

# Adaptive News Access

Daniel Billsus and Michael J. Pazzani

The Adaptive Web

2007

Edited by Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl

550–70

Lecture Notes in Computer Science 4321

Springer Berlin Heidelberg

Presentation by Daniel Steinberg

February 10, 2015

Many Slides Use Text Directly from the Chapter and Papers,  
without Proper Quotations

# Motivation

- Virtually every news organization now has a presence on the World Wide Web
- The web is increasingly evolving into the most powerful news delivery platform of the 21st century
- We need new technology to help leverage the full potential of web-based news distribution
  - Continued growth of online news content is of limited use if we cannot find the personally most relevant and useful information

# Types of Adaptive News Access

- News Content Personalization
  - Personalize content help users find personally relevant news stories
    - based on a model of the user's interests
  - Recommend or automatically rank stories
- Adaptive News Navigation
  - Assists the user in navigating to the most frequently read sections of a news site

# Types of Adaptive News Access

- Contextual News Access
  - Provide users with news content on the basis of currently viewed information
- News Aggregation
  - Automated aggregation and classification of news content
    - Helps users identify ongoing or emerging news topics
    - Assists in accessing coverage of a specific topic by multiple providers
  - Does not necessarily enable personalization
    - But automated aggregation commonly exhibits adaptive behavior
      - dynamically generated aggregator pages

# News Content Personalization

- Customization
  - User-defined news profiles
    - e.g., customize the news categories to be included on the front page
  - Presented in Contrast to Personalization
- Personalization
  - Adaptive Techniques
  - Model the user's interests based on explicit or implicit feedback, and use the resulting user models to personalize news content

# Important Factors for Designing Adaptive Approaches

- Dynamic Content
  - News content is more dynamic than many other content types, such as movies, music or books
  - News stories are released and updated continuously
  - Content-based methods better suited to news personalization than collaborative methods
    - “sparse matrix” problem is particularly noticeable for news access
      - Limited rating overlap
    - “latency problem”
      - It may take some time for news stories to receive enough user feedback to lead to accurate recommendations

# Important Factors for Designing Adaptive Approaches

- Changing Interests
  - User models should be based on learning algorithms that can quickly adjust to changing interests
    - System should discover interest changes quickly
  - Concept Drift
    - The variable being predicted changes over time in unforeseen ways
      - Less accurate predictions as time passes
- Multiple Interests
  - Users are usually interested in a broad range of different news topics
    - k-nearest-neighbor methods are a good choice to address this issue

# Important Factors for Designing Adaptive Approaches

- Novelty
  - A news story is usually considered most interesting if it conveys information the user does not yet know
    - Keeping track of information the user has previously accessed can help select content that is **similar, but not identical to,** previously accessed information
    - Newsjunkie personalizes news for users
      - identifies the novelty of stories in the context of stories they have already reviewed



# Important Factors for Designing Adaptive Approaches

- Avoiding Tunnel Vision
  - Personalization should not prevent the user from finding important novel information or breaking news stories
  - Possible solutions
    - Integrate editorial input into the recommendation algorithm
    - Explicitly boost the diversity of stories presented to the user
- Editorial Input
  - Retaining editorial input is an important feature for news organizations that are interested in deploying personalization technology
  - Loss of control over the content that users will get to see does not appeal to news organizations

# Important Factors for Designing Adaptive Approaches

- **Brittleness**
  - A single action, such as selecting something accidentally or skipping over an article on a topic should not have a drastic or unrecoverable effect on the model of the user's interests
- **Availability of Meta-Tags**
  - News personalization algorithms can usually not rely on the availability of meta-tags
    - Manually adding meta-tags is not feasible with thousands of new items being added daily

# Example Systems

- Krakatoa Chronicle
  - The first newspaper on the World Wide Web to provide a layout similar to that of real-world newspapers
- Anatagonomy
  - Research prototype
  - Tracks user interactions such as selecting and enlarging articles or scrolling through articles, and interprets these interactions as implicit feedback
- Slider
  - When new stories arrive, users can optionally “slide” these stories onto a panel
  - For each panel, a prototype vector is generated by averaging over all document vectors
  - In subsequent sessions, the system attempts to locate news stories that are similar to these prototype vectors

# Findory



- Recommended stories are annotated with a “sun” icon
  - Quote on site suggests that the system is based on a combination of content-based and collaborative methods

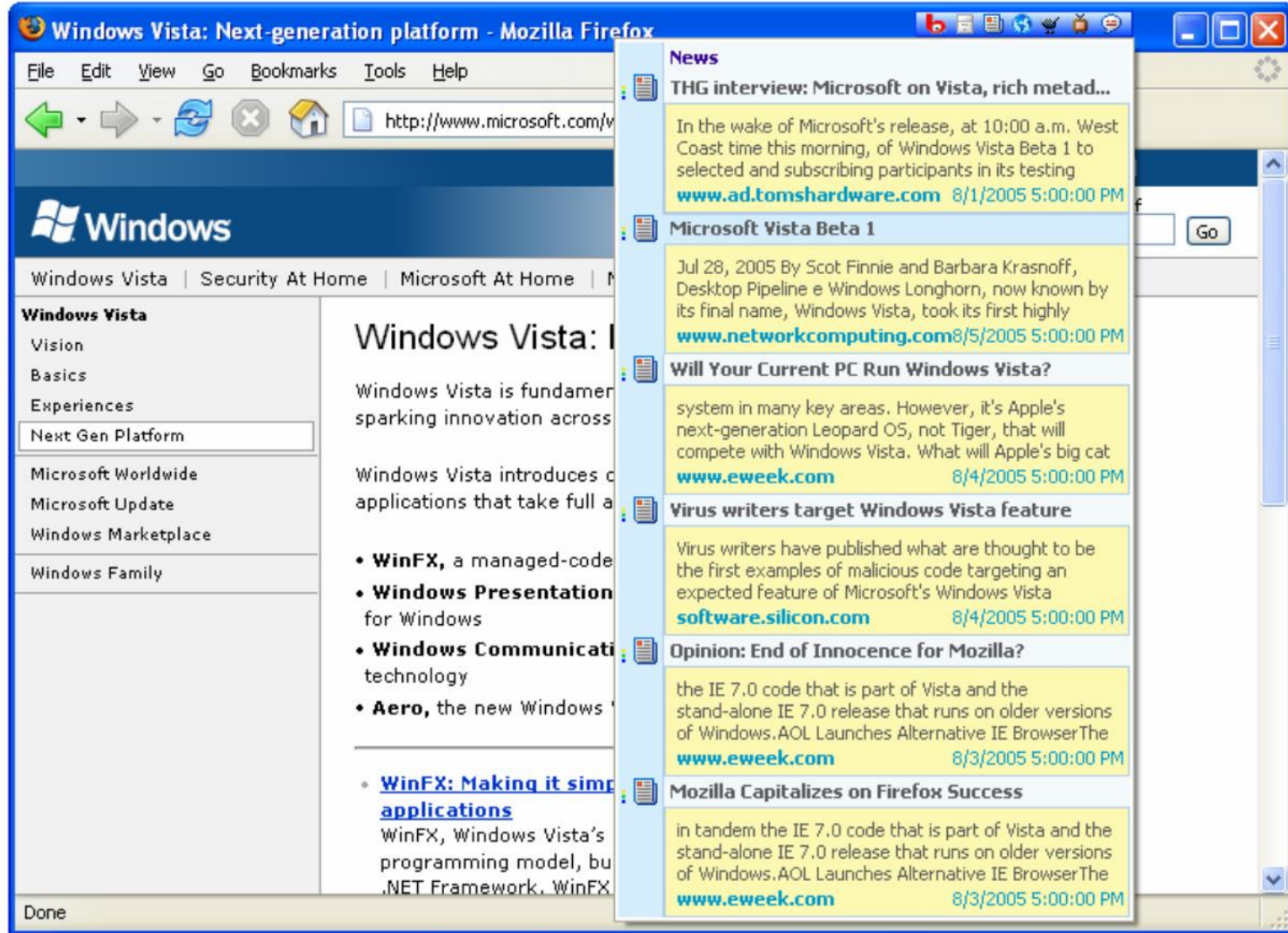
# Adaptive News Navigation

- Analyze the user's access patterns to determine the position of menu items within a menu hierarchy
  - As opposed to finding individual news stories that match the user's interest profile
- Particularly effective for mobile applications on PDAs and cell phones
- Smyth and Cotter Algorithm
  - Estimate the probability that a user will select option  $o$  given that it is included in menu  $m$ , and uses these probabilities to construct menus that are most likely to contain options the user will select

# Contextual Recommendations

- The currently viewed web page or email message is used as an expression of a user's current interests
  - via plug-ins for commonly-used applications
- The system extracts textual information currently displayed on the user's screen
  - Statistical term-weighting techniques are used to identify informative terms
  - Natural Language Processing approaches are also used
    - Named-entity-tagger for identifying the names of companies, people or products

# Blinkx



- Recommends news stories related to a web page the user is viewing

# News Aggregators

- Automatically aggregate content from many different news sources
  - As a result, they adapt to the current news landscape as a whole
- In contrast to inferring users' interests based on the browsing history or currently viewed information
- RSS feeds have significantly contributed to the emergence of news aggregators and other news-related services
- Allow users to compare coverage of a story between different providers, leading to a greater variety of perspectives than one single organization can offer



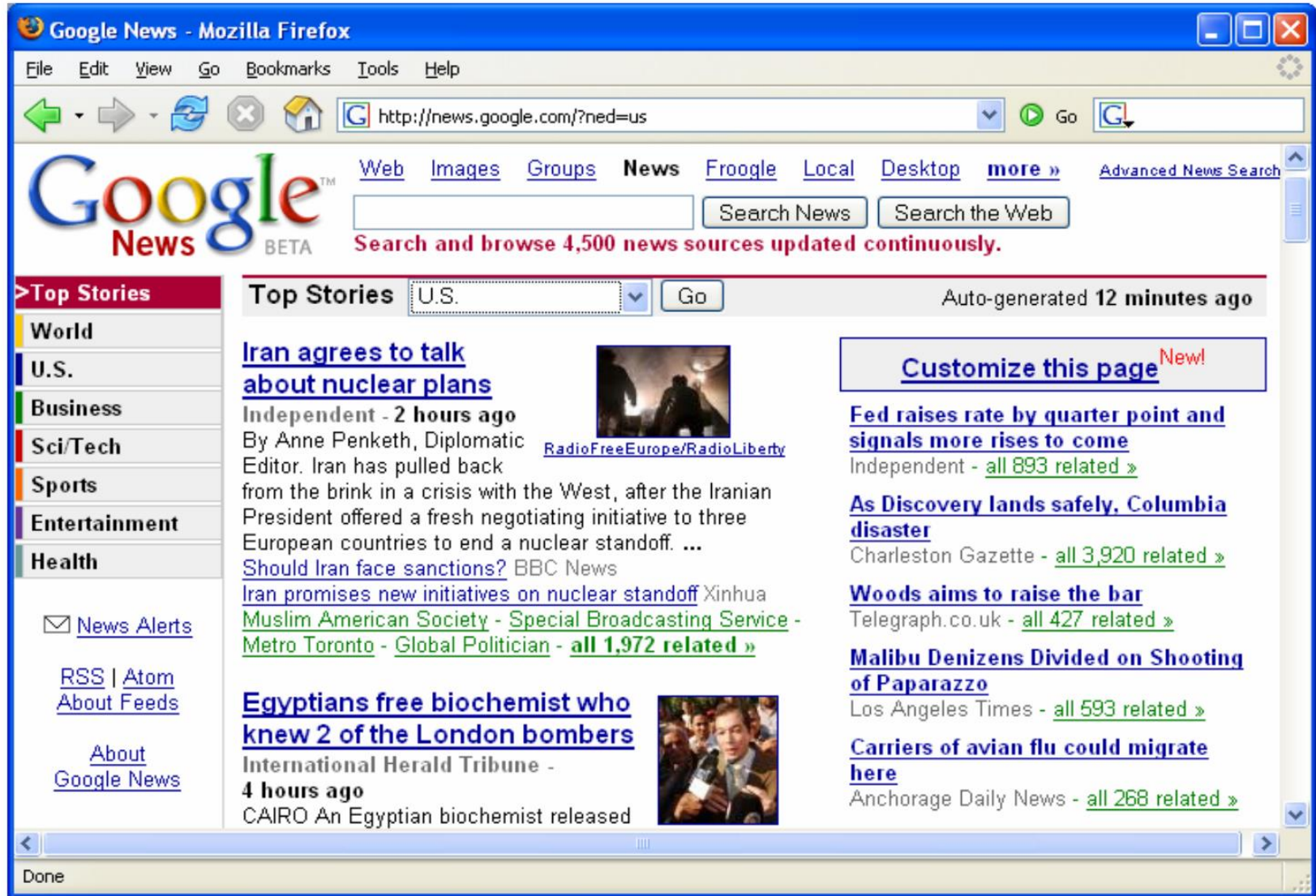
# RSS Feeds

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <title>My Gaming News</title>
    <link>http://mygamingnews.com/</link>
    <description>Gaming News</description>
    <item>
      <title>600,000 Xbox 360 units sold in US</title>
      <link>http://mygamingnews.com/story1.html</link>
      <description>LOS ANGELES -- Microsoft Corp.
        has sold 600,000 of its new XBox 360
        videogame consoles ...
      </description>
    </item>
    <item>
      <title>'Grand Theft Auto' slapped with lawsuit
      </title>
      <link>http://mygamingnews.com/story2.html</link>
    </item>
  </channel>
</rss>
```

# Google News

- A popular news aggregation service
- Adaptively generates a front page without explicit editorial input from a human editor
  - Acts as an unbiased news editor
- Topix is similar, but emphasizes news categorization and geo-coding (e.g., labels stories with the location of events)

# Google News



- Google News' automatically generated front page

# Case Study

- Personalized news service that was first released in 1999, powering the mobile version of various publicly available news services for several years
- The application dynamically generates a user interface that can be rendered on PDAs and cell phones
- The interface displays section names (such as 'Sports'), headlines and articles.
- Uses a machine learning approach to automatically learn a simple model of each user's individual interests
  - Content-based approach specifically designed for news access
    - Similarity-Based methods
    - Bayesian methods

# Case Study

- Multi-strategy learning approach that learns two separate user-models
  - short-term interests
    - contains recently read stories, so that other stories which belong to the same event thread can be identified
    - allows for identification of stories that the user already knows
    - e.g., k-nearest-neighbor algorithm using  $n$  most recent stories
  - long-term interests
    - model a user's general preferences

# Modeling Long Term Interests

- Use tf-idf to select informative words that recur over a long period of time
- The top 50 long-term features selected from a set of 10,000 science news stories
  - drug, cancer, space, cells, patients, women, crops, gene, launched, disease, food, virus, rocket, city, mission, bacteria, infection, children, heart, hiv, satellite, eclipse, blood, genetic, suns, winds, trial, mice, orbit, antibiotics, vaccine, resistance, russian, human, aides, storm, percent, brain, fda, cdc, mosquitoes, energy, test, damage, hurricane, computer, baby, government, hospital, texas.

# Case Study

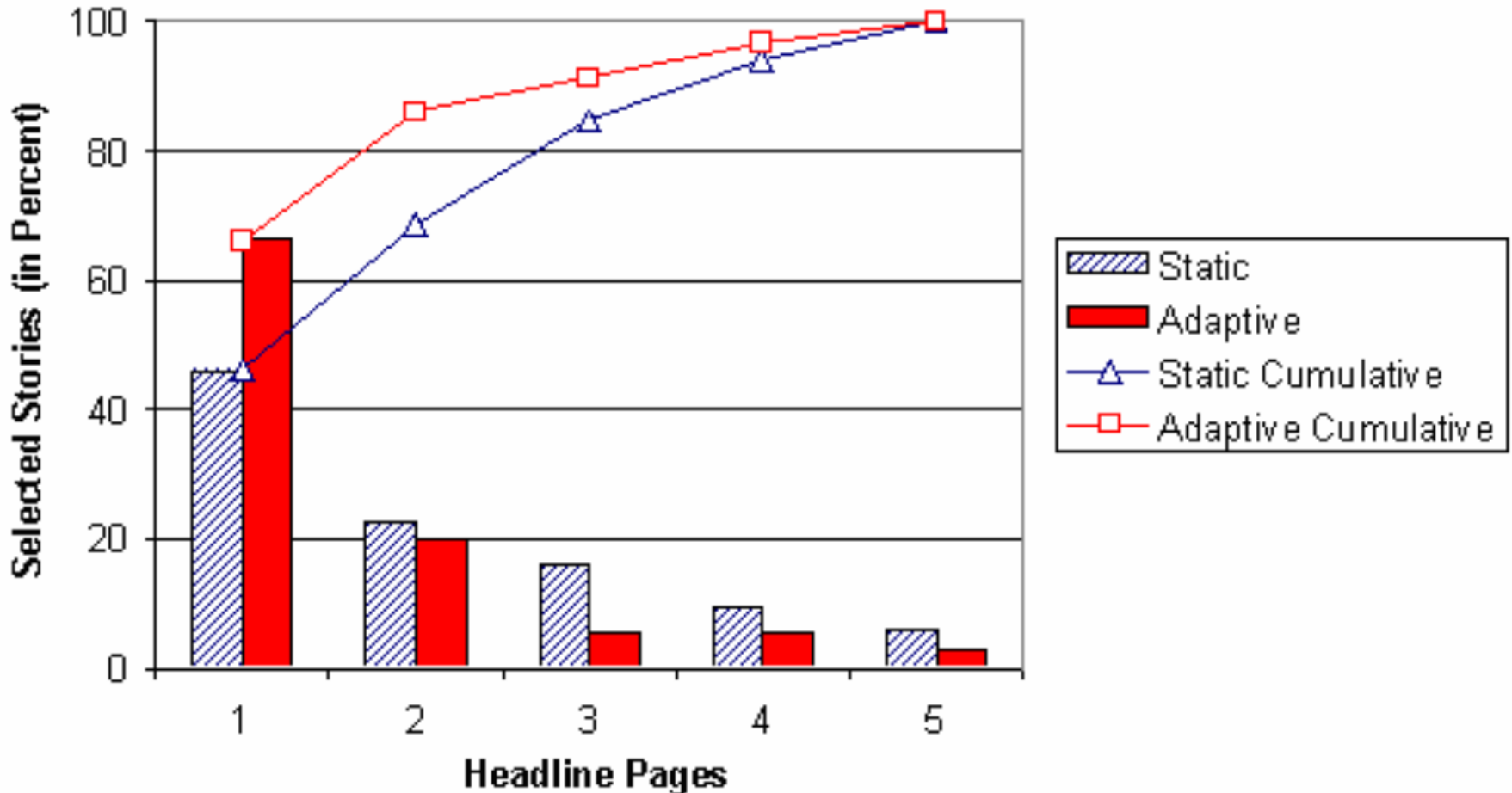
- To predict whether a user would be interested in a news story, the system applies the two models sequentially
  - short-term model first, because it is based on the most recent observations only
  - If a story cannot be classified with the short-term model, the long-term model is used
- Editorial input is incorporated by boosting the priority of lead stories

# Alternating Sessions Experiment

- On odd days, users with odd account registration numbers received news in personalized order and even users received a static order
- On even days, this policy was reversed
- The average display rank of selected stories was 6.7 in the static mode and 4.2 in the adaptive mode
  - Based on 50 users that selected 340 stories out of 1882 headlines



# Alternating Sessions Experiment

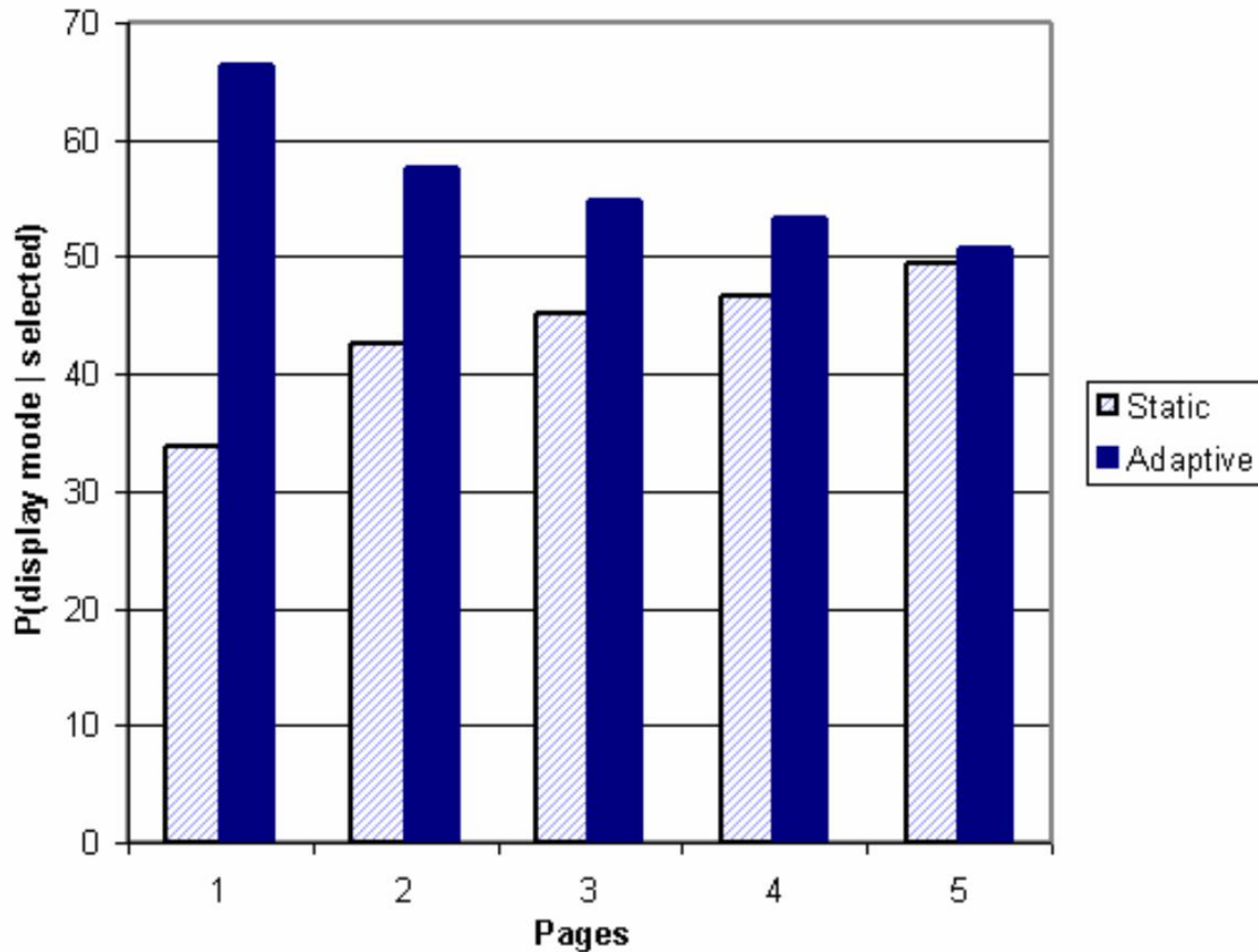


- Distribution of selected stories
  - In the static mode, 68.7% of the selected stories were on the top two headline screens
  - In the personalized mode, 86.7% of the selected stories were on the top two headline screens

# Alternating Stories Experiment

- A drawback of the “alternating sessions experiment” is that users might not see stories they would have seen in the adaptive mode
- During this experiment, the client was configured to display four stories on each screen, with every screen containing two adaptive stories and two static stories
- The difference between the two average display ranks was not as pronounced as in the “alternating sessions” experiment
  - 5.8 for the static mode
  - 5.27 for the adaptive mode
- Users selected 13.26% of all displayed static stories (663 stories), vs 19.02% (951 stories) of all displayed adaptive stories

# Alternating Stories Experiment



- $p(\text{adaptive} \mid \text{selected})$  versus  $p(\text{static} \mid \text{selected})$  for separate headline screens
- On the first headline screen,  $p(\text{static} \mid \text{selected}) = 0.33$ , versus  $p(\text{adaptive} \mid \text{selected}) = 0.66$

# Recent Trends and Systems

- Personalizing Audio and Video News Feeds
  - Future Work
    - Automatically extract text from audio files
    - Enable full-text search
    - Find topic-based segments within audio
    - Use content-based recommendation techniques to assemble personalized podcasts
- Personalization and the Blogosphere
  - NewsGator can recommend news and blog feeds, based on a collaborative approach that uses a user's subscriptions as the basis for personalized feed recommendations
  - [www.technorati.com](http://www.technorati.com) or [www.blogdigger.com](http://www.blogdigger.com) are beginning to incorporate customization features
    - usually static and do not adapt to the user's interests

# Zeitgeist

- A German word that means "the spirit (Geist) of the time (Zeit)"
- Many sites are incorporating Zeitgeist features
- The goal is to automatically identify the most popular or talked about topics
- Daypop
- ([www.daypop.com](http://www.daypop.com)) uses blogs as a Zeitgeist meter for news content, by generating a list of the 40 news stories that are most frequently cited in the blogosphere
- Digg ([www.digg.com](http://www.digg.com))
  - Users submit potentially interesting news articles or blog posts to the site
  - Digg's user community expresses interest in the submitted stories by clicking corresponding "digg it" buttons.
  - The stories with the most "diggings" are then prominently displayed on the site

# News@hand: A Semantic Web Approach to Recommending News

Cantador, Iván, Alejandro Bellogín, and Pablo Castells

Adaptive Hypermedia and Adaptive Web-Based Systems

2008

Edited by Wolfgang Nejdl, Judy Kay, Pearl Pu, and Eelco Herder

279–83

Lecture Notes in Computer Science 5149

Springer Berlin Heidelberg

Slides Use Text Directly from the Chapter and Papers,  
without Proper Quotations

# News@hand

- A news recommender system which applies semantic-based technologies
  - describes and relates news content and user preferences in order to produce enhanced recommendations
  - News items and user profiles are represented in terms of concepts appearing in domain ontologies, and semantic relations among those concepts are exploited to enrich the above representations, and enhance recommendations.

# News@hand

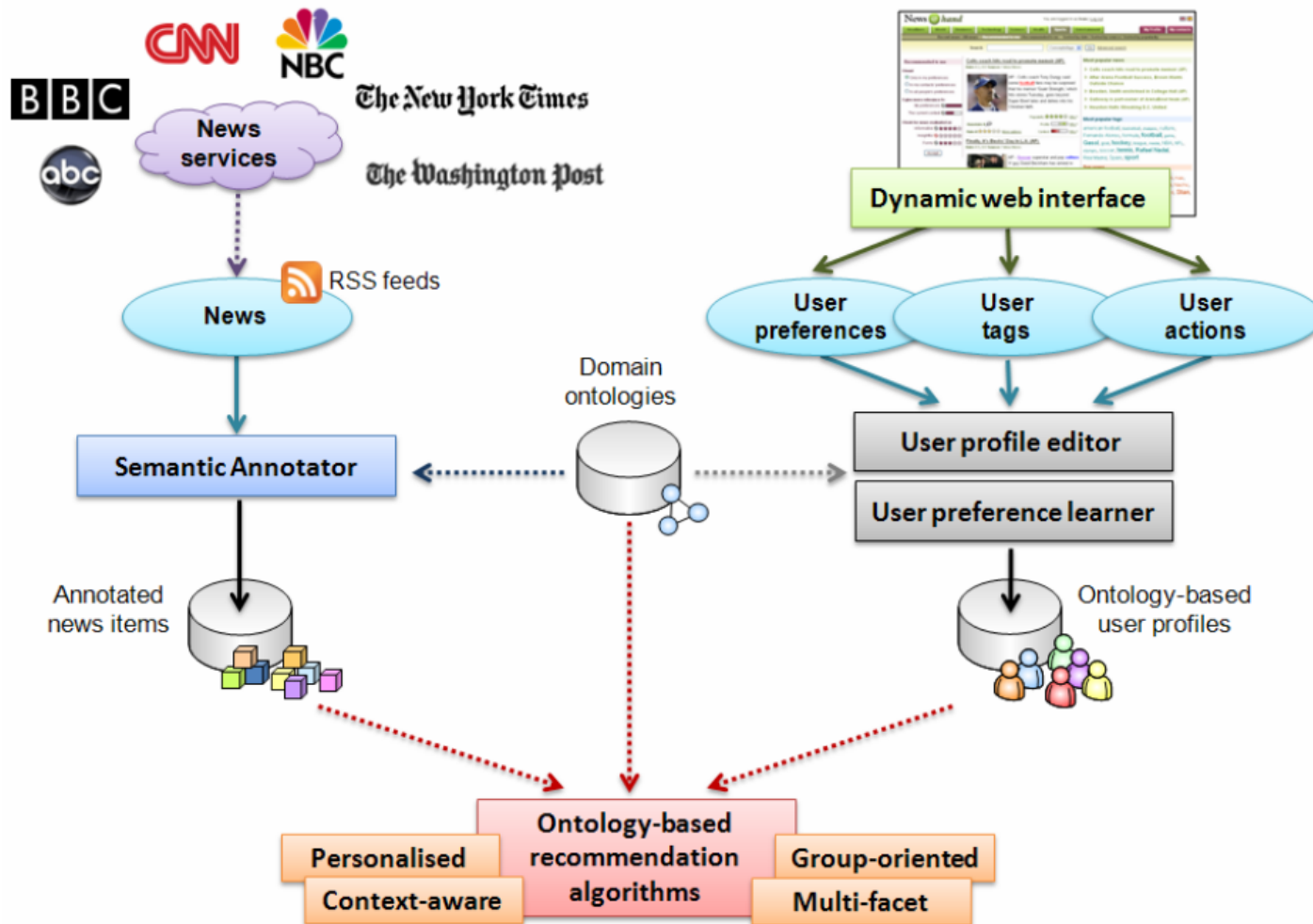
- News are automatically and periodically retrieved from several on-line news services via RSS feeds
- The title and summary of the retrieved news are then annotated with concepts (classes and instances) of the domain ontologies available to the system.
  - A total of 17 ontologies have been used for the first version of the system
  - Adaptations of the IPTC ontology
    - Multiple domains such as education, culture, politics, religion, science, technology, business, health, entertainment, sports, etc.



# News@hand

- Explicit and implicit user preferences are taken into account, via manual preferences, tags and ratings, and via automatic learning from the users' actions
- Uses the semantically annotated news items and the defined ontology-based user profiles to provide personalized, context-aware, group-oriented, and multi-facet recommendations

# News@hand



# News@hand

News @hand

You are logged in as **ivan** | [Log out](#)



Headlines

World

Business

Technology

Science

Health

Sports

Entertainment

My Profile

My contacts

Recent news | All news | Recommended to me | Recommended to us | Sorted by date | Sorted by source | Sorted by popularity

Search

Concepts/tags

[Advanced search](#)

## Recommended to me

### I trust

- Only in my preferences
- In my contacts' preferences
- In all people's preferences

### I give more relevance to

My preferences

The current context

### I look for news evaluated as

Informative  ★★★★★☆

Insightful  ☆☆☆☆☆☆

Funny  ★★★★★☆

## Colts coach hits road to promote memoir (AP)

Date: 6 2, 0:0 Source: Yahoo News



AP - Colts coach Tony Dungy said some **football** fans may be surprised that his memoir 'Quiet Strength,' which hits stores Tuesday, goes beyond Super Bowl tales and delves into his Christian faith.

Popularity ★★★★★☆ Why?

Profile  Why?

Context  Why?

Annotate it

Rate it ★★★★★☆ [More options](#)

## Finally, it's Becks' Day in L.A. (AP)

Date: 6 5, 0:0 Source: Yahoo News



AP - **Soccer** superstar and pop **culture** 'it' guy David Beckham has arrived in Southern California, where he will be introduced as the newest — and by far most the most luminous — star of the Galaxy.

Popularity ★★★★★☆ Why?

Profile  Why?

Context  Why?

Annotate it

Rate it ★★★★★☆ [More options](#)

## Most popular news

- [Colts coach hits road to promote memoir \(AP\)](#)
- [After Arena Football Success, Brown Wants Outside Chance](#)
- [Bowden, Smith enshrined in College Hall \(AP\)](#)
- [Galloway is part-owner of ArenaBowl team \(AP\)](#)
- [Houston Halts Streaking D.C. United](#)

## Most popular tags

american football, basketball, champion, culture, Fernando Alonso, formula, **football**, game, Gasol, goal, **hockey**, league, medal, NBA, NFL, olympic, soccer, **tennis**, **Rafael Nadal**, Real Madrid, Spain, **sport**

## Top users

**Alex**, Barbara, David, Enrico, **Erich**, Farah, Frank, Grady, **Ivan**, John, **Larry**, Michael, **Miriam**, Nacho, Nordin, Pablo, Patrick, Ralph, Rebecca, Richard, **Stan**, Steve, Tom, Vanessa

# News@hand Problems Solved

- Domain dependency
  - Ontologies and Semantic Web standards makes it possible to easily incorporate new domains into the system, and export the obtained knowledge to other applications.
- Restricted content analysis
  - Annotation mechanism allows the extraction of metadata from different multimedia sources, such as texts, videos, or audios
- Content overspecialisation, cold-start, portfolio and sparsity
  - Ontology properties enable the detection of further co-occurrences of interests between users, and finds new interests, available for recommendations

# News@hand Problems Solved

- Gray sheep
  - Hybrid models compare user profiles at different semantic interest layers, enabling further possibilities to find relations between users
- Group-oriented recommendations
  - Vector-based preference description facilitates the combination of multiple profiles to generate a shared profile for groups of users
- Context-aware recommendations
  - Semantic runtime context is applied to provide recommendations according to the live user interests

# Adaptive model for recommendation of news

Matus Medo, Yi-Cheng Zhang, and Tao Zhou

EPL (Europhysics Letters)

2009

88 (3): 38005

doi:10.1209/0295-5075/88/38005

Slides Use Text Directly from the Chapter and Papers,  
without Proper Quotations

# Adaptive Model for Recommendation of News

- Most news recommender systems try to identify users' interests and news' attributes and use them to obtain recommendations
- They propose an adaptive model which combines similarities in users' rating patterns with epidemic-like spreading of news on an evolving network

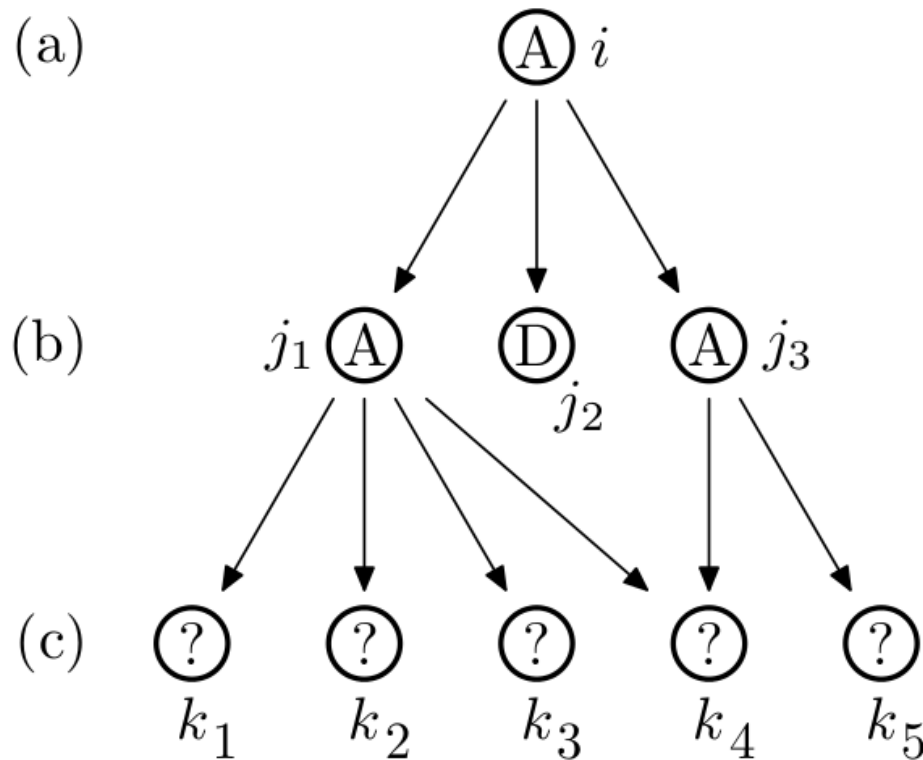
# Description of the Model

- Estimate pairwise user similarity
  - from users' assessments of the news
- Those  $S$  users who are most similar to a given user  $i$  we refer to as authorities of  $i$  and, conversely, those who have user  $i$  as an authority we refer to as followers of  $i$ .
- While the number of authorities for each user is fixed, a highly valued user may have a large number of followers.
- Lacking any prior information, we assume random initial assignment of authorities.



# Description of the Model

- Construct a directed network of authorities and followers



# Description of the Model

- Propagate news through this network
- Updating the assignment of authorities
  - Authorities of user  $i$  should be always those  $S$  users who have the highest rating similarity with  $i$

# Evaluating the Model

- Evaluations were run Using simulations that varied the parameters of the model
- *Approval Fraction* tells us how often are users satisfied with the news they get recommended
- No comparisons with alternative approaches
- Future research
  - Testing with real users

# Adaptive User Profile Model and Collaborative Filtering for Personalized News

Jue Wang, Zhiwei Li, Jinyi Yao, Zengqi Sun,  
Mingjing Li, and Wei-ying Ma

Frontiers of WWW Research and Development – APWeb  
2006

Edited by Xiaofang Zhou, Jianzhong Li, Heng Tao Shen, Masaru Kitsuregawa, and  
Yanchun Zhang

474–85

Lecture Notes in Computer Science 3841  
Springer Berlin Heidelberg

Slides Use Text Directly from the Chapter and Papers,  
without Proper Quotations

# Adaptive User Profile Model and Collaborative Filtering for Personalized News

- In Current content-based filtering approaches, there is a lack of ability to model user's interests at the event level
- Propose a novel approach to user profile modeling
  - User's interests are modeled by a multi-layer tree with a dynamically changeable structure
    - The top layers of which are used to model user interests on fixed categories
    - The bottom layers are for dynamic events

# Adaptive User Profile Model

