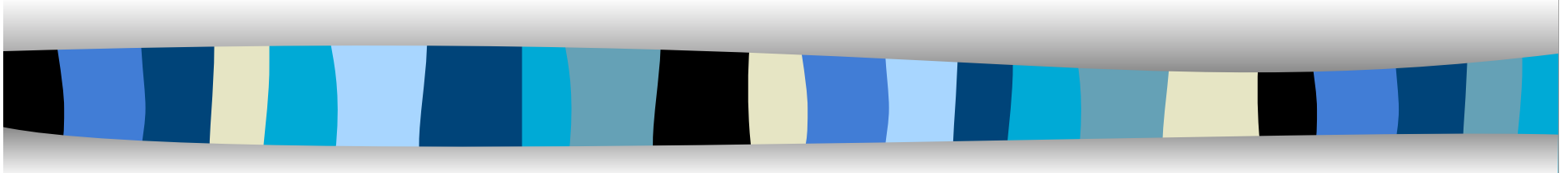


INFSCI 2480

Adaptive Information Systems



# Personalized Web Search



Peter Brusilovsky

<http://www.sis.pitt.edu/~peterb/>

# Where we are?

	Search	Navigation	Recommendation
Content-based			
Semantics / Metadata			
Social			



# Why Search Personalization?

- R. Larsen: With the growth of DL even a good query can return not just tens, but thousands of "relevant" documents<sup>1</sup>
- Personalization is an attempt to find most relevant documents using information about user's goals, knowledge, preferences, navigation history, etc.

<sup>1</sup> Larsen, R.L. Relaxing Assumptions . . . Stretching the Vision: A Modest View of Some Technical Issues. D-Lib Magazine, 3, April (1997), available online at <http://www.dlib.org/dlib/april97/04larsen.html>



# User Profile Where we are?

- Common term for user models in IR/IF
- A user's profile is a collection of information about the user of the system.
- This information is used to get the user to more relevant information
- Views on user profiles in IR community
  - Classic (Korfhage) - a reference point
  - Modern - simple form of a user model



# Core vs. Extended User Profile

## ■ Core profile

- contains information related to the user search goals and interests

## ■ Extended profile

- contains information related to the user as a person in order to understand or model the use that a person will make with the information retrieved



# Simple (vector) Core Profiles

- Primitive profile (any model)
  - A set of search terms (0-1 vector)
- For Boolean model of IR
  - A Boolean query
- For vector model of IR (dominated)
  - A set of terms with their weights (vector)
  - An *overlay* (set of weights) over a simple *domain model* that is just a list of terms that could be of interest to the users



# Advanced Core Profiles

- Domain model is a network of terms
  - A subset of the model (simple overlay)
  - A weighted *overlay* over DM
- Domain model is a hierarchy of *topics*
  - *typically, each topic is a vector of terms*
  - A subset of the model (simple overlay)
  - A weighted *overlay* over DM



# Group Profiles

- A system can maintain a group profile in parallel or instead of user profile
- Could resolve the privacy issue (navigation with group profile)
- Could be use for new group members at the beginning
- Could be used in addition to the user profile to add group “wisdom”





# Extended Profile

- Knowledge: about the system and the subject
- Goals: local and global
- Interests
- Background: profession, language, prospect, capabilities
- Preferences (types of docs, authors, sources...)



# Who Maintains the Profile?

- Profile is provided and maintained by the user/administrator
  - Sometimes the only choice
- The system constructs and updates the profile (automatic personalization)
- Collaborative - user and system
  - User creates, system maintains
  - User can influence and edit
  - Does it help or not?



# Adaptive Search

## ■ Goals:

- Present documents (pages) that are most suitable for the individual user

## ■ Methods:

- Employ user profiles representing short-term and/or long-term interests (Korfhage)
- Rank and present search results taking both user query and user profile into account



# Personalized Search: Benefits

## ■ Resolving ambiguity

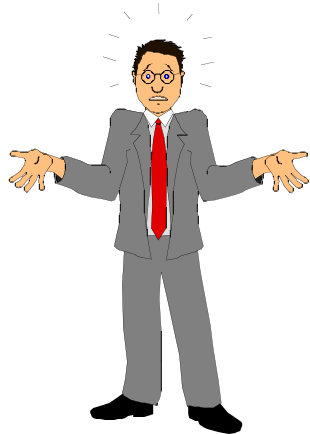
- The profile provides a context to the query in order to reduce ambiguity.
- Example: The profile of interests will allow to distinguish what the user asked about “Berkeley” (“Pirates”, “Jaguar”) really wants

## ■ Revealing hidden treasures

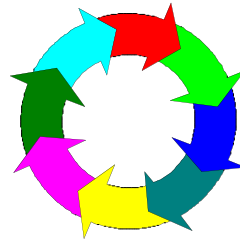
- The profile allows to bring to surface most relevant documents, which could be hidden beyond top results page
- Example: Owner of *iPhone* searches for *Google Android*. Pages referring to both would be most interesting

# The Components of Web Search

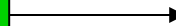
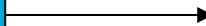
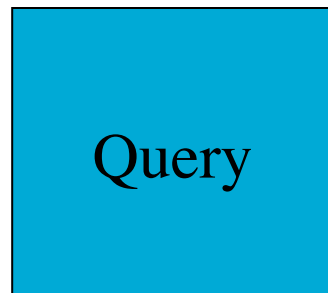
Formulates



Processes



Analyzes

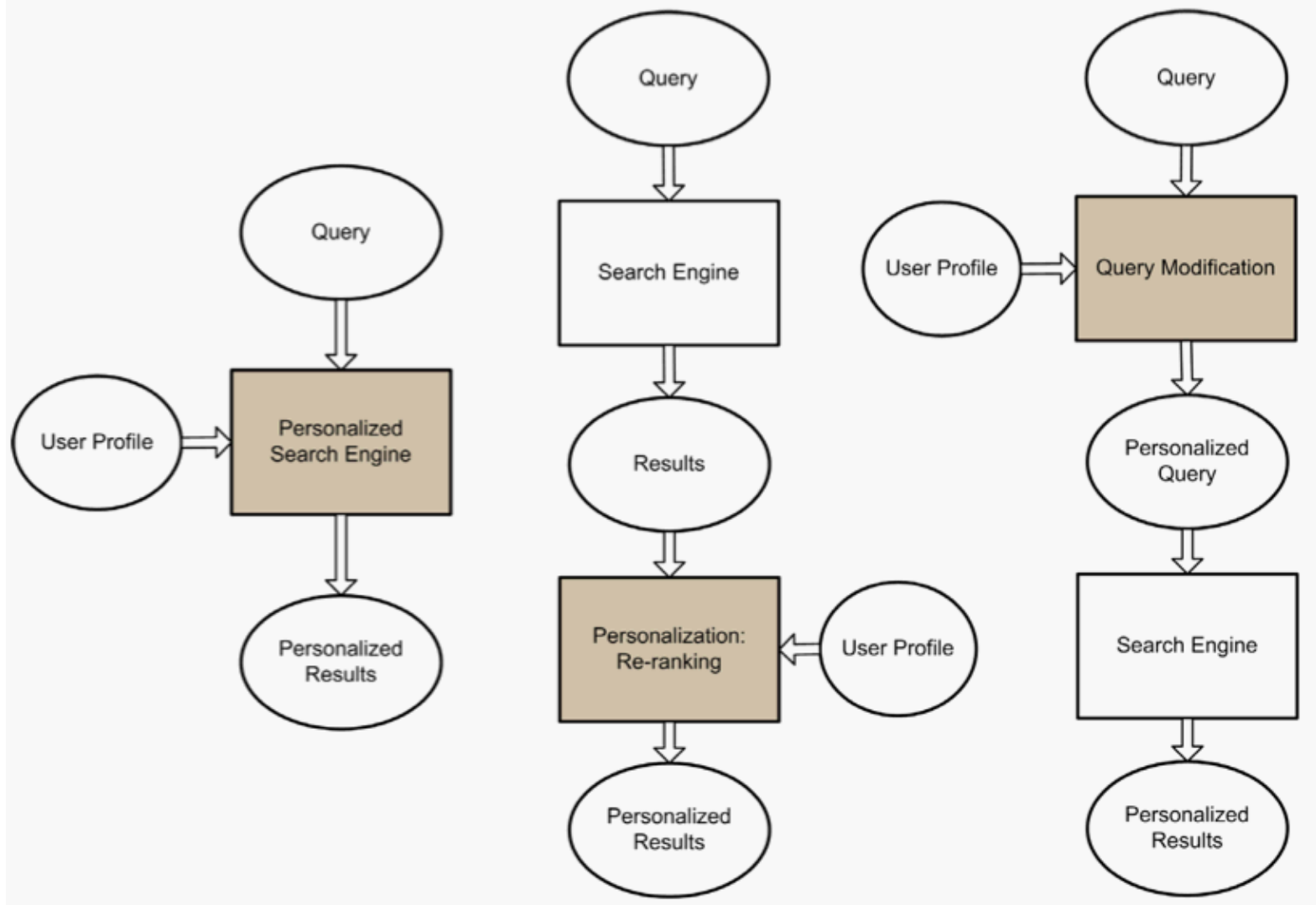




# Where to Apply Profiles ?

- The user profile can be applied in several ways:
  - To modify the query itself (pre-processing)
  - To change the usual way of retrieval
  - To process results of a query (post-processing)
  - To present document snippets
  - Special case: adaptation for meta-search

# Where to Apply Profiles?





# Examples of Systems

- Pre-process with QE - Koutrika, Mobasher, Chirita
- Pre-process with RF : SmartGuide
- Post-process with annotations: Syskill & Webert
- Post-process with re-ranking: Syskill & Webert, WIFS, YourNews, TaskSieve
- Adaptive Snippets: TaskSieve



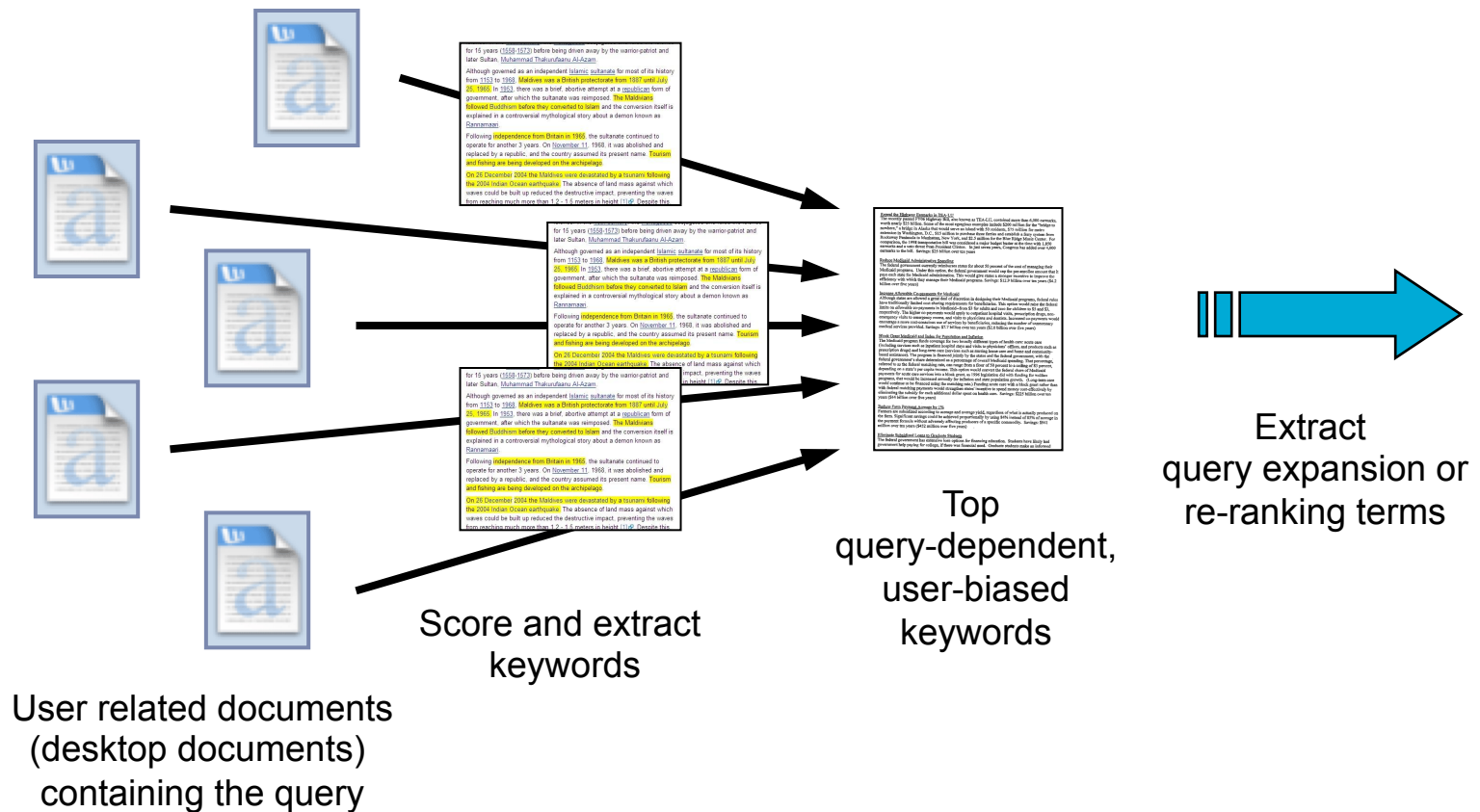


# Pre-Process: Query Expansion

- User profile is applied to add terms to the query
  - Popular terms could be added to introduce context
  - Similar terms could be added to resolve indexer-user mismatch
  - Related terms could be added to resolve ambiguity
  - Works with any IR model or search engine

# Simple Context-based Query Expansion

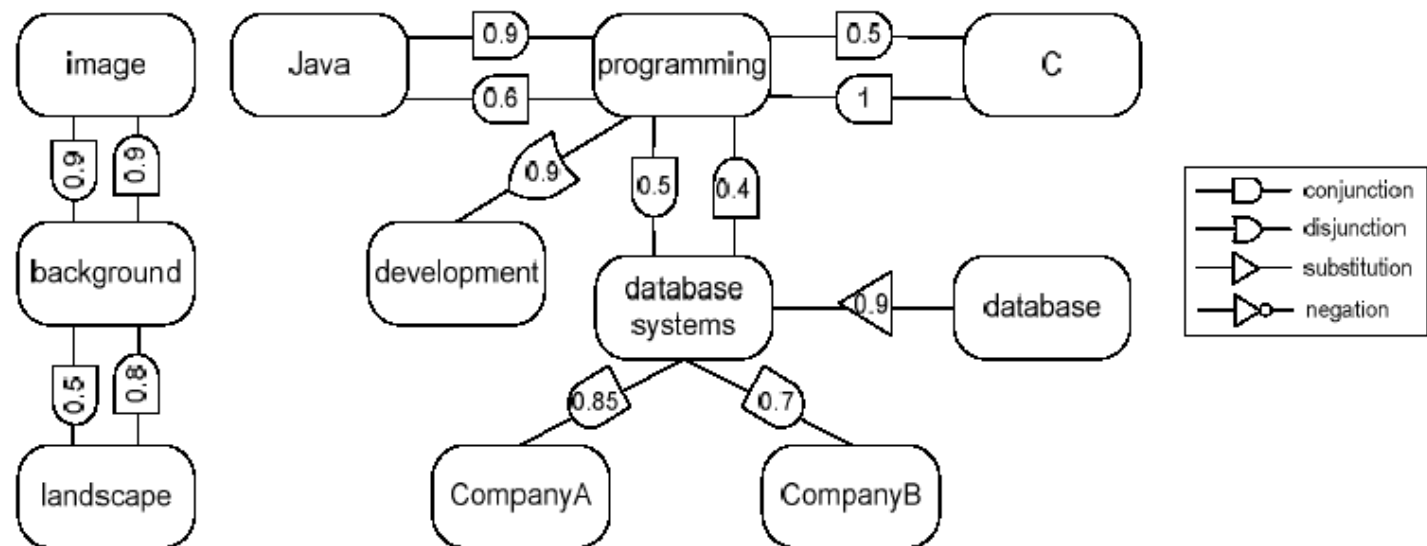
## Expansion: Chirita et al. 2006



[ Chirita, Firan, Nejd. *Summarizing local context to personalize global web search*. CIKM 2006 ]

# Advanced Example: Koutrika & Ioannidis' 05

- Advanced relevance network for query expansion
- java -> java and programming -> java and (programming or development)





# Pre-Process: Relevance Feedback

- In this case the profile is used to “move” the query vector (vector model only)
- Imagine that:
  - the documents,
  - the query
  - the user profileare represented by the same set of weighted index terms



# Pre-filter: Linear Transformation

- The query  $q = q_1, q_2, \dots, q_n$
- The profile  $p = p_1, p_2, \dots, p_n$
- The query modified by the user profile will be something like that:

$$\text{modified } q_i = Kp_i + (1-K)q_i \quad i=1, 2, \dots, n$$



# Pre-process: Linear Transformation

$$\text{modified } q_i = Kp_i + (1-K)q_i$$

- In this case we add the terms of the profile to the query ones, weighted by  $K$

for  $K=0$  *modified*  $q_i = q_i$  the query is unmodified

for  $K=1$  *modified*  $q_i = p_i$  the query is substituted by the profile



# Piecewise Linear Transformation

- if the term appears in the query **and** in the profile then the linear transformation is applied
- if the term appears in the query but not in the profile is left unmodified or diminished slightly
- if the term appears in the profile but not in the query it is not introduced, or introduced with a weight lower than in the profile.



# Example: SmartGuide

- Access to the CIS-like information
- User has a long-term interests profile and current queries
- Information is searched using a combination of both
- Profile is initiated from a stereotype and kept updated
- Increased user satisfaction, decreased navigation overhead

Gates, K. F., Lawhead, P. B., and Wilkins, D. E. (1998) Toward an adaptive WWW: a case study in customized hypermedia. *New Review of Multimedia and Hypermedia* 4, 89-113.





# Post-Processing

- The user profile is used to organize the results of the retrieval process
  - present to the user the most interesting documents
  - Filter out irrelevant documents
- Extended profile can be used effectively
- In this case the use of the profile adds an extra step to processing
- Similar to classic information filtering problem
- Typical way for adaptive Web IR

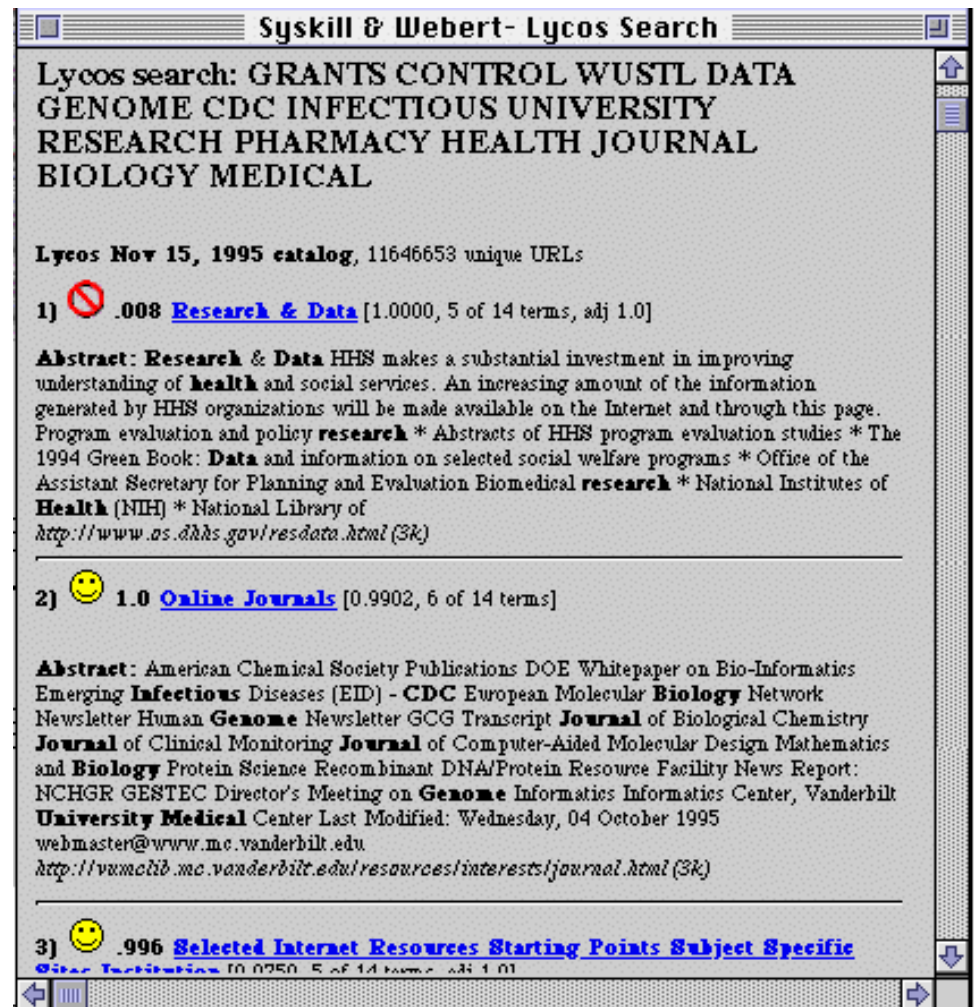


## Post-Filter: Annotations

- The result could be relevant to the user in several aspects. Fusing this relevance with query relevance is error prone and leads to a loss of data
- Results are ranked by the query relevance, but annotated with visual cues reflecting other kinds of relevance
  - User interests - Syskill and Webert, group interests - KnowledgeSea

# Example: Syskill and Webert

- First example of annotation
- Post-filter to Lycos
- Hot, cold, lukewarm
- Pazzani, M., Muramatsu, J., and Billsus, D. (1996) Syskill & Webert: Identifying interesting Web sites. In: Proceedings of the Thirteen National Conference on Artificial Intelligence, AAAI'96, Portland, OR, August, 1996, AAAI Press /MIT Press, pp. 54-61, also available at <http://www.ics.uci.edu/~pazzani/Syskill.html>.





## Post-Filter: Re-Ranking

- Re-ranking is a typical approach for post-filtering
- Each document is rated according to its relevance (similarity) to the user or group profile
- This rating is fused with the relevance rating returned by the search engine
- The results are ranked by fused rating
  - User model: WIFS, group model: I-Spy



## Example: WIFS (Micarelli)

- Adaptive post-filter to AltaVista search engine
- Maintains an advanced stereotype-based user model (Humos subsystem)
- User model is updated by watching the user
- The model is used to filter and re-order the links returned by AltaVista

# YourNews: Adaptive Search and Filtering with Open User Profile



The screenshot displays the 'YourNews' web application interface. At the top, there is a search bar and a navigation menu with categories: National, World, Business, Technology, Sports, Entertainment, Health, Computing, Palm, and YourTab 3. Below the navigation menu, there are links for 'Show all duplicate articles', 'Recent News', and 'Recommended News'. The main content area shows a list of news items, each with a thumbnail, title, and brief description. The sidebar on the right contains a 'Customize YourTab' section with a 'Manage your feeds' link, a list of 'peterb's Interests for Palm News', and a 'Short term' / 'Long term' toggle. Below the interests list, there is a section for 'PALM HANDHELD BIT COLLIGAN PDA TREO NBSP' with a list of related terms. At the bottom of the sidebar, there is a text input field for 'Add your custom keywords' and an 'OK' button, along with a link to 'Read this tab in RSS format'.

**YourNews** [Search]

National World Business Technology Sports Entertainment Health Computing **Palm** YourTab 3

☐ Show all duplicate articles » Recent News | **Recommended News**

**Treocntl** **Palms App Store Called App Catalog** (11 hours ago) ★★  
Palms new Developer Website reveals name of Palms app store...

**PIC1** **Palm Pre Overview and Impressions from CES** (2 days ago) ★★  
In the two days following Palms WebOS and Palm Pre announcement, Ryan and I have spent a great deal of time speaking with Palm representatives, watching demos, and talking at length about the new hardware and software platforms. In this article I take a close look at the Palm Pre hardware and answer many of the questions surrounding Palms new hotly anticipated smartphone. Read on for our in-depth Palm Pre impressions review and a ton of high res photos.

**BrightHand** **Palm Unveils the Pre Smartphone** (4 days ago) ★★  
Palm has announced its new Pre smartphone. A touchscreen device with a sliding keyboard, the Pre will use the just as new Palm webOS.

**PIC1** **Palm Stock Soars on Positive Pre Buzz** (3 days ago) ★★  
After a long down out decline and near dips below the dollar line, Palms stock has make a commanding recovery since the new announcements. After the conclusion of Palms presentation, late-day trading activity responded strongly in Palms favor, sending the stock up 35%. The market seems to be responding positively to the new Pre smartphone and WebOS platform. In trading today price is continuing its climb and is currently trading around \$6. This is a pleasant turn of events from Palms dismal ...

**PIC1** **Palm Pre Hands on Videos Part 2** (3 days ago) ★★  
Next in our series of video updates from CES, we've put together three new hands on clips of the Palm Pre. In the first clip we have a demonstration of the Palm webOS Copy and Paste ability. Following that we have a ~2 minute clip of the Pres web browser visiting everyones favorite Palm site like on the device. You can see the Pre under the standard version PIC and scroll around

**Customize YourTab**  
Manage your feeds

peterb's Interests for Palm News [Hide]

Short term Long term

**PALM HANDHELD BIT COLLIGAN PDA TREO NBSP**  
DEVELOP EOL LEOPARD INTEL POSTERITY CUFF  
ASTRAWARE INDICATE ANNOUNCEMENT SNOW LONG  
PRACTICALLY SALE RESTART NEW APP CONTINUE  
FOUNDATIONS DIVIDE DISPLAY STORE VERSION  
SUFFICIENT INEVITABLE FACEBOOK POTENTIAL  
ENTERPRISE SCREENSHOT MODE PREFERENCE ED  
BLACKBERRY VEGA

Add your custom keywords [OK]

☐ Read this tab in RSS format

<http://amber.exp.sis.pitt.edu/yournews/>

# TaskSieve: User-Controlled Adaptive Search

The screenshot displays the TaskSieve web application interface. At the top, the title "TaskSieve" is shown next to a search bar containing the query "austria train fire" and a "Search" button. The user is logged in as "User 'jahn'". Below the search bar, there are tabs for "TaskSieve Filtering" and "Sort by relevance to Query Terms", "Sort by relevance to Query and Task model", and "Sort by relevance to Task model". The main content area shows "Matching 1000 records" and a list of search results. The first result is titled "AFA20001111.1800.0160" with a rating of 5 stars and a score of (0.89). The text of the result describes a fire in an electric train elevator in Austria. The second result is titled "ZBN20001113.0400.0019" with a rating of 4 stars and a score of (0.80). The text describes a fire in a tunnel in Austria. The third result is titled "ALH20001112.0700.0037" with a rating of 5 stars and a score of (0.74). The text describes a fire in a train in Austria. The fourth result is titled "AFA20001111.1400.0094" with a rating of 5 stars and a score of (0.71). The text describes a fire in a train in Austria. On the right side of the interface, there is a "Task Model Terms" section with a list of terms: TRAIN, FIRE, SKI, KILL, ELECTRIC, ERUPT, PEOPLE, PERSON, PULL, RESULT, ANNOUNCE, SALZBURG, BODY, SEASON, BROKE, ELEVATOR, TRANSPORT, TUNNEL, VICTIM, FRANZ, YESTERDAY, GOVERNOR, YOUNG, MOUNTAIN, OPEN. Below this is a "Task Model Notes" section with a list of notes: 1. a transport train pulling a fire broke out of skiing, ski resulting in at least 172 persons were killed[x]; 2. ecovered the bodies of more than 40 victims of Saturday's train fire in a mountain tunnel[x]; 3. The governor of Salzburg Franz announced that at least 172 people, mostly young people, were killed in a fire erupted yesterday in an electric train[x]; 4. with about 170 persons were killed in a fire erupted in the electric train elevator at the opening of the season skiing[x]. At the bottom of the right sidebar, there are buttons for "Add Selection to Notes" and "Print Notes".

TaskSieve

User "jahn" logged in | Logout | **HELP**

TaskSieve Filtering

Sort by relevance to Query Terms | Sort by relevance to Query and Task model | Sort by relevance to Task model

Matching 1000 records | Pages 1 2 3 4 5 6 7 8 9 10 Next 100»

☐ Query terms ☐ Task model terms (from notes) ☐ Both terms |  ☐ Auto Refresh

1 **AFA20001111.1800.0160** ★★★★★ (0.89)  
Today( Saturday), met with about 170 persons were killed in a fire erupted in the electric train elevator at the opening of the season skiing in Austria in the incident was the worst disasters that have occurred in the Alps. ... It was announced in a statement reported by the governor of Salzburg Franz Austrian news agency that at least 170 persons, mostly young people were killed in a fire erupted on Saturday in an electric train rise() within an expenditure of Austrian Alps. ... He said the governor of Salzburg to rescue the men were able to lift it came only three bodies on fire on the train as a whole.

2 **ZBN20001113.0400.0019** ★★★★ (0.80)  
Salzburg, Austria, Burg introduce the state governor said yesterday that alpine tunnel fire a total of 165 train passengers, of which 12 people escape, and the rest were all killed. ... Death toll was more than 154 young ski lovers who he said that as a driver were killed in fire, bringing the total number of deaths to 154 people. ... Tunnel fire extinguished, Austria rescue personnel entered the tunnel began yesterday in the burning of the train to find the bodies of victims.

3 **ALH20001112.0700.0037** ★★★★★ (0.74)  
Vienna - AFP - The governor of Salzburg Franz announced that at least 172 people, mostly young people, were killed in a fire erupted yesterday in an electric train the rise in one of the Alps Austrian expenditure has not been spared, only six passengers. ... He added during a press conference at the centre of skiing that eight people only managed to escape from a window, adding that the disaster have led to the killing of " At least 172 people, mostly young people who were skiing with a view to the practice of sport ". ... He said that the lifting of the three men were able to rescue the bodies of the fire came on the train as a whole.

4 **AFA20001111.1400.0094** ★★★★★ (0.71)  
The governor of Salzburg Franz announced in a statement reported by Austrian news agency that at least 170 persons, mostly young people were killed in a fire erupted on Saturday in an electric train rise() within an Austrian alpine expenditure has not been spared, only eight passengers. ... He added during a press

Task Model Terms[Hide]

TRAIN FIRE SKI KILL ELECTRIC  
ERUPT PEOPLE PERSON PULL  
RESULT ANNOUNCE SALZBURG BODY  
SEASON BROKE ELEVATOR TRANSPORT  
TUNNEL VICTIM FRANZ YESTERDAY  
GOVERNOR YOUNG MOUNTAIN OPEN

Task Model Notes

1. a transport train pulling a fire broke out of skiing, ski resulting in at least 172 persons were killed[x]
2. ecovered the bodies of more than 40 victims of Saturday's train fire in a mountain tunnel[x]
3. The governor of Salzburg Franz announced that at least 172 people, mostly young people, were killed in a fire erupted yesterday in an electric train[x]
4. with about 170 persons were killed in a fire erupted in the electric train elevator at the opening of the season skiing[x]

완료 노트 열기

<http://amber.exp.sis.pitt.edu/yournews-experiment/>





# TaskSieve: Adaptive Snippets

- The goal of a usual snippet is to show query relevance. TaskSieve applies adaptive snippets to show profile relevance as well
- Selects top 3 relevant sentences combining query relevance and task relevance to sentences
- Applies color coding by query/profile

2 ZBN20001113.0400.0019 ★★★★★ (0.80)  
Salzburg, Austria, Burg introduce the state governor  
passengers, of which 12 people escape, and the rest  
lovers who he said that as a driver were killed in fire,  
... Tunnel fire extinguished, Austria rescue personnel  
the train to find the bodies of victims.





# Knowledge Sea Adaptive Search: Three Reference Points

- Adaptive search in TaskSieve uses linear combinations of two ranks - query-based and profile-based to calculate the final rank
- Query and profile are two reference points for ranking
- What if there are three reference points?
- Knowledge Sea Search allows to do user-adaptive ranking with three reference points: Query, task profile, and lecture

# Knowledge Sea Adaptive Search: Three Reference Points

## KnowledgeSea Search

Query

SEARCH

Stemmed query: *speak user languag* | 1464 of 1545 documents retrieved | Search time: 0.09 seconds  
Removed common words: *the*

Result pages: 1 2 3 4 5 6 7 8 9 10 »Next 100

Rank	Source	Title	Score	State
1	<a href="#">Shneiderman</a>	8.6.1 Natural-language interaction »	0.004	
2	<a href="#">Shneiderman</a>	8.1 Introduction »	0.003	
3	<a href="#">Dix</a>	10.4.1 Designing for users with disabilities »	0.003	
4	<a href="#">Shneiderman</a>	7.8.1 Audio menus »	0.003	
5	<a href="#">Dix</a>	4.2.8 Language versus action »	0.003	
6	<a href="#">Lewis</a>	7.1 Manuals »	0.003	
7	<a href="#">Lewis</a>	7.1 Manuals »	0.003	
8	<a href="#">Newman</a>	15.4.1 General principles »	0.003	
9	<a href="#">Preece</a>	1.6.1 Heuristics and usability principles »	0.003	
10	<a href="#">Dix</a>	10.3.1 Sound in the interface »	0.003	
11	<a href="#">Dix</a>	9.3.2 Heuristic evaluation »	0.003	
12	<a href="#">Dix</a>	4.2.2 Video display units »	0.003	
13	<a href="#">Shneiderman</a>	9.4.1 Discrete-word recognition »	0.003	
14	<a href="#">Shneiderman</a>	9.4.2 Continuous-speech recognition »	0.003	
15	<a href="#">Preece</a>	10.2.2 Why you need to evaluate »	0.003	
16	<a href="#">Lewis</a>	6.4 Windows, the Shared-Code Approach, and Visual Basic »	0.003	
17	<a href="#">Lewis</a>	6.4 Windows, the Shared-Code Approach, and Visual Basic »	0.003	
18	<a href="#">Newman</a>	8.6.4 The method of evaluation »	0.003	

### Task Model

heuristic evaluation

Delete

New

#### Task Description:

Nielsen heuristics golden rules of design usability engineering speak the user language

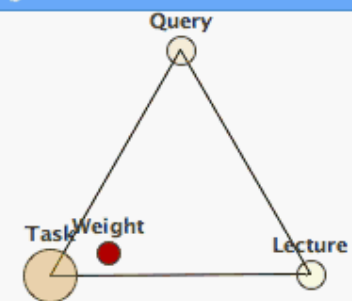
... applic code dde \_ dll ... user heuristic  
interappl language program rules speak uim  
usability user vendor visual window

Reset task model

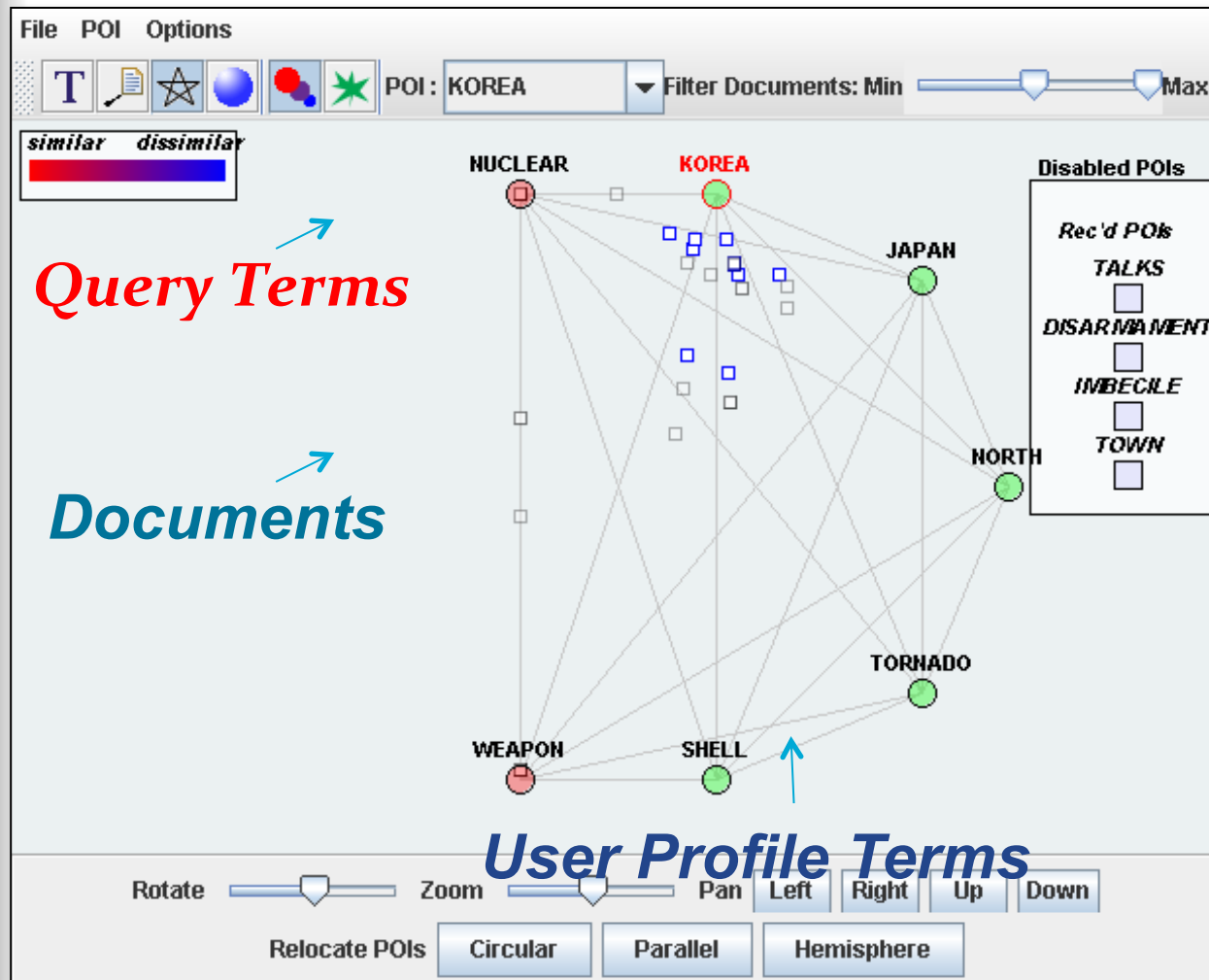
### Lecture Model

Lecture 6 - Heuristic Analysis

### Weights



# Adaptive VIBE: Moving to 2D



Presents  
(1) User profile,  
(2) Query,  
(3) Documents

Allows to  
merge query  
relevance with  
user profile  
relevance

Reveals  
relevance to  
individual  
terms



# Some Ideas for Profile-Adapted Retrieval (Korfhage)

- Query and Profile are considered as *Separate Reference Points*
- In this case documents are retrieved if they are “near” the query or the profile.
- For the following slides, let’s assume that the similarity is measured by distance

where  $D$  is the document  $\|D, Q\|$  and  $Q$  is the query



# Separate Reference Points

- We have different way to integrate query and profile as separate reference points:
  - Disjunctive model of query-profile integration
  - Conjunctive model of query-profile integration
  - Ellipsoidal model
  - Cassini oval model



# Disjunctive Model

- We will take the document if the following condition is satisfied:

$$\min(\|D, Q\|, \|D, P\|) < d$$

- The D document should be “near” the query Q or the profile P



# Conjunctive Model

- We will take the document if the following condition is satisfied:

$$\max(\|D, Q\|, \|D, P\|) < d$$

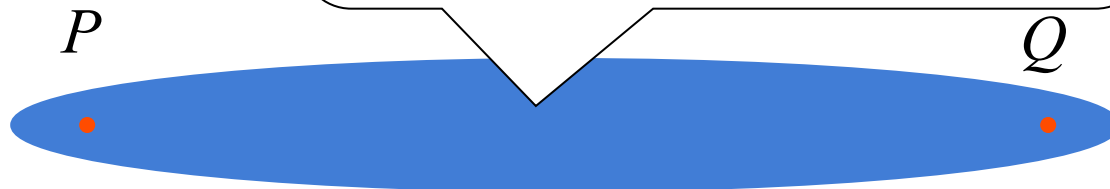
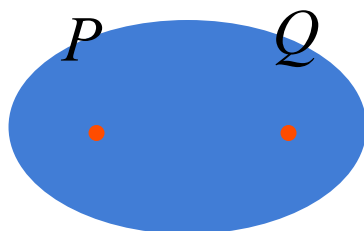
- The D document should be “near” the query *Q* and the profile *P*
- In this case if the profile and the query have little in common very few documents are retrieved

# Ellipsoidal Model

- Condition to satisfy

$$\|D, Q\| + \|D, P\| < d$$

this is the equation of an ellipse.



If the profile and query are  
“far apart” a lot of documents  
not relevant are retrieved



# Cassini Model

- Condition to satisfy

$$\|D, Q\| \times \|D, P\| < d$$

