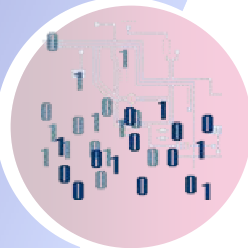
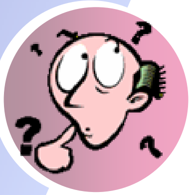


DELOS Summer School

Pisa 2004

Personalization:

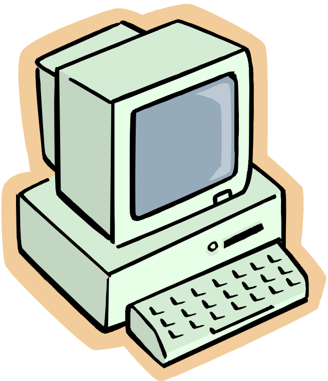
Models and Methods



Y. Ioannidis, G. Koutrika

University of Athens, Greece

Information Access



Find information about java?

Find latest movies?

Find new restaurants?

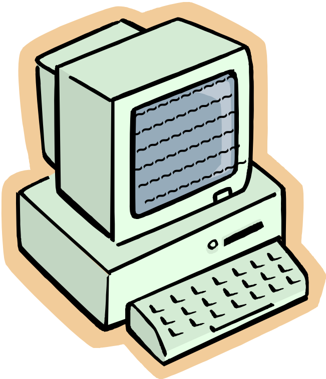
Find publications on AI?

…

Find something I would be interested in?

Information Access

Problems



A user may have to:

- reformulate queries issued several times
- encounter long or empty lists of results
- repeat tedious search tasks for new results
- learn search tricks

Information Access

The truth

Information overload haunts user searches!

It is difficult to find what you are searching for…
It is difficult to keep up with it…

Information Access

A solution?

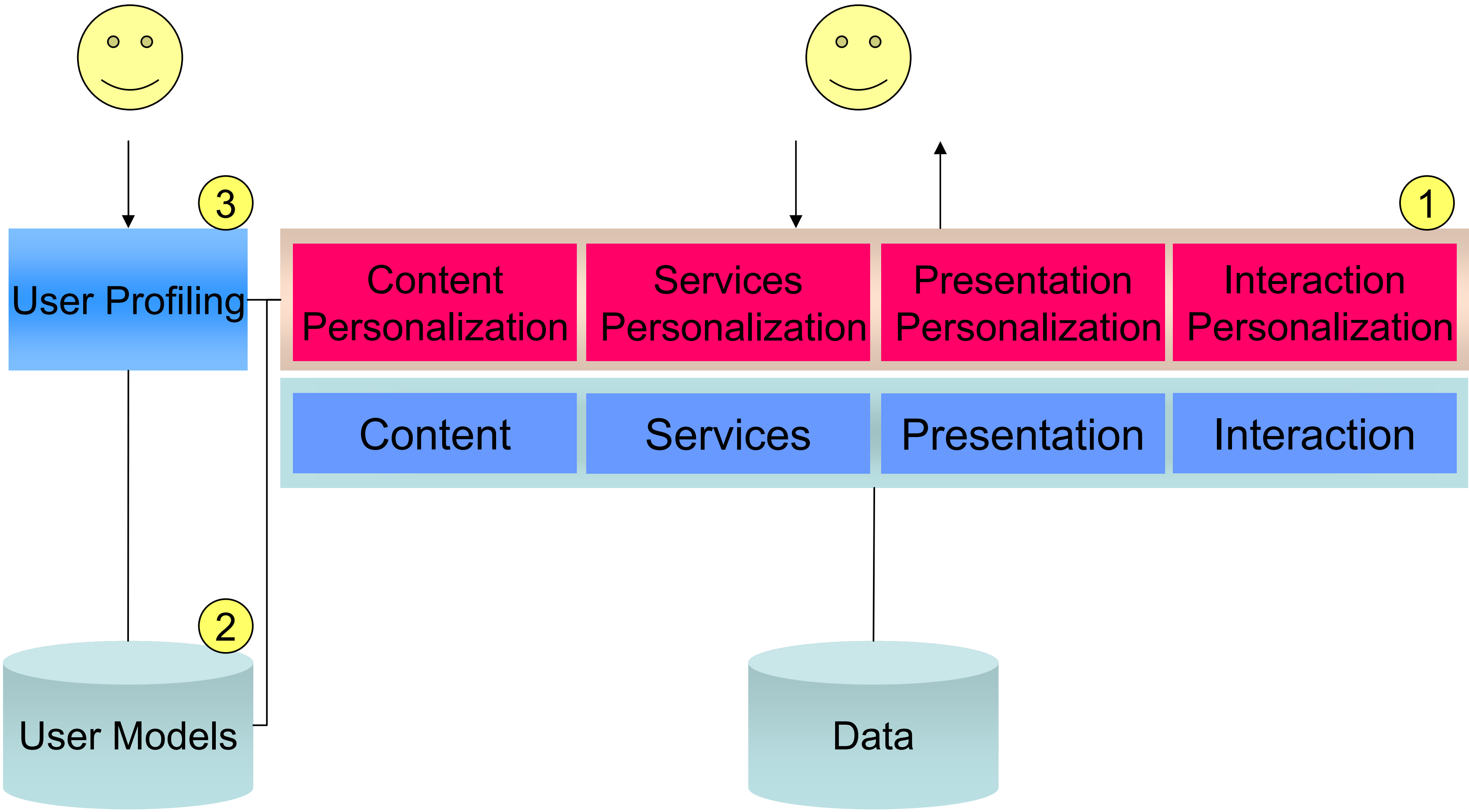
maybe !

- Shift towards a more user-centred information access paradigm

Personalization

Providing an overall customized, individualized user experience by taking into account the needs, preferences and characteristics of a user or group of users.

Personalization



Personalization Methods

- Information Filtering
- Continuous Queries
- Recommenders
- Personalized Search

Personalization Methods

- Information Filtering
- Continuous Queries
- Recommenders
- Personalized Search

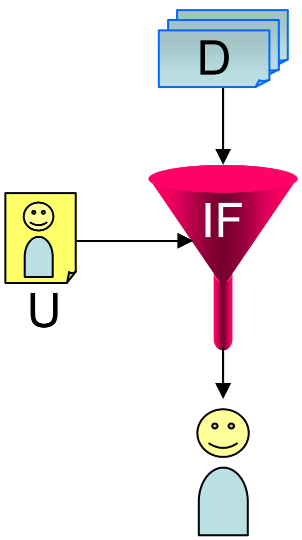
Personalization Methods

Information Filtering

Basic Idea

(slowly changing) long-term interests

(streams of) unstructured or semi-structured data:
textual information, images, video

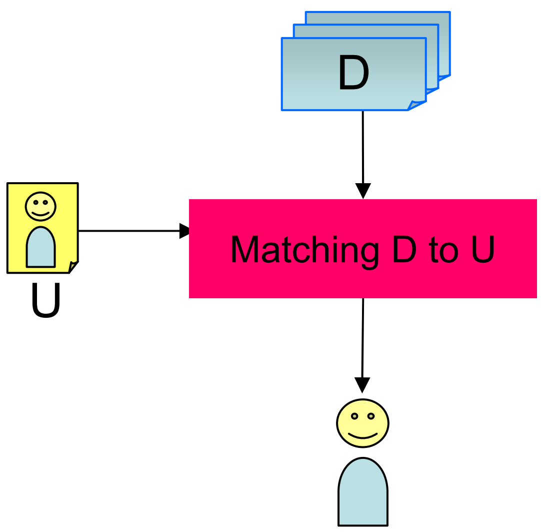


Personalization Methods

Information Filtering

System Model

Matching a user profile towards the representations of items of a collection resulting in the selection of items which are likely to be of interest to a user



1

Personalization Methods

Information Filtering

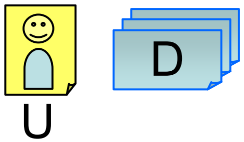
Matching Functions

- Exact-Match
  - Boolean
- Best-Match
  - Vector-space
  - Probabilistic

Personalization Methods

Information Filtering

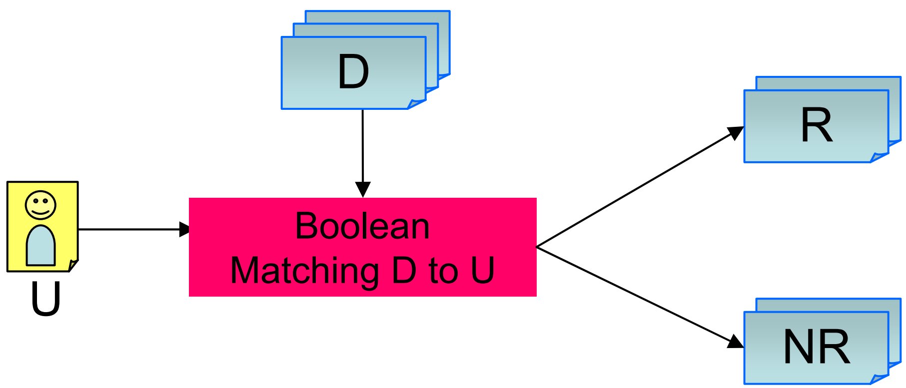
Matching Functions: Exact-Match



Sets of keywords

Boolean Matching D to U

All documents containing U are retrieved
No distinction between them



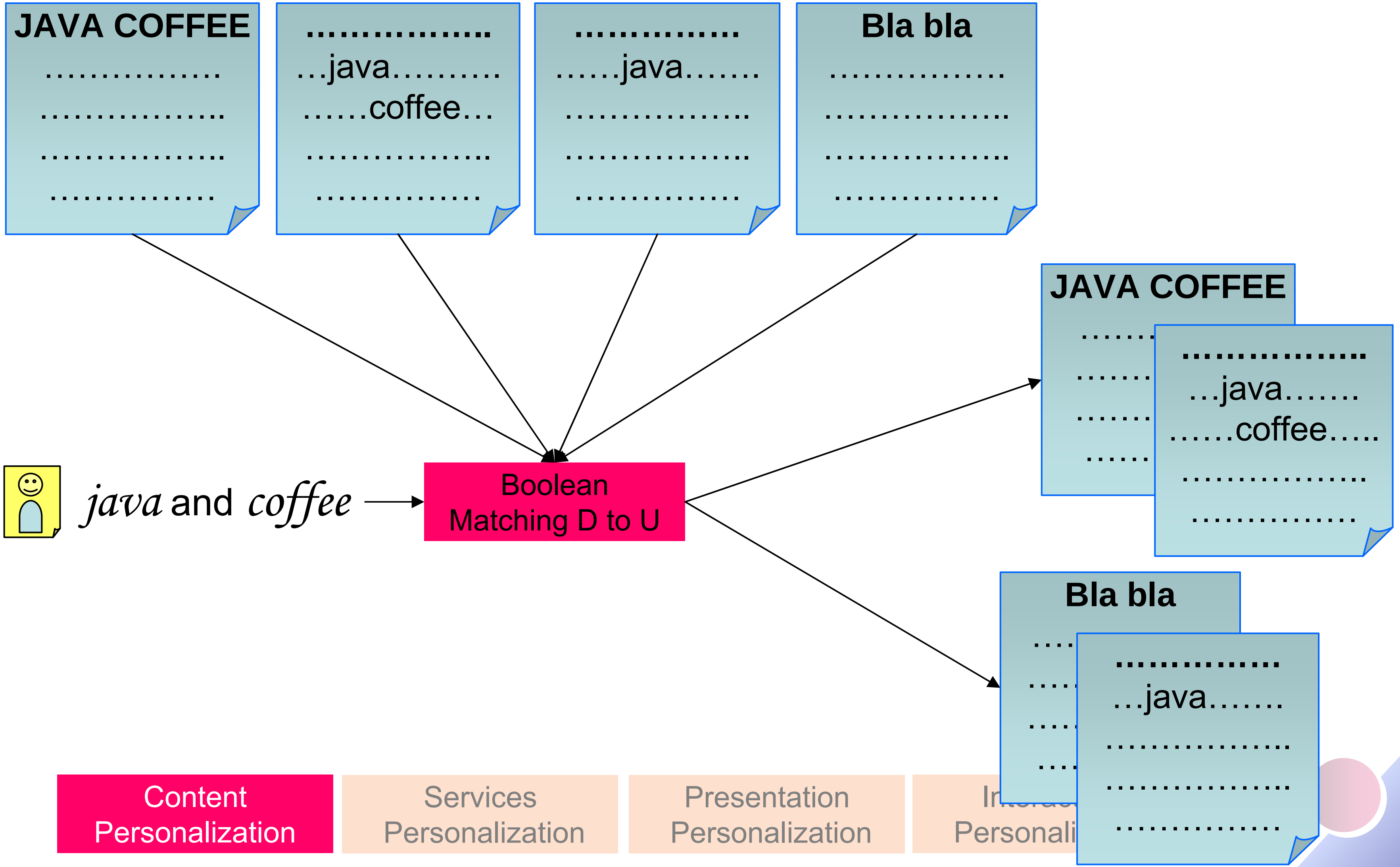
Set of retrieved documents

Set of not-retrieved documents

Personalization Methods

Information Filtering

Matching Functions: Exact-Match



Matching Functions: Exact-Match

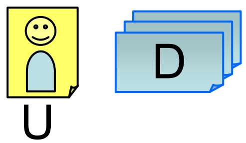
- Some documents are more relevant to a need than others
- Excluding documents that do not precisely match the profile results in lower effectiveness

1

Personalization Methods

Information Filtering

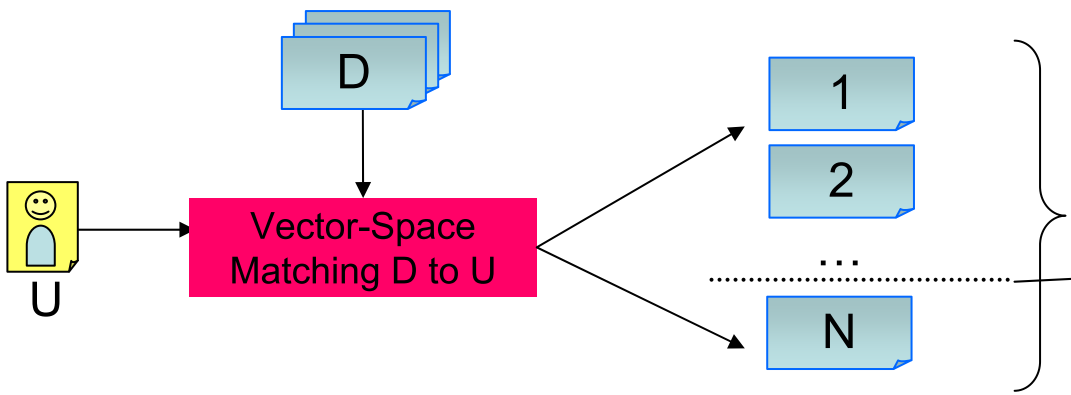
Matching Functions: Best-Match



Vectors of keywords

Vector-Space Matching D to U

Computation of similarity between U and D
Use of threshold



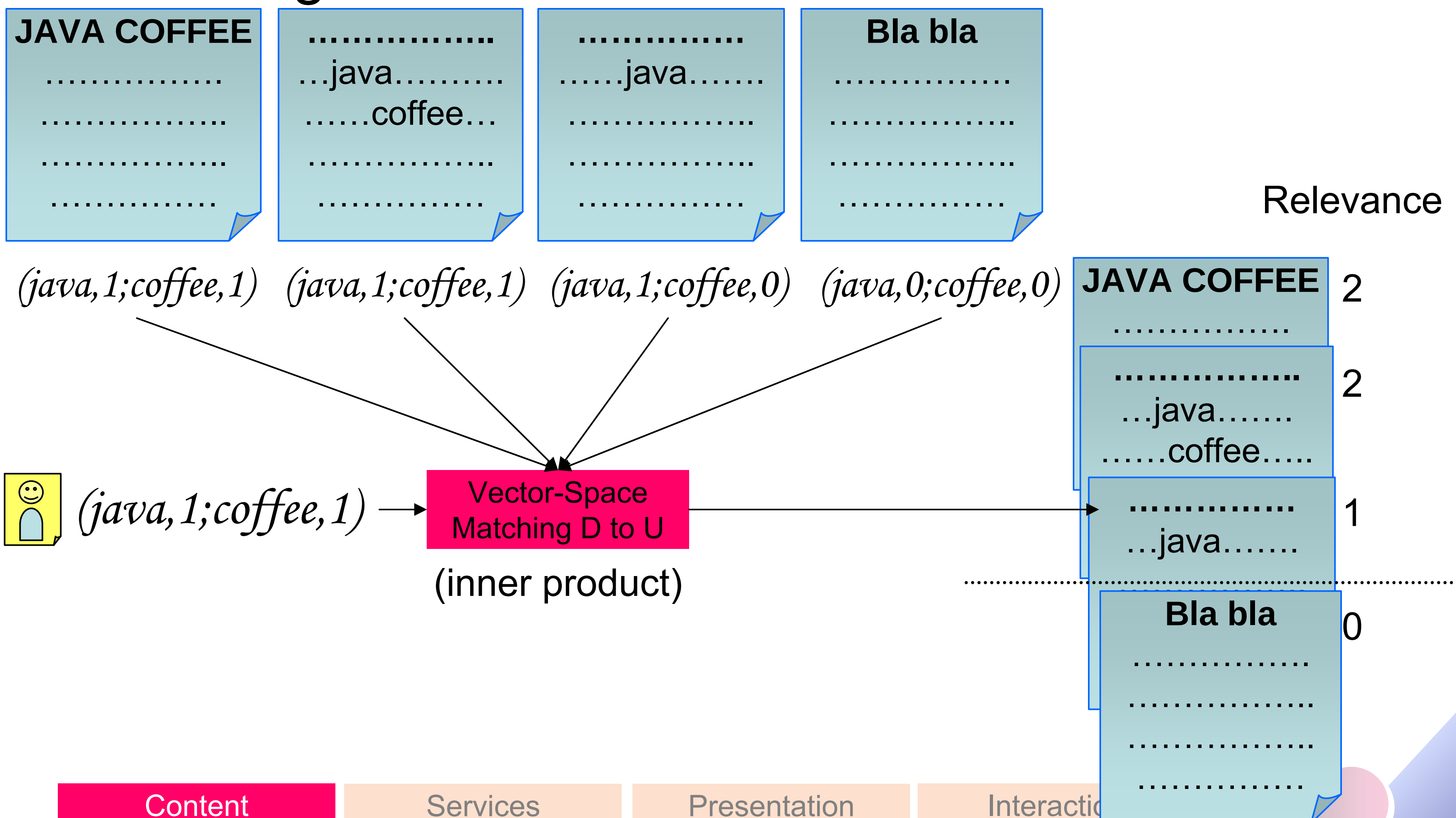
Ranked list
order of relevance

Dissemination Threshold

Personalization Methods

- Information Filtering

Matching Functions: Best-Match

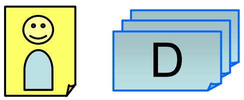


Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

Personalization Methods

Information Filtering

Matching Functions: Best-Match



Use of tf*idf weights

tf (term frequency) : term frequency in a document

idf (inverse document frequency) : term frequency in the universe of documents

Extensions: Latent Semantic Indexing (LSI)

Assumption: there is an underlying “latent” structure in the pattern of word usage across documents that can be exploited

- Result: Reduced dimensional space

Personalization Methods

Comparison to Information Retrieval

IR : *collection and organization* of texts,
IF : *distribution* of texts to groups or individuals.

IR : selection of texts from a relatively *static database*,
IF : selection or elimination of texts from a *dynamic datastream*.

IR : responding to the user's interaction with texts
within a *single information-seeking episode*,
IF : long-term changes over *a series of information-seeking episodes*

Personalization Methods

Adaptive IF systems

Learning:

- Profiles
- Corpus statistics (idf)
- Dissemination thresholds

Personalization Methods

Information Filtering

Systems

E-mail	Sift-Mail, ProcMail
News	SIFT, NewsWeeder
Documents	SIFTER, InRoute
Music	Personal DJ

Personalization Methods

Information Filtering

Filter Delivery Patterns

- Continuous
- Synchronous
- Asynchronous

Personalization Methods

Information Filtering

Information Lifetime

- Minutes: Stock market
- Days: News, Events, Mail
- Decades: Technology Reports
- Centuries: Entertainment

Personalization Methods

- Information Filtering
- Continuous Queries
- Recommenders
- Personalized Search

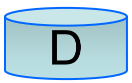
1

Personalization Methods

Continuous Queries

Basic Idea

(slowly changing) long-term interests expressed as queries



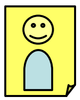
(streams of) structured data

Repeated execution of queries over the entire database is inefficient !

Personalization Methods

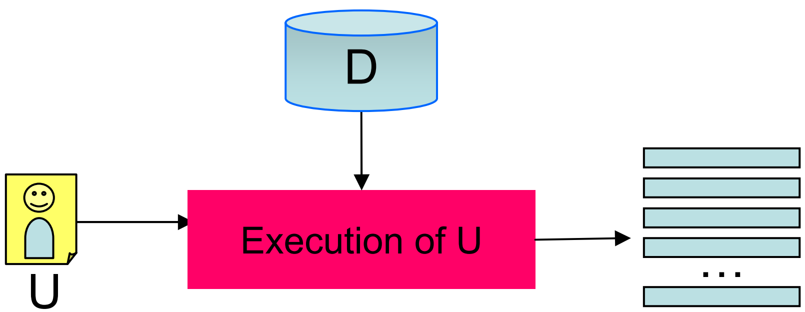
Continuous Queries

System Model



Persistent Query

Execution of U U is executed over the new part of D



New Results

Personalization Methods

Continuous Queries

Query Types

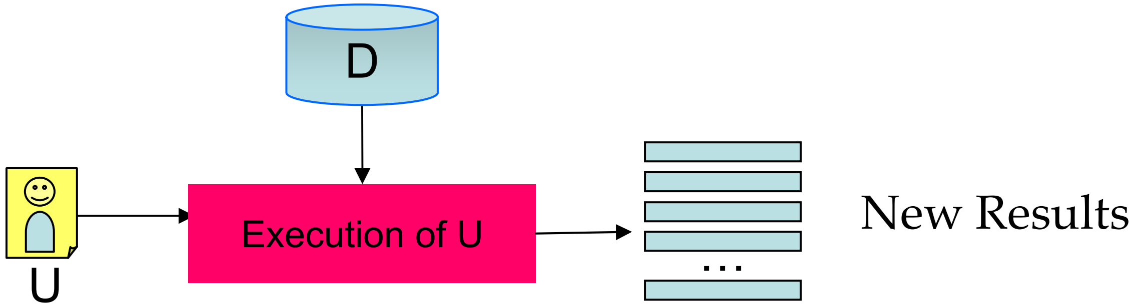
- Change-based — *Whenever the price of MM stock drops by more than 5%*
- Timer-based — *Every Monday*

Personalization Methods

Continuous Queries

Techniques

- Group Optimization
- Adaptive Query Processing
- Online data structures



Systems

- Tapestry
- OpenCQ
- NiagaraCQ
- TelegraphCQ
- CQL
- Oracle
- AdaptiveCQ

Personalization Methods

Applications

- Financial tickers
- Network monitoring and traffic management
- Web tracking
- Sensor applications
- Call detail records in telecommunications

Personalization Methods

Comparison to Triggers

CQ : consist of *millions* of continuous queries,
TR : consist of *limited* number of triggers.

CQ : monitor autonomous and heterogeneous *Internet sources*,
TR : monitor *local* databases.

CQ : support *change-based* and *timer-based* events,
TR : support *change-based* events.

1

Personalization Methods

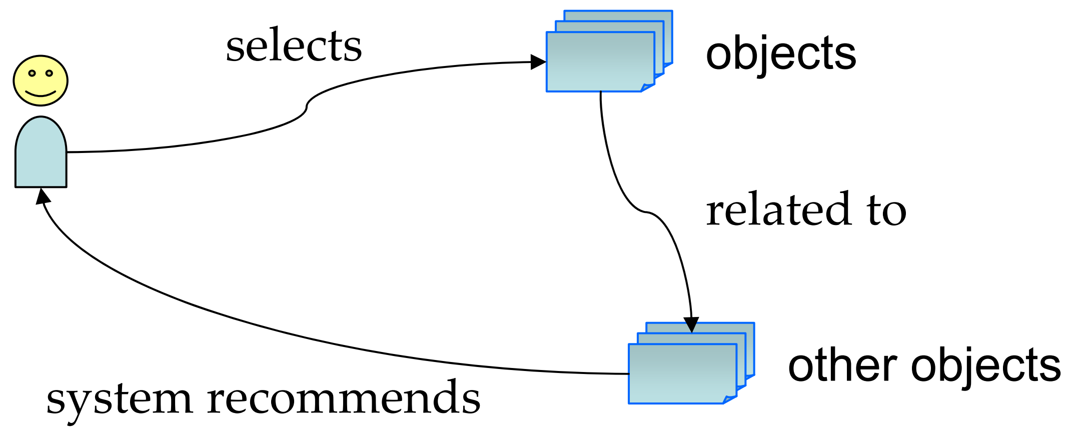
- Information Filtering
- Continuous Queries
- Recommenders
- Personalized Search

Content Personalization	Services Personalization	Presentation Personalization	Interaction Personalization

Personalization Methods

Recommenders

Basic Idea



Personalization Methods

Recommenders

Basic Idea

A recommender system is any system that provides a recommendation, prediction, opinion, list of items that assist a user in evaluating items.

(Schafer, Konstan, Riedl, CIKM 2002)

1

Personalization Methods

Recommenders

Types of Recommenders

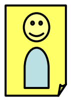
- User - Items
  - Content-based
  - Knowledge-based
  - Utility-based
- User - Users
  - Collaborative Filtering
  - Demographic
  - Community-based
- Hybrid

Personalization Methods

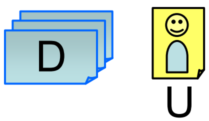
Recommenders

Types of Recommenders: Content-based

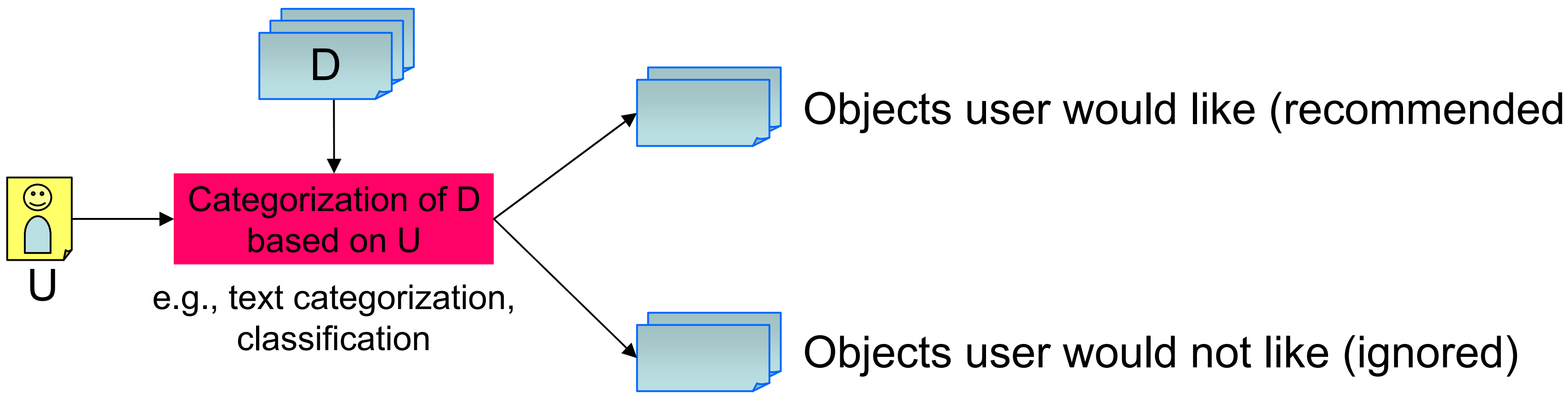
Find me things like those I have liked in the past



Preferences based on things I have liked in the past



Representations as in Information Filtering



Personalization Methods

Recommenders

Types of Recommenders: Knowledge-based

Functional models

Recommendations are decided based on quantitative decision support tools or case-based reasoning

Personalization Methods

Recommenders

Types of Recommenders: Utility-based

Constraints on objects' features

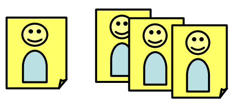
Recommendations are decided by building a utility function for each user across all features of the objects under consideration

Personalization Methods

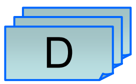
Recommenders

Types of Recommenders: Collaborative Filtering

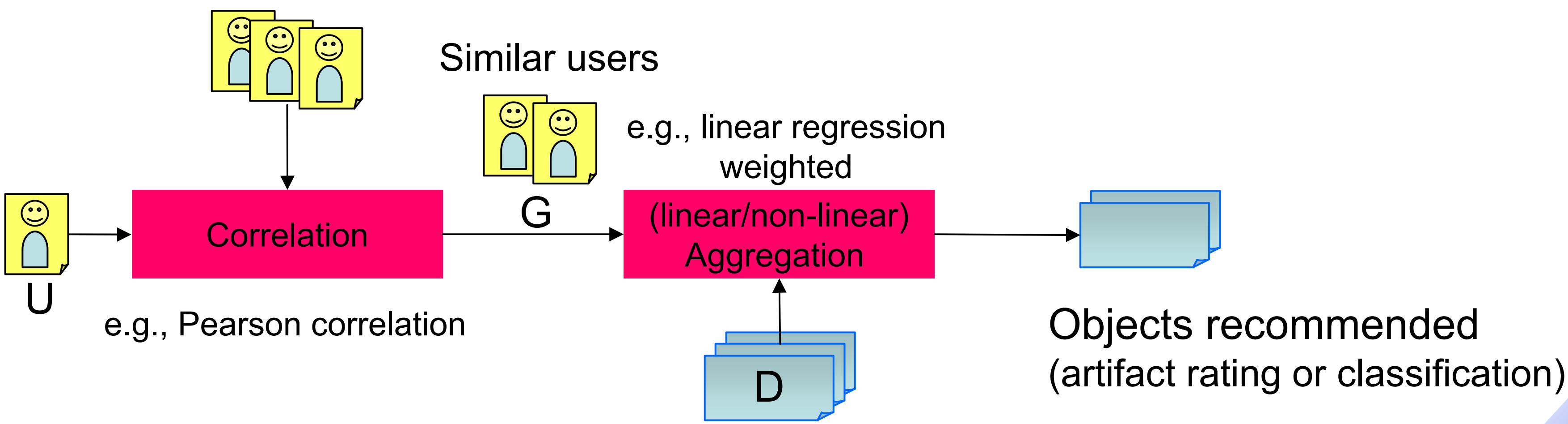
An attempt to facilitate “word of mouth”:
Find (predict) objects like those similar people have liked



Ratings of objects seen in the past by the user



Ids of objects



Personalization Methods

Types of Recommenders: Demographic

It is based on the user's personal attributes and demographic class

Personalization Methods

Types of Recommenders: Community-based

Find and exploit communities of people with same characteristics

Personalization Methods

Recommenders

Comparison

User - Items

- − Require sources of content information
- − Overspecialization
- \+ Do not depend on other users

User - Users

- \+ Any kind of content
- \+ Serendipity
- − Cold-start problems
- − Grey-sheep
- − Sparsity

▶ Solution: Hybrid Systems

Personalization Methods

Recommenders

Systems

Content-based	NewsWeeder, Libra, NewT, Amalthea
Knowledge-based	Entrée, Wasabi
Utility-based	Tete-Tete
Collaborative Filtering	GroupLens, Ringo, Phoaks
Demographic	LifestyleFinder
Community-based	Referral Web, QuickStep
Hybrid	Fab, ProfBuilder, SmartPad, FilterBot

Meta-Recommenders

A meta-recommender system is a system that presents unified and more meaningful recommendations fused from "recommendation data" from multiple information sources *(Schafer, Konstan, Riedl, CIKM 2002)*

E.g. MetaLens

Personalization Methods

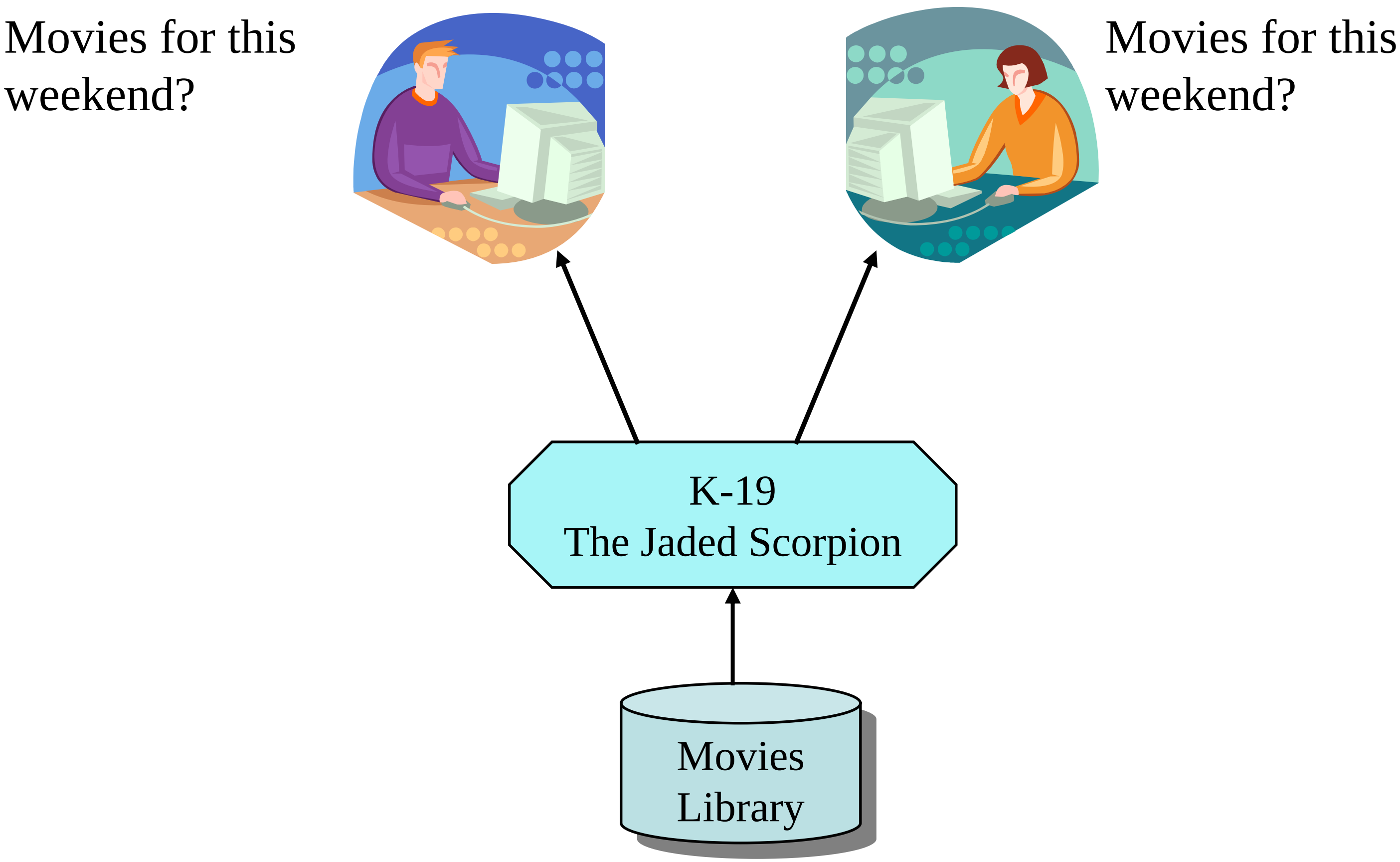
- Information Filtering
- Continuous Queries
- Recommenders
- Personalized Search

Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

Personalization Methods

Personalized Search

Basic Idea

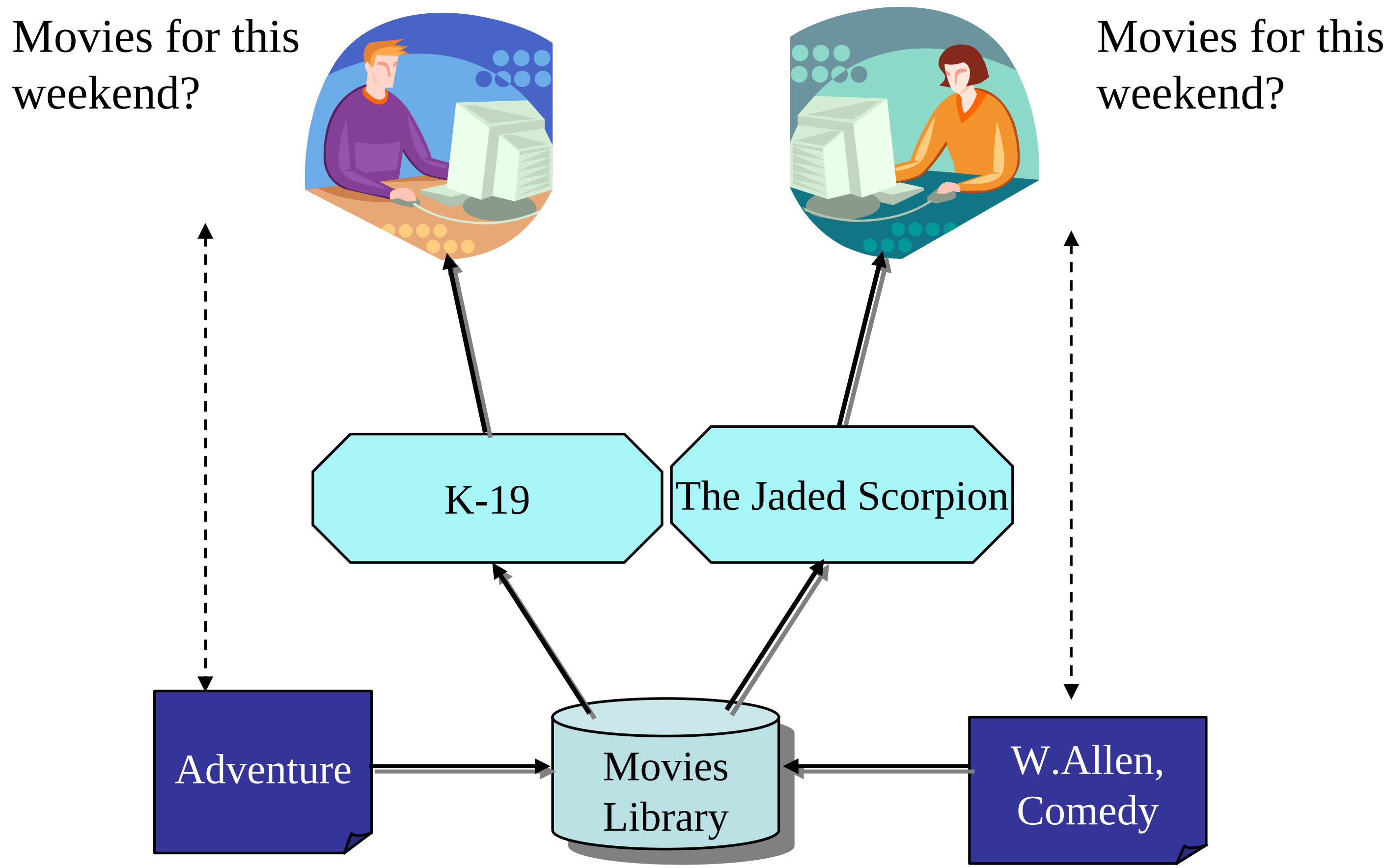


Personalization Methods

Personalized Search

Basic Idea

Different people find different things relevant/interesting



Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

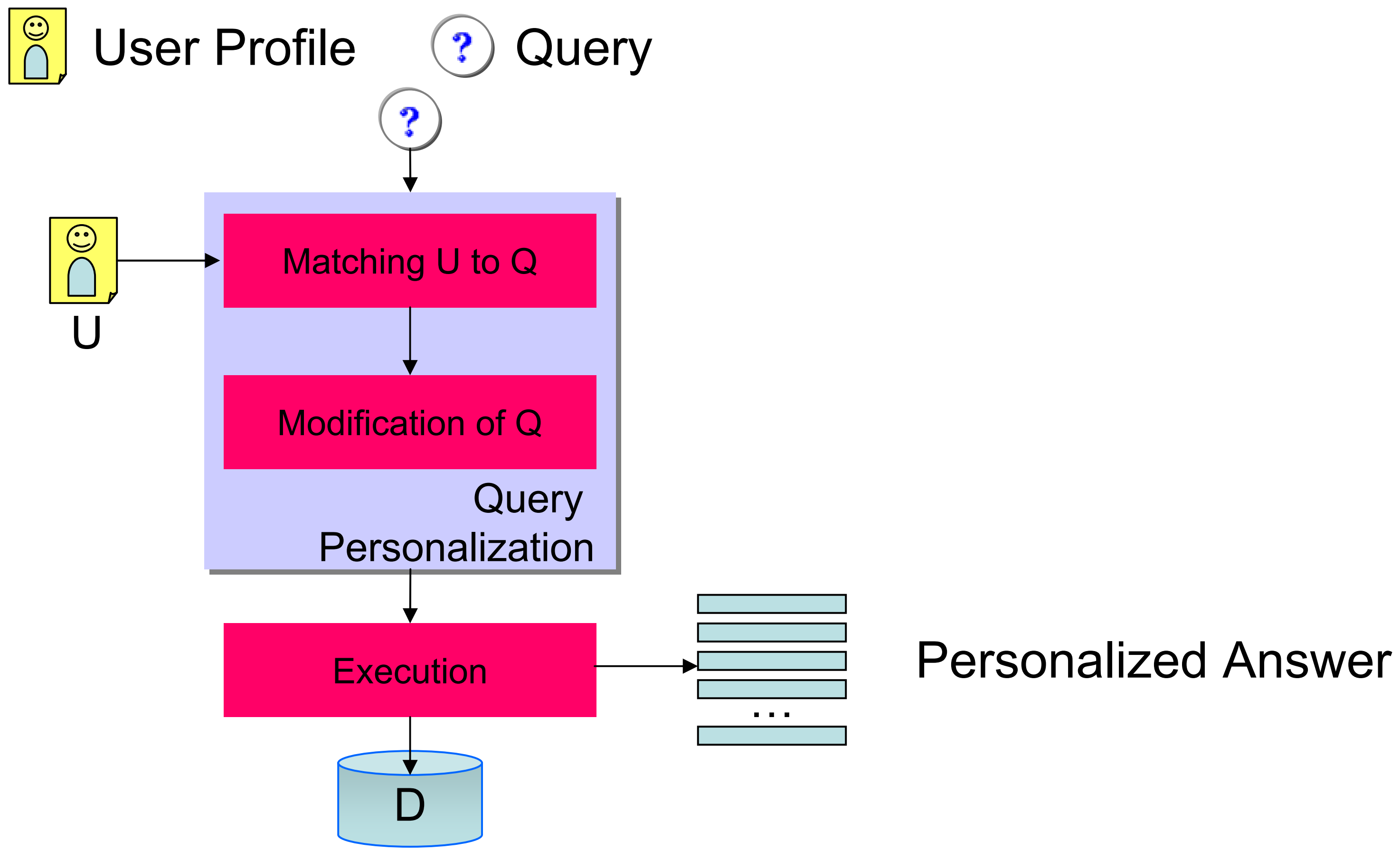
Basic Idea

A shift from ‘*consensus relevancy*’ toward ‘*personal relevancy*’ (*Pitkow et al, Communications of ACM, 45(2)*)

Personalization Methods

Personalized Search

System Model



Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

1

Personalization Methods

Personalized Search

A personalized answer should be:

- Interesting
- Ranked
- Self-Explanatory

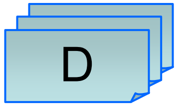
Personalization Methods

Query Personalization

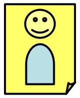
- IR-based
- DB-based

Personalization Methods

Query Personalization: IR-based



Vectors of keywords



Vectors of keywords

Query

Matching U to Q — Vector-space matching techniques

Modification of Q — Query augmentation

Personalization Methods

Personalized Search

Query Personalization: IR-based

java, coffee

ford, car

java

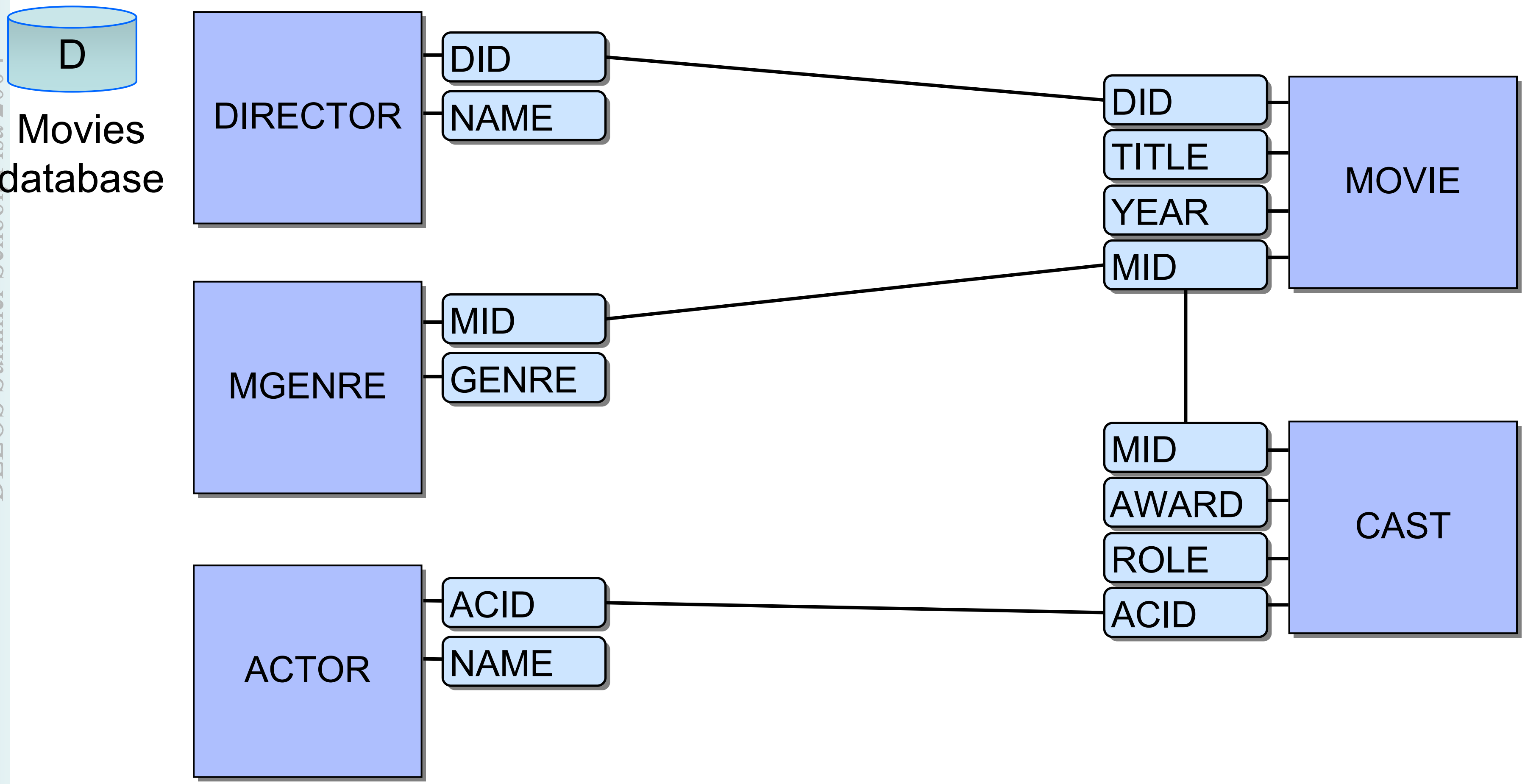
java and coffee

Personalized Query

Personalization Methods

Personalized Search

Query Personalization: DB-based

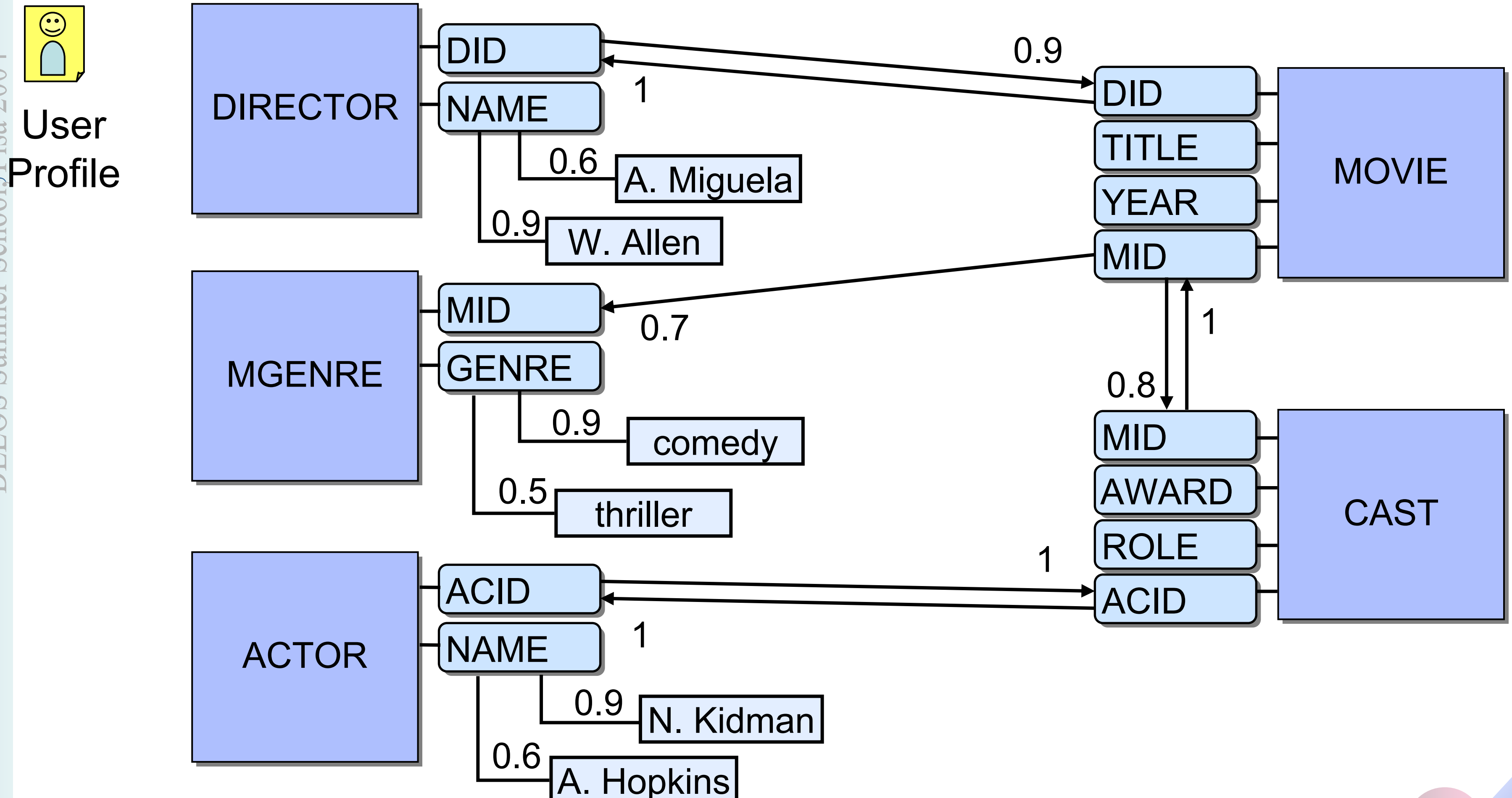


Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

Personalization Methods

Personalized Search

Query Personalization: DB-based



Query Personalization: DB-based

?

```
SELECT MV.title
FROM   MOVIE MV
WHERE  MV.YEAR='2003'
```

Query Personalization: DB-based

Query Personalization Logic

- L of the top K preferences
- L and K are determined by some criterion
  - explicitly given (e.g., 1 of the top 2)
  - related to degree of interest (e.g., …of those with d>0.6)
  - related to each other (e.g., half of the top …)

e.g., satisfy my top 3 preferences

Query Personalization: DB-based

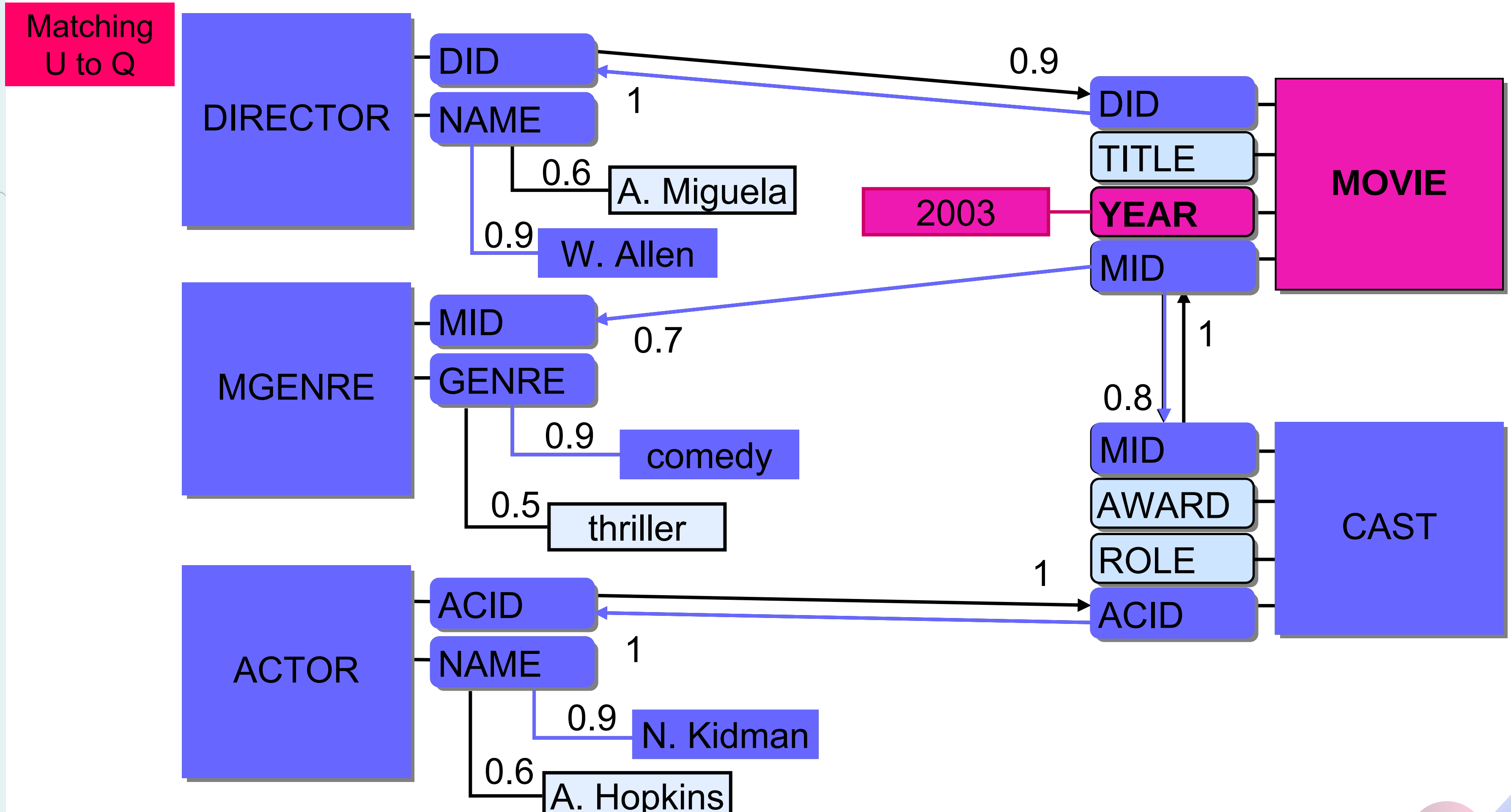
Matching U to Q

- Selection of top K preferences
- Best-first traversal of the personalization graph
- Path construction in decreasing order of degree of interest

Personalization Methods

Personalized Search

Query Personalization: DB-based



Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

Query Personalization: DB-based

Matching U to Q

MOVIE.did=DIRECTOR.did **and** DIRECTOR.name='W. Allen'

MOVIE.mid=CAST.mid **and** CAST.aid=ACTOR.aid **and** ACTOR.name='N. Kidman'

MOVIE.mid=MGENRE.mid **and** MGENRE.genre='comedy'

Query Personalization: DB-based

Modification of Q

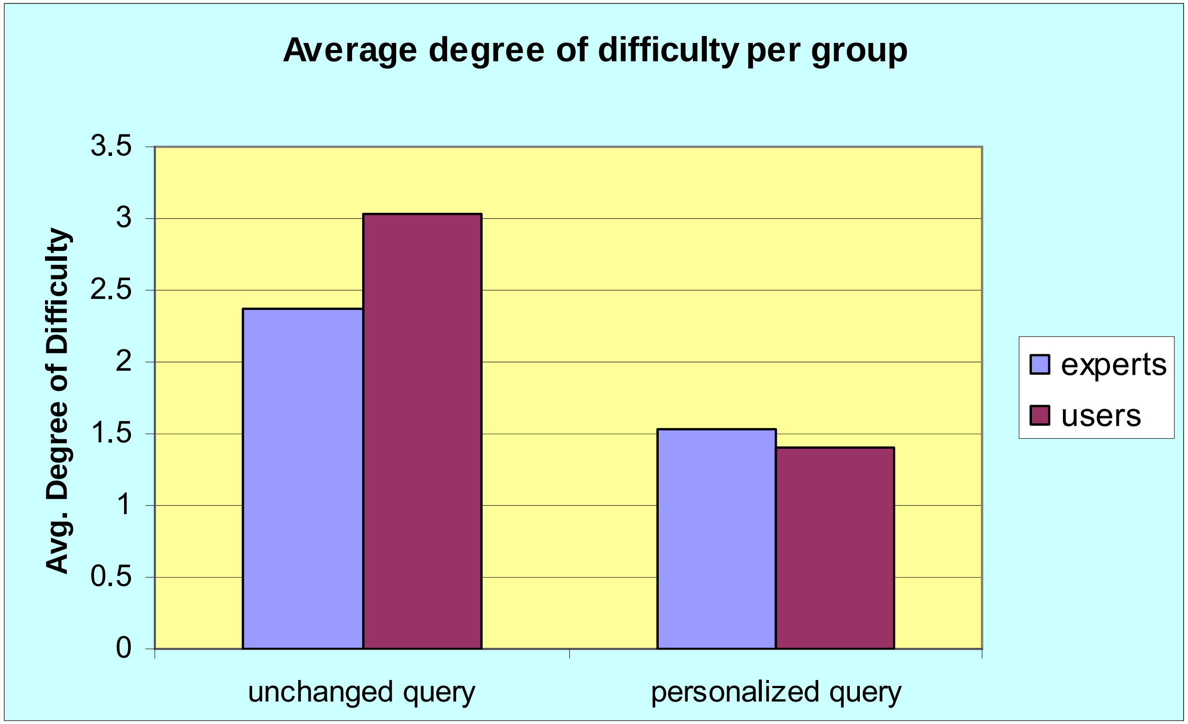
Query Rewriting: Personalized Query

SELECT MV.title
FROM MOVIE M,
CAST C, ACTOR A, MGENRE G, DIRECTOR D
WHERE MV.YEAR='2003' and
(M.MID=G.MID and GENRE='Comedy') and
(M.DID=D.DID and D.NAME='W.Allen') and
(M.MID=C.MID and C.ACID=A.ACID and
A. NAME='N.Kidman')

Benefits

Personalized vs. Unchanged Queries

(*G. Koutrika, Y. Ioannidis, 2004*)



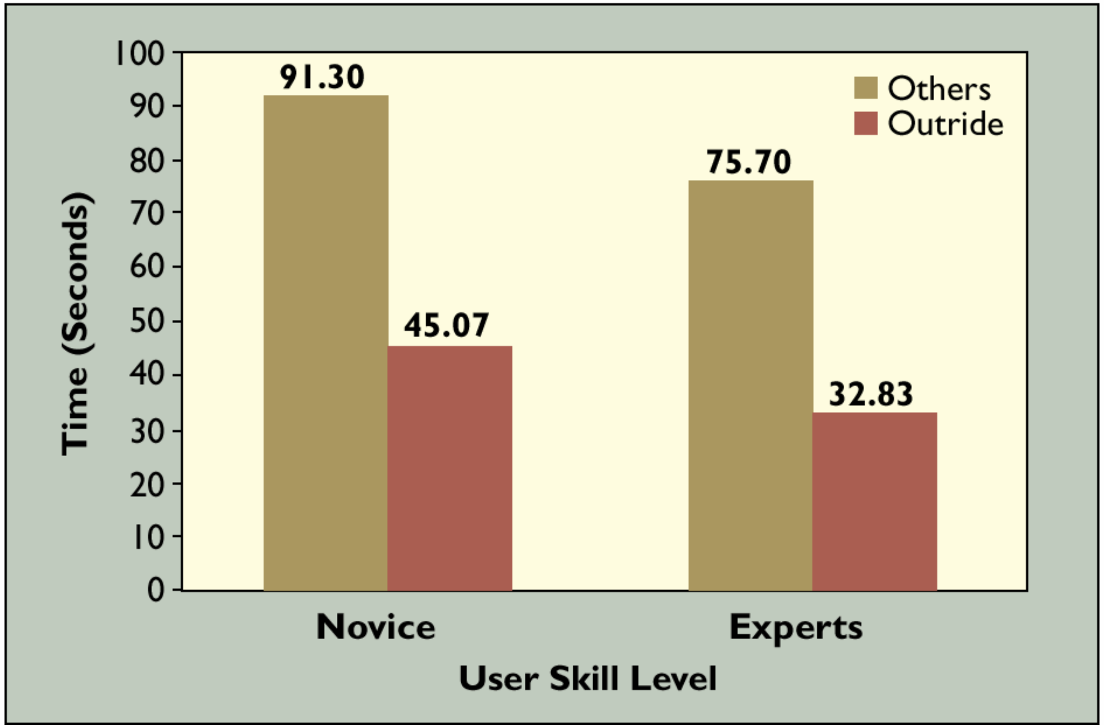
Personalization Methods

Personalized Search

Benefits

Personalized vs. Unchanged Queries

(*Pitkow et al, Communications of ACM, 45(2)*)



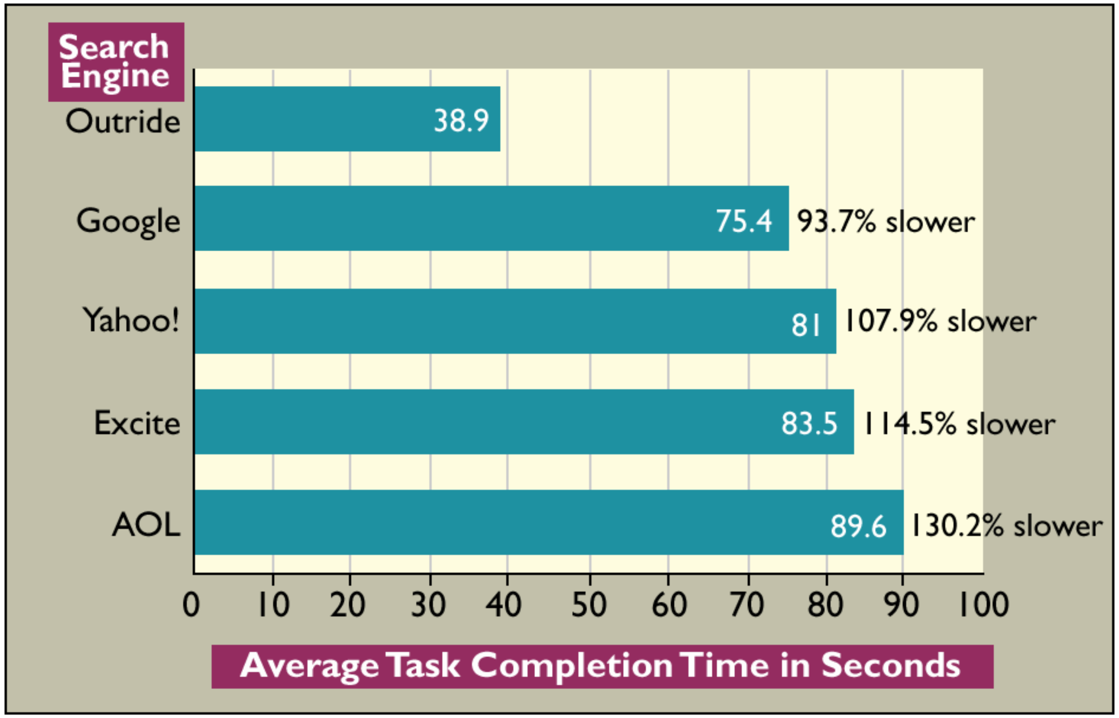
Personalization Methods

Personalized Search

Benefits

Personalized vs. Unchanged Queries

(*Pitkow et al, Communications of ACM, 45(2)*)



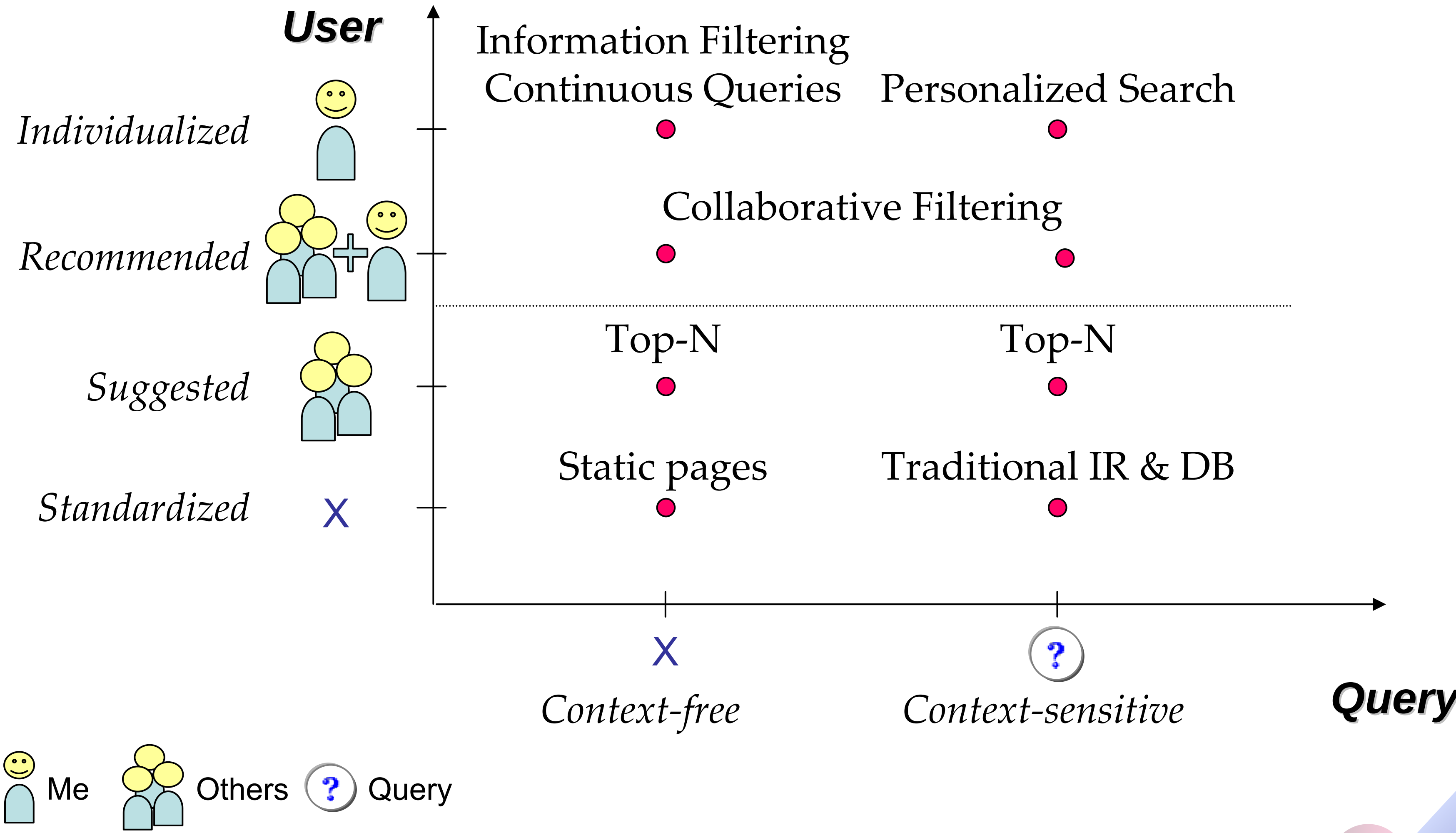
Personalization Methods

Systems

- IR-based — Outride
- DB-based — Personalized Database System

1 Personalization Methods

A Map



1 Personalization Methods

- Service Properties
- Special Services

1

Personalization Methods

Special Services

- Personalized Errands
- Personalized Negotiations
- Alert services

1 Personalization Methods

- Content Presentation
- Multimedia Presentation

Personalization Methods

Content Presentation

Forms

- Personalized descriptions
- Personalized links
- Personalized layout

Personalization Methods

Content Presentation

Examples

- Web catalogs

  (e.g., SETA)

- My Portals

  (e.g., myYahoo)

1 Personalization Methods

Multimedia Presentation

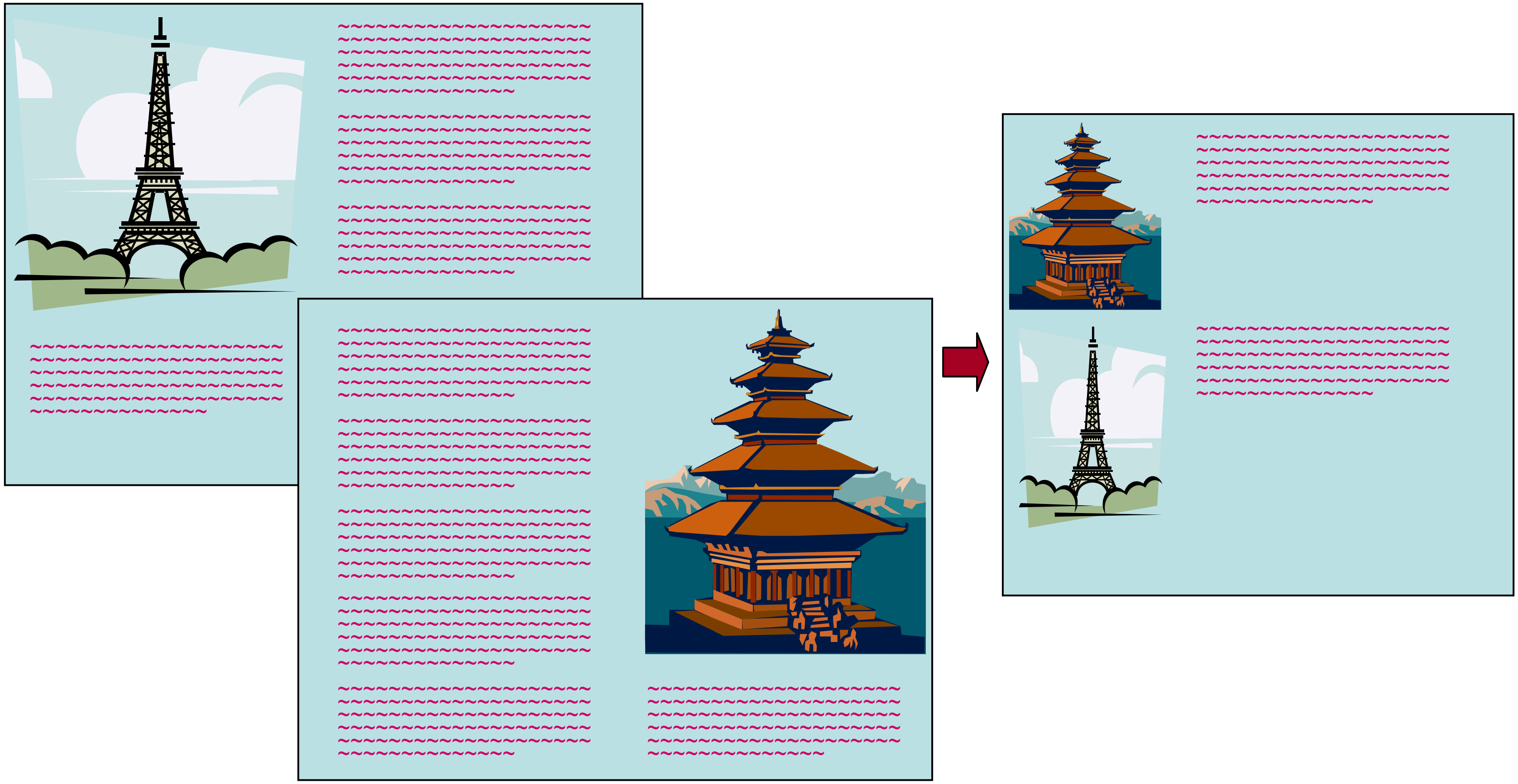
Forms

- File size
- Synchronization
- Transcoding

Personalization Methods

Multimedia Presentation

Example of multimedia presentations



Personalization Methods

Interaction Personalization:

optimising the way in which users access content and services based on user preferences as well as capabilities (universal access)

Content Personalization	Services Personalization	Presentation Personalization	Interaction Personalization

Personalization Methods

- Navigation Shortcuts
- Guided Tours
- Entry Points
- Web Companions

Personalization Methods

Navigation shortcuts

Make frequently-visited destinations easier to find based on frequent navigational user patterns

Personalization Methods

Guided Tours

Personalized superimposed navigation structures

Personalization Methods

Web companions

Embodied conversational characters

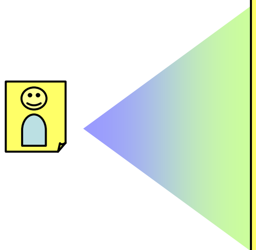
- Teachers
- Sales assistants (e.g., MIHU, COSIMA)
- Web chauffers

Companies

www.artificial-language.com
www.extempo.com
www.haptek.com
www.vperson.com

2 User Models

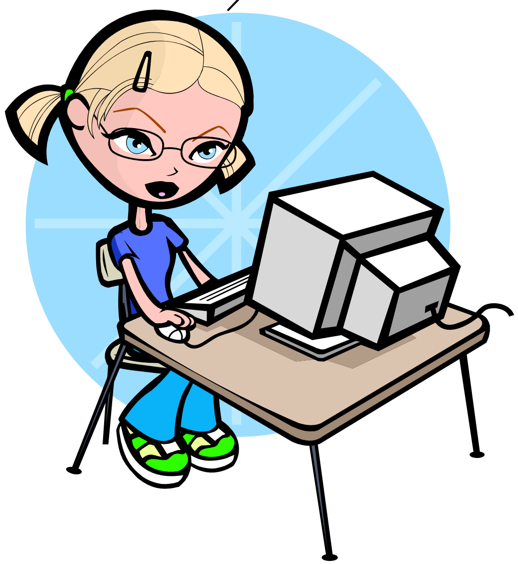
A user model



Personal data	Id, name, …
Demographic data	Age, marital status, …
Behavioral data	Behavior patterns, …
Preferences	Interests, …
Accessibility	Disabilities, …
Transient	Time, platform, connectivity, location

User Preference Models

Preferences



I *like* W. Allen *very much*

I *like* N. Kidman *better than* J. Roberts

I *like* adventures only *a little*

I *don't like* thrillers *at all*

I *prefer* movies *around* 2 hours

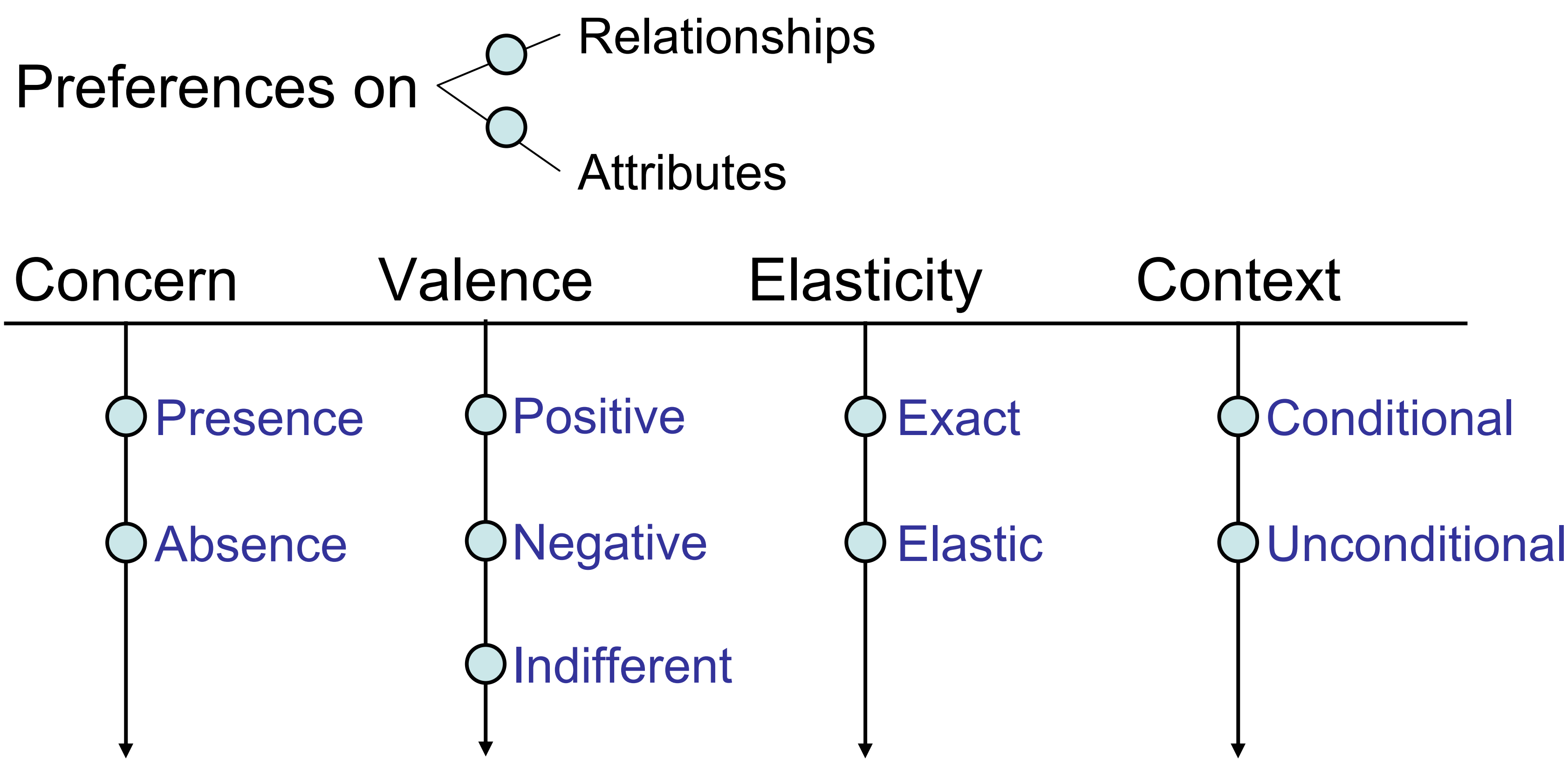
I *like* movies *without* violence

I'*m interested* in the director of a movie more than the cast

…

User Preference Models

A taxonomy of preferences



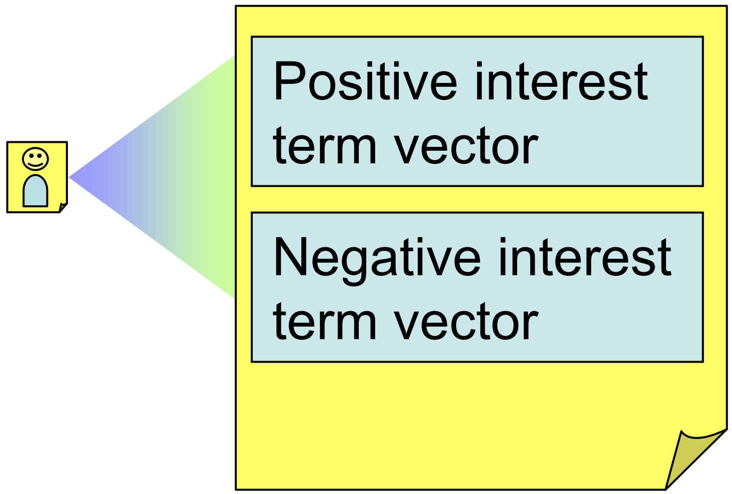
2 User Preference Models

- IR-based
- DB-based

User Preference Models

IR-based

Binary Representation

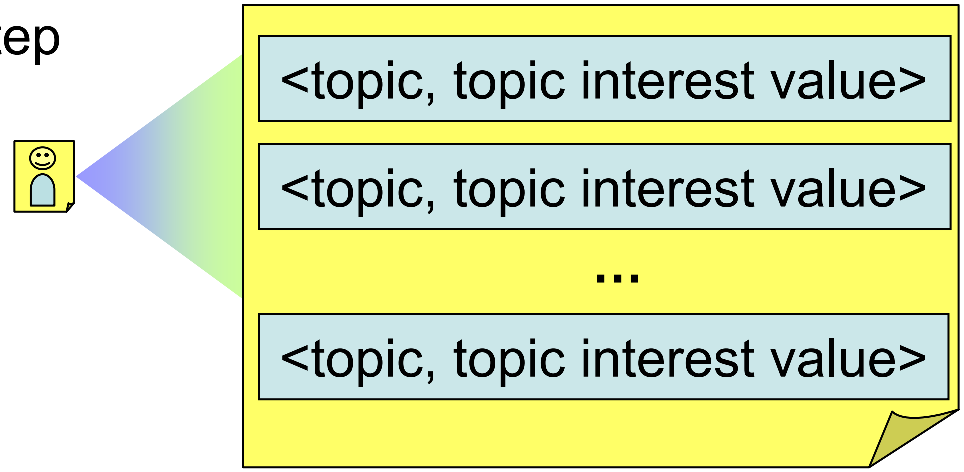


2 User Preference Models

IR-based

Multi-class Representation

QuickStep



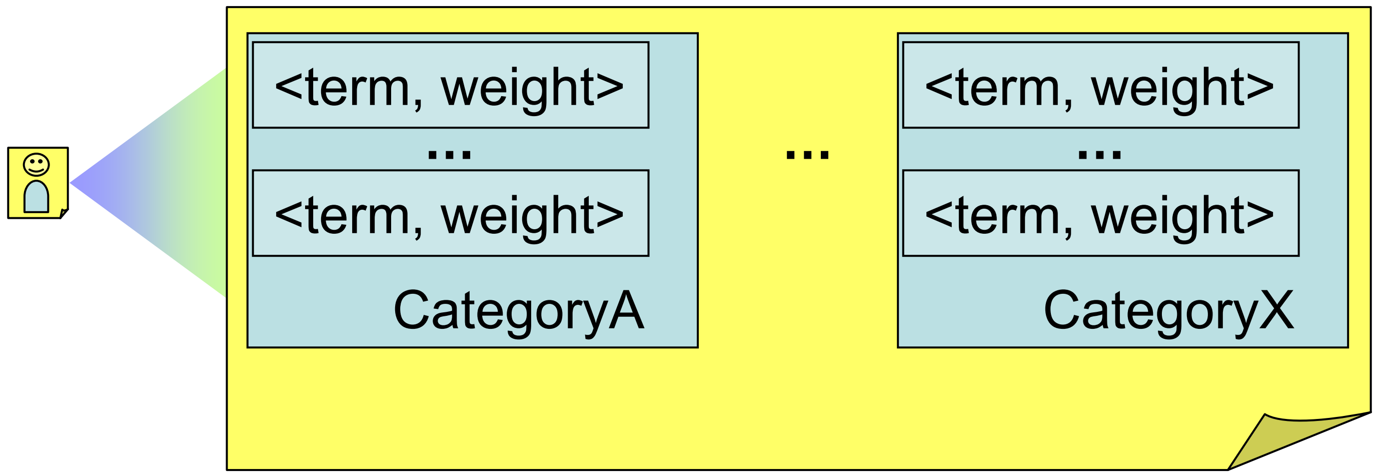
2

User Preference Models

IR-based

Multi-class Representation

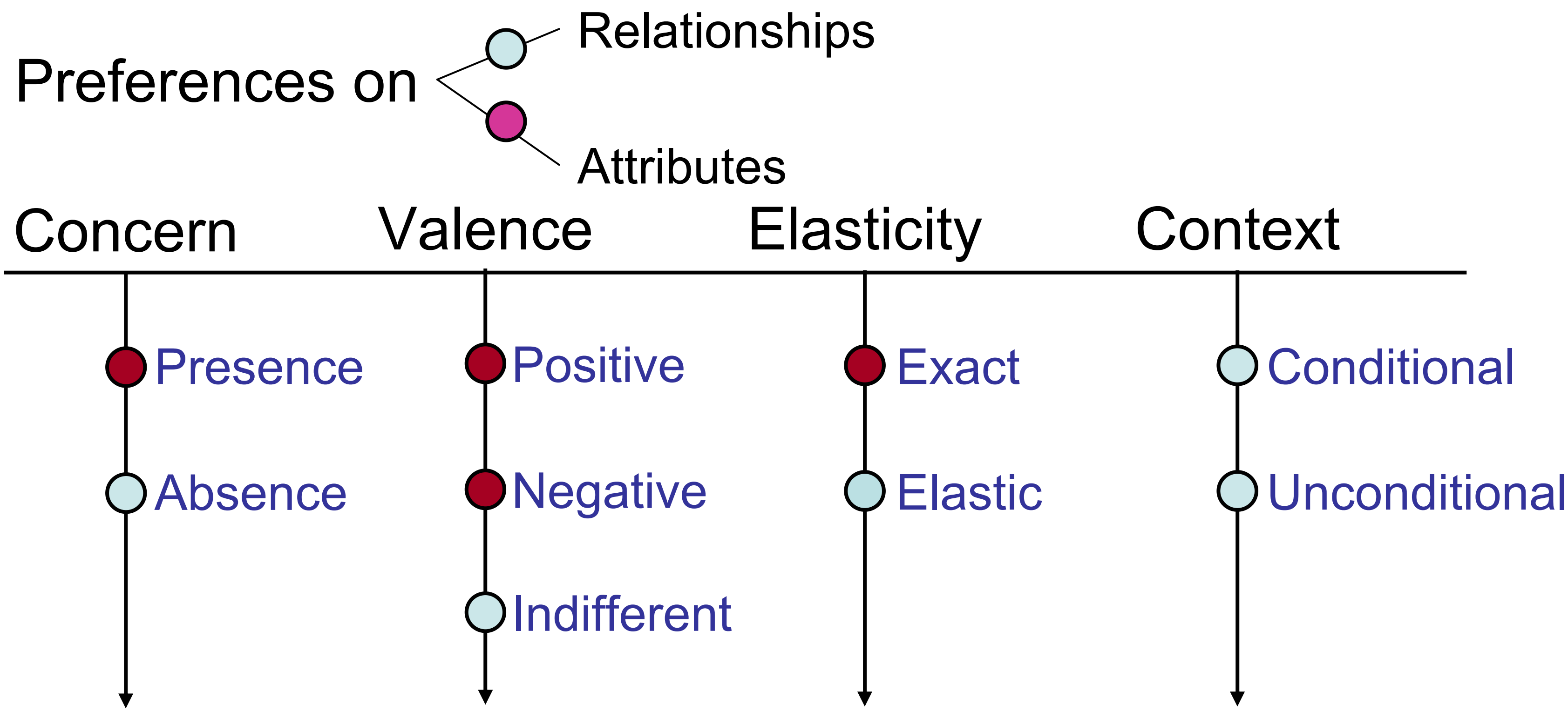
Liu, Yu, Meng, CIKM 2002



Cat. \ term	apple	recipe	pudding	football	fifa
COOKING	1	0.37	0.37	0	0
SOCCER	0	0	0	1	0.7

User Preference Models

IR-based



User Preference Models

DB-based

Qualitative Approaches

I like *A* better than *B*

Two frameworks

- Chomicki

  Chomicki, J.
  Preference Formulas in Relational Queries. ACM TODS, 28(4), 2003

- Kiessling

  (Kießling, W. Foundations of preferences in database systems. VLDB 2002)

User Preference Models

DB-based

Qualitative Approaches

Preferences between tuples in the answer to a query are specified directly using binary preference relations

- Chomicki logical formulas

Relation *Book(Title,Vendor,Price)*.

Preference :

$$(i, v, p) >_C (i', v', p') \equiv i = i' \wedge p < p'$$

User Preference Models

● DB-based

Qualitative Approaches

Preferences between tuples in the answer to a query are specified directly using binary preference relations

- Kiessling special preference constructors

Preference : P = (A, <P)

Some constructors

base HIGHEST(A)

{x <P_new y iff x < y};

base AROUND(A, z)

{x <P_new y iff abs(x – z) > abs(y – z)};

base POS/NEG(A, POS-set, NEG-set)

{x <P_new y iff $(x \in \text{NEG-set} \wedge y \notin \text{NEG-set}) \vee (x \notin \text{NEG-set} \wedge x \notin \text{POS-set} \wedge y \in \text{POS-set})$}

User Preference Models

Qualitative Approaches

- Kiessling special preference constructors

Preferences :

POS(transmission, {automatic})

NEG(make, {Ferrari})

POS/NEG(color, {yellow}; {gray})

POS/POS(category, {cabriolet}; {roadster})

EXP(color, {(green, yellow), (green, red), (yellow, white)})

User Preference Models

Qualitative Approaches

Preference relations are embedded into relational query languages through a relational operator that selects from its input the set of the most preferred tuples

- Chomicki winnow
- Kiessling BMO

User Preference Models

Quantitative Approaches

I (*do not*) like *A that much*

Two frameworks

- Agrawal, Wimmers

  (Agrawal, R., Wimmers, E.
  A Framework for Expressing and Combining Preferences. SIGMOD 2000)

- Koutrika, Ioannidis

  (Koutrika, G., Ioannidis, Y.
  Personalization of Queries in Database Systems. ICDE 2004)

User Preference Models

Quantitative Approaches

- Agrawal, Wimmers

<tuple, score>

score $\in$ [0, 1], $\perp$

Example

Relation *Book(Title,Vendor,Price*).

Preference: <*, *, 10, 0.8>

User Preference Models

DB-based

Quantitative Approaches

- Koutrika, Ioannidis

$<q_s, d_T(u), d_F(u)>$

...

$< q, d >$

...

q join

$d \in [0,1]$

q_s selection, *u* values satisfying q

$d_T(u)$: presence $d_F(u)$: absence

$d_T(u), d_F(u) \in [-1,1]$

where [-1,0) negative preference

0 indifference

(0, 1] positive preference

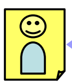
2

User Preference Models

DB-based

Quantitative Approaches

- Koutrika, Ioannidis



< DIRECTOR.name='W. Allen', **0.9**, **0** >

< GENRE.genre='adventure', **0.4**, **0** >

< GENRE.genre='thriller', **-0.9**, **0** >

< THEATRE.region='downtown', **0.7**, **-0.5** >

<MOVIE.mid=MGENRE.mid, **0.7** >

<MOVIE.did=DIRECTOR.did, **0.9** >

<DIRECTOR.did=MOVIE.did, **1** >

User Preference Models

DB-based

Quantitative Approaches

- Koutrika, Ioannidis

A preference $< q, d_T(u), d_F(u) >$ is satisfied if:

- q evaluates to true and $d_T(u) \geq 0$ or
- q evaluates to false and $d_F(u) \geq 0$

Example

< GENRE.genre='thriller', **-0.9**, **0** >

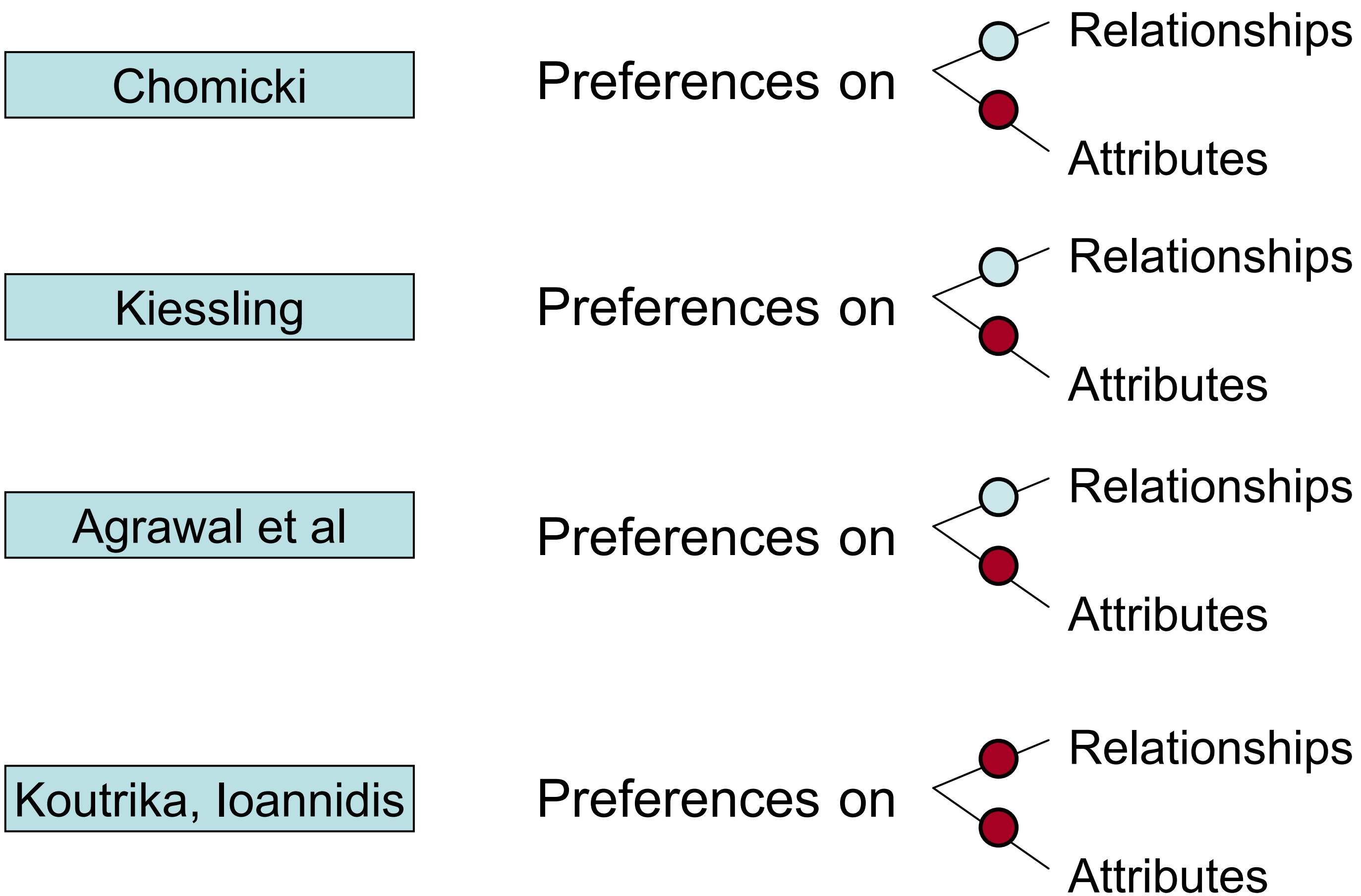
e.g., movies that are not thrillers satisfy this preference

< THEATRE.region='downtown', **0.7**, **-0.5** >

e.g., theatres located downtown satisfy this preference

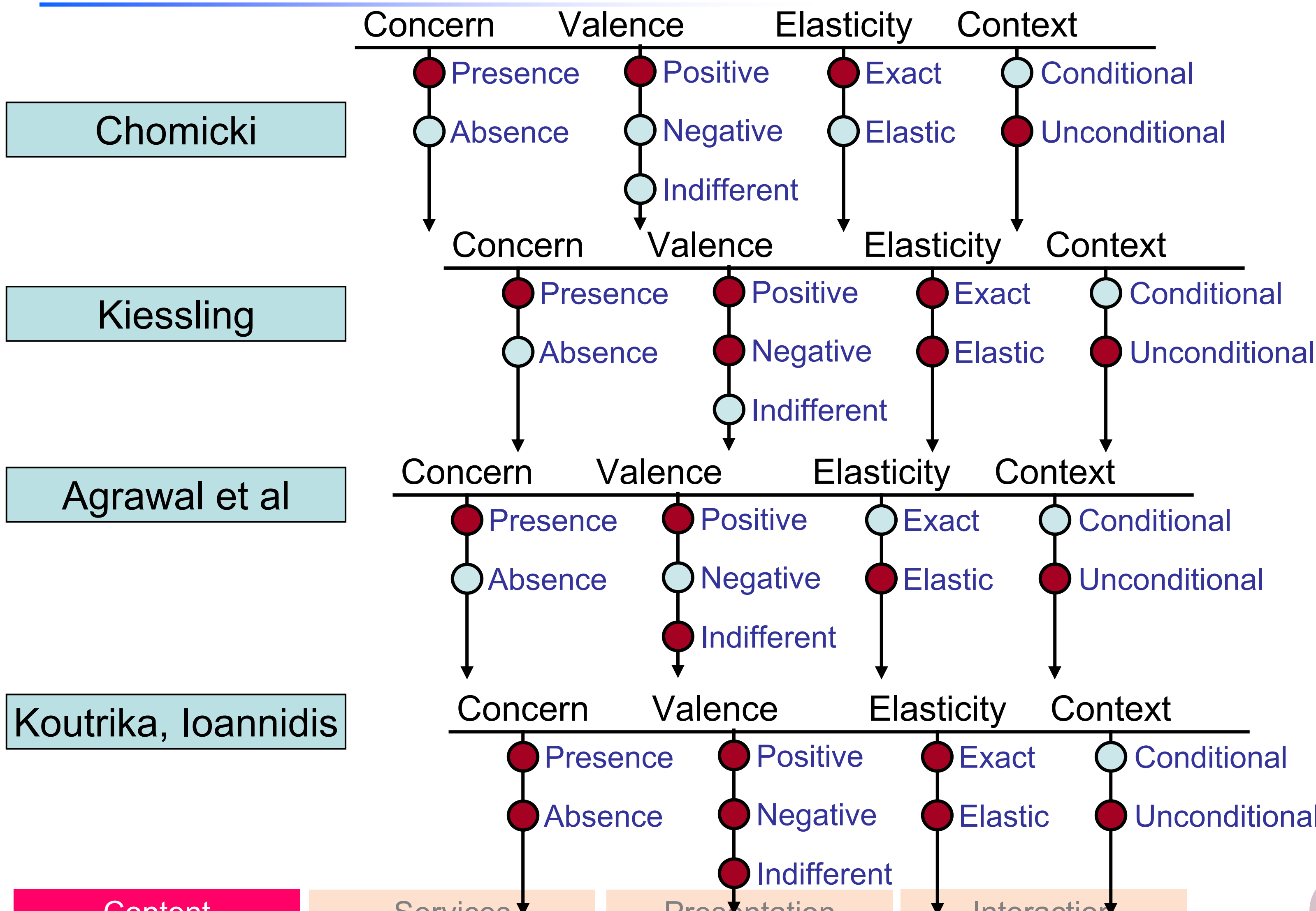
User Preference Models

DB-based



User Preference Models

- DB-based



Content Personalization | Services Personalization | Presentation Personalization | Interaction Personalization

Quantitative vs. Qualitative Approaches

Qualitative models

- − Provide an abstract, generic way to talk about priority and importance
- − Hard evaluation of preference queries
- + More intuitive

Quantitative models

- + Provide an ordering of all the answers
- + Capture preference intensity
- + Can be implemented using SQL3

Unified Approach?

User Preference Models

DB-based vs. IR-based models

DB-based models are defined for structured data

- They are domain-independent
- They are more expressive

On the other hand:

IR-based models are defined for unstructured data

- They are subject to all limitations stemming from unstructured data

Hybrid Models ?

2

User Preference Models

Critique on Models

- Diversity
- Domain/application-dependence
- Low expressivity (IR-based models)
- Preference expiration policy
- Testing

2 User Preference Models

Directions

- Specialized research and collaboration between different disciplines (*)
- Increased Expressivity
- Cross-Application Independence
- Declarative expression of preferences
- Multiple profiles per user

(*) Dagstuhl-Seminar 04271:Preferences: Specification, Inference, Applications

Content Personalization	Services Personalization	Presentation Personalization	Interaction Personalization

User Profiling

Types of Users

- anonymous
- session
- tracked
- identified

User Profiling

User Feedback

Feedback: Positive or negative

- Explicit
  - specifying keywords
  - selecting and marking documents
  - answering questions about their interests
  - providing ratings
- Implicit
  - reading time
  - saving
  - printing
  - selecting
  - search history
  - navigation history
  - physical activity

3

User Profiling

User Feedback

Sources of Implicit feedback

- ClickStream Analysis
- Web Logs
- Sensors

User Profiling

User Feedback

	positive	negative	explicit	implicit
Fab	√	√	√	√
WebMate	√		√	
Amalthea	√	√	√	
NewT	√	√	√	

User Profiling

User Feedback

- Explicit
  - + Easier implementation of profiling
  - + User control

On the other hand:

Users engage in additional activities beyond their normal searching behavior

High cost to the user

Benefits not always apparent

- – Out-of-date profiles
- – Sparse profiles

User Profiling

User Feedback

- Implicit
  - + No user burden

On the other hand:

- − Lower confidence
- − Privacy

User Profiling

User Feedback

Classification (*Oard and Kim*)

Minimum Scope

Behavior Category

	Segment	Object	Class
Examine	View Listen Scroll Find Query	Select	Browse
Retain	Print	Bookmark Save Delete Purchase	Subscribe
Reference	Copy-Paste Quote	Forward Link Cite	
Annotate	Mark up	Rate Publish	Organize
Create	Type Edit	Author	

3 User Profiling

User Feedback

Fundamental questions:

Which observable behaviors can be used as implicit measures of interest?

What should the weight of each one be?

3

User Profiling

User Feedback

Studies

Claypool, Le, Waseda, and Brown (IUI2001)

Time spent on a page

Amount of scrolling on a page
(all scrolling measures combined)

Combination of time and scrolling

strong positive correlation with the explicit ratings

Number of mouse clicks

Individual scrolling measures

ineffective in predicting explicit ratings

User Profiling

User Feedback

QuickStep

	topic interest value
Paper browsed	1
Recommendation followed	2
Topic rated interesting	10
Topic rated uninteresting	-10

$$\text{Topic interest} = \sum_{n}(\text{interest\_value}(n))/\text{days\_old}(n)$$

User Profiling

User Profiling Techniques

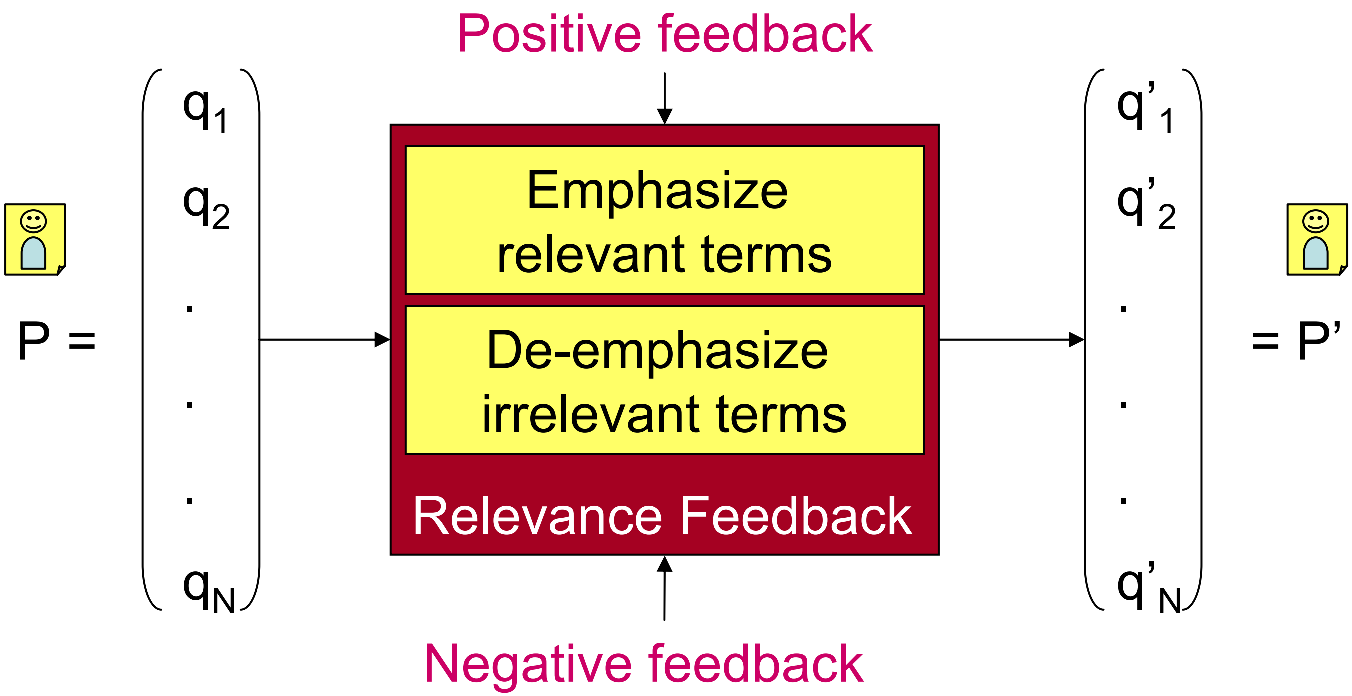
- Relevance Feedback
- Machine Learning
- Mining

User Profiling

Relevance Feedback

A history over 30 years in Information Retrieval

Basic Idea



Techniques

Vector Processing Methods

Ide $$P' = P + \sum_{\text{relevant}} D_i - \sum_{\text{non-relevant}} D_i$$

Rocchio $$P' = P + \beta \sum_{n_1 \text{ relevant}} D_i / n_1 - \gamma \sum_{n_2 \text{ non-relevant}} D_i / n_2$$

Probabilistic Retrieval Methods

conventional $$P' = \log[p_i(1-u_i)/u_i(1-p_i)]$$

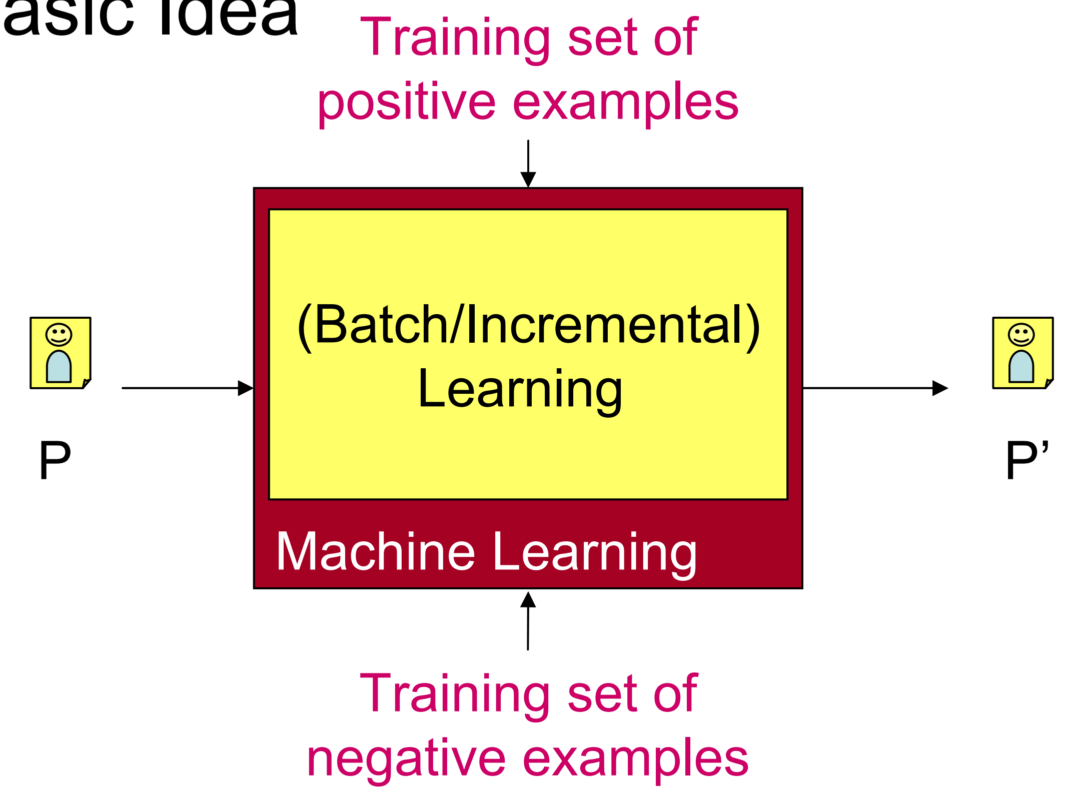
$$p_i = P(x_i | rel),\ u_i = P(x_i | nonrel),$$

User Profiling

Machine Learning

Building personal profiles

Basic Idea



User Profiling

Machine Learning

Building personal profiles

Its form depends on the ML approach applied
(e.g., rules, predictive model)

E.g. a Bayesian model is used to predict the class of new content

User Profiling

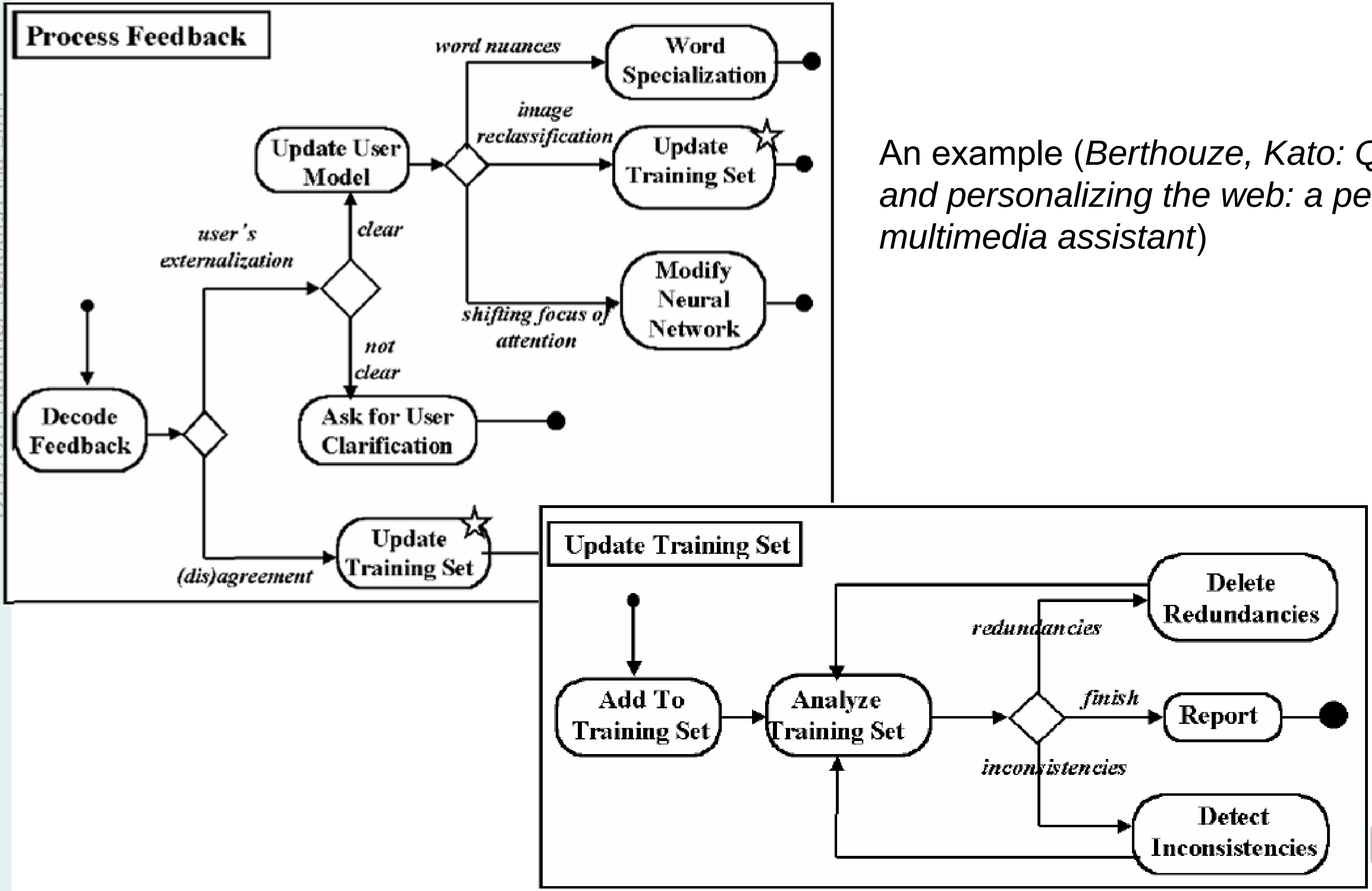
Machine Learning

Techniques

Neural nets	ARAM
Rule learners	Ripper, HCV, CDL4
Decision Trees	C4.5, ID3
Probabilistic Classification	Naïve Bayes

User Profiling

Machine Learning



An example (*Berthouze, Kato: Querying and personalizing the web: a personal multimedia assistant*)

Data Mining

The semi-automatic discovery of

- patterns,
- classes,
- associations,
- statistically significant structures

Data Mining Techniques

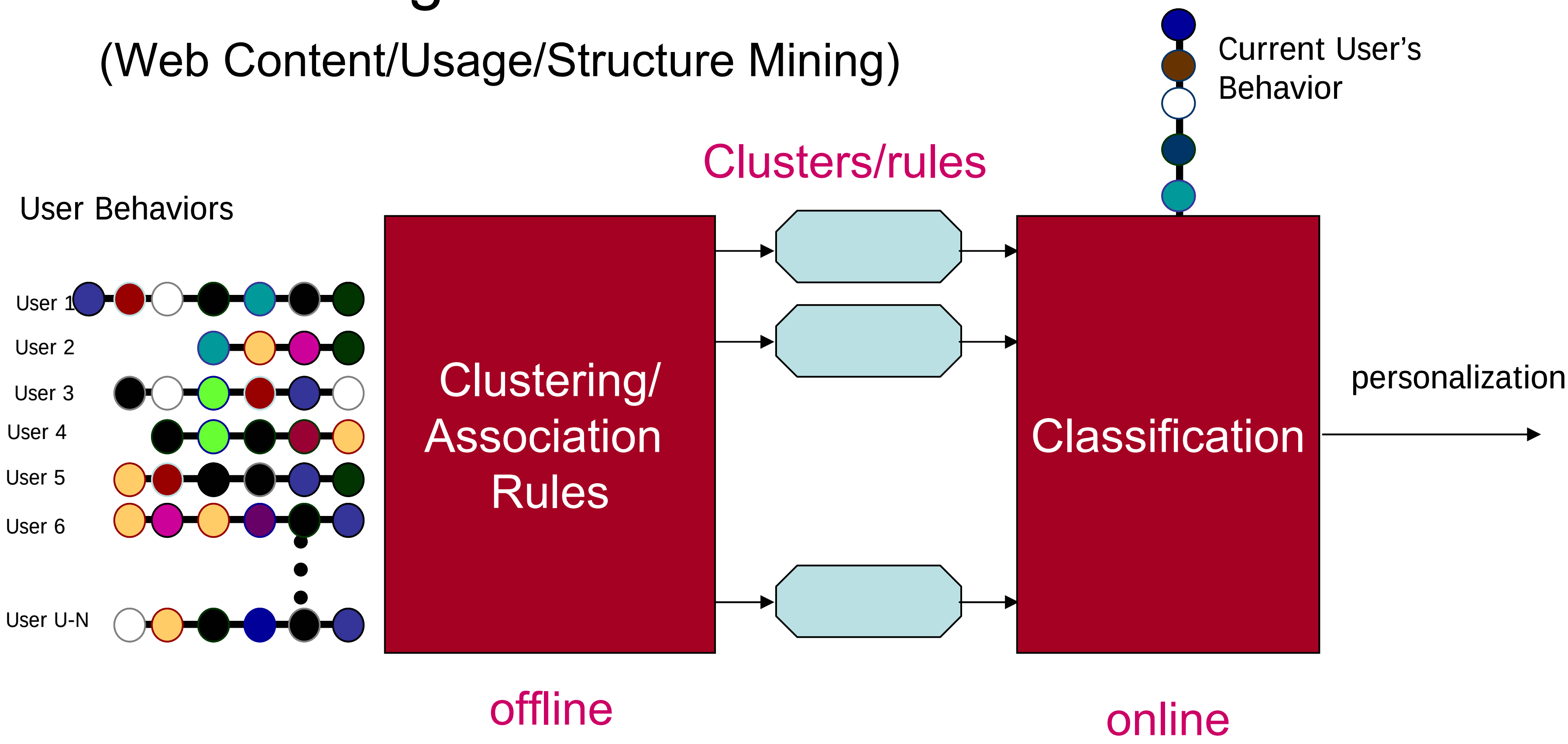
- Clustering
- Classification
- Association Rules

User Profiling

Mining

Web Mining

(Web Content/Usage/Structure Mining)



Web Mining: Association Rules

- Action Rules

  Action_1, Action_2, … $\text{Action}_N \rightarrow \text{Action}_R$; confidence= C, support = S

- Market Basket Rules

  Item_1, Item_2, … $\text{Item}_N \rightarrow \text{Item}_R$; confidence= C, support = S

E.g. APriori

Web Mining: Clustering

- *Partioning methods*:
  create k groups of a given data set, where each group represents a cluster. (e.g., PageGather, EM)

- *Hierarchical methods*:
  decompose a given data set creating a hierarchical structure of clusters. (e.g., BIRCH)

- *Model-based methods:*
  find the best fit between a given data set and a mathematical model (e.g., COBWEB, Autoclass, ITERATE)

(*Han and Kamber 2001*)

User Profiling

Directions

- Handle all preference types
- Obtain negative examples
- Distinguish dislike from indifference
- Capture changes in user interests
- Distinguish between long-term and short-term preferences

User Profiling

Directions

- Scalability
- Privacy
- Batch and incremental construction of profiles
- Users should be able to inspect their personal profiles
- Integration of user temporal characteristics

Evaluation of Personalization

A question

It is a very reasonable question to ask whether or not user models and personalization will actually improve information access?

Evaluation of Personalization

- Adding a user model to any system $\rightarrow$ more complex, less predictable system
- A personalized configuration may actually be slower or more error-prone than a conventional configuration
- Different configurations make it difficult for users in a group to cooperate.
- A common adaptation for user models is information filtering that seems to be helpful
- On the other hand, eliminating seemingly irrelevant information can confuse users

Evaluation of Personalization

- Empirical evaluations to determine which users are helped or hindered by user-adapted interaction
- Insufficient empirical evaluations, but an encouraging upward trend.

Evaluation of Personalization

Some Hints

- The user interface needs to provide a way to explain what the system is doing to personalize the experience as well as to undo the personalization.
- Allowing users control the extent of the personalization can also help alleviate inaccurate personalization.

REFERENCES (1)

Smyth, B., Bradley, K., and Rafter R. (2002). Personalization Techniques for Online Recruitment Services. *Communications of the ACM*, Vol. 45(5), 39-40.

Glover, E., Lawrence, S., Birmingham, W., and Lee Giles, C. (1999): Architecture of a Metasearch Engine That Supports User Information Needs. *In Proceedings of the ACM I nternational Conference on Information and Knowledge Management* (*CIKM*), 210-216.

Foltz, P., and Dumais, S. (1992). Personalized Information Delivery: An Analysis of Information Filtering Methods. Communications of the ACM, Vol 35(12), 51-60.

Mooney, R.J., Roy, L.: Content-Based Book Recommending Using Learning for Text Categorization Proceedings of the Fifth ACM Conference on Digital Libraries, San Antonio, TX, June (2000) 195–204

Sakagami, H., Kamba T., Sugiura A., and Koseki Y. (1997). Learning personal preferences on online newspaper articles from user behaviours. In Proceedings of the 6th International World Wide Web Conference, 291-300.

Belkin, N., and Croft W.B. (1992). Information Filtering and Information Retrieval: Two sides of the same coin?. *Communications of the ACM*, Vol.35(12), 29-38.

Towle, B., Quinn, C.N.: Knowledge Based Recommender Systems Using Explicit User Models. Knowledge-Based Electronic Markets Workshop at AAAI 2000, Austin, TX.

Goldberg, D., Nichols, D., Oki, B.M., Terry, D.: Using Collaborative Filtering to Weave an Information Tapestry. Communications of the ACM, Vol. 35, No. 12, December (1992) 61–70

Krulwich, B,: Lifestlye Finder: Intelligent User Profiling Using Large-Scale Demographic Data Artificial Intelligence Magazine 18, vol. 2. 37–45, 1997

Konstan, J.A., Riedl, J., Borchers, A., Herlocker, J.L.: Recommender Systems: A Group-Lens Perspective. In Recommender Systems: Papers from the 1998 Workshop (AAAI Technical Report WS-98-08) Menlo Park, (1998) 60–64

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M.: Combining content-based and collaborative filters in an online newspaper. In Proceedings of ACM SIGIR Workshop on Recommender, August (1999)

Chen, J., DeWitt, D., Tian, F., and Wang, Y. (2000). NiagaraCQ: A Scalable Continuous Query System for Internet Databases. *In Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 379 - 390.

REFERENCES (2)

Liu F., Yu C., and Meng W. (2002). Personalized Web Search by Mapping User Queries to Categories. *In Proceedings of the ACM International Conference on Information and Knowledge Management* (*CIKM*), 558-565.

Koutrika, G., and Ioannidis, Y. (2004b). Personalization of Queries in Database Systems. *In Proceedings of the International Conference on Data Engineering (ICDE)*, 597-608.

Self-Adaptive User Profiles for Large-Scale Data Delivery Ugur Cetintemel, Michael J. Franklin, C. Lee Giles ICDE 2000

Chomicki, J. Preference Formulas in Relational Queries. ACM TODS, 28(4), 427–466, 2003.

Kießling, W., and Köstler, G. (2002a). Foundations of Preferences in Database Systems. In Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), 311-322.

Kießling, W., and Köstler, G. (2002b). Preference SQL-Design, Implementation, Experiences. *In Proceedings of the 28th International Conference on Very Large Data Bases* (*VLDB*), 990-1001.

Torlone, R., Ciaccia, P. Which Are My Preferred Items? Workshop on Recommendation and Personalization in E-Commerce, 2002.

Chomicki, J. Semantic Optimization of Preference Queries. Proc. of the Int'l Sym. on Applications of Constraint Databases, 133-148, 2004.

Borzsonyi, S., Kossmann, D., Stocker, K. The Skyline Operator. Proc. of ICDE, 421–430, 2001.

Papadias, D., Tao, Y., Fu, G., Seeger:, B. An Optimal and Progressive Algorithm for Skyline Queries. Proc. of ACM SIGMOD, 467–478, 2003.

Kossmann, D., Ramsak, F., Rost, S. Shooting stars in the sky: An online algorithm for skyline queries. Proc. of VLDB, 2002.

Agrawal, R., and Wimmers, E. (2000). A Framework for Expressing and Combining Preferences. *In Proceedings of the ACM International Conference on Management of Data (SIGMOD)*, 297-306.

REFERENCES (3)

Hristidis, V., Koudas, N., and Papakonstantinou Y. (2001). PREFER: A System for the Efficient Execution of Multi-parametric Ranked Queries. In Proceedings of the ACM *International Conference on Management of Data (SIGMOD)*, 259-270.

Bruno, N., Chaudhuri, S., Gravano, L. Top- k Selection Queries over Relational Databases: Mapping Strategies and Performance Evaluation. ACM TODS, 27(2), 153-187, 2002.

Zhu, L., Meng W. Learning-Based Top-N Selection Query Evaluation over Relational Databases. Proc. of WAIM, 2004.

Chang, K., Hwang, S. Minimal Probing: Supporting Expensive Predicates for Top-k Queries. Proc. of ACM SIGMOD, 2002.

Ilyas, I, Aref W., Elmagarmid, A. Supporting Top-k Join Queries in Relational Databases. Proc. of VLDB, 2003.

Ilyas, I., Shah, R. Aref, W., Vitter, J., Elmagarmid, A. Rank-aware Query Optimization. Proc. of ACM SIGMOD, 2004.

Balke, W-T, Guntzer, U. Multi-objective Query Processing for Database Systems. Proc. of VLDB, 2004.

Srivastava, J., Cooley, R., Deshpande, M., & Tan, P. (2000). Web usage mining: Discovery and applications of usage patterns from web data. SIGKDD Explorations, 1(2), 12-23.

Holland, S., Ester, M., and Kießling, W. (2003). Preference Mining: A Novel Approach on Mining User Preferences for Personalized Applications. PKDD, LNAI 2838, 204–216.

James Allan. Incremental Relevance Feedback for Information Filtering. SIGIR 1996.

Liu, L., Pu, C., and Tang, W. (1999). Continual Queries for Internet Scale Event-Driven Information Delivery. *IEEE Transactions on Knowledge and Data Engineering*, Vol. 11(4), 610-628.

Pitkow, J., Schutze, H., Cass, T., Cooley, R., TurnBull, D., Edmonds, A., Adar, E., and Breuel, T. (2002). Personalized Search. *Communications of the ACM*, Vol. 45(9), 50-55.

Semeraro, G., Degemmis, M., Lops, P., Thiel, U., and L'Abbate, M. (2003). A Personalized Information Search Process Based on Dialoguing Agents and User Profiling. *ECIR, Lecture Notes in Computer Science 2633*, 613-621.

André E., Rist, T. (2002). From adaptive hypertext to personalized web companions. *Comm. of the ACM*, 45(5), 43-46.

REFERENCES (4)

Tan, P.N, and V. Kumar: 2002, Discovery of Web Robot Sessions Based on their Navigational Patterns. *Data Mining and Knowledge Discovery*, 6(1), pp. 9-35.

Tauscher, L. and S. Greenberg: 1997, How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies*, Special issue on World Wide Web Usability, 47(1), pp. 97-138.

Theodoridis, S. and Koutroubas, K: 1999, *Pattern Recognition*. Academic Press.

Utgoff, P. E: ID5: 1988, An incremental ID3. In Proceedings of the 5th International Conference on Machine Learning, pp. 107--120, San Mateo, CA, Morgan Kaufman

Webb, G.I , M. J. Pazzani and D. Billsus: 2001, Machine Learning for User Modeling. *User Modeling and User-Adapted Interaction,* 11, pp. 19-29

Wexelblat, A., and P. Maes: 1997, Footprints: History-rich Web browsing. *In Proceedings Conference Computer-Assisted Information Retrieval (RIAO).* pp. 75-84

Widmer G. and M. Kubat: 1996. Learning in the presence of concept drift and hidden contexts. Machine Learning, 23(2), pp 69-101

Wu, K., P. S. Yu, and A. Ballman: 1998, Speedtracer: A Web usage mining and analysis tool. *IBM Systems Journal,* 37(1)

Wu, X.: 1993, The HCV Induction Algorithm. *In Proceedings of the 21st ACM Computer Science Conference.* ACM Press, pp. 169-175

Yan, T. W., M. Jacobsen, H. Garcia-Molina, and U. Dayal: 1996, From User Access Patterns to Dynamic Hypertext Linking. WWW5 / Computer Networks 28(7-11), pp. 1007-1014.

Zhang, T., R. Ramakrishnan, and M. Livny: 1996, BIRCH: an efficient data clustering method for very large databases. *In Proceedings ACM-SIGMOD International Conference in Management of Data.* Montreal-Canada, pp. 103-114

Zhu, T: 2001, Using Markov Chains for Structural Link Prediction in Adaptive Web Sites. *UM 2001, LNAI 2109*, pp. 298-300

Zukerman, I, D.W. Albrecht, and A.E. Nicholson: 1999, Predicting users' requests on the www. In Proceedings of the Seventh International Conference on User Modeling, Banff, Canada, pp. 275-284.

Pitkow, J., and K. Bharat: 1994, WEBVIZ: A Tool for World-Wide Web Access Log Visualization. In Proceedings of the 1st International World-Wide Web Conference. Geneva, Switzerland.

Pitkow, J.: 1997, In search of reliable usage data on the WWW. In Proceedings of the 6th Int. World Wide Web Conference, Santa Clara, CA.

Pitkow J. and P. Pirolli: 1999, Mining longest repeating subsequences to predict WWW surfing. In Proceedings of the 1999 USENIX Annual Technical Conference.

REFERENCES (5)

Schafer, J. B., J. A. Konstan, and J. Riedl: 2001, E-commerce Recommendation Applications, Data Mining and Knowledge Discovery, 5(1-2), pp 115-152

Schwarzkopf, E: 2001, An adaptive Web site for the UM2001 conference. In Proceedings of the UM2001 Workshop on Machine Learning for User Modeling, pp 77-86.

Shahabi, C., A. M. Zarkesh, J. Abidi, and V. Shah: 1997, Knowledge discovery from user's Web-page navigation. In Proceedings of the 7th IEEE Intl. Workshop on Research Issues in Data Engineering (RIDE), pp. 20-29.

Shahabi, C., F. Banaei-Kashani, and J. Faruque: 2001, A Reliable, Efficient, and Scalable System for Web Usage Data Acquisition. In WebKDD'01 Workshop in conjunction with the ACM-SIGKDD 2001, San Francisco, CA, August

Spiliopoulou, N., and L. C. Faulstich: 1998, WUM: A Web Utilization Miner. In International Workshop on the Web and Databases. Valencia, Spain.

Spiliopoulou, M: 1999, Tutorial: Data Mining for the Web. PKDD' 99. Prague, Czech Republic.

Spiliopoulou, M., L. C. Faulstich, and K. Wilkler: 1999a, A data miner analyzing the navigational behavior of Web users. In Proceedings of the Workshop on Machine Learning in User Modeling of the ACAI99. Greece.

Spiliopoulou, M., C. Pohle, and L. C. Faulstich: 1999b, Improving the effectiveness of a web site with Web usage mining. In Proceedings of the 1999 KDD Workshop on Web Mining, San Diego CA. Springer-Verlag

Srivastava, J., R. Cooley, M. Deshpande, and P. T. Tan: 2000, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. SIGKDD Explorations, 1 (2).

Paliouras, G., C. Papatheodorou, V. Karkaletsis, P. Tzitziras, and C. D. Spyropoulos: Large-Scale Mining of Usage Data on Web Sites. AAAI Spring Symposium on Adaptive User Interfaces. Stanford, California.

Paliouras, G., C. Papatheodorou, V. Karkaletsis, and C. D. Spyropoulos: 2000 Clustering the Users of Large Web Sites into Communities. In Proceedings of Intern. Conf. on Machine Learning (ICML), Stanford, California, pp. 719-726

Pei, J., J. Han, B. Mortazavi-Asl, B., and H. Zhu: 2000, Mining Access Pattern efficiently from Web logs. In Proceedings 2000 Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'00), Kyoto, Japan, pp. 396-407.

Pennock, D., E. Horvitz, S. Lawrence, and C. Lee Giles: 2000, Collaborative Filtering by Personality Diagnosis: A Hybrid Memory and Model-Based Approach.UAI-2000: The 16th Conference on Uncertainty in Artificial Intelligence. Stanford University, Stanford, CA, pp. 473-480.

Perkowitz, M. and O. Etzioni: 1998, Adaptive sites: Automatically synthesizing Web pages. In Proceedings of the 15th National Conf. on Artificial Intelligence. Madison, Wisconsin, pp. 727-732.

Perkowitz, M. and O. Etzioni: 2000, Adaptive Web Sites. Communications of the ACM. 43 (8), pp. 152-158

REFERENCES (6)

Mobasher, B., R. Cooley, and J. Srivastava: 1999b, Creating Adaptive Web Sites Through Usage-Based Clustering of URLs. In Proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99).

Mobasher, B., H. Dai, T. Luo, Y. Sung, and J. Zhu: 2000a, Integrating Web Usage and Content Mining for More Effective Personalization. In Proceedings of the International Conference on E-Commerce and Web Technologies (ECWeb2000). Greenwich, UK, pp. 165-176

Mobasher, B., R. Cooley, and J. Srivastava: 2000b,.Automatic Personalization based on Web Usage Mining. Communications of the ACM. 43 (8), pp. 142-151.

Kobsa, A., J. Koenemann and W. Pohl: 2001, Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. The Knowledge Engineering Review 16 (2), pp. 111-155

Kobsa, A: 2001, Tailoring Privacy to User's Needs. Invited Keynote, 8th International Conference on User Modeling, Sonthofen, Germany

Ardissono, L., A. Goy, R. Meo, G. Petrone, L. Console, L. Lesmo, C. Simone and P. Torasso: 1999. A configurable system for the construction of adaptive virtual stores. World Wide Web (WWW), 2(3), pp. 143-159

Ardissono, L. and P. Torasso: 2000, Dynamic user modeling in a Web store shell. In Proceedings of the 14th Conference ECAI. Berlin, Germany. Pp. 621-625

Anderson, C. R., P. Domingos, and D. S. Weld: 2001b, Personalizing Web Sites for Mobile Users. In Proceedings of the 10th World Wide Web Conference (WWW10).