for like

projects > waverunner

waverunner

KEYWORDS:

Actions...

wave_test tagtest

PROJECT FILES:

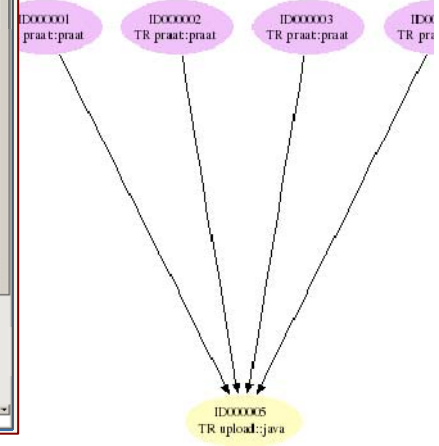
his1.wav

his2.wav

his3.wav

his4.wav

The screenshot shows the Praat software interface. The main window is titled 'Sound Praat1.n' and contains a 'Sound editor' with a waveform and a spectrogram. Below the editor are controls for 'edit', 'in', 'out', 'set', and 'Group'. A 'Praat picture' window is also visible, showing a red box on a white background. The bottom status bar indicates 'Praat 4.1.28 Feb 2004'.



Export range from to (Ranges from None to None milliseconds)

Process Data (select script and new experiment name)

New Experiment Name:

Script:

Tags

- **What we have**

- Tags
- Attributes (keyword-value)
- Query against these

- **What we want**

- Tag map
- My tags
- Suggest tags
- Show tags available, in use, others, mine
- Regular expressions in query
- Include context
 - tag qualifiers (AUTOSCANv1.0_blue, AUTOSCANv1.0:blue)
 - Avoid trashing namespace

Next Steps

- **Automation: Seeding the metadata...?**
 - Encourage experimentation
 - Generate some kinds of tags automatically
 - ...but don't poison the metadata
- **Presentation techniques to support finding what you want**
 - SOM
 - Tags, keywords, date, media types, other attributes
 - Thumbnail clusters
 - Exemplars
- **Personalization with "My View"**
 - Personal choice, taste, priorities
 - Avoid poisoning everybody
 - Agent processes
 - Clustering
 - Alerts (new data, changes to experiment set)

http://del.icio.us/tag/

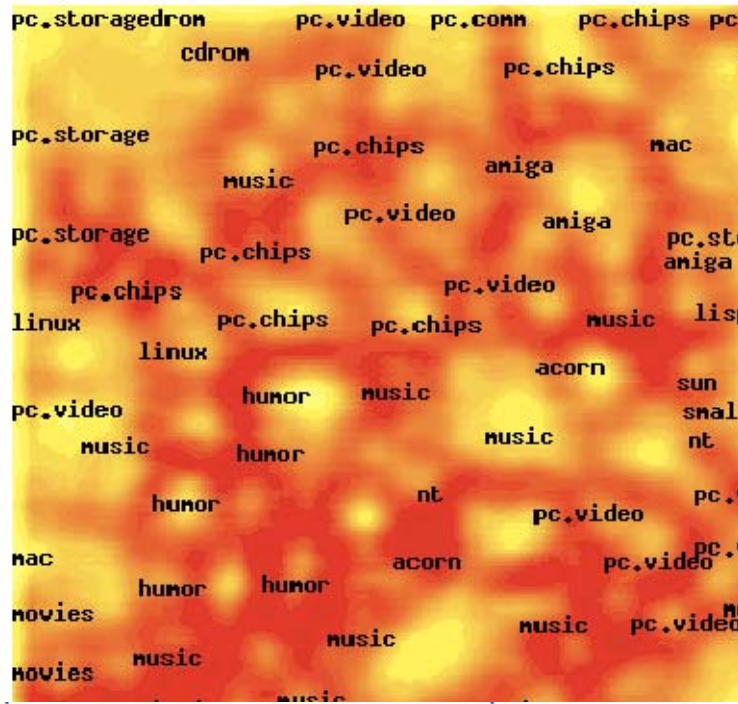
del.icio.us / tag /

your bookmarks | your network | subscriptions | links

Popular tags on del.icio.us

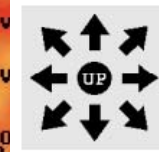
WEBSOM map - Million documents

[Instructions](#)



Explanation of the symbols on the map

- acorn - comp.sys.acorn.hardware
- amiga - comp.sys.amiga.hardware
- books - rec.arts.books
- cdrom - comp.publish.cdrom.hardware
- compilers - comp.compilers
- fuzzy - comp.ai.fuzzy
- genetic - comp.ai.genetic
- hp - comp.sys.hp.hardware
- humor - rec.humor
- lang.eiffel - comp.lang.eiffel
- lang.ml - comp.lang.ml



Click arrows to move to neighboring areas on the map, and to

This is a tag cloud

.net advertising ajax apple archite

book books business code com

design development diy

fashion finance firefox flash flickr

games gifts google graphics gtd t

icons illustration images impo

javascript language library lifehac

s m

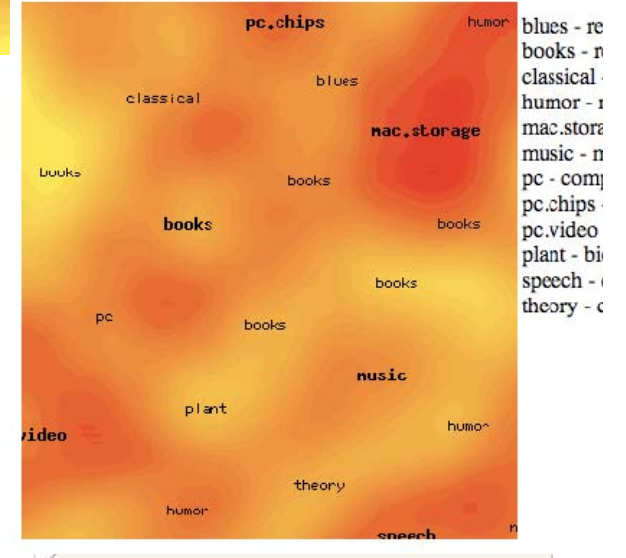
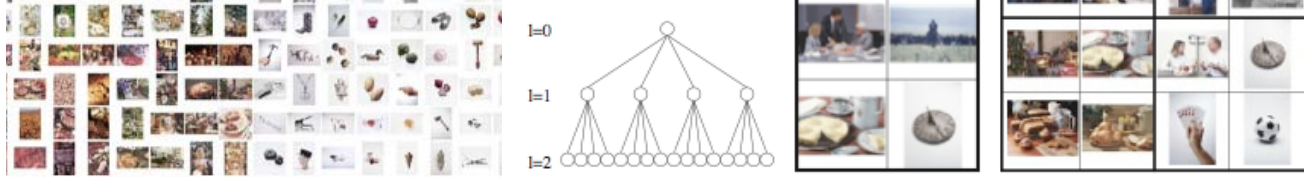
photos photoshop php podcast politics pom portfolio

python radio rais recipes reference research resources

arch security seo sex shopping social software

Similarity Pyramids for Browsing and Organization of Large Image Databases [†]

Jau-Yuen Chen[†], Charles A. Bouman[†] and John C. Dalton[‡]



SIDGrid

- **Multimodal data**
 - Collection and archiving
 - Collaborative annotation
 - Analysis
- **Widely applicable architecture**
 - Social and behavioral studies at many scales
 - Computer science
 - HCI, dB, pattern recognition, MM, Grid Multimodal Data
- **Sharing Better Resources**
- **Integrating Data, Computation, and Tools**
- **New Research Modalities**
- **New Analysis Techniques**

Questions?



STOP