Multimedia Web Analysis Framework towards Development of Social Analysis Software

Project granted by MEXT, Japan
(NII, Univ. of Tokyo, Waseda Univ.)

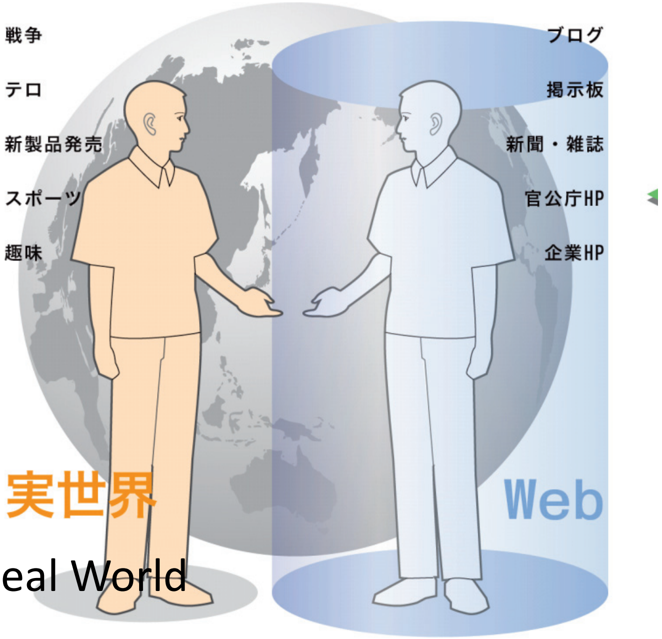
Masashi Toyoda ← **Shin'ichi Satoh**

IIS, University of Tokyo   National Institute of Informatics

Web as a projection of the real world

- Web has been reflecting various social events in the real and virtual world
- Web could be used as social sensor

War
Terro
Sports
Hobbies
Books
Movies

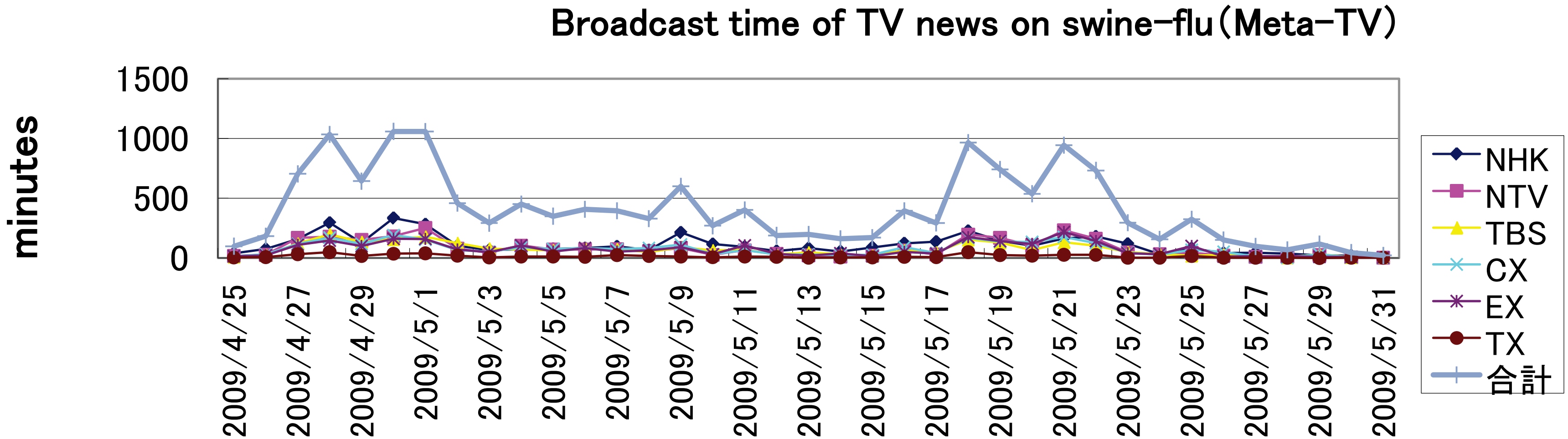


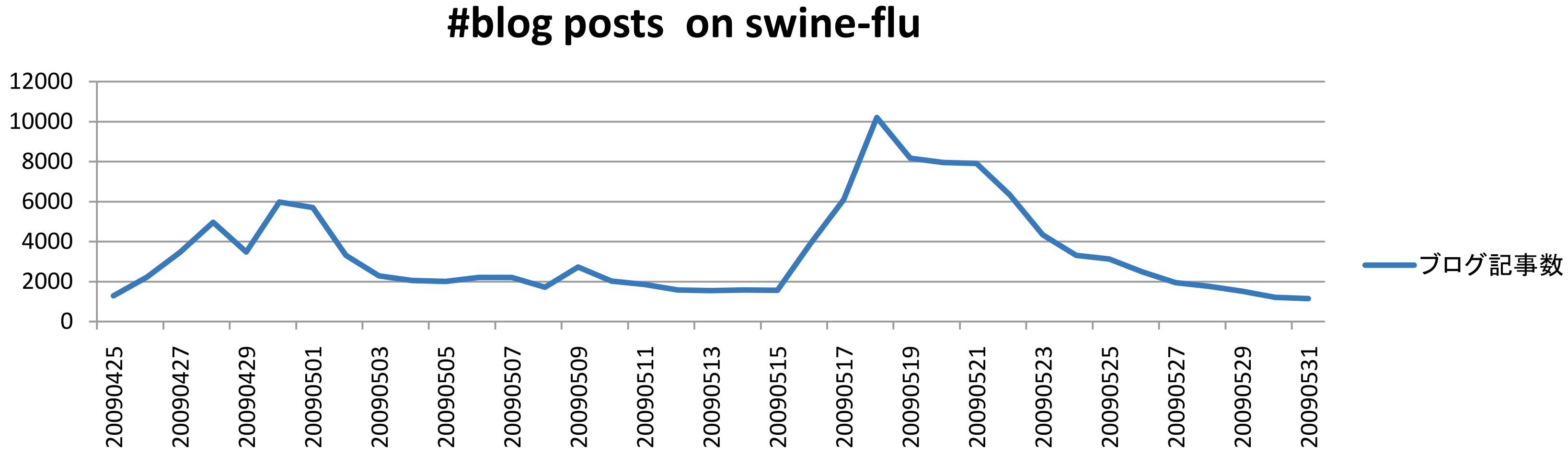
Real World

Online news/magazines
Blogs
SNS
BBS
Youtube

Correlation between Web and TV

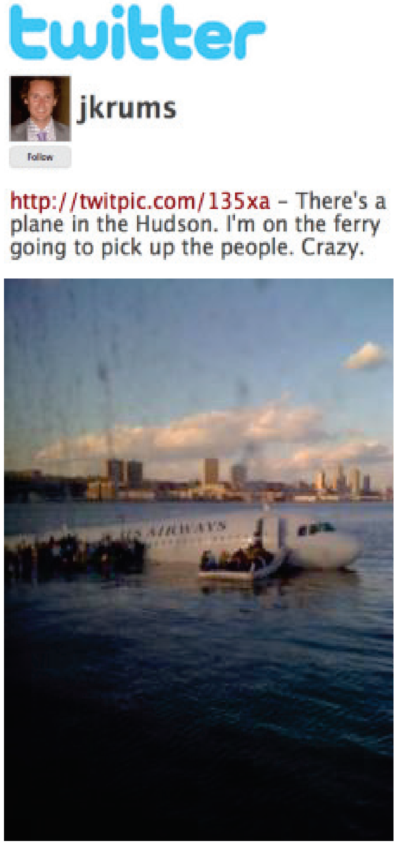
- In many case, trends in Web and TV are similar
  - Example: broadcast time of TV news and #blog posts on swine-flu





Cross-media evolution of topics

- The first photo of an accident uploaded to twitpic, and used in TV News

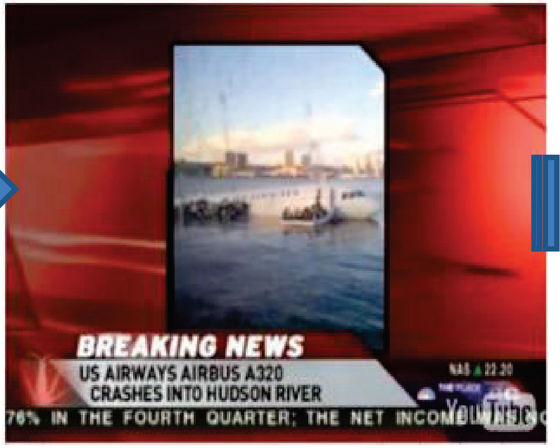


twitter

jkrums

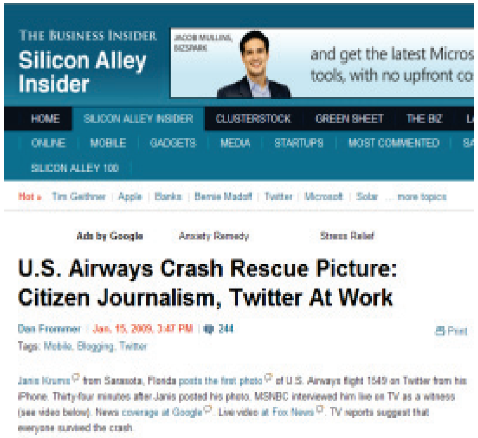
Follow

http://twitpic.com/135xa - There's a plane in the Hudson. I'm on the ferry going to pick up the people. Crazy.



BREAKING NEWS
US AIRWAYS AIRBUS A320
CRASHES INTO HUDSON RIVER
76% IN THE FOURTH QUARTER; THE NET INCOME
NAS ▲ 22.20
YouTube

MSNBC News (TV)



THE BUSINESS INSIDER
Silicon Alley Insider

JACOB MULLINS, BIZSPARK
and get the latest Micros
tools, with no upfront co

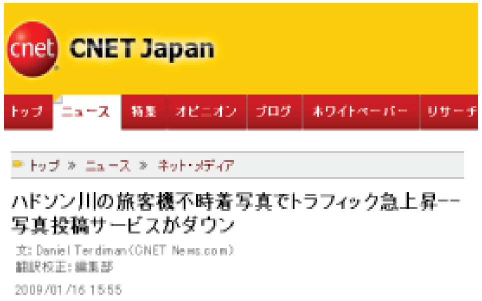
HOME | SILICON ALLEY INSIDER | CLUSTERSTOCK | GREEN SHEET | THE BIZ
ONLINE | MOBILE | GADGETS | MEDIA | STARTUPS | MOST COMMENTED
SILICON ALLEY 100
Hot » Tim Geithner | Apple | Banks | Bernie Madoff | Twitter | Microsoft | Solar ... more topics
Ads by Google Anxiety Remedy Stress Relief

U.S. Airways Crash Rescue Picture: Citizen Journalism, Twitter At Work

Dan Frommer | Jan. 15, 2009, 3:47 PM | 244 Print

Tags: Mobile, Blogging, Twitter

Janis Krums from Sarasota, Florida posts the first photo of U.S. Airways flight 1549 on Twitter from his iPhone. Thirty-four minutes after Janis posted his photo, MSNBC interviewed him live on TV as a witness (see video below). News coverage at Google. Live video at Fox News. TV reports suggest that everyone survived the crash.



CNET Japan

トップ | ニュース | 特集 | オピニオン | ブログ | ホワイトペーパー | リサーチ
トップ » ニュース » ネット・メディア

ハドソン川の旅客機不時着写真でトラフィック急上昇－－写真投稿サービスがダウン

文：Daniel Terdiman（CNET News.com）
翻訳校正：編集部

2009/01/16 15:55

Topic evolution from a photo appeared in twitter to TV and news broadcasting

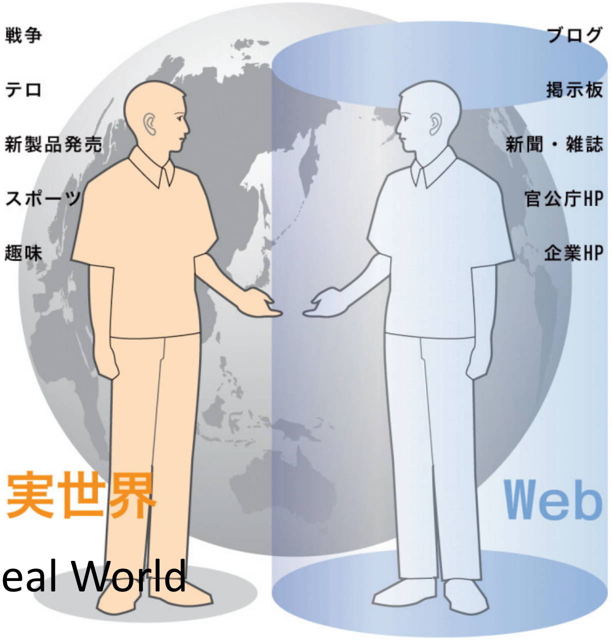
http://www.businessinsider.com/2009/1/us-airways-crash-rescue-picture-citizen-jouralism-twitter-at-work

Challenge: Tracking cross-media events

Web and media (TV, news papers, …) dynamically interact reflecting the real world events

Can we track the evolution of topics over multiple media?

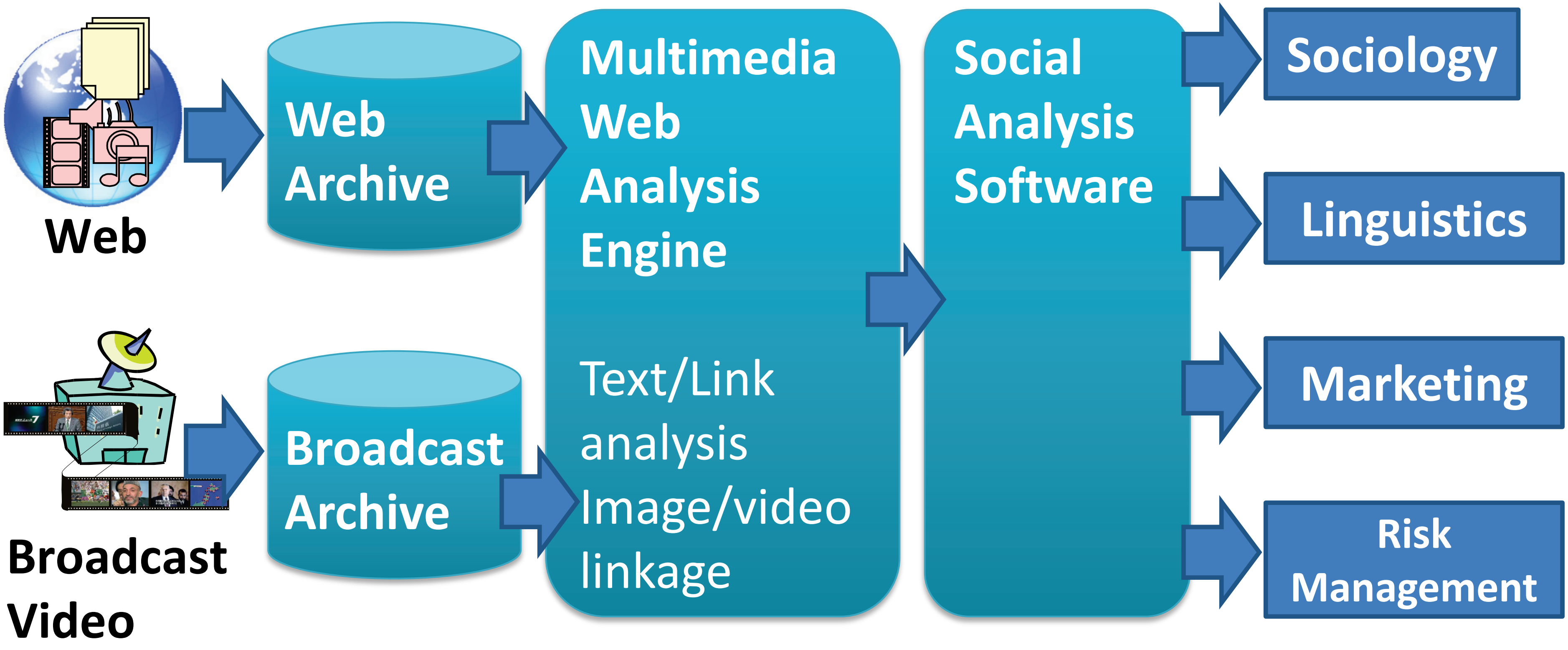
War
Terrorism
Sports
Hobbies



Real World

Online news/magazines
Blogs
SNS
BBS
Youtube

Multimedia Web Analysis Framework for Social Analysis



Web

Web Archive

Broadcast Video

Broadcast Archive

Multimedia Web Analysis Engine

Text/Link analysis

Image/video linkage

Social Analysis Software

Sociology

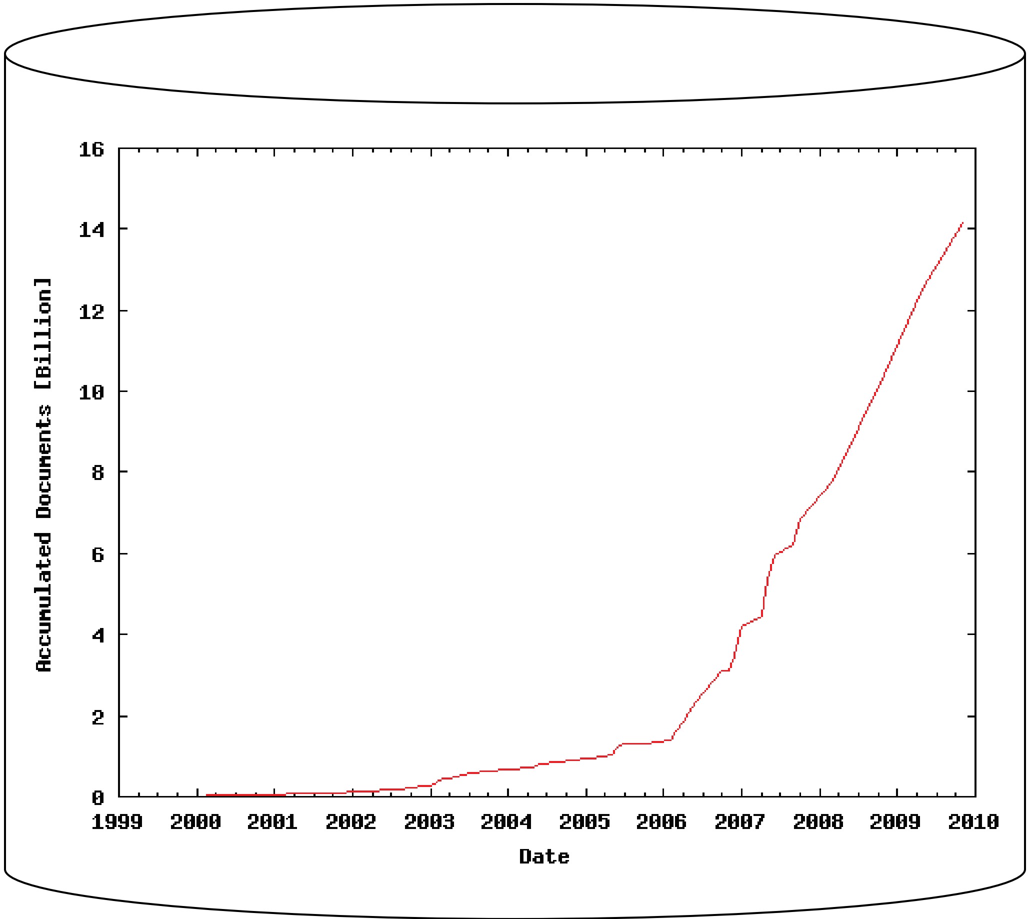
Linguistics

Marketing

Risk Management

10 Year Web Archive in Univ. of Tokyo

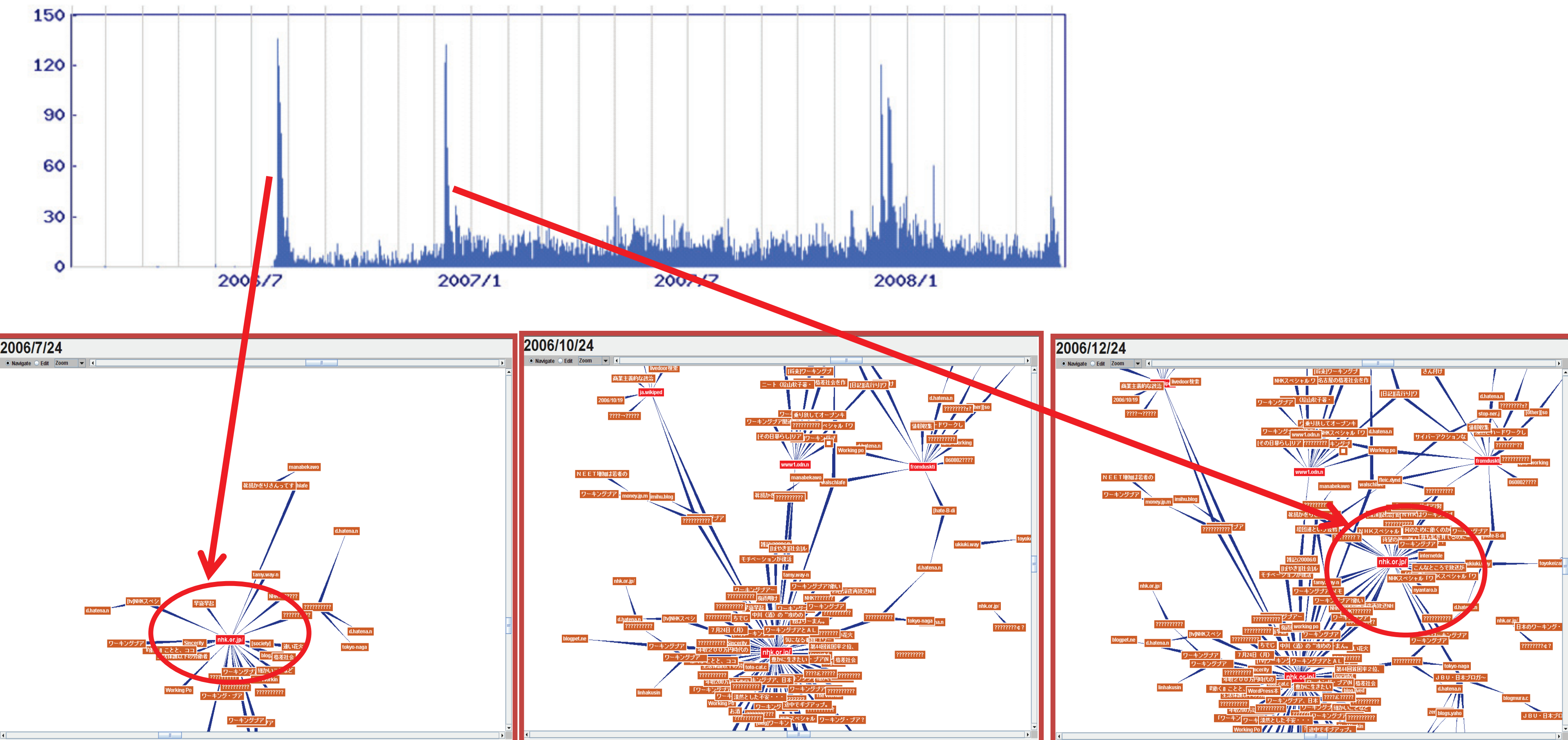
- One of the biggest Web archive in Asia
  - Focused on Japanese pages in any domain (1999～)
- Contents of **14 billion URLs** are stored in total including HTML texts and images (Dec. 2009)



Visual Analysis of Evolving Web

- The first page on a topic
- The most influential page

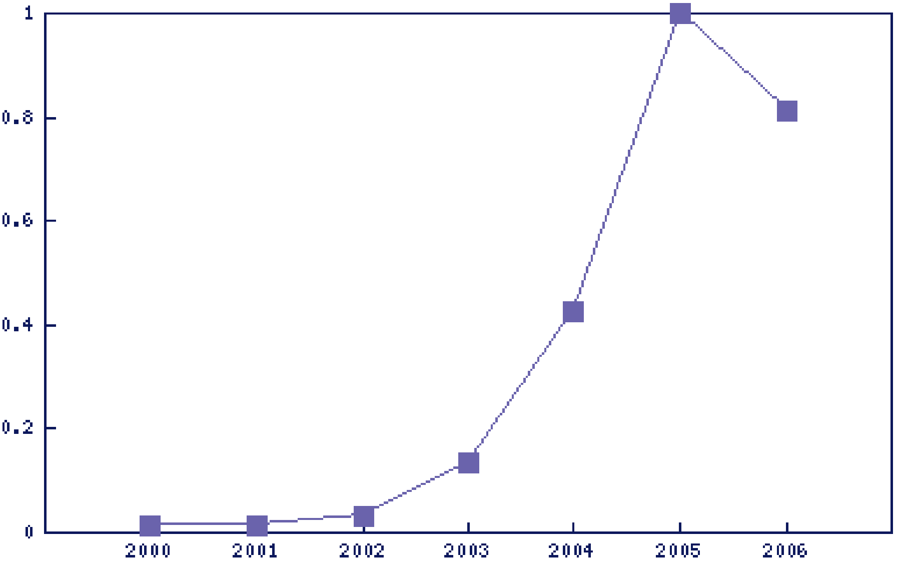
Example: diffusion of the “working poor” problem in blogosphere



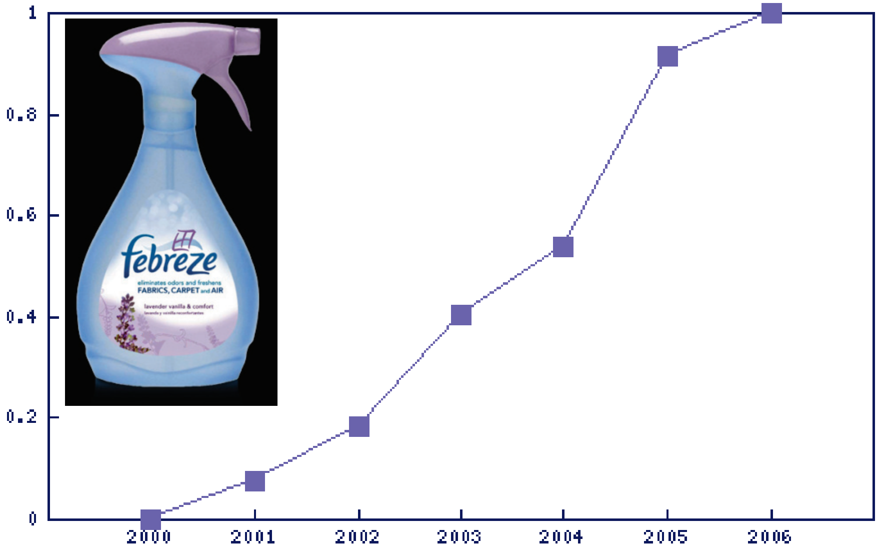
New Word Extraction for linguistics

- Words not in dictionary but used as noun, verb,…
  - Learning the context of known words, and predict Part-of-Speech tags of unknown words
  - Filtering by the growth ratio of occurrences

GUGU-RU (Verb: Search by Google)

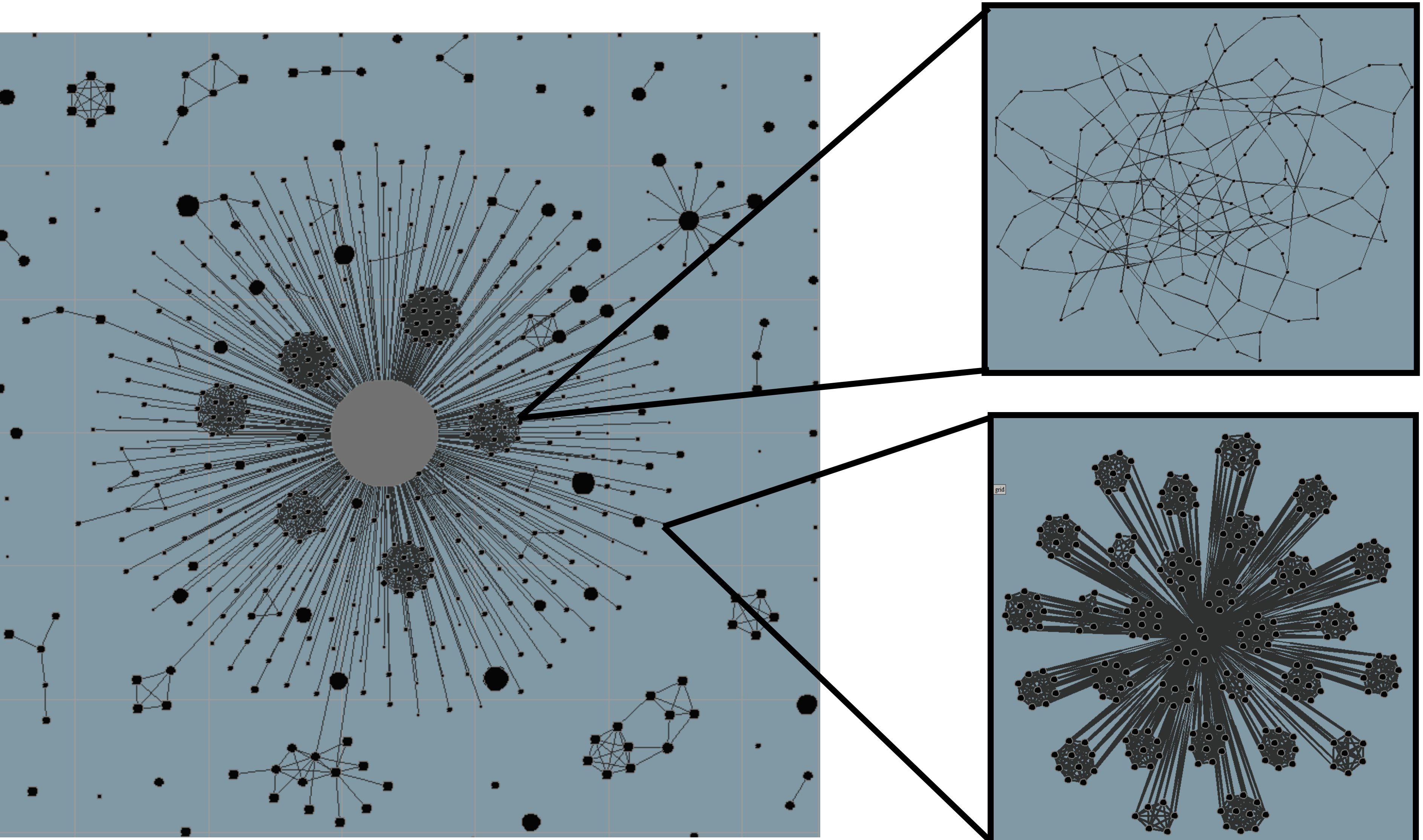


FABU-RU (Verb: Use Febreze, a fabric & air refresher by P&G)



Search Engine SPAM [Chung. 09, 10]

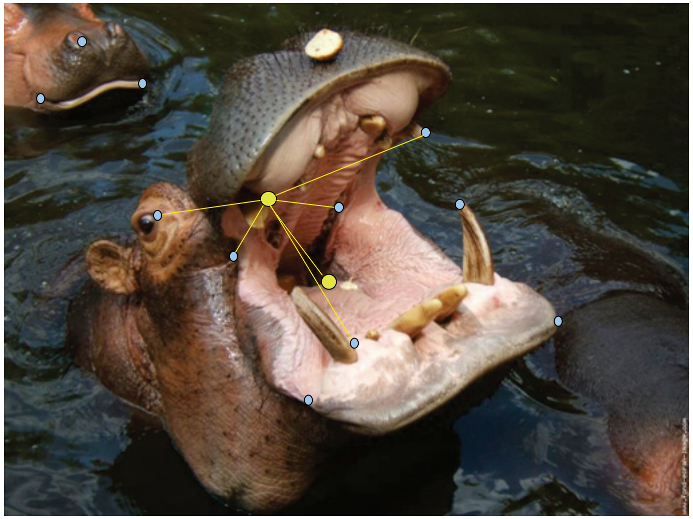
- Many web sites intend to deceive SEs and boost their ranking



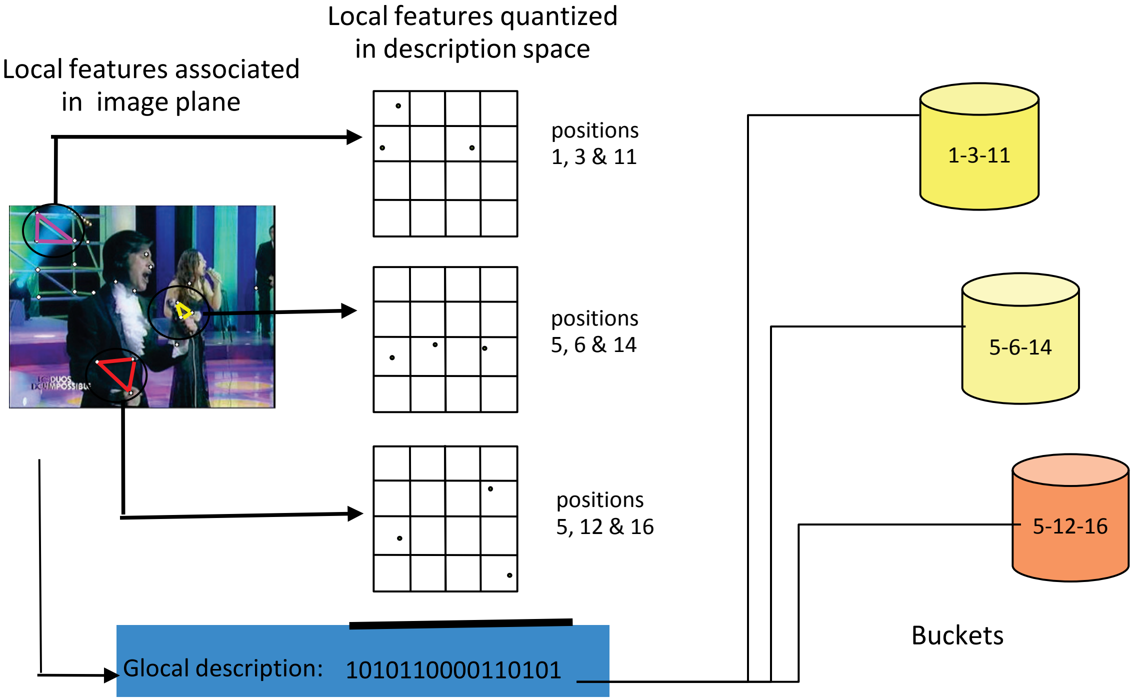
Broadcast Archive in NII

- Capture and store Japanese TV broadcast
  - 7 channels
  - About 1 year (60,000 hours)

Content-based image search in the video archives [Poullot et. al. 09]



1. Construct a database of the video archives (offline)
- Extract local description in KeyFrames
- Associate them in image place → define buckets in an index (spatial hashing)
- Built a global description using the local ones and insert it in this buckets



Local features quantized in description space
Local features associated in image plane
positions 1, 3 & 11
positions 5, 6 & 14
positions 5, 12 & 16
1-3-11
5-6-14
5-12-16
Buckets
Glocal description: 1010110000110101

2. Search (online)
- Same extraction process on query picture
- Explore matching buckets
- Rank result by similarity of global description

Summary

- Tracking cross-media events is an important issue for social analysis
- Enabler: Large-scale Web and broadcast archive
- Analysis Engine:
  - Text/link based tracking techniques
  - Scalable video indexing technique
- Next step: Integration of Web/image analysis
  - Text/link/image based topic tracking
  - CF filter, face recognizer