DAI-Labor
Distributed Artificial Intelligence Laboratory

Personalization and Context-awareness in Retrieval and Recommender Systems

Ernesto William De Luca

DAI-Lab / Berlin Institute of Technology

CC IRML
Information Retrieval
& Machine Learning

Outline

- About Me
- Information Management
- Personalized and Context-aware Information Management
- Conclusions

About me

Ernesto William De Luca

- Berlin Institute of Technology (TUB)
- DAI-Lab
- Head of the Competence Center for Information Retrieval and Machine Learning.
  (http://www.dai-labor.de)

- E-mail: ernesto.deluca@dai-labor.de

Berlin Institute of Technology (TUB)

Distributed Artificial Intelligence Lab (DAI-Lab)

DAI-Labor
Distributed Artificial Intelligence Laboratory

Prof. Sahin Albayrak

- Chair of Agent Technologies in Business Applications (AOT)
- 10 Post-docs
  50+ Phd Students
  50+ Students

CC ACT Agent Core Technologies
CC IRML Information Retrieval & Machine Learning
CC NEMO Network & Mobility
CC NGS Next Generation Services
CC SEC Security

CC IRML Information Retrieval & Machine Learning

Intelligent Information Management

- 3 PostDocs and 9 Full-time PhD students
  - Profile Management
  - Semantic Search
  - Personalization of search results, filtering, visualization
  - Information Classification
  - Topic Detection
  - Recommender Systems and Social Network Analysis

Outline

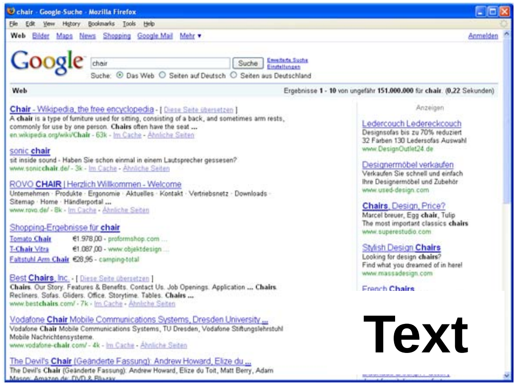
- About me
- Information Management
- Personalized and Context-aware Information Management
- Conclusions

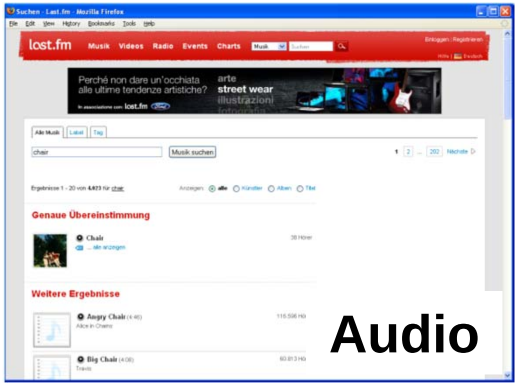
Information Management

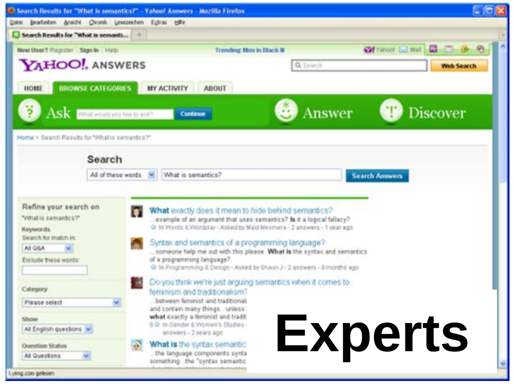
- Definition:
  - It is the **collection** and **management** of information
    - from one or more sources
    - and the **distribution** of that information to one or more audiences.
  - It means the **organization** of and **control** over the **structure**, **processing** and **delivery** of information.

  - Examples: Recommender and Retrieval Systems

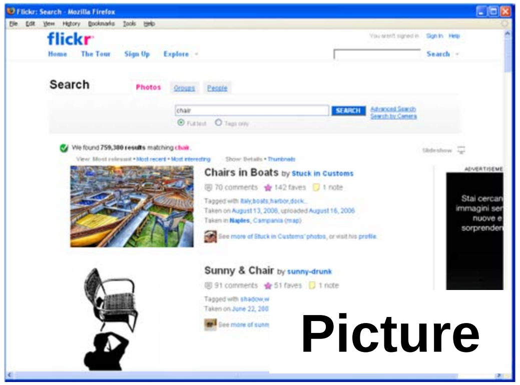
Information

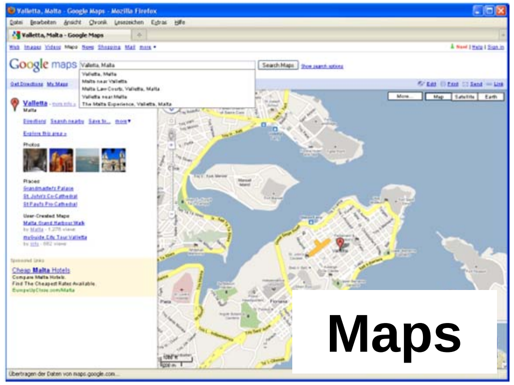






...

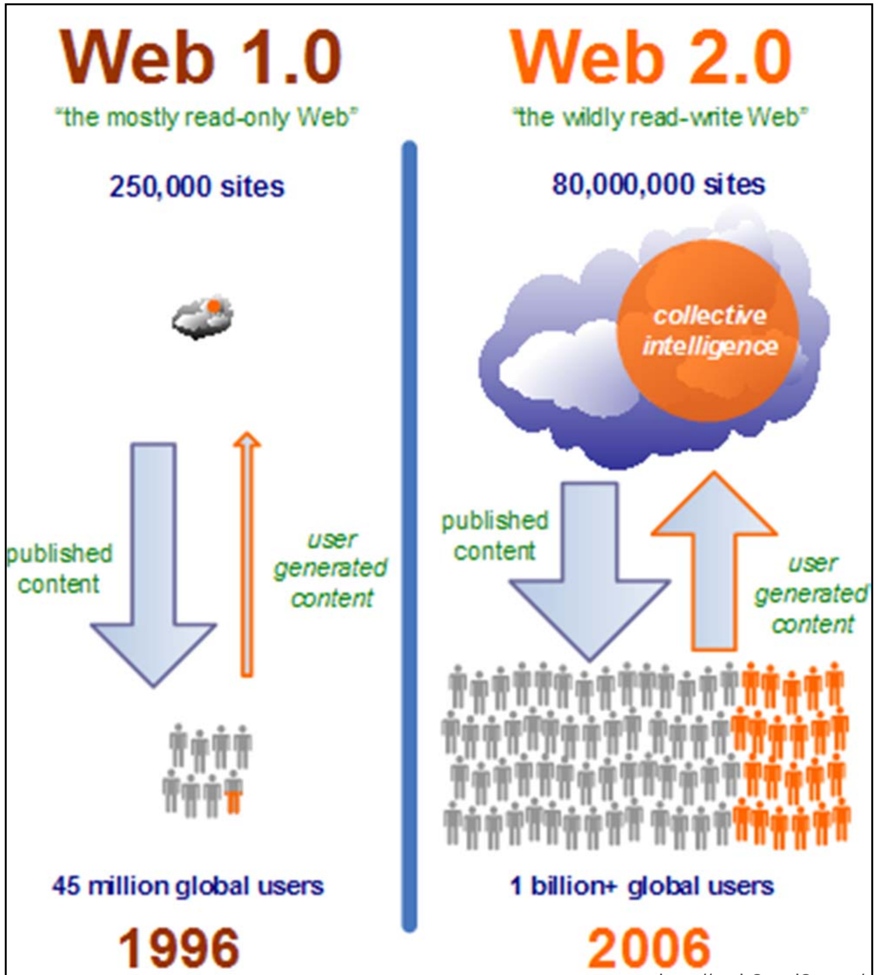




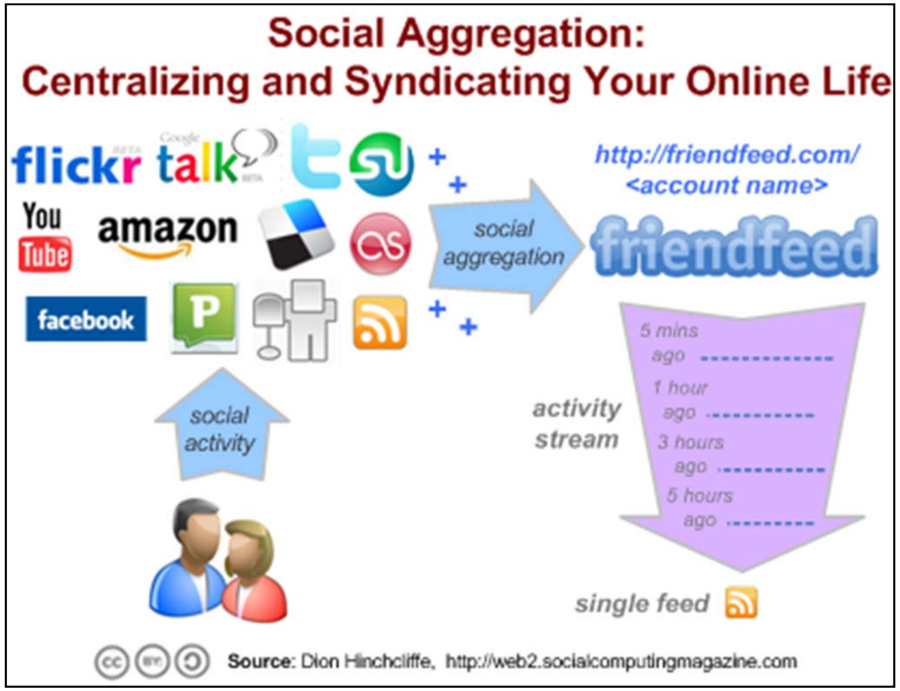
User Profiles

...

Structured Interaction and Retrieval



source: http://web2.wsj2.com/



How can we manage it?

Outline

- About me
- Information Management
- Personalized and Context-aware Information Management
  - Goals
  - Personalization
  - Context-awareness
- Conclusions
- Future Work

Outline

- About me
- Information Management
- Personalized and Context-aware Information Management
  - Goals
  - Personalization
  - Context-awareness
- Conclusions
- Future Work

Goals

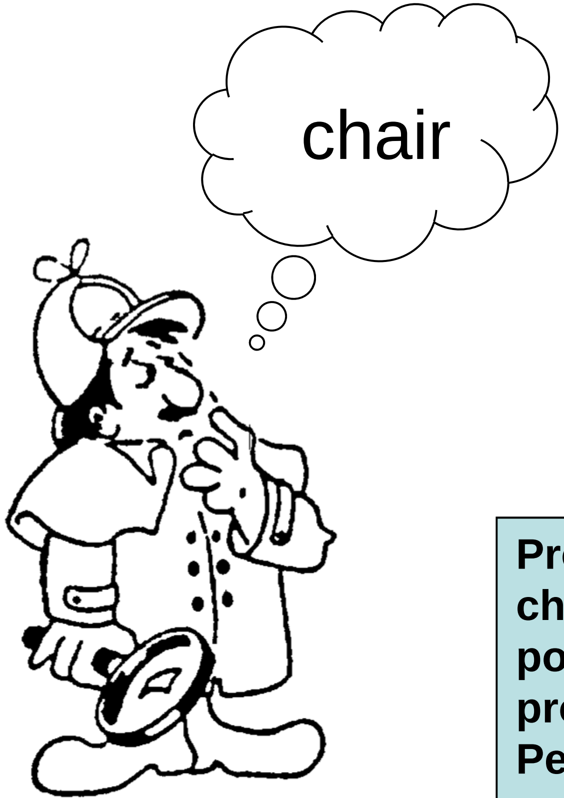
In Recommendation and Retrieval

- **Personalization**
  - to **understand** the **user**
  - to understand the **user needs**
  - to **find semantically-related content**
- **Context-awareness**
  - To identify features that influence the user’s (or item’s) **current situation**

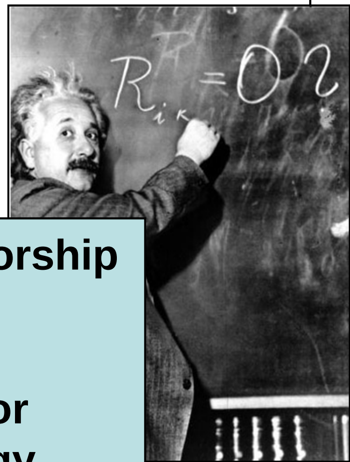
Outline

- About me
- Information Management
- Personalized and Context-aware Information Management
  - Goals
  - Personalization
    - Personalization view
    - SERUM Project
    - SPIM Workshop
  - Context-awareness
- Conclusions

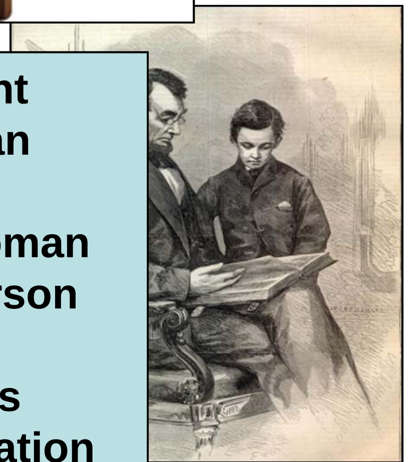
Semantic view



chair
Furniture
Support
back
armchair
…

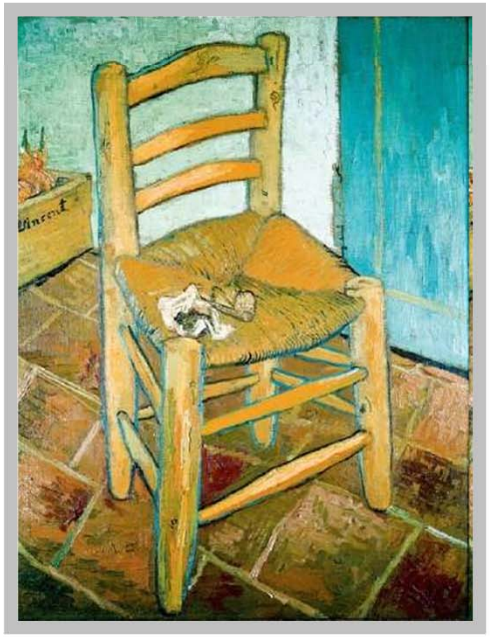


Professorship
chair
position
professor
Pedagogy
…



President
Chairman
Person
Chairwoman
chairperson
officer
meetings
Organization
…

User view



Personalization Views

Multilingual view

DAI-Labor
Distributed Artificial Intelligence Laboratory

Problems and Challenges

- Retrieval
  - How can we provide **individual** experience?
  - How can we help users in finding only **relevant** information?
- Recommendation
  - How can we give **personal** recommendations?
  - How can we **recognize** what the user wants to be recommended?

SERUM Project

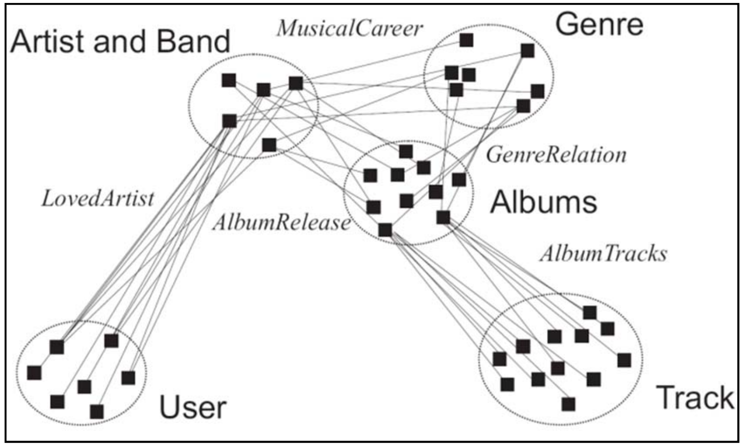
SEmantic Recommendation and Unstructured data Management (SERUM)

- Goal:
  - **Recommend news articles** based on the previous behavior of a user
- How:
  - **User behavior** is analyzed
  - **Semantic knowledge** is being linked to current news articles.
  - Algorithms were developed to **analyze** semantic information and user behavior

Supported by:
Federal Ministry of Economics and Technology
on the basis of a decision by the German Bundestag

Music and News Recommendation

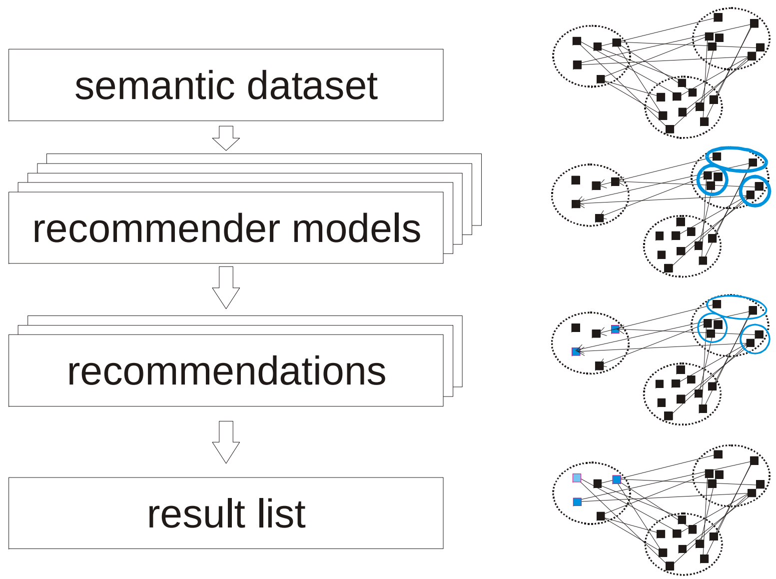
- We recommend
  - **semantically related entities of interest**, like similar artists or genres
  - best matching news article
  - music domain



- We combine
  - semantic/encyclopedic knowledge graph (5.5 mio. entities)
  - a large news dataset (7.2 mio. news)

Recommender Approach

- Recommendations are calculated by a **meta-recommender** combining results from different recommender agents.
  - Different models for different recommendation scenarios
  - For each request the best matching recommender models are selected
  - **Path-based/latent-model based recommenders**



User Profile Management (UPM)

serum

DAI-Labor
Distributed Artificial Intelligence Laboratory

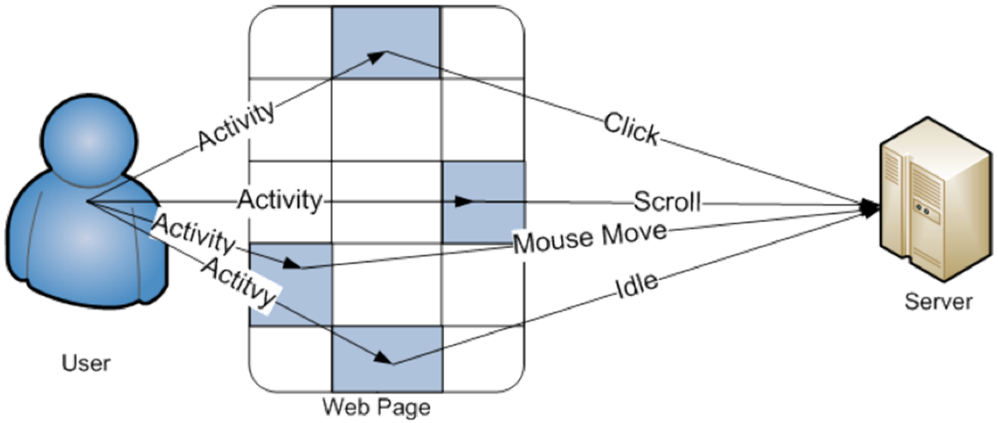
Our Goals:

- Understand **user interests / needs**
  - Tracking **user behaviour** on **dynamic websites**
  - Recommend **user-relevant information**

- Create **user profiles**
  - Tracking and understanding **relations** between content on dynamic websites
  - Including user interests / needs
  - Aggregation of different user profiles from different applications

UPM: Techniques Behind

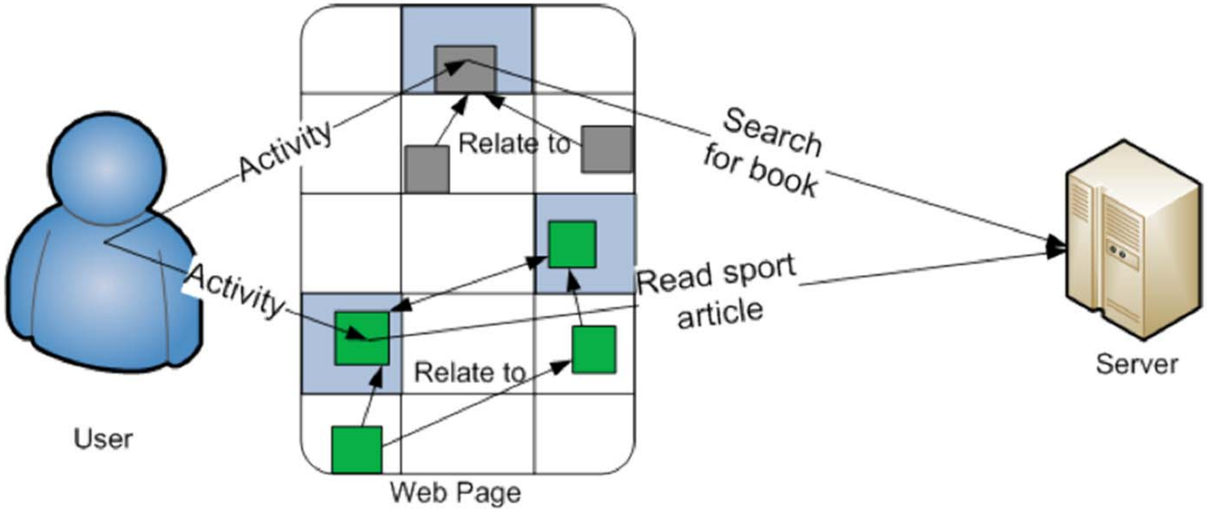
- Using Events allows us to track:
  - Where the user is active
  - What is he doing (scrolling, typing)
  - Is he active
  - Mouse – Eye correlation



- Next step is to understand WHY the user is doing this?

UPM: Ontology-based UPM

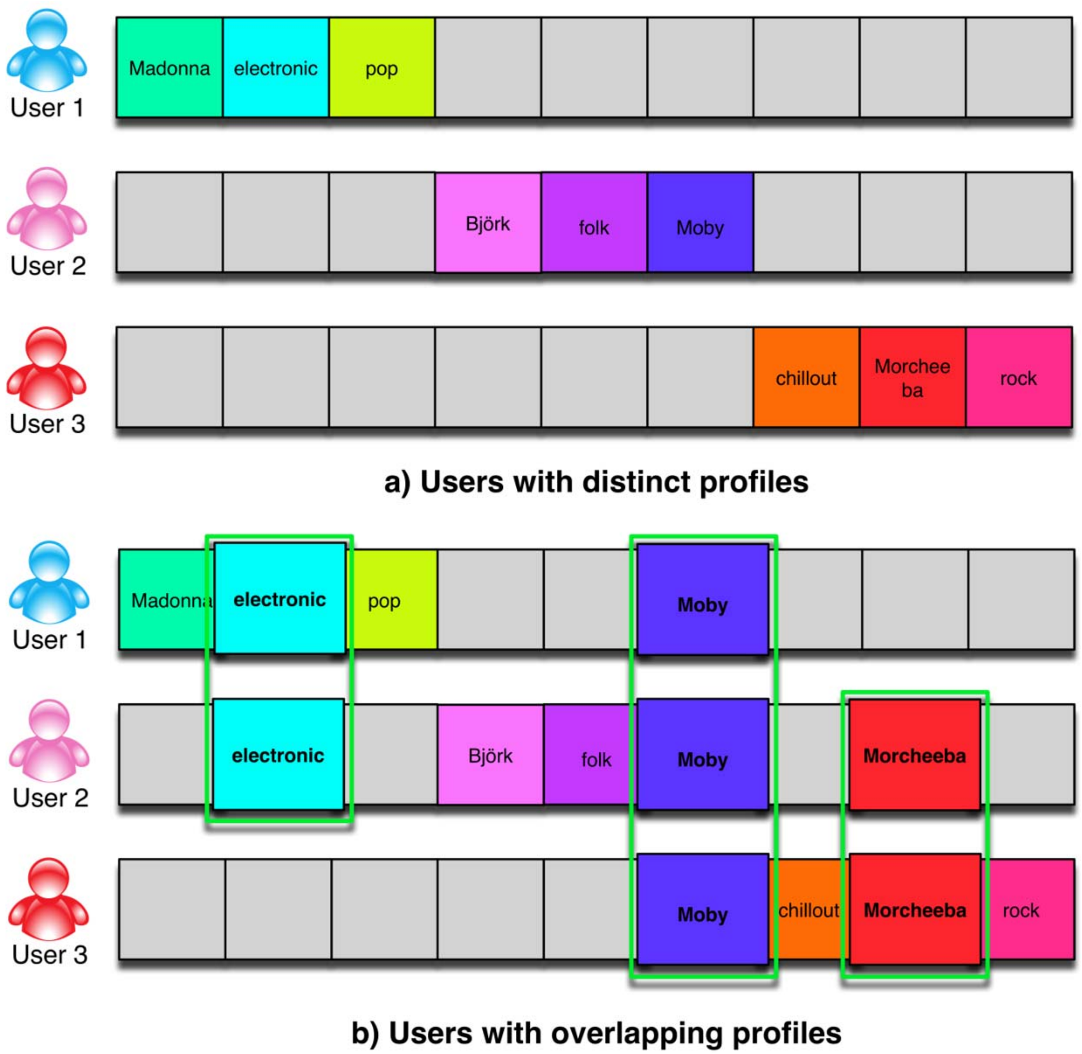
- Collecting meta information
- Defining relations between different parts and content of the website



→ Can we improve recommendation quality using semantically enriched user profiles?

UPM: Semantic Enrichment

DAI-Labor
Distributed Artificial Intelligence Laboratory



Simplified visualization of the initial cold start problem.
a) Before the enrichment, there is no overlap between the different user profiles and collaborative filtering is not possible.
b) After the enrichment, the user profiles overlap and collaborative filtering is possible

Lesson Learned

serum DAI-Labor Distributed Artificial Intelligence Laboratory

New semantic approach

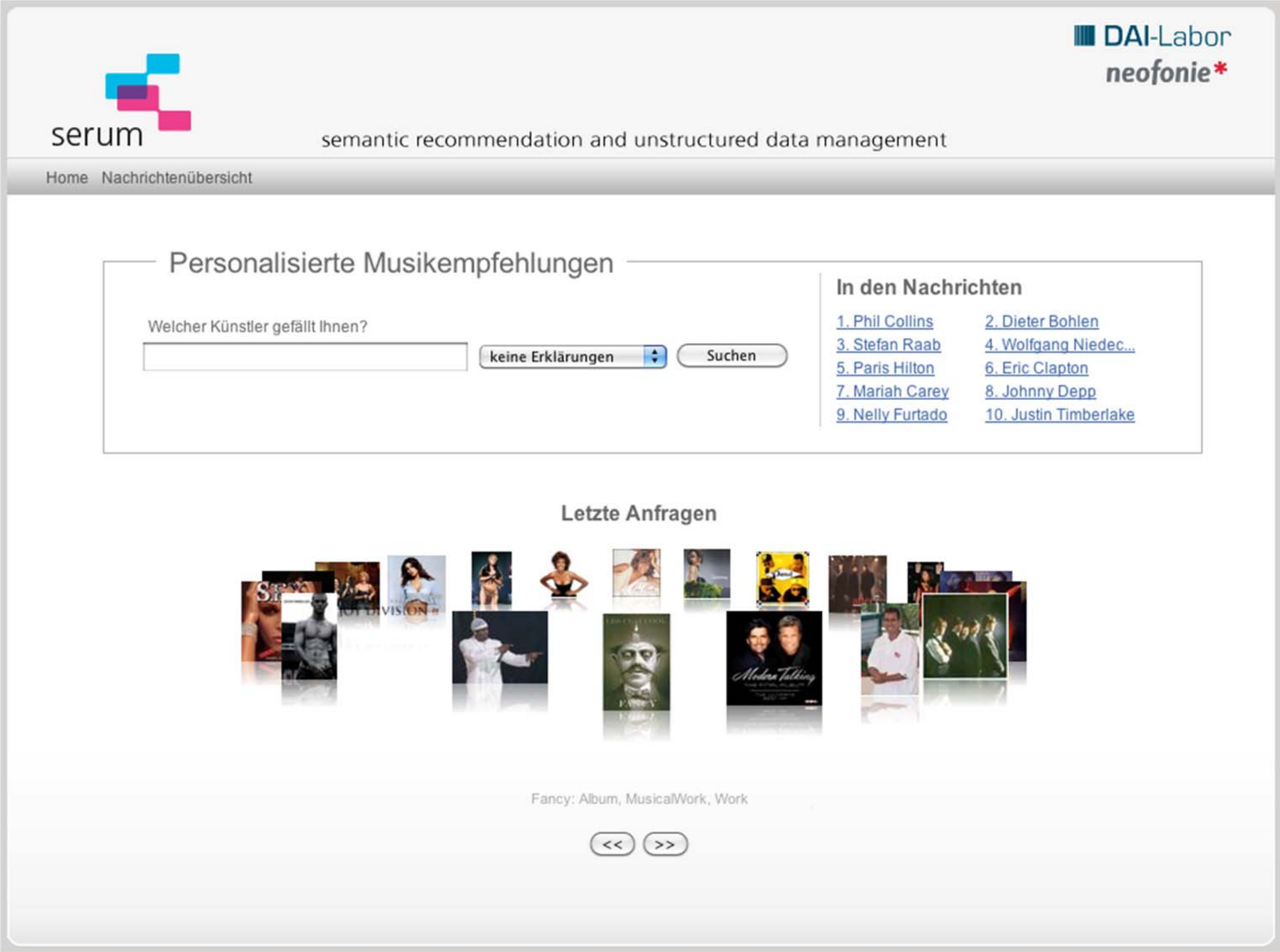
- to overcome the cold start problem
- Enriched user profiles with data from **semantic encyclopedic datasets**.
- Depending on the scenario, the profile enrichment **improves** the recommendation quality.

Semantically Enriched Profile

- works very well for users with an **unusual music taste**
- **not helpful for users with large profiles or a popular music taste** (Freebase data contains general domain information, and for users with a common taste more or less universal knowledge is added).

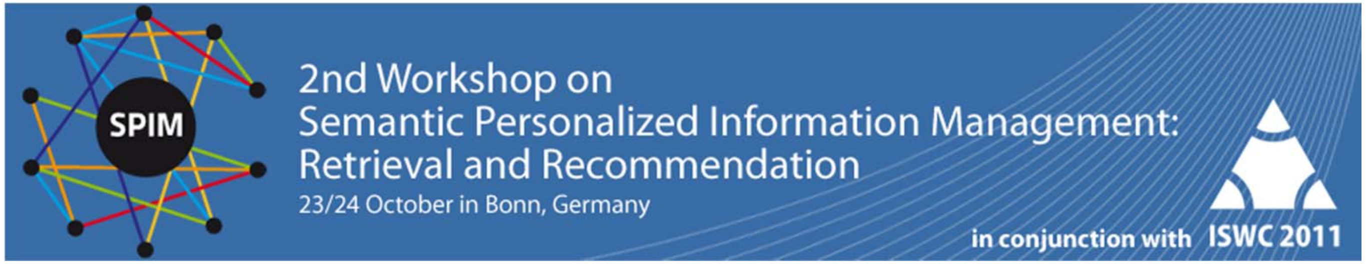
Personalization - SERUM

Demo



Serum Demo

2nd SPIM Workshop



Topics of interest include, but not limited, to following aspects:

- Exploiting Linked Open Data for PIM
- Semantic search, Semantic exploratory browsing, Semantic recommenders
- Ontology-based user modeling and user model aggregation
- Use of semantic technologies in UI/HCI
- Hybridizing Semantic Web and Machine Learning techniques for PIM
- New personalization algorithms for large semantic data sets
- Applications for SPIM

Outline

- About me
- Information Management
- Personalized and Context-aware Information Management
  - Goals
  - Personalization
  - Context-awareness
    - Overview
    - KMulE Project
    - CAMRa Challenge
- Conclusions

Context Definition

- Context is [Dey 2001]
  “any information that can be used to characterise the situation of entities”
- Interpreting this definition in a recommender systems scenario
  - context can be seen as any feature that affects the **user’s**, or the **item’s** current **situation**, i.e. time of day, location, weather, company, etc., etc.

Why do we need Context?

DAI-Labor
Distributed Artificial Intelligence Laboratory

+

- Filters relevant information
- Ad hoc recommendations
- Aware of changes

-

- What is context?
- Where do we find it?

Context-aware information

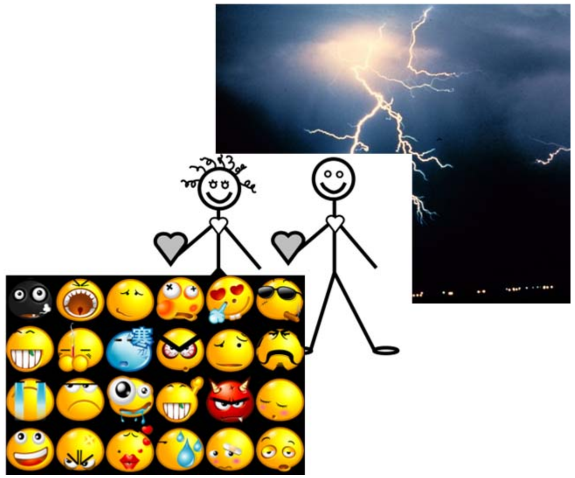
- It is widely available in **various ways**
  - interaction patterns, location, devices, annotations, query suggestions and user profiles
- It is becoming more and more important for
  - **enhancement** of retrieval performance and recommendation results
  - **personalization** and **adaptation** of information

Context-aware information

Recommendation

- Item context
  - Seasonal (Christmas, Oscar's)
  - Relation (movie sequel, director, actor)

- User context
  - Surroundings (weather, location)
  - Company (alone, with friends)
  - Mood/emotions
  - any user related factor



Context-aware Recommenders

- Adomavicius and Tuzhilin [2011] divide context-aware recommender systems into three types:
  - **Contextual Pre-Filtering**, where context directs data selection
  - **Contextual Post-Filtering**, where context is used for filtering recommendations computed by traditional approaches.
  - **Contextual Modeling**, where context is directly integrated into the model
  - Variants and combinations are possible

Kmule Project

Context-aware Multimedia Recommender Systems [Kontextbasiertes Multimedia-Empfehlungssystem] (KMulE)

Supported by:

Federal Ministry of Economics and Technology

on the basis of a decision by the German Bundestag

- Goal :
  - implicit identification of **context-related preferences** based on analysis of users' **interaction histories** and current **usage contexts**
- How:
  - Key contextual and metadata features are identified and used for the creation of several sets of **user-specific** and **context-aware recommendations**.

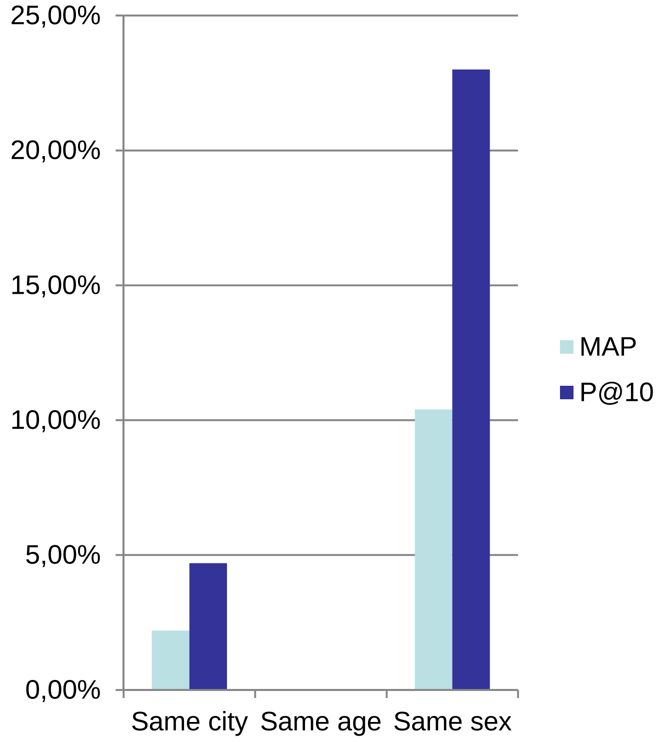
Demographic Context

- We evaluated different **demographic user features** for improving movie recommendations.
- Utilizing **user-user relations** in a recommendation scenario one could improve the quality of recommendations.
- Said et al. [2010] showed that different **social groups** have difference in taste when it comes to movies as well.

- Similar Work:
  - Weber and Castillo [2010] used demographic information like average income, race, etc. to find difference between groups in a search engine scenario. **Demographic data boost the quality of different IR tasks**.

Demographic Context

- Users in the **same demographic group** are ”more important” than those who are most similar to the user
  - Simply: making the number of a user’s peers smaller
  - Potentially high improvements
  - More data = better results



Home/Cinema Context

Inferred Contextual User Model (CUP)

- A user profile, similar to the "micro-profile" concept by Baltrunas and Amatriain [2009].
- based on a combination of **movie meta-data** and the creation **time of the rating** (i.e. the time when the movie was rated by a user)

- Assumption:
  - movies rated within two months of their cinema premiere date → seen in cinema
  - movies rated later → seen at home.

→ context-aware sub-profiles

Home/Cinema Context

- Recommending movies based on where they are to be seen (cinema or at home)
- User-movie and user-movie-context (CUP) matrix example:

	u_i	u_j	u_k	u_m	u_l
m_a	1	3		5	
m_b		4			4
m_c			5	2	
m_d	5	3		3	
m_e	3	4	1		1

(a) The original, uncontextualized user-rating structure.

	u_i		u_j		u_k		u_m	u_l
	home	cinema	home	cinema	home	cinema	cinema	home
m_a	1				3		5	
m_b				4				4
m_c						5	2	
m_d	5		3				3	
m_e		3		4		1		1

(b) The same set of users, after having been divided into two CUPs where applicable. Some users will only have one CUP (i.e. they have only seen movies at home, or at the cinema).

Improvement: Recommendations using **contextualized user profiles** between **40%** (P@1, K=100) and **66%** (P@20, K=200) in **precision**.

Lesson Learned

Demographic Context

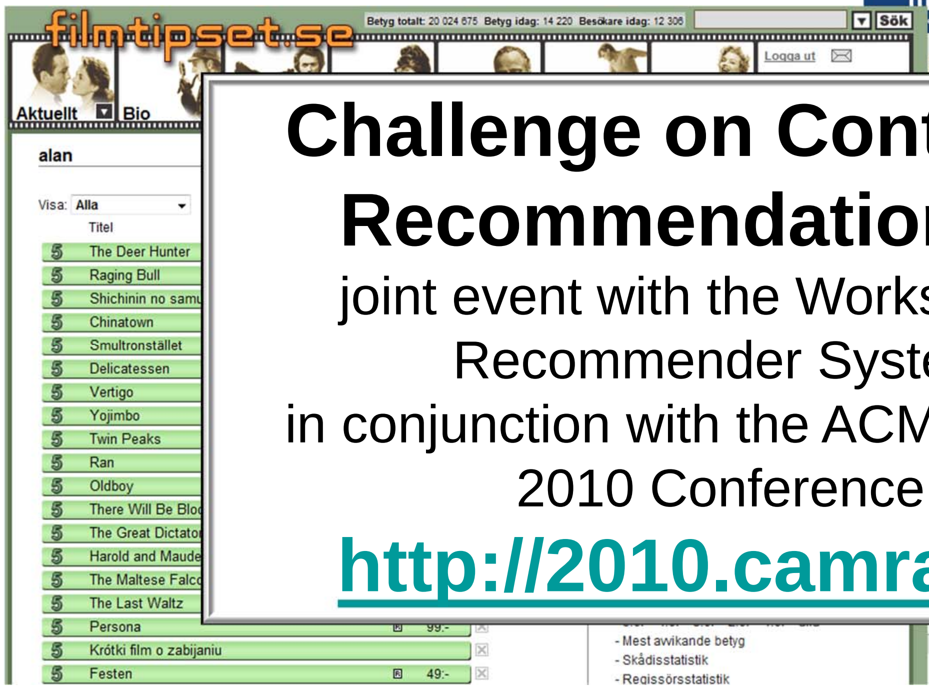
- Demographic and social-related information help to give better recommendations

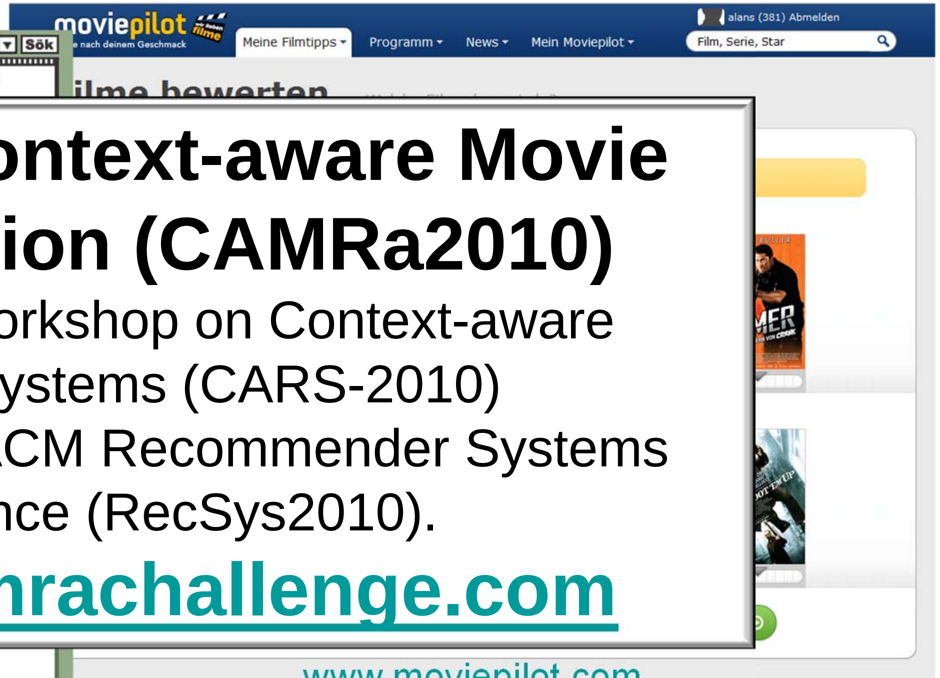
Contextual User Profile

- The results confirm the notion of "**situated action**", where users have separated rating profiles depending on the combination of where, how and when (context) the movie is seen.

CAMRa 2010 Motivation

Current SOA systems are generally **context-ignorant**, CAMRa focused on identifying and utilizing contextual data in <u>real-life</u> data from <u>Moviepilot</u> and <u>Filmtipset</u>.





Challenge on Context-aware Movie Recommendation (CAMRa2010)

joint event with the Workshop on Context-aware Recommender Systems (CARS-2010) in conjunction with the ACM Recommender Systems 2010 Conference (RecSys2010).

http://2010.camrachallenge.com

www.moviepilot.com

www.filmtipset.se

CAMRa 2010 Organizers

Co-organized by Industry :

CAMRa 2010 Goals

- CAMRa2010 aimed at boosting the research of **context-awareness** in recommender system.
- CAMRa focused on
  identifying contextual features in datasets
  and **generating context-aware movie recommendations**.
- Two datasets,
  gathered by the **Moviepilot** and **Filmtipset** online **movie recommendation** communities,
  were released exclusively for the challenge.

CAMRa 2010 Datasets

Filmtipset Statistics

- > 95, 000 users
- > 75, 000 movies
- > 20 Million ratings
- user generated lists/topics
- user assigned similiarities
- Rating scale 1-5

Moviepilot Statistics

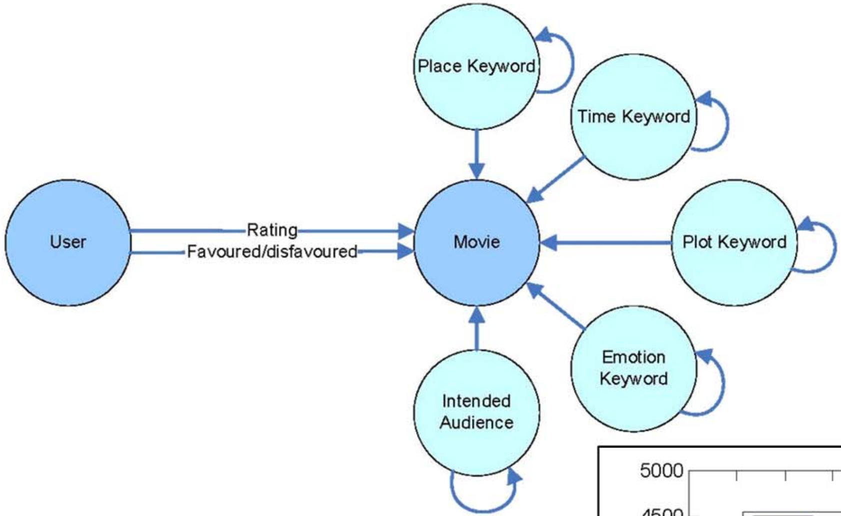
- > 100, 000 users
- > 40, 000 movies
- > 6 Million ratings
- tags (emotions, intended audiences, etc)
- listings (cinema, tv, etc)
- Rating scale 1-10

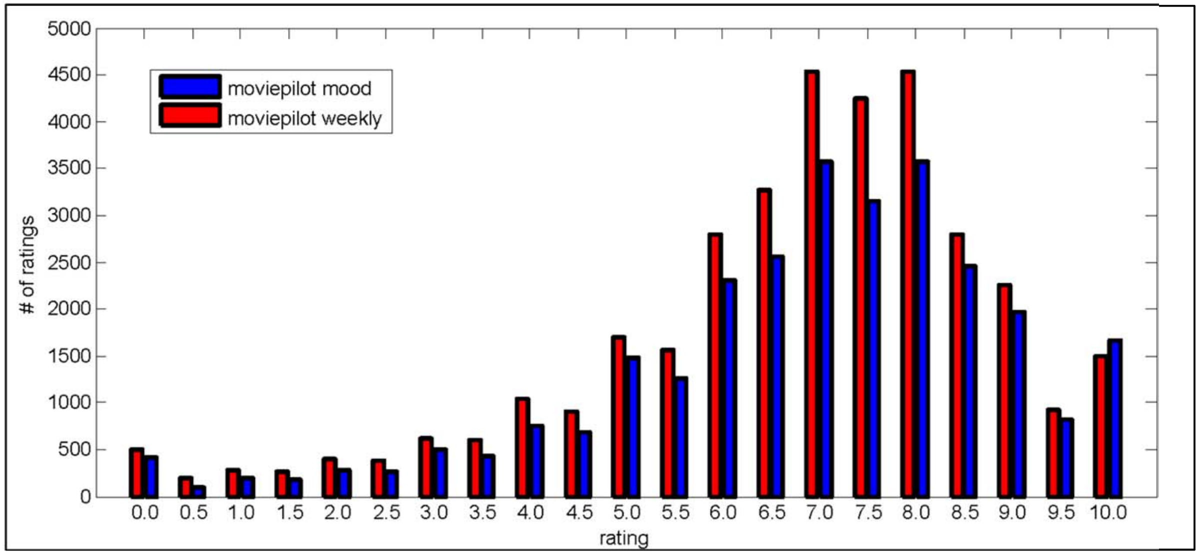
44

Moviepilot Dataset

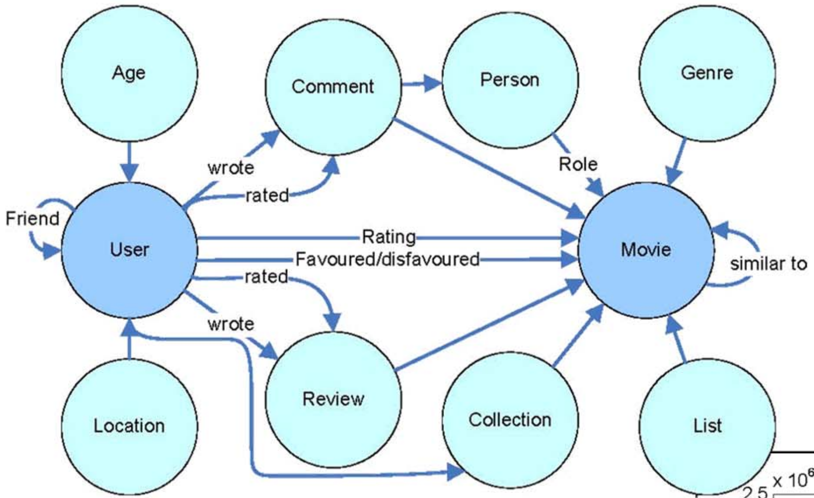
DAI-Labor
Distributed Artificial Intelligence Laboratory

Moviepilot
Entity-Relations &
Rating Distribution

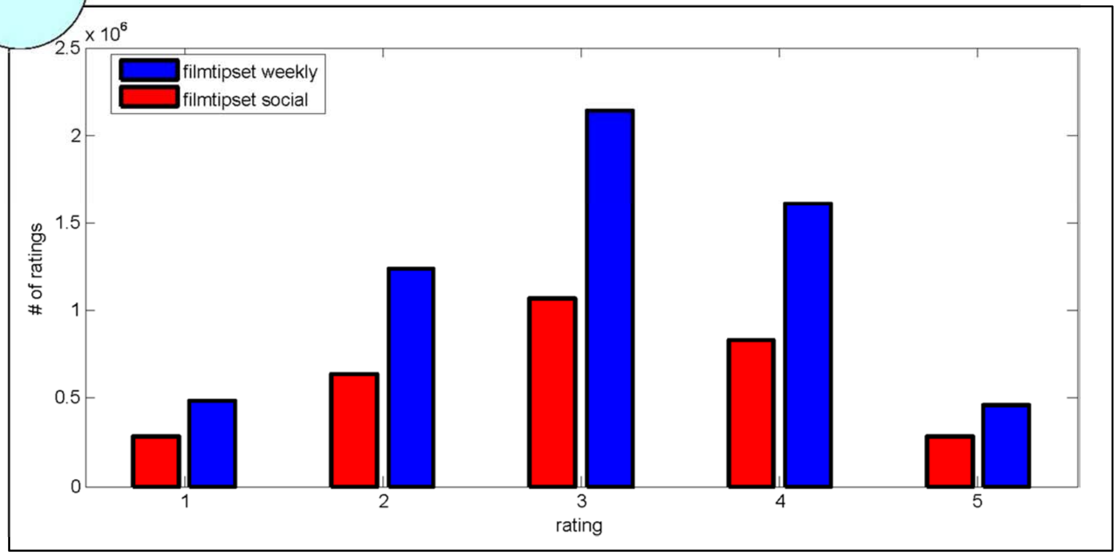




Filmtipset Dataset



Filmtipset Entity-Relations & Rating Distribution



Recommendation Tracks

- The participants could participate in three recommendation tracks:
  - recommend movies based on
    - (1) the **time of the year** and **special events**,
    - (2) **social relations** of users, and
    - (3) a user's (implicit) **mood**.

Recommendation Tracks

CAMRa offers four datasets for three different tracks

- **Track 1 – Moviepilot and Filmtipset Weekly recommendation**
  - One dataset from each website
  - Recommendations for calendar week 52 2009 (Christmas)
  - Recommendation for calendar week 9 2010 (Oscar's)
  - Live Evaluation of best performing teams with real users
- **Track 2 – Moviepilot Mood**
  - Dataset from Moviepilot
  - Recommendations for certain users and one certain mood
- **Track 3 – Filmtipset Social**
  - Dataset from Filmtipset
  - Recommendations for certain users based on their social relations

Week Track (Moviepilot & Filmtipset)

DAI-Labor
Distributed Artificial Intelligence Laboratory

- Approaches
  - **[Gantner et al. 2010]** used an approach from **tag recommendation. Pairwise Interaction Tensor Factorization (PITF)** where weeks were used to form **user-movie-weeks tensors** (Moviepilot data).
  - [Liu et al. 2010] implemented a time-aware collaborative filtering model using matrix factorization (Both tracks and datasets).
  - [Campos et al, 2010] presented a time-based kNN recommender (Filmtipset data).
  - [Brenner et al, 2010] presented a regression models-based approach (Filmtipset data).

→ It is better to **recommend the movies that were most popular in the 10 days before the Oscar ceremonies** than to use item-based collaborative filtering on the full dataset!

Mood Track (Moviepilot)

- Approaches
  - **[Shi et al, 2010]** an extended **matrix factorization model** that included mood information.
  - [Wang et al, 2010] used a mood and user-based hybrid kNN weighted mean
  - [Wu et al, 2010] a k-nearest-neighbor collaborative filtering algorithm utilizing expert users.

→ It is better to use **one specific mood** tag than general mood tags. It particularly helps to learn latent movie features with respect to the specified mood.

→ The **general mood-based** similarity **only gives general closeness** measurement of two movies in terms of all their mood properties.

Social Track (Filmtipset)

- Approaches
  - **[Liu et al, 2010] was, similarly to the weekly approach covered by this paper, based on matrix factorization.**
  - [Díez et al, 2010], was a random-walk model utilizing the implicit information in friendships
  - [Liu and Yuan 2010] presented and extension of traditional collaborative filtering where social data was taken into consideration,
  - [Rahmani et al, 2010] presented two approaches: a kNN approach based on linear combinations of similarity measures between users, and one approach based on inductive logic programming.

→ incorporating the **social network** (**similar socio-demographic** or **behavioral characteristics** - homophily principle) between users **as additional matrix** into the matrix factorization model increases performance.

General Information

in conjunction with the 2011 ACM Recommender Systems Conference, Chicago, IL, USA, October 27th, 2011
http://www.camrachallenge.com

- The challenge consists of two tracks:
  - in the first track, the participants are requested to **generate recommendations for households**,
  - in the second the focus lies on **identifying** which **member of a household** performed a **specific rating**.

Outline

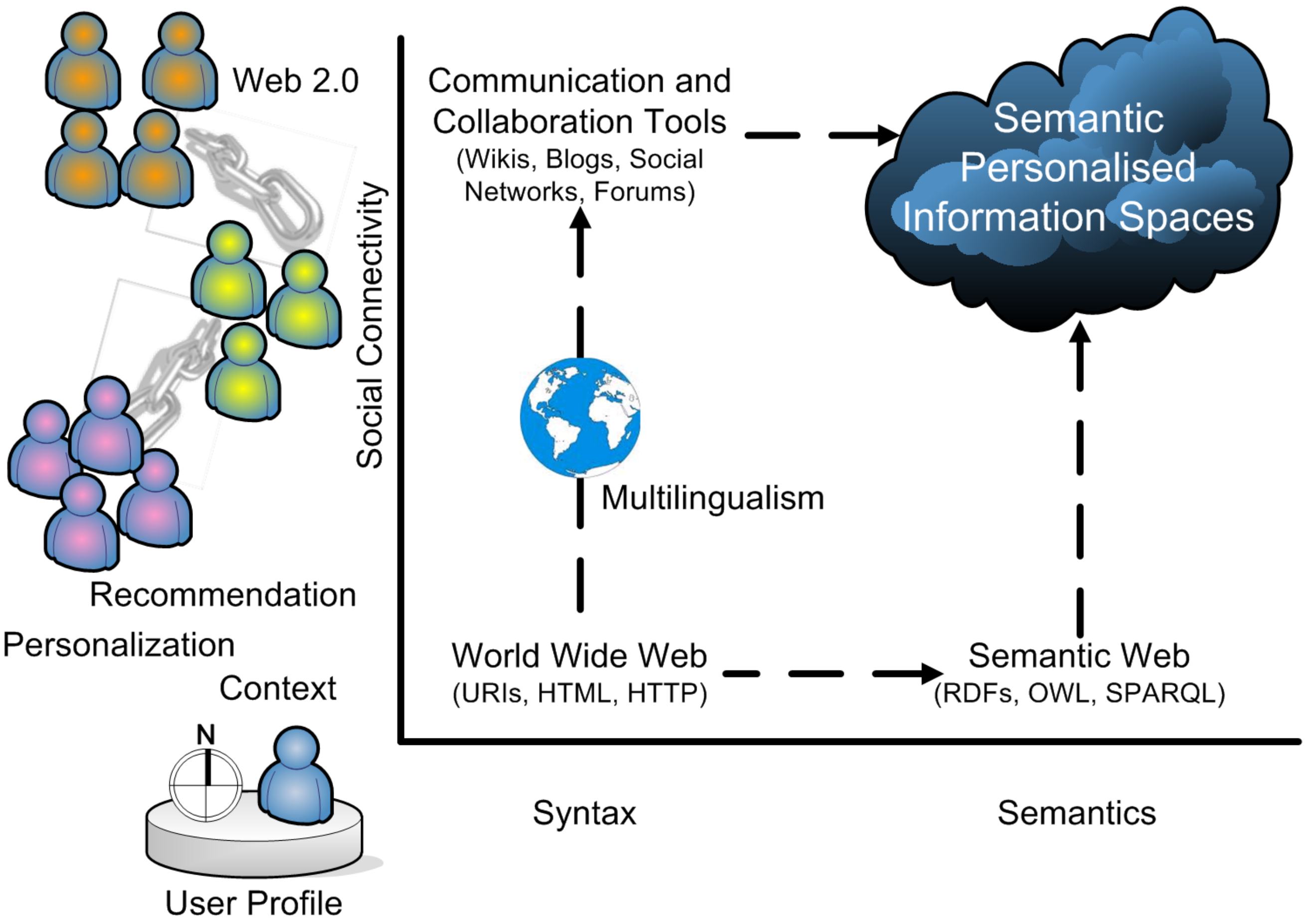
- Information Management
- Personalized and Context-aware Information Management
- Conclusions

Conclusions

- Problems not covered sufficiently:
  - Personalization
  - Bad Recommendation
  - Not relevant retrieval results
- Solutions:
  - We can find and recommend semantically-related content
  - We can use context to give better recommendations
  - We can better support users (user profiling and recommendation)

Conclusions

Is this the digital future?



Questions?

DAI-Labor
Distributed Artificial Intelligence Laboratory