OTTO VON GUERICKE
UNIVERSITÄT
MAGDEBURG

INF

FAKULTÄT FÜR
INFORMATIK

DKE
Data &
Knowledge
Engineering
Group

Adaptivity in Audio and Music Retrieval

Course Overview

Andreas Nürnberger and Sebastian Stober

Data & Knowledge Engineering Group, Faculty of Computer Science
Otto-von-Guericke-Universität Magdeburg, Germany

Email: andreas.nuernberger@ovgu.de, sebastian.stober@ovgu.de

Instructors

- **Andreas Nürnberger** and **Sebastian Stober**
- Data & Knowledge Engineering Group, Faculty of Computer Science
- Otto-von-Guericke-Universität Magdeburg, Germany
- Emails:
  - andreas.nuernberger@ovgu.de
  - sebastian.stober@ovgu.de
- Web: http://www.dke.ovgu.de/

Outline

- Day 1: Adaptation and Personalization: Concepts and Challenges
- Day 2: Adaptive Music Retrieval: An Overview
- Day 3: Adaptive Hierarchies: Constrained Clustering and Utility
- Day 4: Adaptive Similarity
- Day 5: User Interfaces and Gamification: Design and Evaluation

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF FAKULTÄT FÜR INFORMATIK

DKE Data & Knowledge Engineering Group

Adaptivity in Audio and Music Retrieval

Course Overview

Andreas Nürnberger and Sebastian Stober

Data & Knowledge Engineering Group, Faculty of Computer Science
Otto-von-Guericke-Universität Magdeburg, Germany

Email: andreas.nuernberger@ovgu.de, sebastian.stober@ovgu.de

Instructors

- **Andreas Nürnberger** and **Sebastian Stober**
- Data & Knowledge Engineering Group, Faculty of Computer Science
- Otto-von-Guericke-Universität Magdeburg, Germany
- Emails:
  - andreas.nuernberger@ovgu.de
  - sebastian.stober@ovgu.de
- Web: http://www.dke.ovgu.de/

Outline

- Day 1: Adaptation and Personalization: Concepts and Challenges
- Day 2: Adaptive Music Retrieval: An Overview
- Day 3: Adaptive Hierarchies: Constrained Clustering and Utility
- Day 4: Adaptive Similarity
- Day 5: User Interfaces and Gamification: Design and Evaluation

OTTO VON GUERICKE UNIVERSITÄT MAGDEBURG

INF

FAKULTÄT FÜR INFORMATIK

DKE
Data &
Knowledge
Engineering
Group

Adaptivity in Audio and Music Retrieval

Adaptation and Personalization: Concepts and Challenges

Andreas Nürnberger and Sebastian Stober

Data & Knowledge Engineering Group, Faculty of Computer Science
Otto-von-Guericke-Universität Magdeburg, Germany

Email: andreas.nuernberger@ovgu.de, sebastian.stober@ovgu.de

Outline

- **Day 1: Adaptation and Personalization: Concepts and Challenges**
- Day 2: Adaptive Music Retrieval: An Overview
- Day 3: Adaptive Hierarchies: Constrained Clustering and Utility
- Day 4: Adaptive Similarity
- Day 5: User Interfaces and Gamification: Design and Evaluation

Overview (Day 1)

- **Motivation**
- Systems: Digital Libraries and Multimedia
  - Current Visions (based on DELOS)
  - General Comments
- User Profiles and Profiling
- Applications and Algorithms

General Motivation

- Overview of adaptation and personalization approaches in more general

- Point out relations to research in
  - Data Bases and Digital Libraries
  - Machine Learning
  - Human Computer Interaction (HCI)

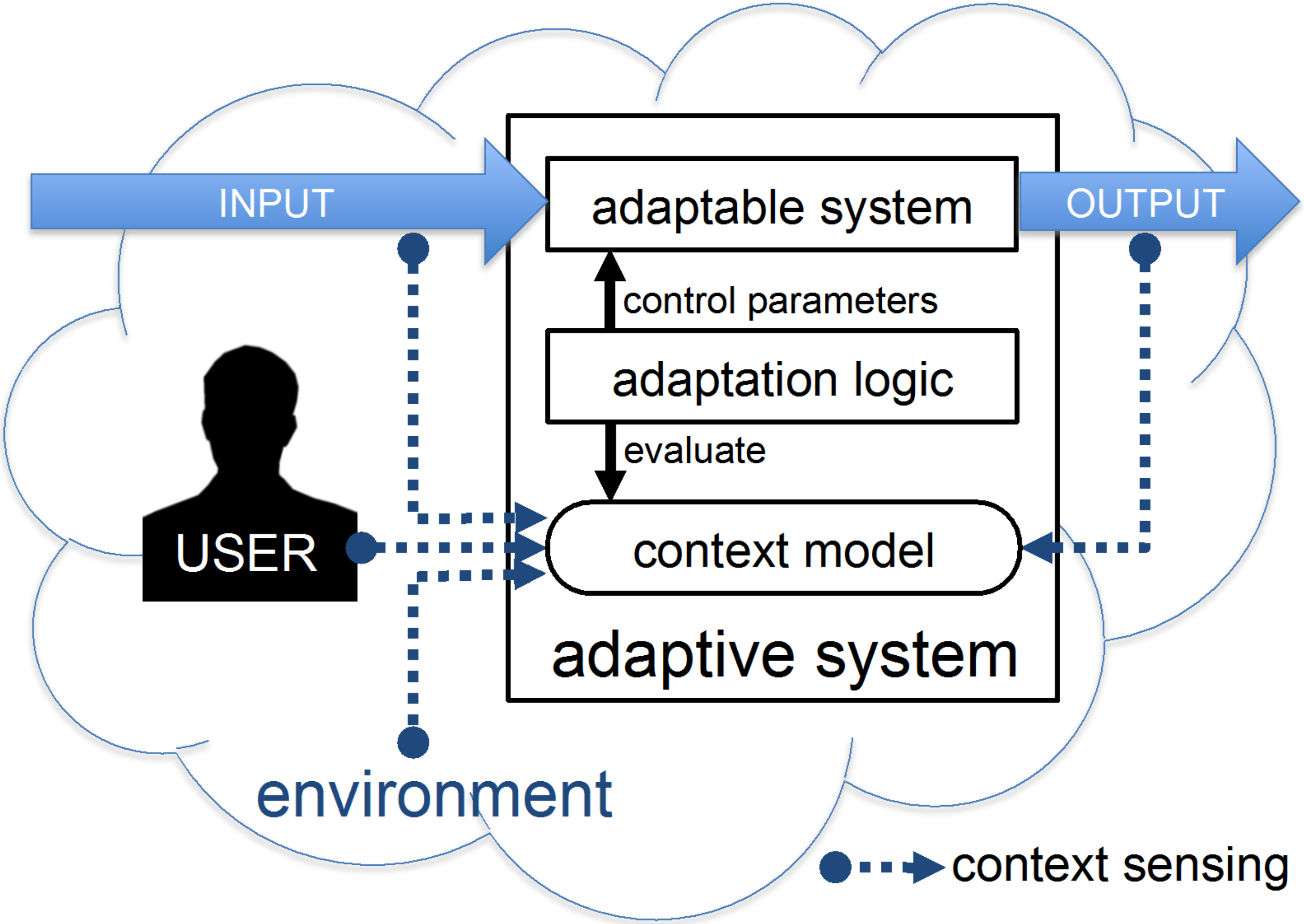
- Give specific examples of
  - possible applications
  - concrete applications and algorithms

  in multimedia (retrieval) systems

Adaptive Systems – Definitions

- *behavior*:
  - (set-valued) input/output (I/O) function of a system
  - does not require knowledge about system internals
- *adaptable system*:
  - provides means to change its behavior
- *adaptation*:
  - change of internal system structure (invisible) and behavior (visible)
- *context*:
  - (operational) environment,
  - user context
  - data (i.e., input/output values and their characteristics)

Adaptable → Adaptive System



Adaptive Systems

A system is (user and/or context) *adaptive* iff

1) it behaves different in different contexts given the same input [based on Broy et al. ‘09]

AND

2) the respective adaptation (i.e., the difference in behavior) is goal-driven in that it aims to optimize the system’s behavior in the given context according to some pre-defined measure.

Personalization (Definition)

- What does “personalization” mean?
  - “tailoring a consumer product, electronic or written medium to a user based on personal details or characteristics they provide“ (Wikipedia)
- In HCI:
  - Adapt a (software) interface and the information exchange between the user and the computer in order to improve the efficiency and effectiveness for a specific task.
- Different methods:
  - manual adaptation / configuration (*customization*)
  - automatic adaptation

Personalization (Facets)

- Different facets:
  - Adaptation of presentation (change style/ format)
  - Adaptation of structure (navigation and/or orientation support)
  - Adaptation of content (show/hide/filter content)
  - Query disambiguation (with respect to usage context)
  - Result/content processing (structuring and/or visualization)

Personalization (HCI approach)

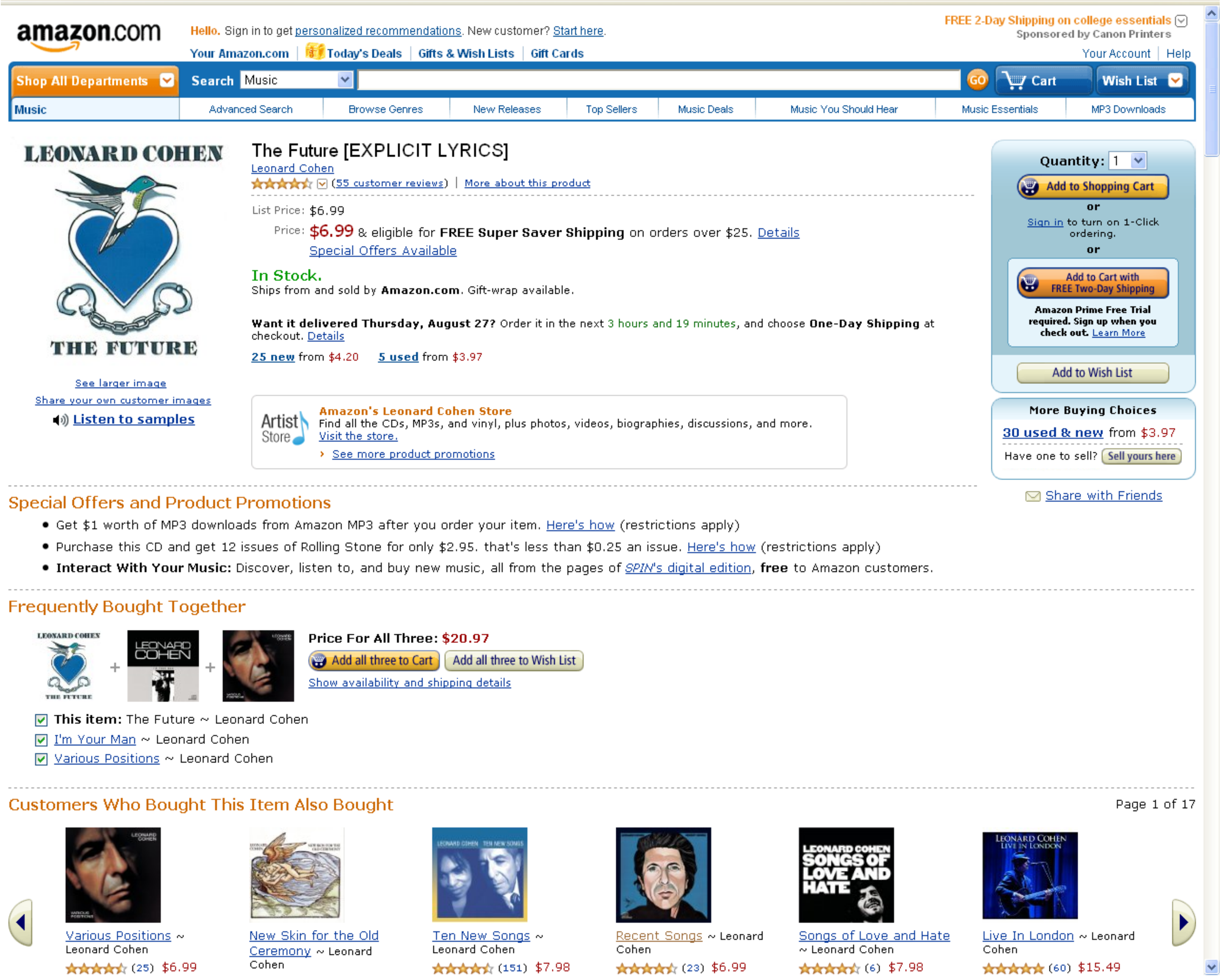
- In HCI automatic adaptation usually performed by some
  - agent that
  - observes the user during
  - interaction with the system and that
  - adapts (parts of) the interface based on
  - (learned or pre-defined) adaptation rules that use information extracted from
  - context,
  - visualized/accessed data and/or
  - user feedback.

Personalization (Example: Web shops)

- Goals: Attract consumers to ecommerce websites by
  - personalized and less obtrusive interactions with consumers
  - minimize user interactions by reduced number of search steps when searching for products
  - personalized recommendations

- Amazon: Recommendation on several levels, e.g.:
  - Entry page: Recommendation based on prior purchases
  - Product page (varies depending on product):
  - “Customers Who Bought This Item Also Bought”
  - “Frequently Bought Together”
  - “What Do Customers Ultimately Buy After Viewing This Item?”

Personalization (Example: Web shop Amazon)



Hello. Sign in to get personalized recommendations. New customer? Start here.

FREE 2-Day Shipping on college essentials

Sponsored by Canon Printers

The Future [EXPLICIT LYRICS]

Leonard Cohen

★★★★☆ (55 customer reviews) | More about this product

List Price: $6.99

Price: $6.99 & eligible for **FREE Super Saver Shipping** on orders over $25. Details

Special Offers Available

In Stock.

Ships from and sold by **Amazon.com**. Gift-wrap available.

Want it delivered Thursday, August 27? Order it in the next 3 hours and 19 minutes, and choose **One-Day Shipping** at checkout. Details

25 new from $4.20   5 used from $3.97

See larger image

Share your own customer images

Listen to samples

Amazon's Leonard Cohen Store

Find all the CDs, MP3s, and vinyl, plus photos, videos, biographies, discussions, and more. Visit the store.

› See more product promotions

or

Sign in to turn on 1-Click ordering.

or

Amazon Prime Free Trial required. Sign up when you check out. Learn More

More Buying Choices

30 used & new from $3.97

Have one to sell?

Share with Friends

Special Offers and Product Promotions

- Get $1 worth of MP3 downloads from Amazon MP3 after you order your item. Here's how (restrictions apply)
- Purchase this CD and get 12 issues of Rolling Stone for only $2.95. that's less than $0.25 an issue. Here's how (restrictions apply)
- **Interact With Your Music:** Discover, listen to, and buy new music, all from the pages of *SPIN*'s digital edition, **free** to Amazon customers.

Frequently Bought Together

Price For All Three: $20.97

Show availability and shipping details

☑ **This item:** The Future ~ Leonard Cohen

☑ I'm Your Man ~ Leonard Cohen

☑ Various Positions ~ Leonard Cohen

Customers Who Bought This Item Also Bought

Various Positions ~ Leonard Cohen ★★★★☆ (25) $6.99

New Skin for the Old Ceremony ~ Leonard Cohen

Ten New Songs ~ Leonard Cohen ★★★★☆ (151) $7.98

Recent Songs ~ Leonard Cohen ★★★★☆ (23) $6.99

Songs of Love and Hate ~ Leonard Cohen ★★★★☆ (6) $7.98

Live In London ~ Leonard Cohen ★★★★★ (60) $15.49

Personalization (Examples)

- Search engine rankings
  - User specific result set ranking
  - e.g. Google personalized search
    - Requires Google login and toolbar in order to collect information about user queries and pages visited
- Personalized News
  - Filter news based on personal interests
  - Allow to customize the news provided *on the web site*
    - e.g. Findory, reddit.com
  - Filter *arbitrary* RSS feeds
    - e.g. SearchFox, LeapTag
- Personalized Start Pages
  - Feed filtration for customized page
    - e.g. Netvibes, Pageflakes

Personalized Interaction (Summary)

- User specific adaptation of
  - user interface (e.g. layout)
  - information structure
  - data visualization

  before, during and after interaction of the user with the system

- Required:
  - *User profiling*:
    - logging and analyzing user interactions
    - analyzing structuring behavior of individual user
  - Information about usage context
  - Expressive features and metadata of multimedia objects

Multimedia Retrieval

- Searching for text works well for most ad hoc queries (using standard search engines)
- What about searching for images, sound, video?
- For example, assume the following search interest:

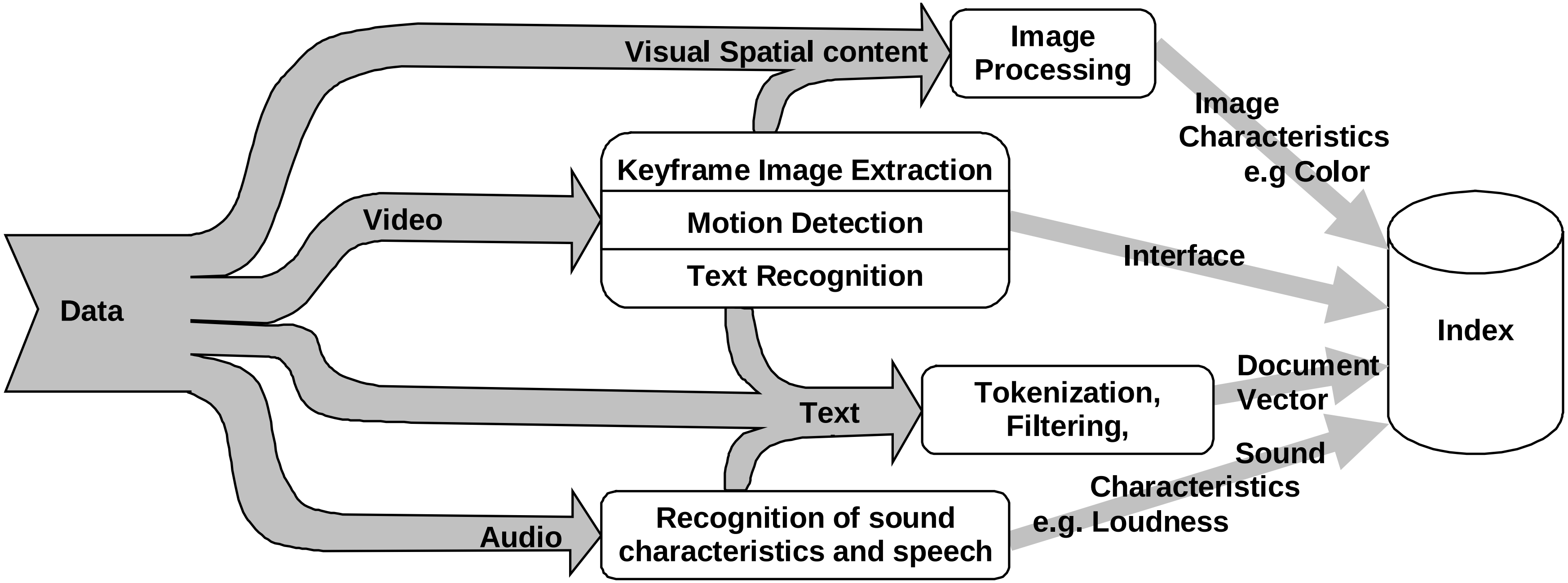
  „I would like to get the video sequence with a surfer at the beach in Lisbon I have seen some days ago in some news channel.“
- Required:
  - Semantic content description of video sequences
  - Meta-data about time of broadcast and category
  - Appropriate query language

Multimedia Indexing

- Goals: Assign or extract descriptive features allowing for retrieval, navigation and browsing
- Major types of multimedia data:
  - Text
  - Still Image
  - Sound (Music, Voices, Noise, …)
  - Video
    - Soundtrack (possibly in different languages)
    - Series of images (usually compressed data)
    - Subtitles (possibly in different languages)

Example: Video Indexing

- Segment video stream into shots
- Extract from each shot
  - Descriptive keyframe (→ still image indexing)
  - Sound track, subtitles, …



Multimedia Personalization

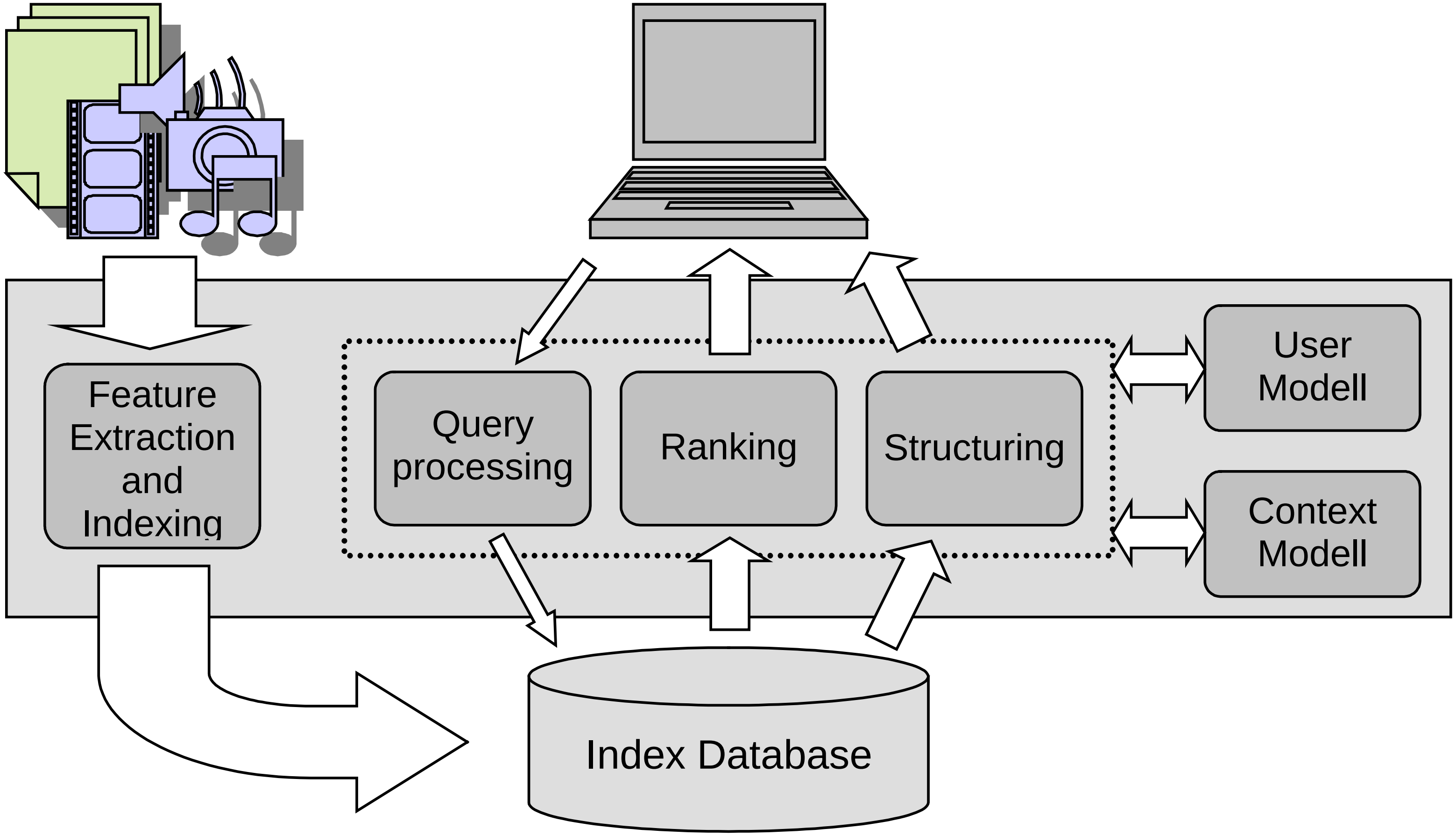
- Integrated user support for
  - search,
  - exploration
  - organization

  of digital (multimedia) collections

- Current approaches are usually limited to individual aspects like
  - search (possibly personalized ranking), or
  - rudimentary exploration support (identical for all users!)
  - simple visualization

Multimedia Retrieval Systems

- Basic components of a personalized retrieval system



Overview (Day 1)

- Motivation
- **Systems: Digital Libraries and Multimedia**
  - **Current Visions (based on DELOS)**
  - **General Comments**
- User Profiles and Profiling
- Applications and Algorithms

DELOS

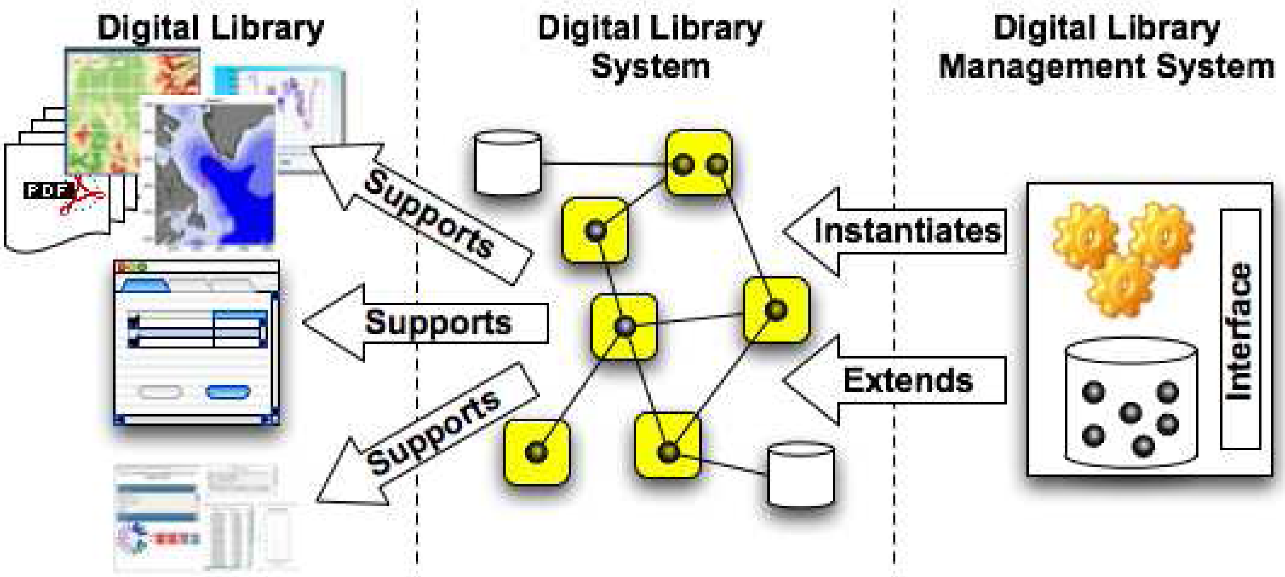
- Network of Excellence on Digital Libraries that was partially funded by the European Commission (http://www.delos.info/)
  - Follow ups: DL.org Digital Library Interoperability, Best Practices and Modelling Foundations (http://www.dlorg.eu/)
- Joint program of activities aimed at integrating and coordinating the ongoing research efforts of the major European teams working in Digital Library-related areas.
- Main objective and goal was to develop the next generation of Digital Library technologies, based on sound, comprehensive theories and frameworks for the life-cycle of Digital Library information.

DL as seen by DELOS

- Digital Library (DL)
  - Organisation (possibly virtual) that collects, manages and preserves for the long term **rich digital content**, and offers to its **user** communities specialised **functionality** on that content, of measurable **quality** and according to codified policies.
- Digital Library System (DLS)
  - Software system based on a defined (possibly distributed) **architecture** providing all functionality required by a particular Digital Library.
  - Users interact with a Digital Library through it.
- Digital Library Management System (DLMS)
  - Generic software system that provides the appropriate software infrastructure
    - to produce and administrate a DLS incorporating the functionalities considered fundamental for Digital Libraries and
    - to integrate additional software offering, more refined, specialized or advanced functionality.

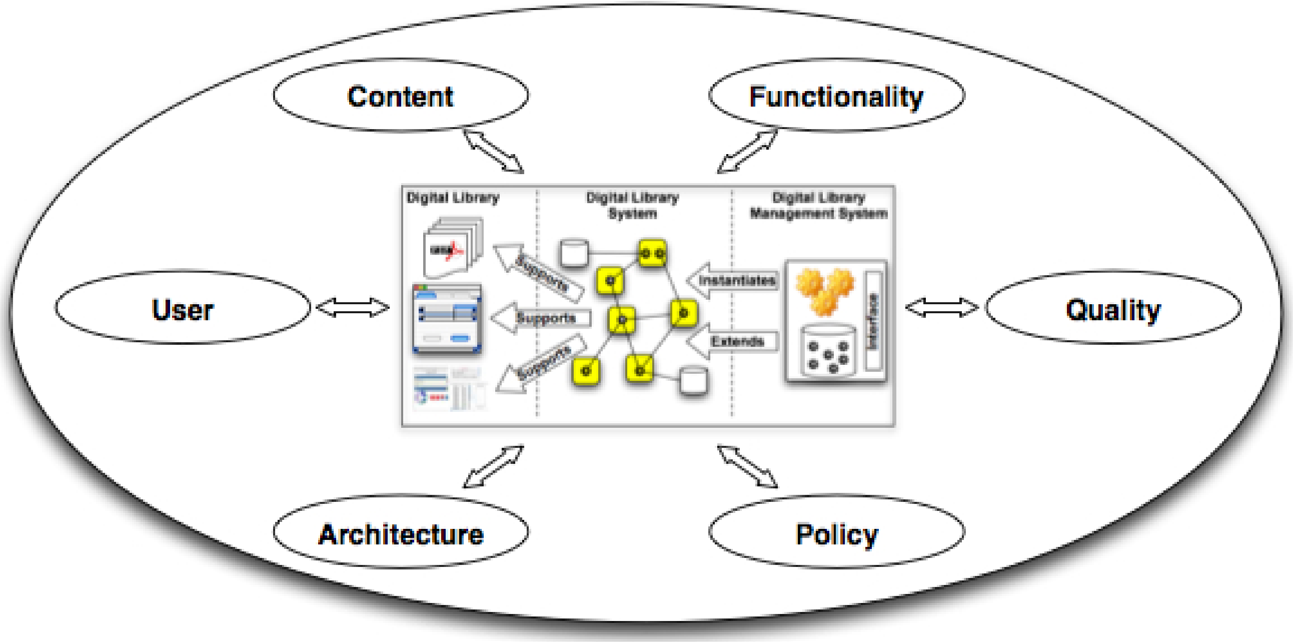
DELOS Tasks

- Development of a Digital Library Reference Model that is designed to meet the needs of the next-generation systems. (More details in the following…)
- Development of a globally integrated prototype implementation of a Digital Library Management System, which should serve as a concrete partial implementation of the reference model.



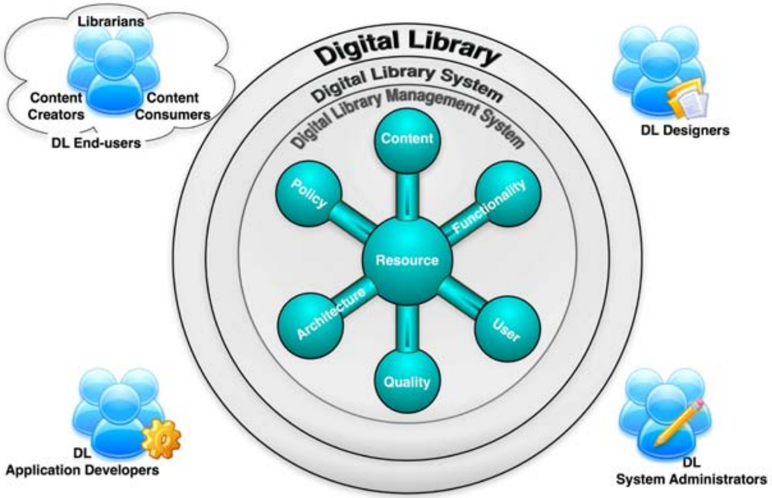
DELOS Reference Model

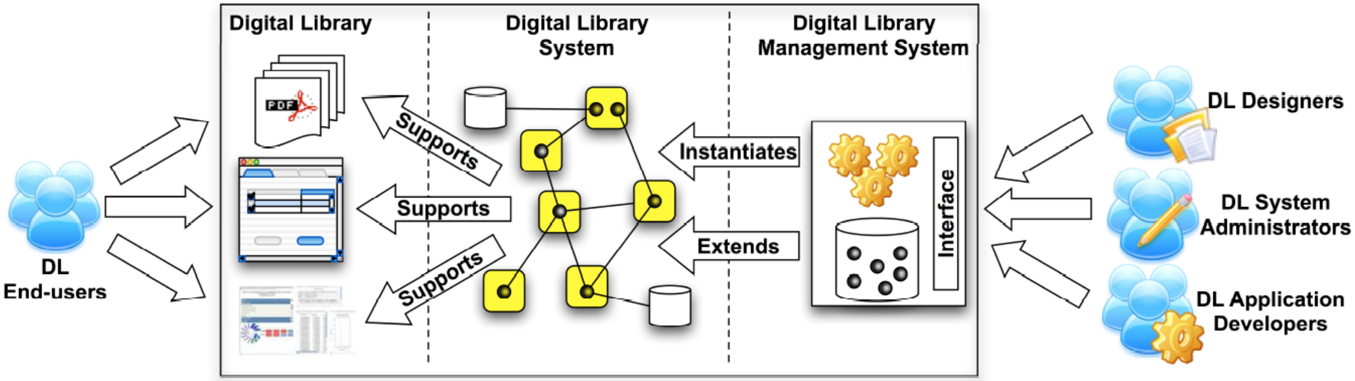
- A formal and conceptual framework describing the characteristics of this particular type of information system.
  - The model exploits the understanding of the architecture and functionality expected from an operational DLMS.
  - Model identifies and characterizes key concepts of a DLMS, such as the information space, documents handled, user profile, services, architecture, etc.



DELOS Reference Model: User

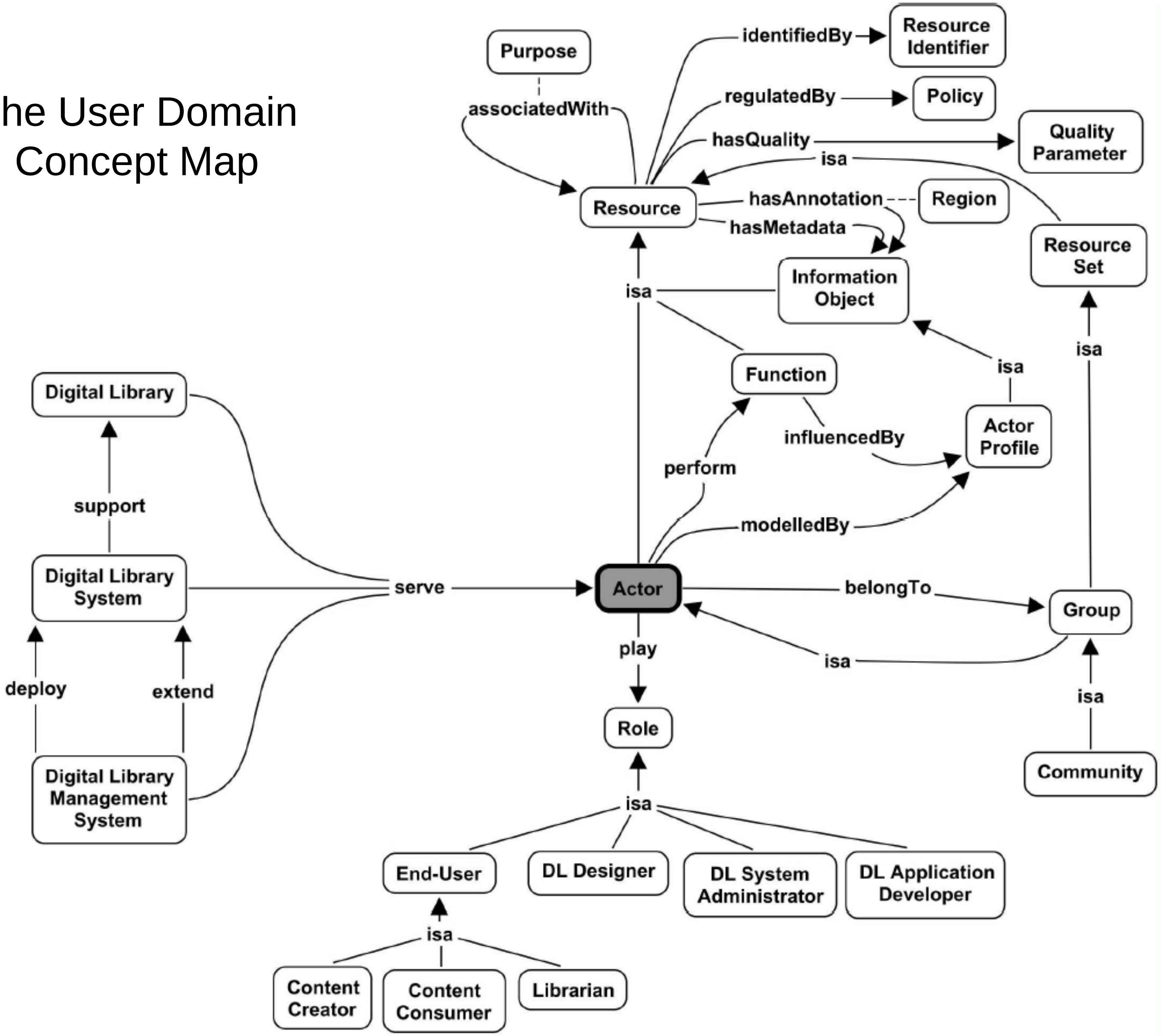
- User is the root for concepts like roles, communities, profiles, etc., that represent aspects of DL users
- Different actors:
  - DL End-users
    - Content creators
    - Librarians
    - Content Consumers
  - DL Designers
  - DL System Administrator
  - DL Application Developers





DELOS Reference Model: User

The User Domain
Concept Map



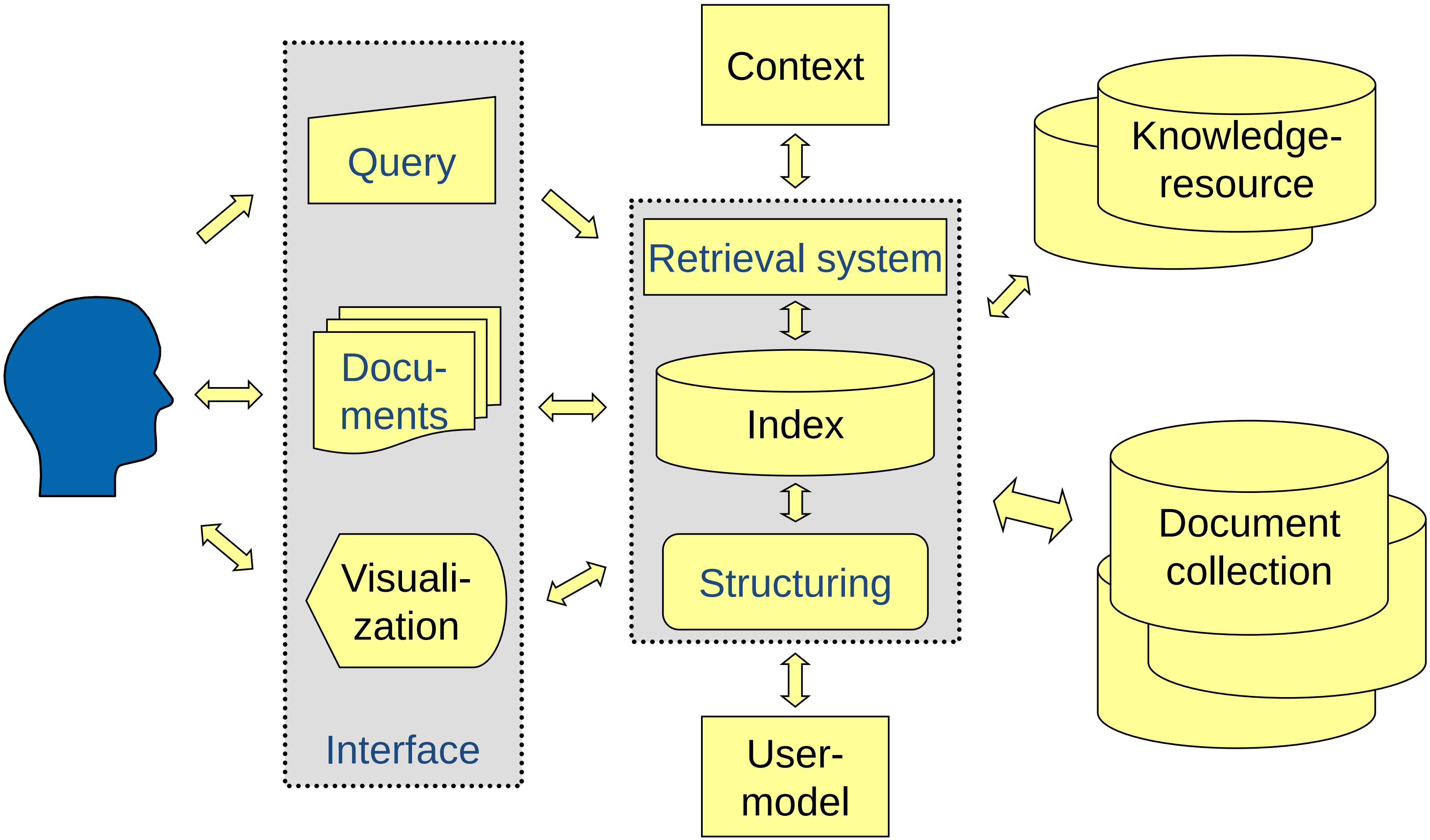
DELOS Reference Model (Summary)

- Reference Model provides nice overview of
  - DL concepts
  - DL components
  - DL actors
    - actors with quite different requirements!
  - Relations between all these elements
    - allows to derive facets and dependencies of a user model
- Missing:
  - What information is really necessary to describe the needs and interests of users?
  - How do we obtain and store this information?
- In the following we focus on content consuming end users in order to discuss these aspects in more detail…

Overview (Day 1)

- Motivation
- Systems: Digital Libraries and Multimedia
- **User Modeling: Profiles and Profiling**
  - Definitions
  - Profile types
  - Profile content
  - Profile structure
  - Profile acquisition
  - Profile classification and grouping
- Applications and Algorithms

Structure of considered systems



User Modeling (Definitions)

- User Modeling
  - Sub-area of human-computer interaction, in which cognitive models of human users are developed, including modeling of their skills and declarative knowledge.
- User Profile
  - The result of a user modeling process may be stored in a user profile.
- User Profiling
  - The content of a user model may be obtained/extracted via user profiling methods:
    - e.g. logging user behaviour and analyzing log files and related objects/ressources (using statistical and machine learning approaches) to derive user characteristics

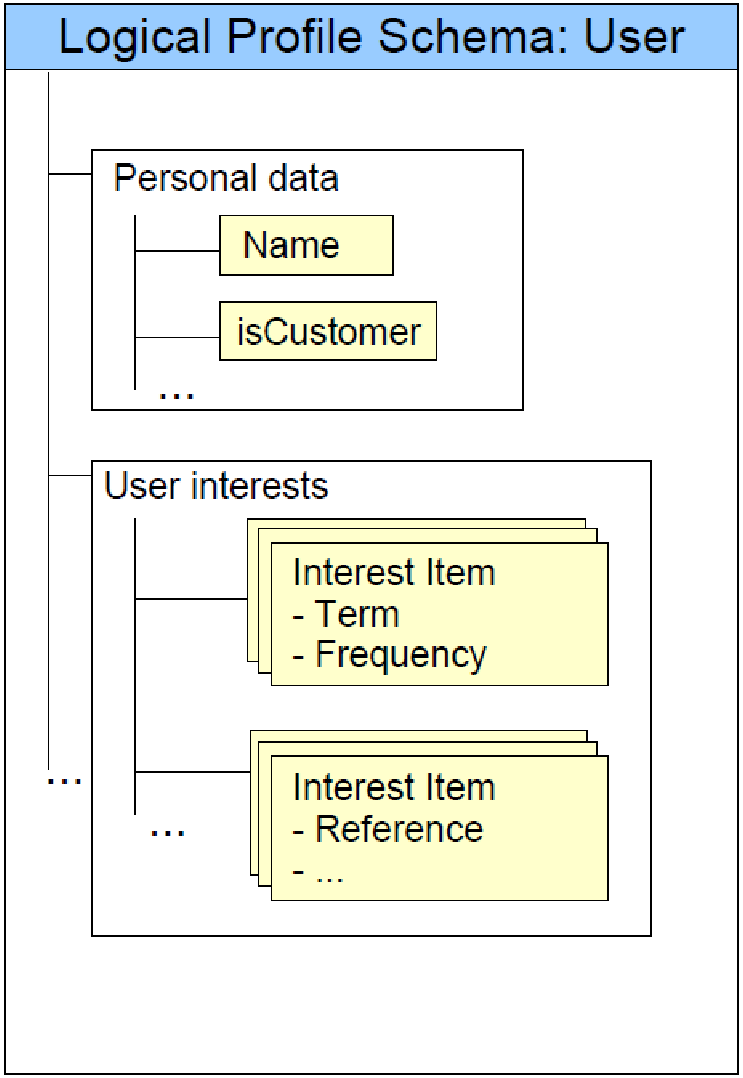
(Definitions based on: http://en.wikipedia.org/wiki/User_modeling)

User profiles (Profile types)

- Individuality versus generality
  - User profile
    - profile represents single user (most specific)
  - Group profile
    - profile represents group of (similar) users
  - Additional knowledge resources
    - “general” knowledge
    - usable for all users (e.g. ontologies, category hierarchy for query disambiguation, multilingual dictionaries, …)

User profiles (Profile content)

- The content of a user profile can be split in different sub categories, e.g.:
  - Personal data (most specific)
  - User / group interests
  - Browsing behavior
  - Additional knowledge resources



User profiles (Personal Data)

- “Personal data” of a user profile usually contains the following information:
    - Static data
        - User identification
        - User properties (e.g. age, employee/customer, languages, etc.)
        - Access rights
        - …
    - Dynamic data
        - Context information
            - User location
            - Interface
                - type (e.g. PC or PDA)
                - Language of interface (browser, search language)
                - …
        - Login count
        - …

User profiles (User/Group interests)

- Information about User/Group interests covers information about queries and accessed objects:
  - Query terms (+ usage frequency)
  - Items / Item references (+ ratings)
  - Extracted content from items (+ ratings)
    - e.g. a list of words extracted from visited documents
  - long-term and/or short-term interests
  - Ontology of terms and their possible term rewritings [KouIoa05]
  - …

User profiles (Browsing behavior)

- Browsing behavior related information is especially important to analyze the order (and dependency) of objects accessed by the user.
- Basic elements are:
  - Visited pages (+rating, e.g. page interest value)
  - Graph-structure of accessed pages like Web access graph [Cha00]
    - Pages as vertices containing the access frequency
    - Page change frequency or page associations as edges between two vertices

User profiles (Additional knowledge)

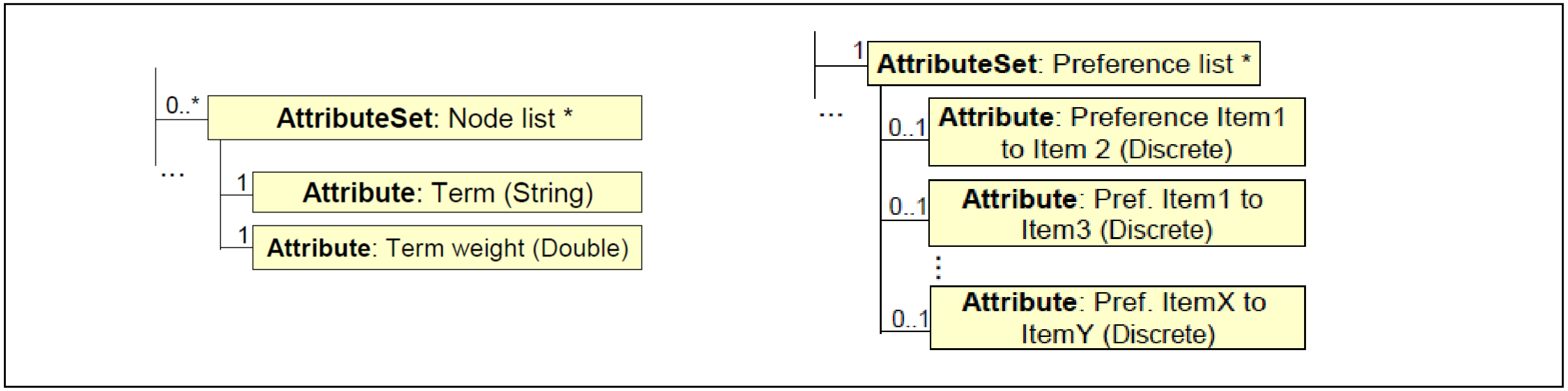
- Additional knowledge resources can provide valuable knowledge in order to disambiguate queries or categories information objects (domain specific information for a user/group).
- Basic elements are:
  - Category hierarchies
    - e.g. Open Directory Project, Wikipedia
  - Semantic ontologies
    - e.g. linguistic/lexical resources like WordNet, EuroWordNet
  - Linked Data
    - e.g. Linking Open Data (LOD) Project

User profiles (Profile structure)

- Information can be stored in different data structures
  - Flat profiles
  - Hierarchical profiles
  - Graph based profiles
  - Semantic profiles
- The structure selected for an application depends on the complexity of the information that needs to be stored and might consists of a combination of different structures.
- In the following, some examples…

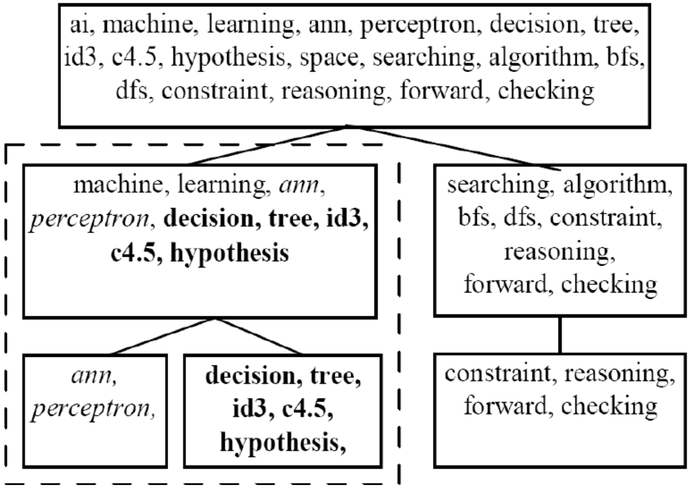
User profiles (Profile structure: Flat profile)

- Flat profiles store information of a user in a set like structure:
  - Attribute-value pairs
    - List of user-preferred items (+ranking)
  - Vector-based profile
    - Vector of weighted item or terms
    - Representing single document [Cha00]
    - Representing cluster centers of documents [SiMoGa04]
  - Preference Lists
    - Attributes define preference ranking of a user, e.g.


0..*
AttributeSet: Node list *
...
1
Attribute: Term (String)
1
Attribute: Term weight (Double)
1
AttributeSet: Preference list *
...
0..1
Attribute: Preference Item1 to Item 2 (Discrete)
0..1
Attribute: Pref. Item1 to Item3 (Discrete)
⋮
0..1
Attribute: Pref. ItemX to ItemY (Discrete)

User profiles (Profile structure: Hierarchical profile)

- Hierarchical interest profile (1)
  - Hierarchical clustered user interest profile (UIH) [KimCha03, Chi04]
  - Input: Several user relevant documents
  - Structure:
    - Interests at different abstraction levels (the higher-level interests are more general, the lower-level more specific).
  - Algorithm: Divisive hierarchical clustering (DHC)



ai, machine, learning, ann, perceptron, decision, tree, id3, c4.5, hypothesis, space, searching, algorithm, bfs, dfs, constraint, reasoning, forward, checking

machine, learning, *ann*, *perceptron*, **decision, tree, id3, c4.5, hypothesis**

searching, algorithm, bfs, dfs, constraint, reasoning, forward, checking

ann, *perceptron*,

decision, tree, id3, c4.5, hypothesis,

constraint, reasoning, forward, checking

Example of a hierarchical cluster structure of user interests

User profiles (Profile structure: Hierarchical profile)

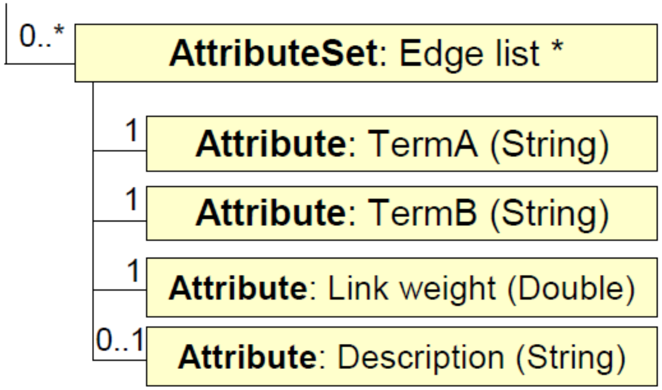
- Hierarchical interest profile (2)
  - Hierarchical structured user interest
  - Input:
    - word phrases (ordered sequence of one or more words) that are common to set of documents
  - Algorithm:
    - Suffix Tree Clustering (STC) [ZamEtz98], a linear time clustering algorithm that is based on identifying the phrases that are common to groups of documents.
    - At first, an inverted word index is build structured as a suffix tree (compact trie) strings of words and the related documents
    - Each node of the suffix tree represents a group of documents and a phrase that is common to all of them
    - Quite similar sub-trees are combined to a cluster

User profiles (Profile structure: Hierarchical profile)

- Hierarchical interest profile (3)
  - Build ontology-based user profile [TraGau04]
  - Input:
    - several user relevant documents and Open Directory Project (ODP) structure
  - Structures:
    - Algorithm structures documents according to ODP hierarchy
  - Algorithm:
    - The documents (tf×idf vector) and the concept (vector) are matched based on similarity

User profiles (Profile structure: Graph based profile)

- Graph structure
  - of accessed documents
    - [Cha00] Web Access Graph (WAG)
    - Structure:
      - nodes represent web pages and store, e.g., access frequency
      - edges represent association degree between two pages
    - Algorithm evaluates server logs
  - of term relations
    - e.g. in order to model user specific similarity, relation or dependency


0..*
AttributeSet: Edge list *
1
Attribute: TermA (String)
1
Attribute: TermB (String)
1
Attribute: Link weight (Double)
0..1
Attribute: Description (String)

User profiles (Profile structure: Semantic profile)

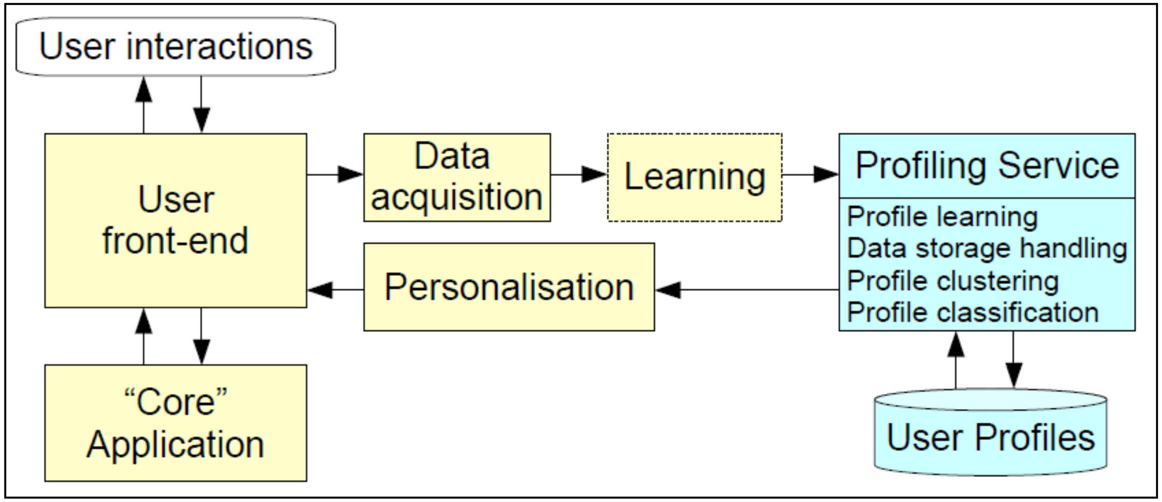
- RDF description
  - Example: Friend-of-a-friend (FOAF) Ontology
  - Describes persons, their activities and their relations to other people and objects.
  - Can be considered the first Social Semantic Web application, in that it combines RDF technology with 'Social Web' concerns.

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/"
   xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
   xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
   <foaf:Person rdf:about="#JW">
      <foaf:name>Jimmy Wales</foaf:name>
      <foaf:mbox rdf:resource="mailto:jwales@bomis.com" />
      <foaf:homepage rdf:resource="http://www.jimmywales.com/" />
      <foaf:nick>Jimbo</foaf:nick>
      <foaf:depiction rdf:resource="http://www.jimmywales.com/aus_img_small.jpg" />
      <foaf:interest rdf:resource="http://www.wikimedia.org" rdfs:label="Wikipedia" />
      <foaf:knows>
         <foaf:Person>
            <foaf:name>Angela Beesley</foaf:name>
         </foaf:Person>
      </foaf:knows>
   </foaf:Person>
</rdf:RDF>
```

Example of a FOAF profile (XML format)

User profiles (Profile acquisition)

- How to acquire the information?
  - Explicit
    - User has to provide information actively
  - Implicit
    - Information are automatically gathered during the interaction of a user with the system)
  - In most systems a combination of explicit and implicit methods are used.



Example of a general profiling architecture [Otto06]

User profiles (Profile acquisition: Explicit)

- User registration
  - User provides personal information, interests and knowledge level(s)
- User states search “objective”/search keywords at beginning of a session (information need)
- Explicit relevance feedback
  - Interest indicator (interesting or not) [Teo03]
  - Page evaluation (marks for relevance, comments: relevant, not relevant (known), no opinion, irrelevant) [BueDav01]
- Chat-robot for communication with customer
  - E.g. chat-robot from eBrain was used in the EU project COGITO which aimed at an agent-based interface for B-to-C applications

User profiles (Profile acquisition: Implicit)

- Log more general user (-system) interaction:
  - pages/objects accessed
  - printing or saving
  - bookmarking
  - widgets used
  - …
- Implicit relevance feedback
  - Clicked or not (decision bases on title and snippet/ summary) e.g. Toogle [Ruv03]
  - Time duration as indication of interest [WhRuJo02]
- Mining of log files to obtain “higher level” properties
  - Analysis of search paths (navigation)
  - Mining / grouping of objects accessed in order to derive topics of interest

User profiles (Profile acquisition: Technical Issues)

- Server-side:
  - Evaluation of Server logs, identify sessions, evaluate surf path, duration, etc.

- Client-side:
  - Modify interface, Browser with ActiveX Controls, JavaApplets, Plugins or a modified Browser
  - Logging of visited sites, duration per site, mouse movements, page scrolling, cut-and-paste operations, saving/printing of pages
  - Browser logs [Cha00]
  - Browsing history (Time + URL)
  - Bookmark lists
  - …

Finding user groups (1)

- Clustering on user type (user category, usage frequency, etc.)
- Clustering on user interests:
  - On user relevant items (e.g. movies)
  - Input:
    - list of user relevant documents (with rating)
  - Similarity:
    - correlation between item lists and ratings (Collaborative Filtering)
    - measure on fully/partially specified preference orders (Case Based recommendation) [HaHad03]
  - Clustering:
    - Hierarchical agglomerative (bottom-up) clustering (HAC)
    - Divisive (top-down) hierarchical clustering (DHC)
    - K-means (initiated with “stereotype” clusters)

Finding user groups (2)

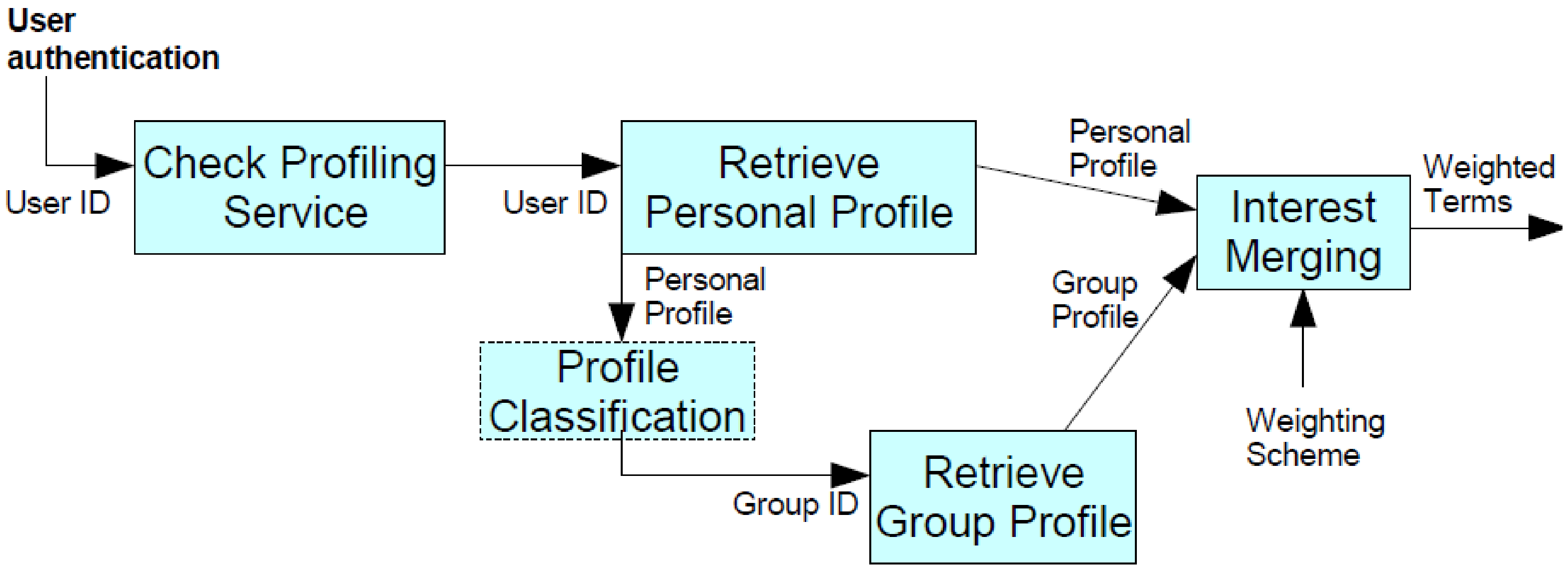
- Clustering on user interests (cont.):
  - On extracted terms (Content-based)
  - Input:
    - List containing weighted terms representing user interests
  - Similarity:
    - Similarity between term vectors
  - Clustering:
    - Hierarchical agglomerative (bottom-up) clustering (HAC)
    - Divisive (top-down) hierarchical clustering (DHC)
    - K-means (initiated with “stereotype” clusters)
- Clustering on the web access graph (page access frequency) [Cha00]
  - Cluster equal sub-graphs
  - Input: User’s WAG
  - Similarity: overlapping degree of graph
  - Clustering: HAC, DHC

User Classification

- Assignment of users to user groups
  - Collaborative Filtering (CF)
    - e.g. Pearson r correlation coefficient
- Case-based reasoning (CBR) / CB recommendation on partially preference orders [Had97]
  - Euclidean distance
  - Probabilistic distance
- Knowledge-based recommendation
  - Use general knowledge (like category hierarchy to find similar users)
- Rule-based approaches

Profile extension

- Group profiles could be used in order to derive additional (possibly relevant) information for a specific user
- Even if user is not yet assigned to a group, additional information could be derived by exploiting existing group (or user) profiles, e.g. [Otto06]



User Profiles (References 1)

User Profiling

[BueDav01] “METIORE: A Personalized Information Retrieval System“; David Bueno, Amos A. David; 2001

[KouIoa05] “A Unified User-Profile Framework for Query Disambiguation and Personalization”; Georgia Koutrika, Yannis Ioannidis; 2005

[LiYuMe02] “Personalized Web Search by Mapping User Queries to Categories“; Fang Liu, Clement Yu, Weiyi Meng; 2002

[PiScCaCoTu EdAdBr02] “Personalized Search”; J. Pitkow, H. Schütze, T. Cass, R. Cooley, D. Turnbull, A. Edmonds, Eytan Adar, Th. Breuel; 2002

[Ruv03] “Adapting to the User's Internet Search Strategy”; Jean-David Ruvini

[SeoZha00] “Learning User's Preferences by Analyzing Web-Browsing Behaviours”; Young-Woo Seo, Byoung-Tak Zhang; 2000

[ZhGrHa03] “Learning a Model of a Web User’s Interests”; Tingshao Zhu, Russ Greiner and Gerald Häubl; 2003

[Otto06] “Design of a Generic User Profiling Architecture and its Application in Customer Support”, Steffen Otto, Master’s Thesis, 2006.

User Profiles (References 2)

Clustering

[BaBiMo04] “Semi-supervised Clustering for Intelligent User Management“; Sugato Basu, Mikhail Bilenko, Raymond J. Mooney; 2004

[Cha00] “Clustering web user profiles: A non-invasive approach”; Philip K. Chan; 2000

[Chi04] “Transient User Profiling”; Ed H. Chi; 2004

[HaHad98] “Toward case-based preference elicitation: Similarity measures on preference structures”; Vu Ha, Peter Haddawy; 1998

[KimCha03] “Learning Implicit User Interest Hierarchy for Context in Personalization“; Hyoung R. Kim, Philip K. Chan; 2003

[SiMoGa04] “Inferring User’s Information Context: Integrating User Profiles and Concept Hierarchies“; A. Sieg, B. Mobasher, R. Burke; 2004

[TraGau04] “Improving Ontology-Based User Profiles”; Joana Trajkova, Susan Gauch; 2004

[ZamEtz98] “Web Document Clustering: A Feasibility Demonstration”; Oren Zamir, Oren Etzioni; 1998

Overview (Day 1)

- Motivation
- Systems: Digital Libraries and Multimedia
- User Modeling: Profiles and Profiling
- **Applications and Algorithms**
  - **Relevance feedback / query reformulation**
  - **Collaborative filtering**
  - **Personalized Structuring**

Query Reformulation

- Problem: How can a query be reformulated?
- Thesaurus Extension:
  - Terms are suggested that are similar to the query term
- Relevance Feedback:
  - Terms (and documents) are suggested based on documents that are marked as relevant

Assumptions to Relevance Feedback

- Given is a query *A*.
- Documents that are rated as relevant have similar properties (e.g., contain similar text or belong to similar genre).
- Documents that are rated as not relevant differ from relevant documents in some properties.
- Based on this properties, the query can be reformulated such that it narrows the document space.

Relevance Feedback

- Idea:
  - Modify existing query based on relevance judgments
    - Extract terms from relevant documents and add them to the (new) query

    and/or

    - re-weight the terms that are already in the query
  - Two main approaches for relevance judgments:
    - Automatic (pseudo-relevance feedback)
    - User selects relevant documents
  - User/System selects terms from a generated list that is based on relevance feedback

Relevance Feedback

- Usually
  - queries are expanded by new terms and
  - query terms are re-weighted

- Many strategies are possible
  - Usually terms from relevant document get positive weights and
  - Terms from non-relevant documents get negative weights.
  - Removed are only terms from non-relevant documents (! A document not marked as relevant need not to be irrelevant! May be a user was simply unable to decide.)

Rocchio Method

- Rocchio
  - Automatically re-weights terms
  - automatically inserts terms (from relevant documents)
  - We have to take care of negative weights!

- Rocchio is not a machine learning approach, but a heuristic
  - improves ranking (proved by evaluations)

- Most „new“ feedback-methods are based on this ideas

Relevance Feedback Summary

- Relevance feedback is an effective method for user-driven query modifications.
- Modification can be done based on direct or indirect user input.
- Modifications can be done based on previous inputs from individuals or groups.

Alternative Notions of Feedback

- Find people whose taste is similar to yours.
  - Will you like what they like?

- Follow a user’s actions.
  - Can this be used to predict what the user will want to see next?

- Track the behavior of many people.
  - Does this directly indicate what a good action is and what not?

Alternative Notions of Feedback

- Several different criteria should be considered:
  - Implicit vs. explicit judgments
  - Individual vs. group judgments
  - Static vs. dynamic topics
  - Similarity of items being judged vs. similarity of the judges themselves

Collaborative Filtering (social filt.)

- If Paul liked the book, I will like the book
- If you liked Star Wars, you will like Independence Day
- Rating is based on ratings of similar people
  - Ignores the content and therefore works with text, music, pictures etc.
  - But: initial users may bias ratings of future users!

	Sally	Bob	Chris	Lynn	Karen
Star Wars	7	7	3	4	7
Jurassic Park	6	4	7	4	4
Terminator II	3	4	7	6	3
Independence Day	7	7	2	2	?

Collaborative Filtering

- Example: Users rate musical artists from “like” to “dislike”, e.g..
  - 1 = dislike
  - 4 = ambivalent
  - 7 = like him/her very much
- Results in a normal distribution around 4
- However, what matters are single events!

- Nearest Neighbors Strategy: Find similar users and determine the (weighted) average of their ratings

Nearest Neighbor

- **Definition**:

  Given are a case base *CB*, a similarity measure *sim* and an object (problem) $p \in M$. $C \in CB$ (with mit $C=(m,c)$, *m* is an attribute and *c* the corresponding category) is a *nearest neighbor to p* iff:

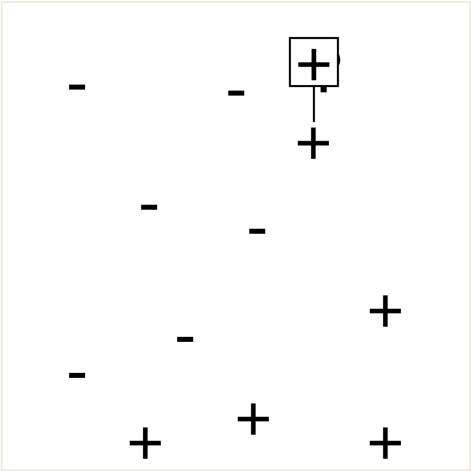
  $$\forall\ (m',c') \in CB : sim(p,m) \geq sim(p,m')$$

- The pair (*CB*,*sim*) defines by the *principle of the nearest neighbor* a classificator: the category of the nearest neighbor is assigned to the object $p \in M$.

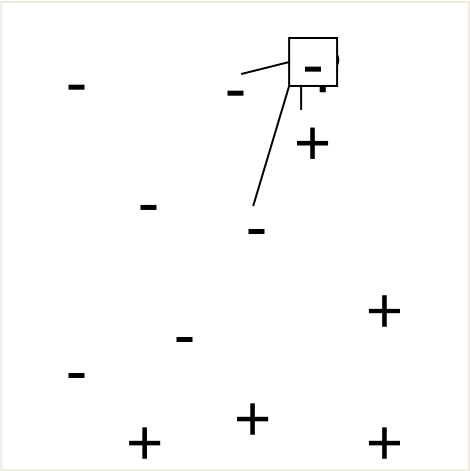
k-nearest Neighbors Algorithm

Variation of nearest neighbor approach:

- Use *k* nearest neighbors to improve categorization
- Problem: Determination of category is not always unambiguous.



„1-nearest neighbors“



„3-nearest neighbors“

k-nearest Neighbors Algorithm

Determination of category

Different approaches are possible, e.g.

- Choose most frequent category
- If an order is defined over the categories:
  - Determine category based on weighted average over all neighbor categories.
  - Weighting according to frequency or the respective similarity to the neighbor (using distance *d*, e.g. *sim*=1/(1+*d*)).
  - Example: *sim* is a similarity measure, *p* the object under observation, m_i the neighbors, c_i the categories and *k* the number of nearest neighbors. For *c** we define:

$$c^* = \frac{1}{\sum_{i=1}^{k} sim(p,m_i)} \sum_{i=1}^{k} c_i \cdot sim(p,m_i)$$

Ringo Collaborative Filtering

- Determine the similarity of users based on the Pearson r correlation:

  Weight is determined on the basis of the correlation between user *x* and user *y*:

$$sim(x,y) = \frac{\sum_i (R(x,i) - R(x))(R(y,i) - R(y))}{\sqrt{\sum_i (R(x,i) - R(x))^2 \sum_i (R(y,i) - R(y))^2}}$$

  while *R*(*x*,*i*) is the judgment of user *x* for the attribute *i* and *R*(*x*) is the average over all judgments of *x*.

  You get

  - 1 for very similar users,
  - 0 for no correlation,
  - -1 if user have “opposed” interest.

Social Filtering

- Ignores the content and only looks who judges objects similarly
- Works well on data related to “taste”
  - People are sometimes good at predicting about each others taste
- Does it work for IR?
  - Depends on specific application…

IR Concepts, Methods, Algorithms (Fundamentals)

[BaeRib99] R. Baeza-Yates, B. Ribeiro-Neto: Modern Information Retrieval, New York, NY: ACM Press; 1999.

[ManRagSch08] C. D. Manning, P. Raghavan, H. Schütze: Introduction to Information Retrieval, Cambridge University Press, 2008.

[Man Sch02] C.D. Manning, H. Schütze: Foundations of Statistical Natural Language Processing, The MIT Press, 2002.

[Hea09] Marti Hearst, Search User Interfaces, Cambridge University Press, 2009.

[FraBae92] Information Retrieval: William B. Frakes and Ricardo Baeza-Yates, Data Structures and Algorithms, Prentice-Hall, 1992.

The End

THANKS A LOT FOR LISTENING!