SWAP researchgroup

**S**emantic
**W**eb
**A**ccess and
**P**ersonalization
research group
http://www.di.uniba.it/~swap

UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

# Content-based Recommender Systems
## problems, challenges and research directions


Giovanni Semeraro & the SWAP group
http://www.di.uniba.it/~swap/
semeraro@di.uniba.it

Department of Computer Science
University of Bari "Aldo Moro"



UMAP 2010 – 8° Workshop on
INTELLIGENT TECHNIQUES FOR WEB PERSONALIZATION
& RECOMMENDER SYSTEMS (ITWP 2010)
BIG ISLAND OF HAWAII, JUNE 20 2010

# Outline

❶ Content-based Recommender Systems (CBRS)

- ✓ Basics
- ✓ Advantages & <u>Drawbacks</u>

❷ Drawback 1: Limited content analysis

- ✓ Beyond keywords: Semantics into CBRS
- ✓ Taking advantage of Web 2.0: Folksonomy-based CBRS

❸ Drawback 2: Overspecialization

- ✓ Strategies for diversification of recommendations

# Content-based Recommender Systems (CBRS)

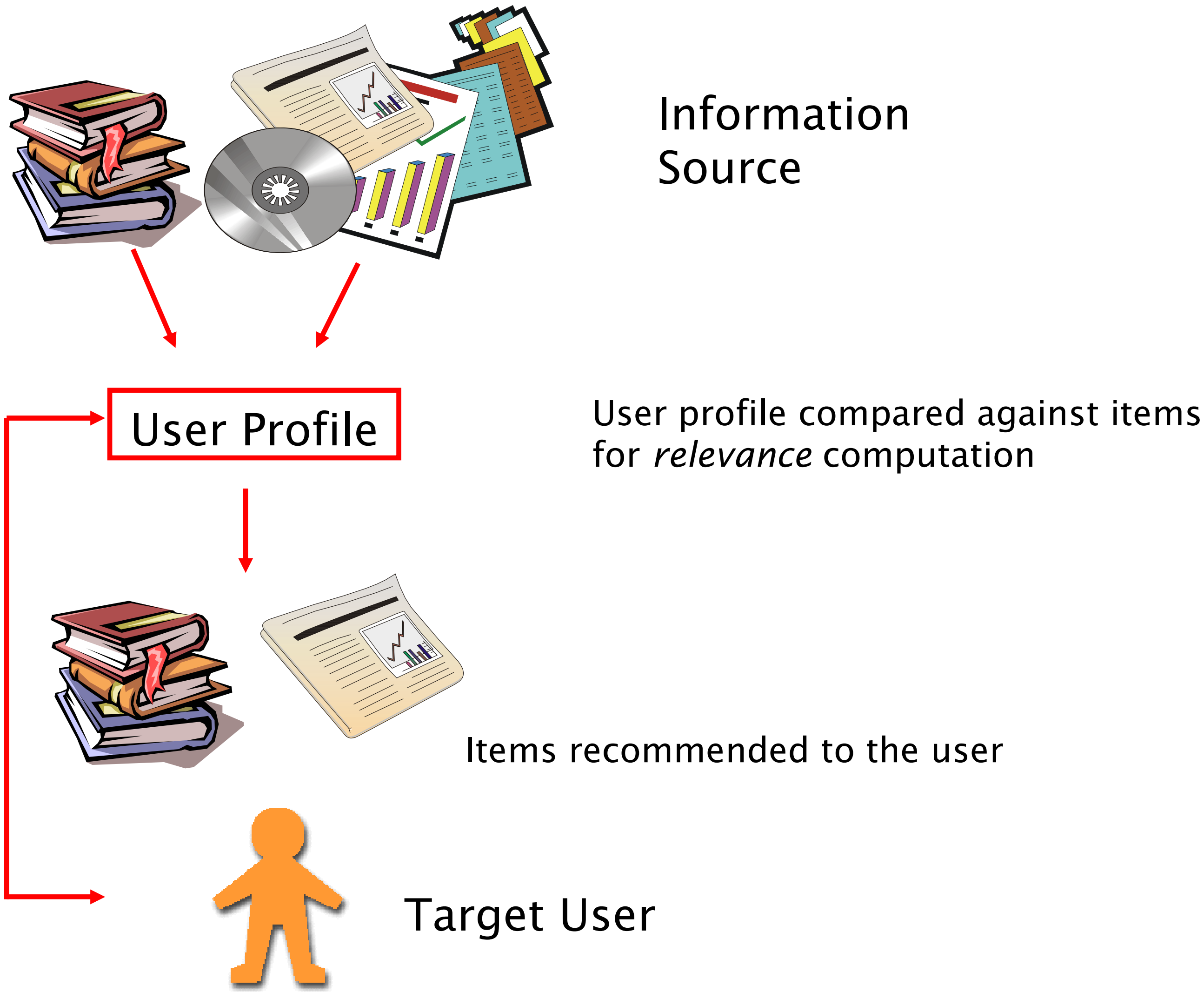
➊ *Recommend an item to a user based upon a description of the item and a profile of the user's interests*

➋ Implement strategies for:

- ✓ representing items
- ✓ creating a user profile that describes the types of items the user likes/dislikes
- ✓ comparing the user profile to some reference characteristics (with the aim to predict whether the user is interested in an unseen item)

[Pazzani07] Pazzani, M. J., & Billsus, D. Content-Based Recommendation Systems. *The Adaptive Web*. Lecture Notes in Computer Science vol. 4321, 325-341, 2007.

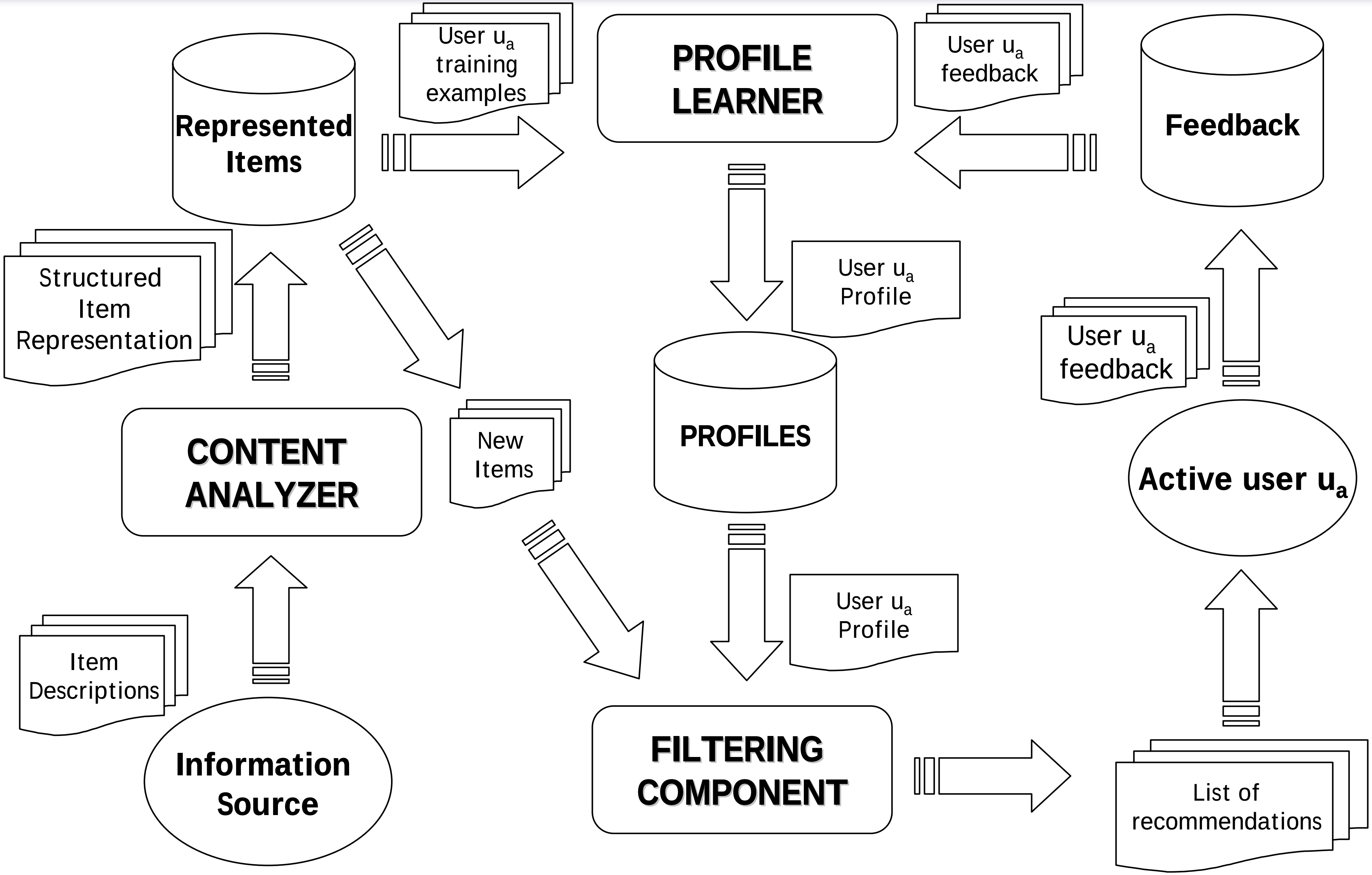
# Content-based Filtering



# Content-based Filtering

1. Each user is assumed to operate independently
2. Items are represented by some features
   - ✓ Movies: actors, director, plot, …
3. The profile is often created and updated automatically in response to feedback on the desirability of items that have been presented to the user
   - ✓ Machine Learning for automated inference
   - ✓ Relevance judgment on items, e.g. ratings
   - ✓ Training on rated items → user profile
4. Filtering based on the comparison between the *content (features)* of the items and the user preferences as defined in the user profile
   - ✓ Keyword-based representation for content and profiles → string matching or text similarity

# General Architecture of CBRS



# Advantages of CBRS

❶ USER INDEPENDENCE

- ✓ CBRS exploit solely ratings provided by the active user to build her own profile
- ✓ No need for data on **other users**

❷ TRANSPARENCY

- ✓ CBRS can provide explanations for recommended items by listing content-features that caused an item to be recommended

❸ NEW ITEM (Item not yet rated by any user)

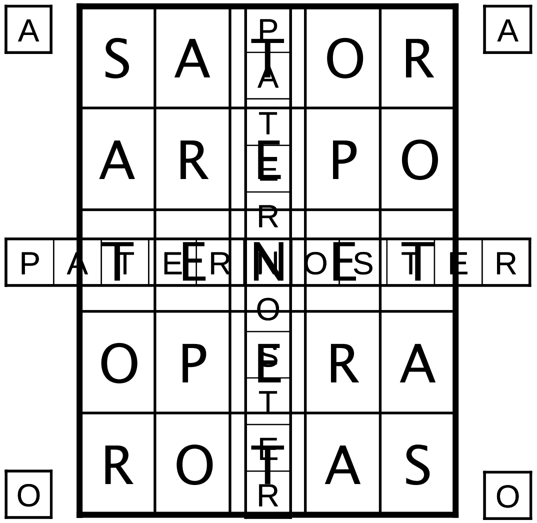
- ✓ CBRS are capable of recommending new and unknown items
- ✓ No first-rater problem

# Drawbacks of CBRS: LIMITED CONTENT ANALYSIS

❶ No suitable suggestions if the analyzed content does not contain enough information to discriminate items the user likes from items the user does not like

❷ Content must be encoded as meaningful **features**

- ✓ automatic/manually assignment of features to items might be insufficient to define distinguishing aspects of items necessary for the elicitation of user interests
- ✓ keywords not appropriate for representing content, due to polysemy, synonymy, multi-word concepts (*homography*, *homophony*,...) – "Sator arepo eccetera" [Eco07]

# Keyword-based Profiles

doc1
**AI** is a branch of computer science

doc2
the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3
**apple** launches a new product…

USER PROFILE

| | |
|---|---|
| artificial | 0.02 |
| intelligence | 0.01 |
| apple | 0.13 |
| AI | 0.15 |
| … | |

MULTI-WORD CONCEPTS

# Keyword-based Profiles

doc1

**AI** is a branch of computer science

doc2

the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3

**apple** launches a new product...

USER PROFILE

| | |
|---|---|
| artificial | 0.02 |
| intelligence | 0.01 |
| apple | 0.13 |
| AI | 0.15 |
| ... | |

SYNONYMY

# Keyword-based Profiles

doc1
**AI** is a branch of computer science

doc2
the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3
**apple** launches a new product...

USER PROFILE

| | |
|---|---|
| artificial | 0.02 |
| intelligence | 0.01 |
| apple | 0.13 |
| AI | 0.15 |
| ... | |

POLYSEMY

NLP methods are needed for the elicitation of user interests

# Drawbacks of CBRS: OVERSPECIALIZATION

❶ CBRS suggest items whose scores are high when matched against the user profile

- ✓ the user is going to be recommended items similar to those already rated

❷ No inherent method for finding something unexpected

❸ Obviousness in recommendations

- ✓ suggesting “STAR TREK” to a science-fiction fan: *accurate* but *not useful*
- ✓ users don’t want algorithms that produce better ratings, but *sensible* recommendations

❹ The Serendipity Problem

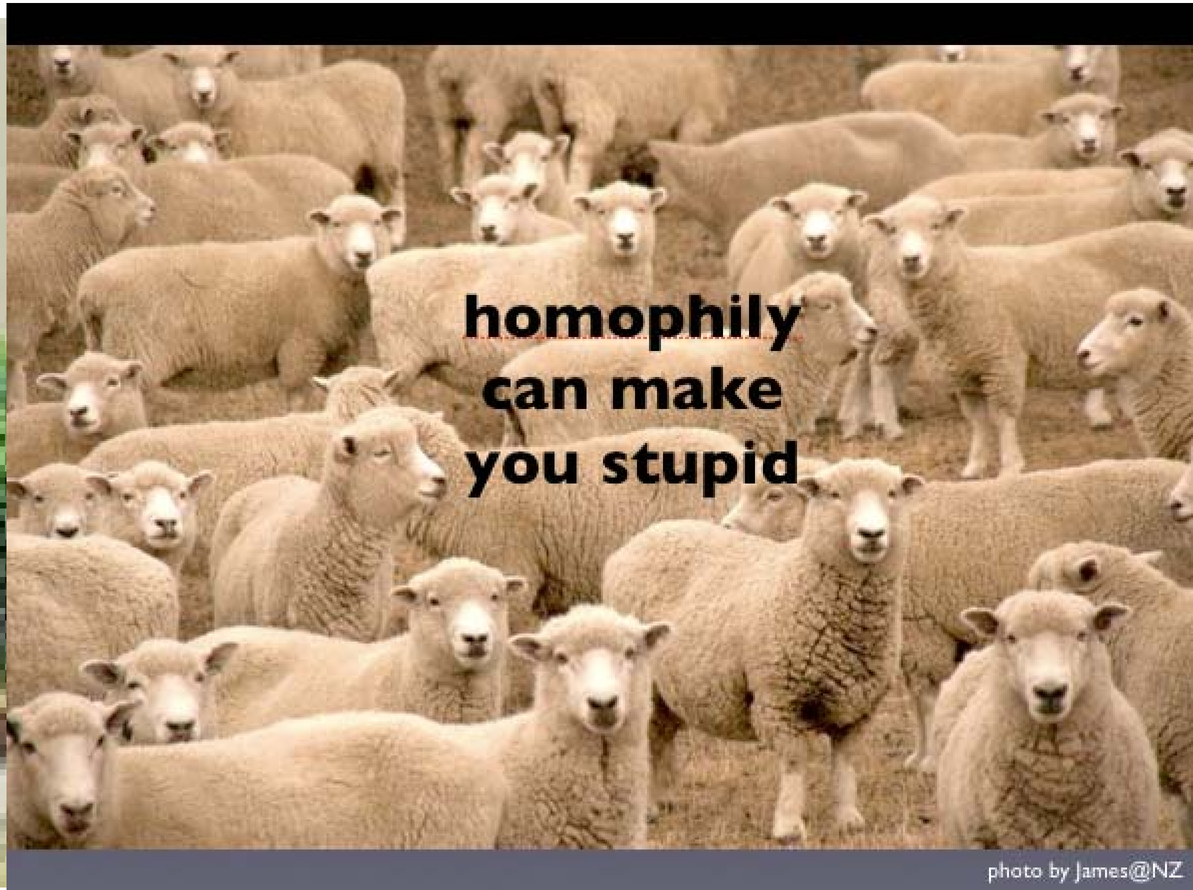
[McNee06] S.M. McNee, J. Riedl, and J. Konstan. Accurate is not always good: How accuracy metrics have hurt recommender systems. In *Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems*, pages 1-5, Canada, 2006.

# The serendipity problem: mind cages

- Homophily: the tendency to surround ourselves by like-minded people

opinions taken to extremes

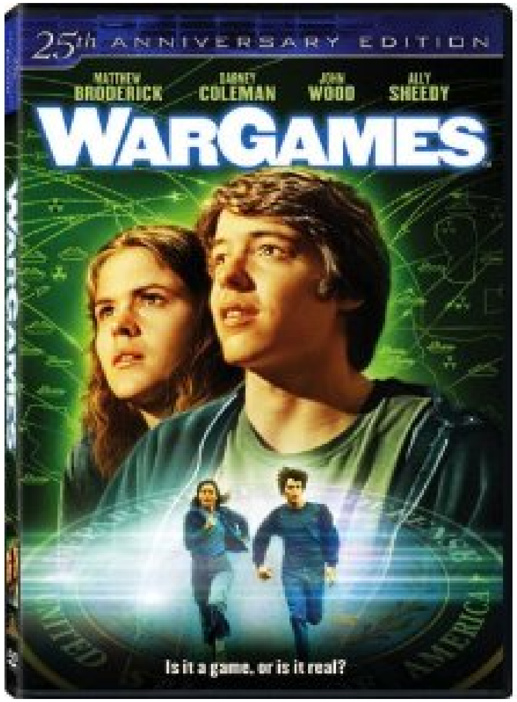
cultural impoverishment




biodiversity?

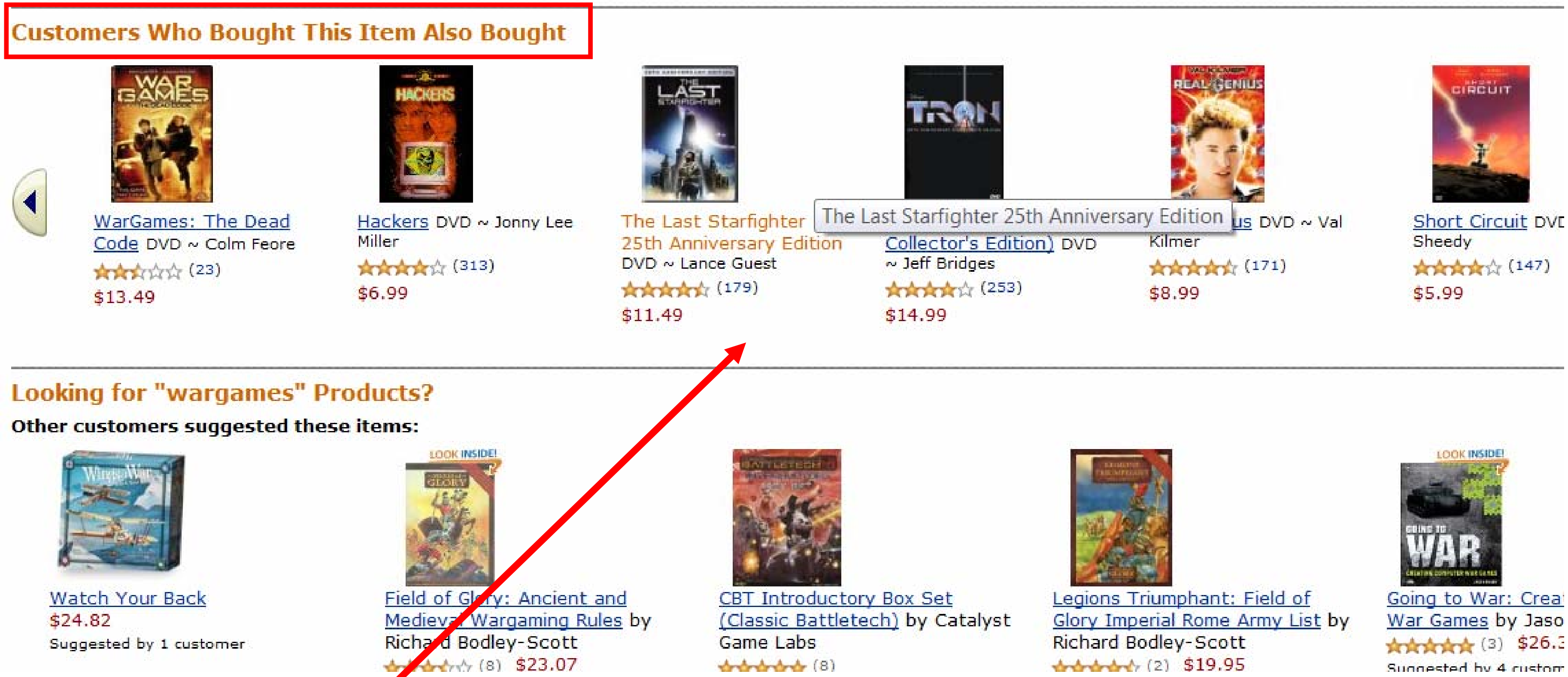
# The homophily trap

- Does homophily hurt RS?
  - ✓ try to tell Amazon that you liked the movie "War Games"…




[Zuckerman08] E. Zuckerman. Homophily, serendipity, xenophilia. April 25, 2008. www.ethanzuckerman.com/blog/2008/04/25/homophily-serendipity-xenophilia/

# The homophily trap




Recommendations by other (ageing?)
COMPUTER GEEKS!

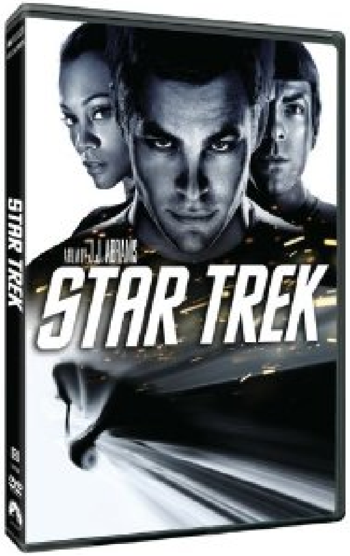
# “Item-to-Item” homophily...
## Harry Potter for ever?

# Novelty vs Serendipity

- Novelty: A novel recommendation helps the user find a surprisingly interesting item she might have autonomously discovered
- Serendipity: A serendipitous recommendation helps the user find a surprisingly interesting item she might not have otherwise discovered
- How to introduce serendipity in (CB)RS?

[Herlocker04] Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, 22(1): 39-49, 2004.

# "Computational" serendipity? A motivating example






for Star Trek fans: Did you try
"Star Trek – The experience"
in Las Vegas?

# Putting Intelligence into CBRS: Challenges & Research Directions

| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ "computational" serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# Putting Intelligence into CBRS: Challenges & Research Directions

| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ “computational” serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# Semantic Analysis: beyond keywords

Semantic Analysis =

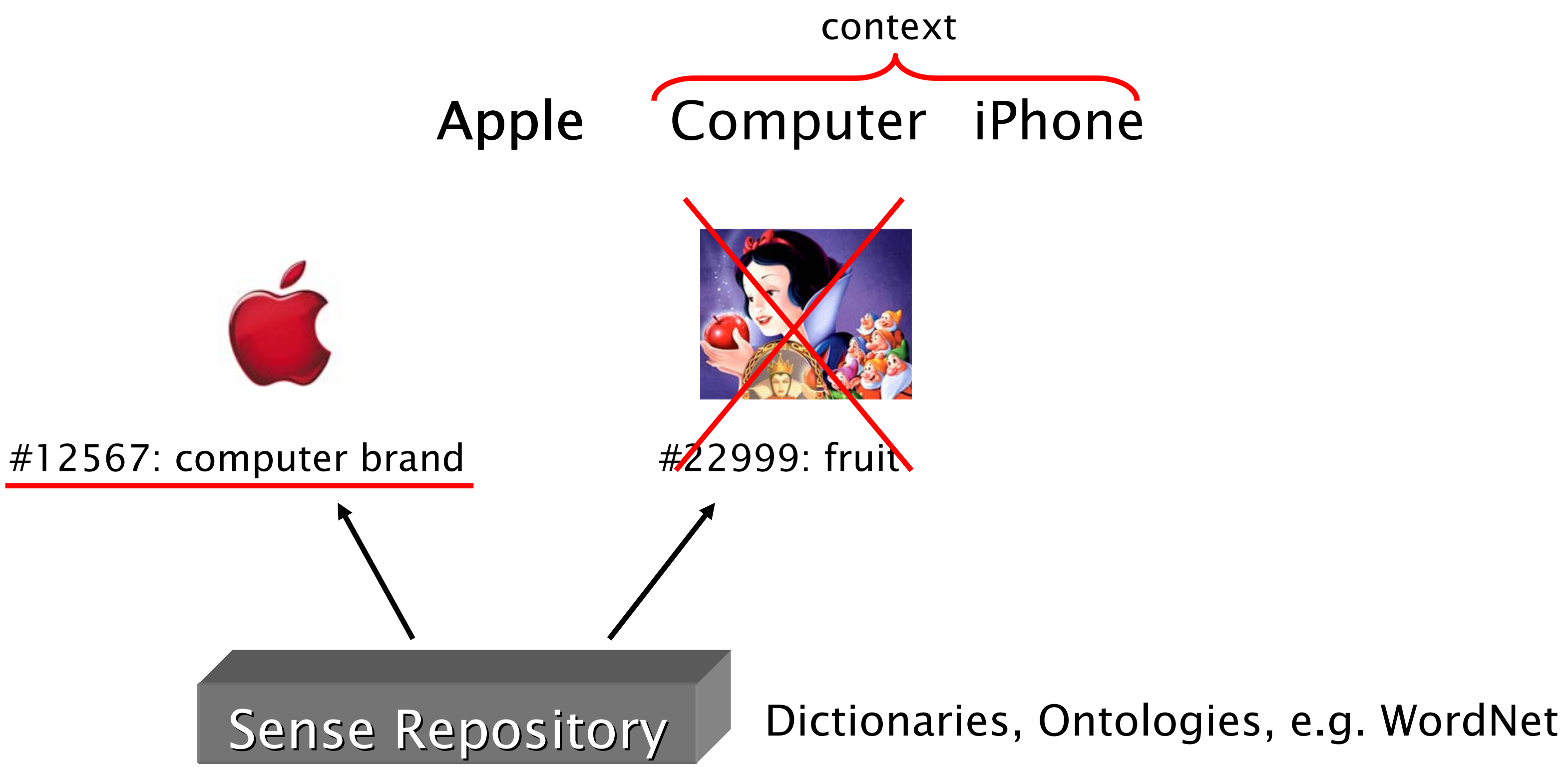
1. Semantics: concept identification in text-based representations through advanced NLP techniques → "*beyond keywords*"

+

2. Personalization: representation of user information needs in an effective way → "*deep (high-accuracy) user profiles*"

# *Beyond keywords*: Word Sense Disambiguation (WSD) - from words to meanings

- WSD selects the proper meaning (*sense*) for a word in a text by taking into account the context in which that word occurs




[Basile07] P. Basile, M. Degemmis, A. Gentile, P. Lops, and G. Semeraro. UNIBA: JIGSAW algorithm for Word Sense Disambiguation. In *Proceedings of the 4th ACL 2007 International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic*, pages 398–401, Association for Computational Linguistics, June 23-24, 2007.

# ITR (ITem Recommender)

## Sense-based Profiles

doc1
**AI** is a branch of computer science

doc2
the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3
**apple** launches a new product…

USER PROFILE

| | |
|---|---|
| #12387 | 0.03 |
| apple | 0.13 |
| AI | 0.15 |
| … | |

MULTI-WORD CONCEPTS

# ITR (ITem Recommender)

## Sense-based Profiles

doc1
**AI** is a branch of computer science

doc2
the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3
**apple** launches a new product…

USER PROFILE

| | |
|---|---|
| #12387 | 0.18 |
| apple | 0.13 |
| #12387 | 0.15 |
| … | |

SYNONYMY

# ITR (ITem Recommender)

## Sense-based Profiles

doc1
**AI** is a branch of computer science

doc2
the 2011 International Joint Conference on **Artificial Intelligence** will be held in Spain

doc3
**apple** launches a new product…

SEMANTIC USER PROFILE
sense identifiers rather than keywords

USER PROFILE

| | |
|---|---|
| #12387 | 0.18 |
| #12567 | 0.13 |
| … | |

POLYSEMY

[Degemmis07] M. Degemmis, P. Lops, and G. Semeraro. A Content-collaborative Recommender that Exploits WordNet-based User Profiles for Neighborhood Formation. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI)*, 17(3):217–255, Springer Science + Business Media B.V., 2007.

[Semeraro07] G. Semeraro, M. Degemmis, P. Lops, and P. Basile. Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. In M. M. Veloso, editor, *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007* , pages 2856–2861. Morgan Kaufmann, 2007.

# Advantages of Sense-based Representations

❶ Semantic matching between items and profiles

- ✓ computing semantic relatedness [Pedersen04] rather than string matching (e.g., by using similarity measures between WordNet synsets)

❷ Senses are inherently multilingual

- ✓ Concepts remain the same across different languages, while terms used for describing them in each specific language change

❸ Improving transparency

- ✓ matched concepts can be used to justify suggestions

❹ Collaborative Filtering could benefit too

- ✓ finding better neighbors: similar users discovered by looking at profile overlap even if they did not rate the same items
- ✓ semantic profiles succeed where Pearson's correlation coefficient fail

[Pedersen04] Pedersen, Ted and Patwardhan, Siddharth, and Michelizzi, Jason. WordNet::Similarity - Measuring the Relatedness of Concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence* (AAAI-2004), pp. 1024-1025, San Jose, CA, July, 2004.

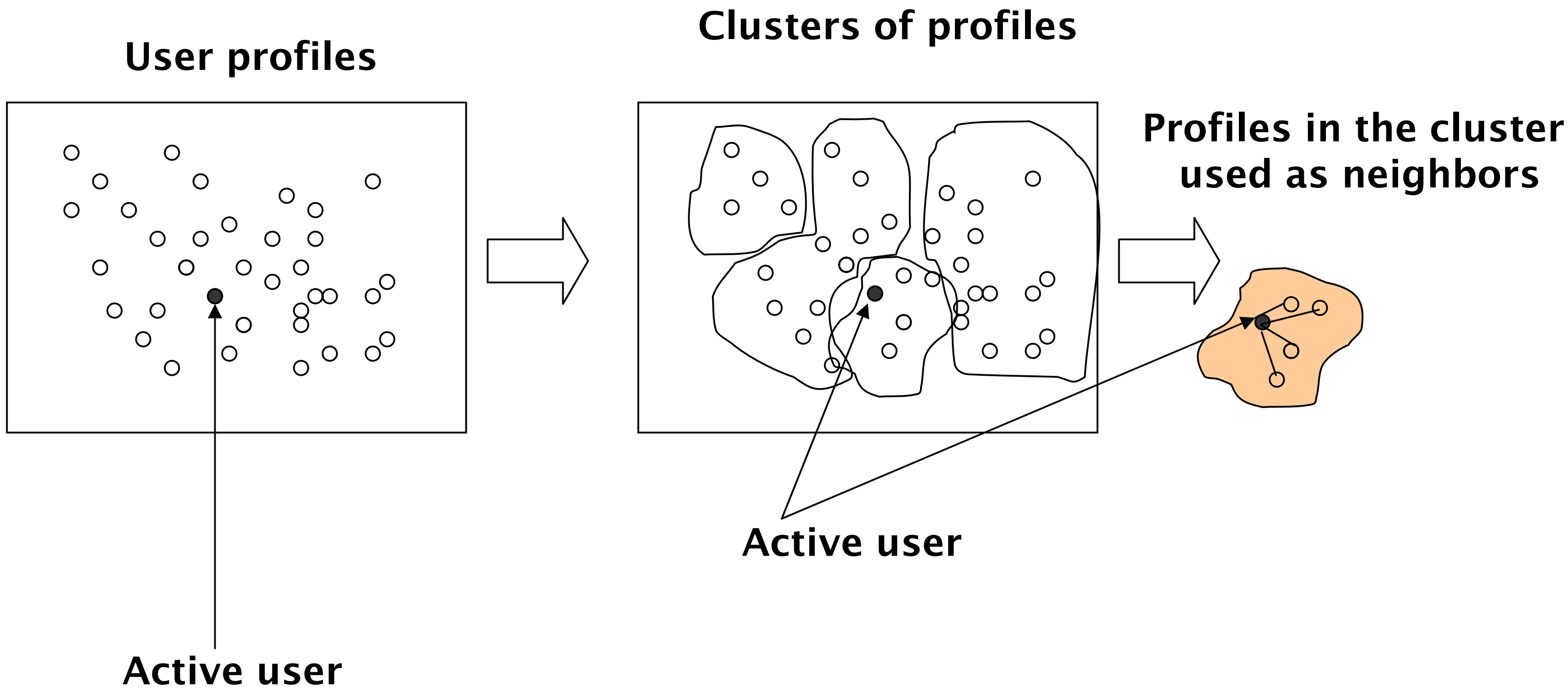
# Sense-based profiles in a hybrid CB-CF recommender

- Sense-based profiles obtained by applying WSD on textual description of items
  - ✓ WordNet as sense repository
  - ✓ Synset-based user profiles
- Hybrid CB-CF RS

[Degemmis07] M. Degemmis, P. Lops, and G. Semeraro. A Content-collaborative Recommender that Exploits WordNet-based User Profiles for Neighborhood Formation. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI)*, 17(3):217–255, Springer Science + Business Media B.V., 2007.

# Sense-based profiles in a hybrid CB-CF recommender

Clustering of sense-based profiles

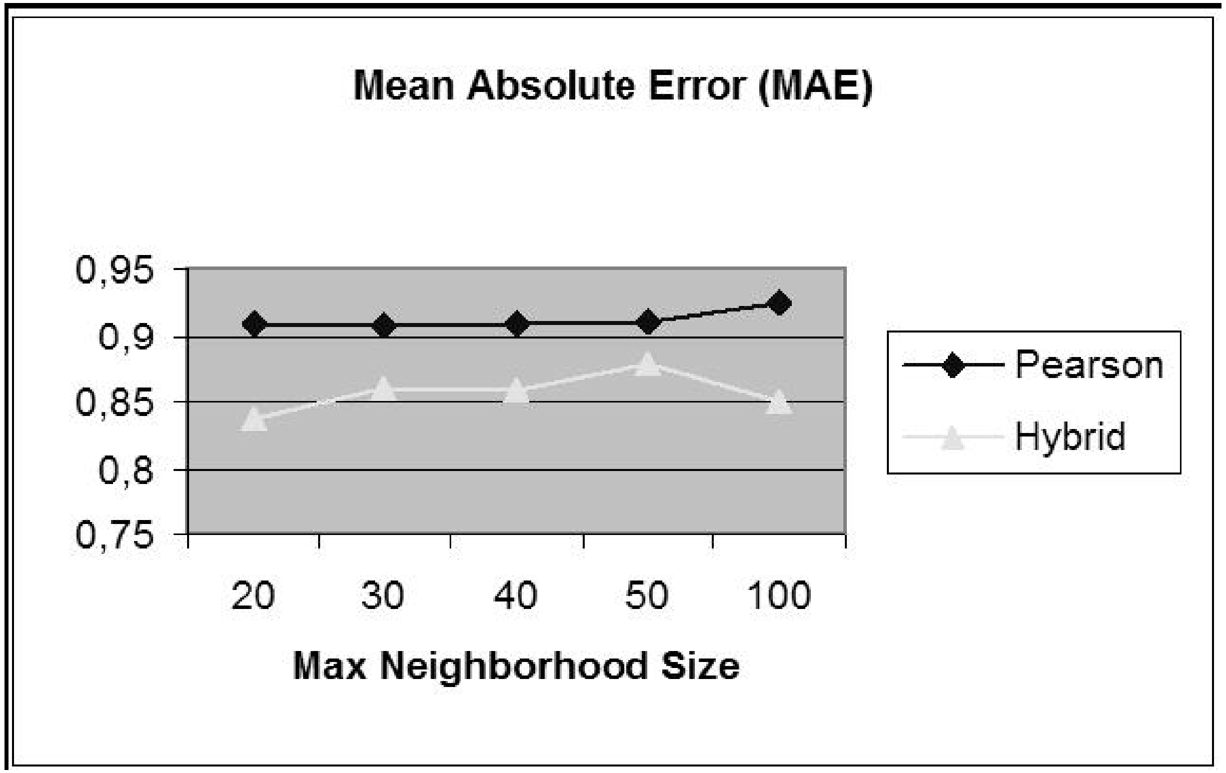


# Experimental Evaluation on EachMovie dataset

- 835 users selected from EachMovie dataset*
  - ✓ 1,613 movies grouped into 10 categories, 180,356 ratings, user-item matrix 87% sparse
  - ✓ Each user rated between 30 and 100 movies
  - ✓ Discrete ratings between 0 and 5
  - ✓ Movie content crawled from the Internet Movie Database (IMDb)
- CF algorithm using Pearson's correlation coefficient vs. CF algorithm integrating clusters of semantic user profiles

*2,811,983 ratings entered by 72,916 users for 1628 different movies. As of October, 2004, HP/Compaq Research (formerly DEC Research) retired the EachMovie dataset. It is no longer available for download

# Sense-based profiles improve recommendations




Rating scale: 0-5

# Semantic Analysis: Ontologies in CBRS

| SYSTEM | DESCRIPTION |
|---|---|
| SEWeP (Semantic Enhancement for Web Personalization) [Eirinaki03] | Manually built domain-specific taxonomy of categories for the automated annotation of Web pages<br>WordNet-based word similarity used to map keywords to categories<br>Categories of interest discovered from navigational history of the user |
| Quickstep & Foxtrot [Middleton04] | Recommendation of on-line academic research papers<br>Research paper topic ontology based on the computer science classification of the DMOZ open directory project<br>K-NN classification used to associate classes to previously browsed papers |

[Lops10] P. Lops, M. de Gemmis, G. Semeraro. Content-based Recommender Systems: State of the Art and Trends. In: P. Kantor, F. Ricci, L. Rokach and B. Shapira (Eds.), *Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners*, Chapter 3, pages 73-105, BERLIN: Springer, 2010.

# Semantic Analysis: Ontologies in CBRS

| SYSTEM | DESCRIPTION |
| --- | --- |
| Informed Recommender [Aciar07] | Consumer product reviews to make recommendations<br>Ontology used to convert consumers' opinions into a structured form<br>Text-mining for mapping sentences in the reviews into the ontology information structure<br>Search-based recommendations |
| RS for Interactive Digital Television [Blanco-Fernandez08] | OWL ontology for representing TV programs and user profiles<br>OWL representation allows reasoning on preferences and discovering new knowledge<br>Spreading activation for matching items and preferences |
| News@hand [Cantador08] | Ontology-based news recommender<br>17 ontologies adapted from the IPTC ontology (`http://nets.ii.uam.es/neptuno/iptc/`)<br>Items and user profiles represented as vectors in the space of concepts defined by the ontologies |

# Semantic Analysis: Wikipedia

- Do we really need only ontologies?
  - ✓ What about encyclopedic knowledge sources available on the Web?
- Is Wikipedia potentially useful for CBRS? How?
  - ✓ It is free
  - ✓ It covers many domains
  - ✓ It is under constant development by the community
  - ✓ It can be seen as a multilingual corpus
  - ✓ Its accuracy rivals that of Encyclopaedia Britannica [Giles05]

[Giles05] J. Giles. Internet Encyclopaedias Go Head to Head. *Nature*, 438:900–901, 2005.

# Explicit Semantic Analysis (ESA)

Technique able to provide a fine-grained semantic representation of natural language texts in a high-dimensional space of comprehensible concepts derived from Wikipedia [Gabri06]


[Egozi09] O. Egozi. *Concept-Based Information Retrieval using Explicit Semantic Analysis*. M.Sc. Thesis, CS Dept., Technion, 2009.

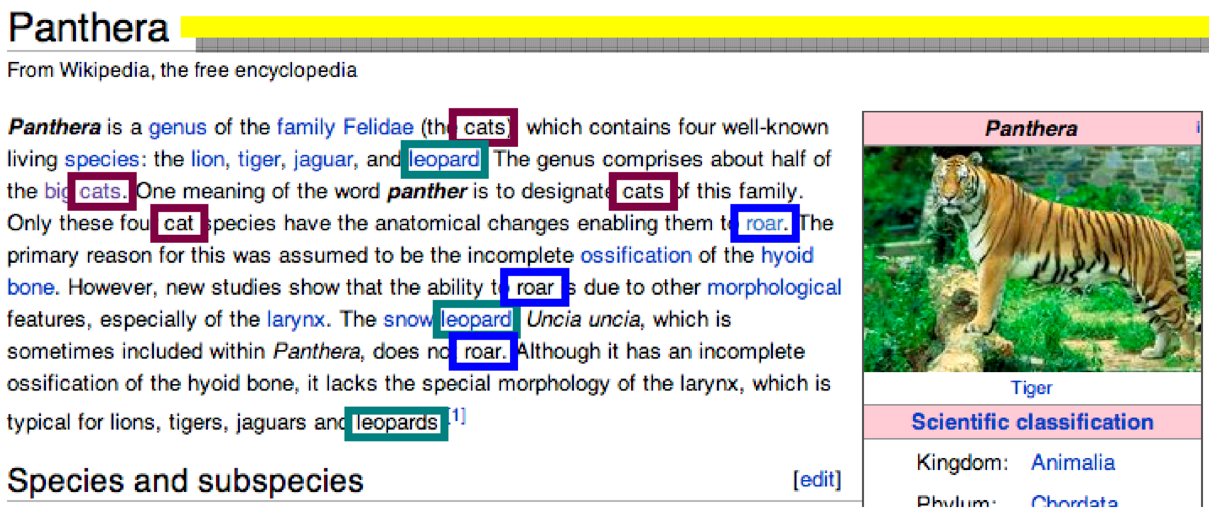
Wikipedia viewed as an **ontology** = a collection of ~**1M** concepts

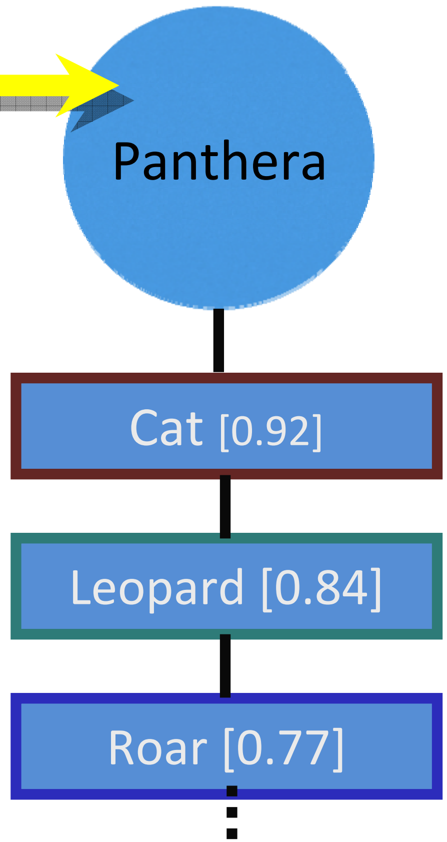
[Gabri06] E. Gabrilovich and S. Markovitch. Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge. In *Proceedings of the 21th National Conf. on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conference*, pages 1301–1306. AAAI Press, 2006.

# Explicit Semantic Analysis (ESA)

Wikipedia is viewed as an **ontology** - a collection of ~**1M** concepts

Every Wikipedia article represents a **concept**



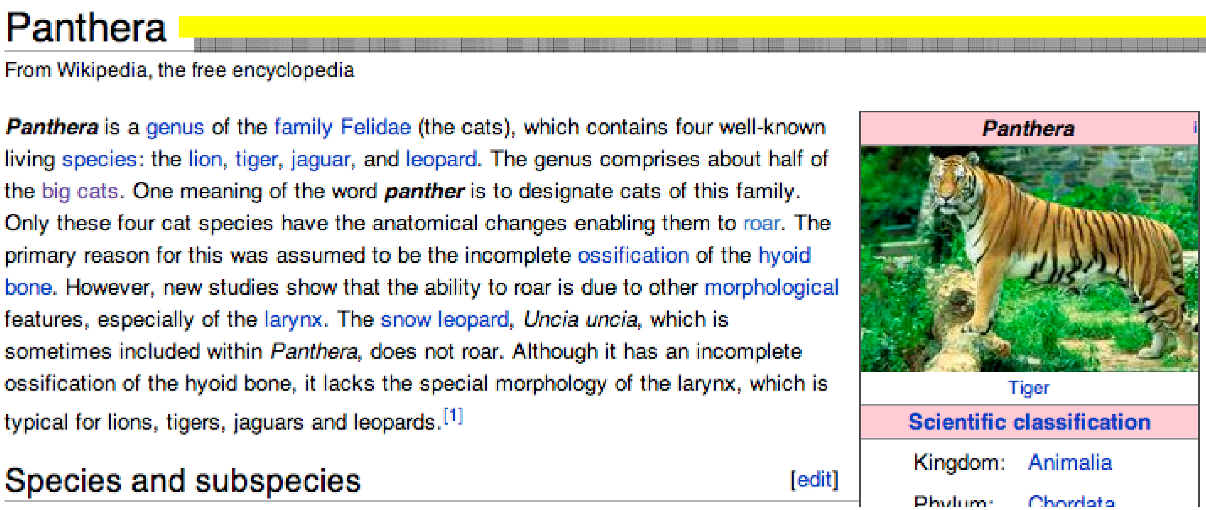




Article words are associated with the concept (TF-IDF)

# Explicit Semantic Analysis (ESA)

Wikipedia is viewed as an **ontology** - a collection of ~**1M** concepts

Every Wikipedia article represents a **concept**


Panthera

From Wikipedia, the free encyclopedia

*Panthera* is a genus of the family Felidae (the cats), which contains four well-known living species: the lion, tiger, jaguar, and leopard. The genus comprises about half of the big cats. One meaning of the word *panther* is to designate cats of this family. Only these four cat species have the anatomical changes enabling them to roar. The primary reason for this was assumed to be the incomplete ossification of the hyoid bone. However, new studies show that the ability to roar is due to other morphological features, especially of the larynx. The snow leopard, *Uncia uncia*, which is sometimes included within *Panthera*, does not roar. Although it has an incomplete ossification of the hyoid bone, it lacks the special morphology of the larynx, which is typical for lions, tigers, jaguars and leopards.[1]

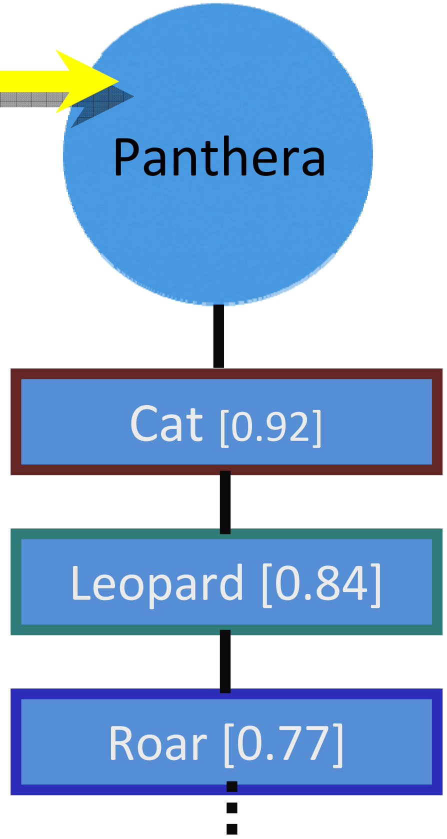
Species and subspecies [edit]

*Panthera*

Tiger

Scientific classification

Kingdom: Animalia




Article words are associated with the **concept** (TF-IDF)

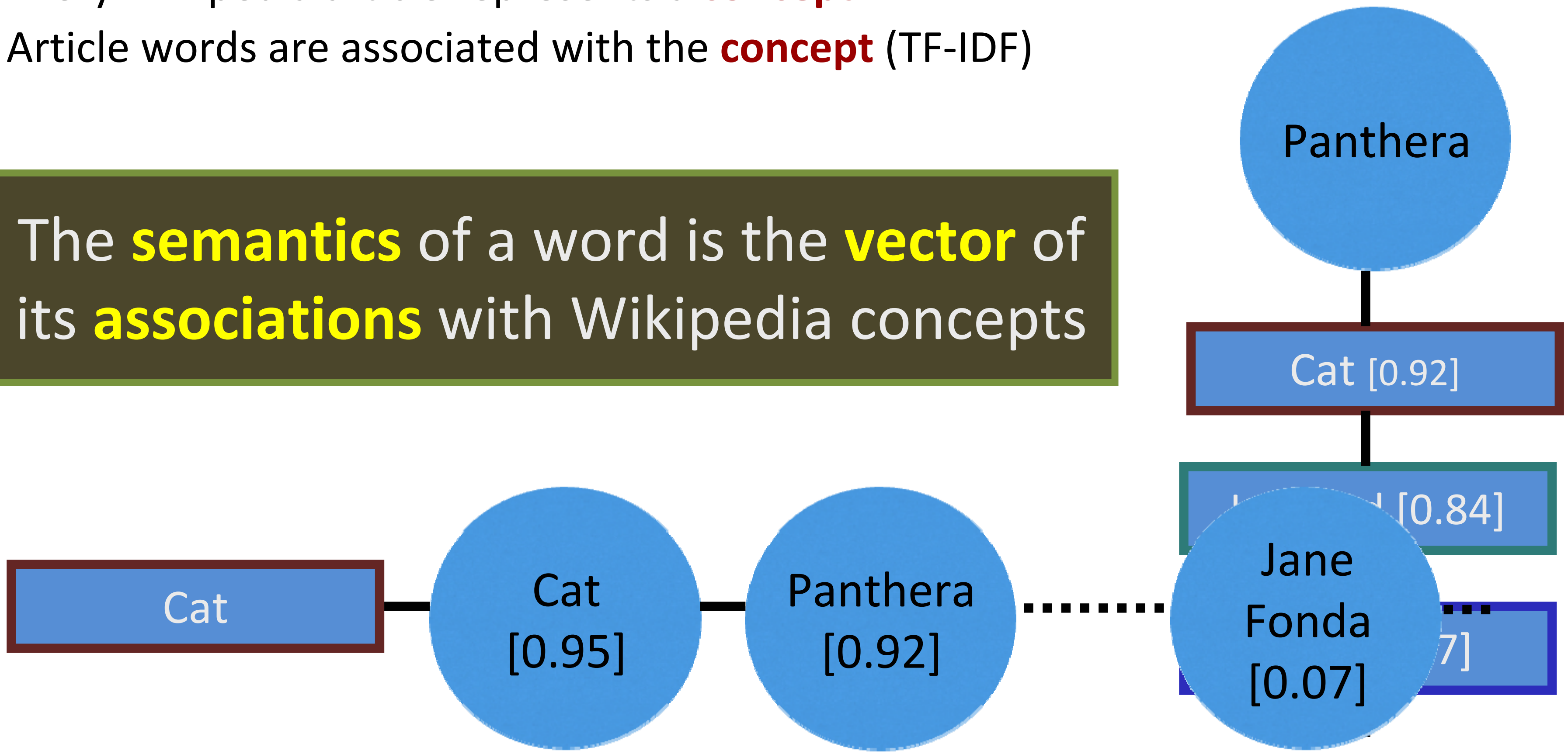
# Explicit Semantic Analysis (ESA)

Wikipedia is viewed as an **ontology** - a collection of ~**1M** concepts

Every Wikipedia article represents a **concept**

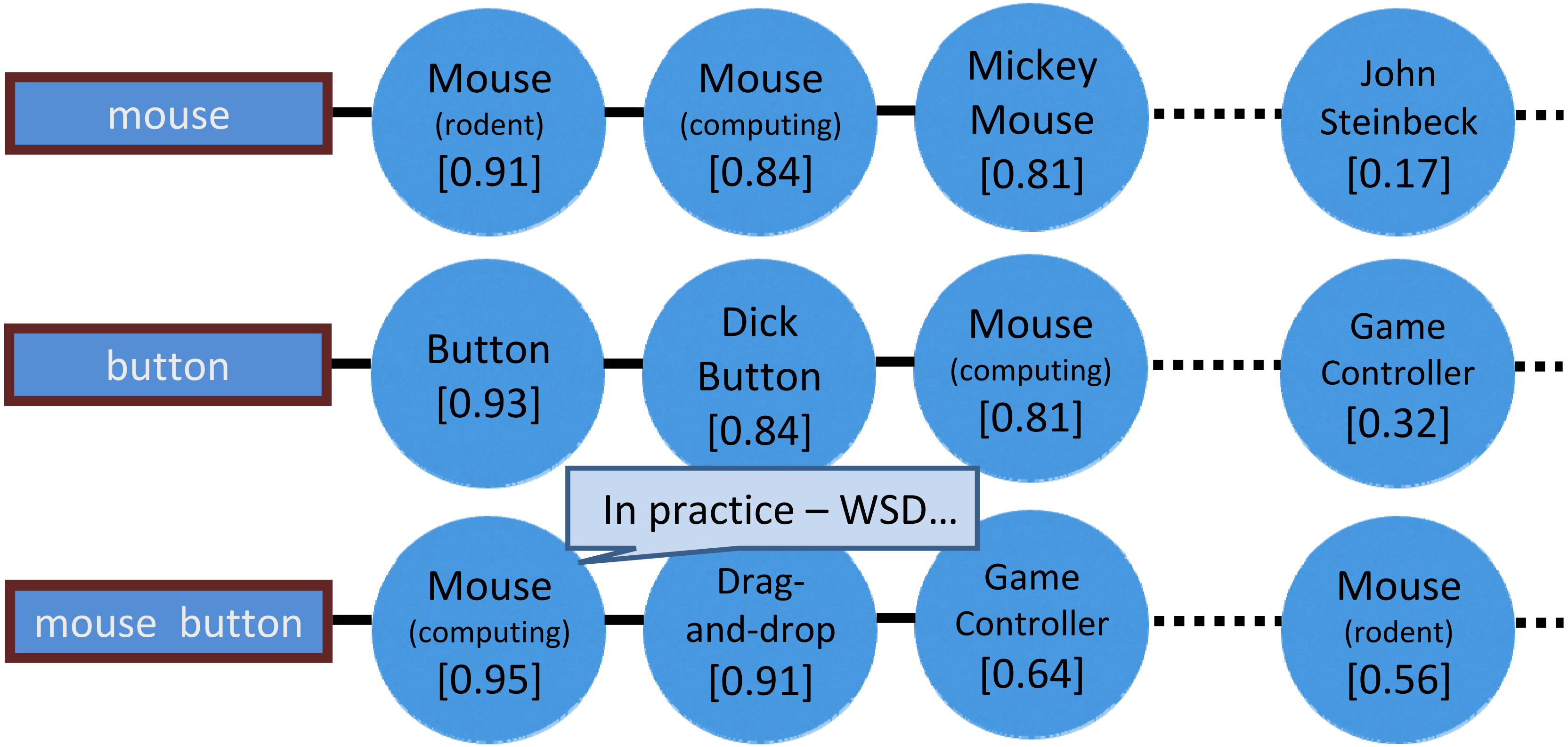
Article words are associated with the **concept** (TF-IDF)

The **semantics** of a word is the **vector** of its **associations** with Wikipedia concepts



# Explicit Semantic Analysis (ESA)

The semantics of a text fragment is the average vector (centroid) of the semantics of its words

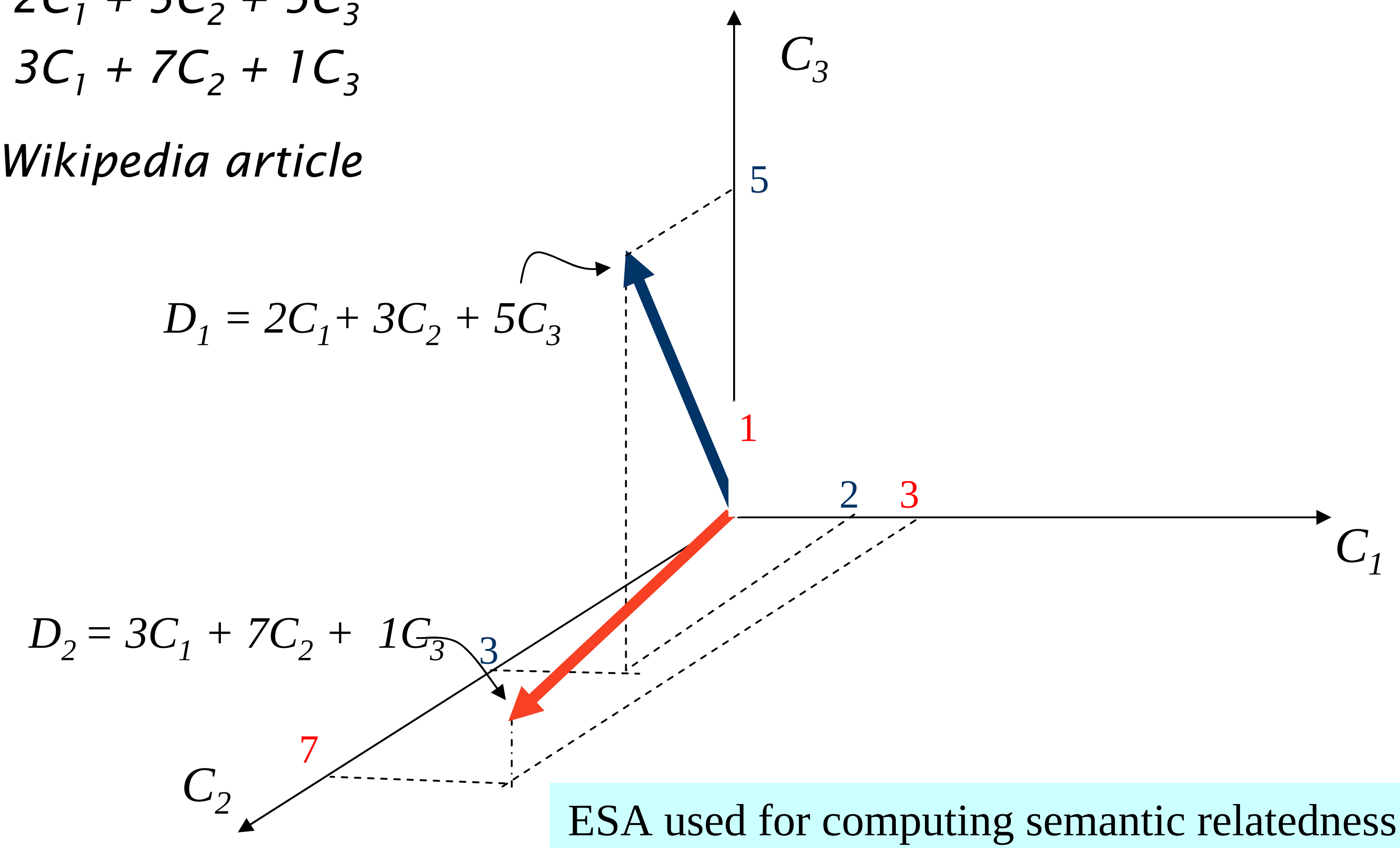


# ESA: concept space

$D_1 = 2C_1 + 3C_2 + 5C_3$

$D_2 = 3C_1 + 7C_2 + 1C_3$

$C_i$ = *Wikipedia article*




ESA used for computing semantic relatedness [Gabri07]

[Gabri07] E. Gabrilovich and S. Markovitch. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In Manuela M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1606–1611, 2007.

# Wikipedia and CBRS: recent ideas

- Wikipedia used for computing the similarity between movie descriptions for the Netflix prize competition [Lees08]
- ESA used for user profiling, spam detection and RSS filtering [Smirnov08]


- Wikipedia included in a Knowledge Infusion process for recommendation diversification [Semeraro09a]

[Lees08] J. Lees-Miller, F. Anderson, B. Hoehn, and R. Greiner. Does Wikipedia Information Help Netflix Predictions? *Proceedings of the Seventh International Conference on Machine Learning and Applications* (*ICMLA*), pages 337–343. IEEE Computer Society, 2008.

[Smirnov08] A. V. Smirnov and A. Krizhanovsky. Information Filtering based on Wiki Index Database. *CoRR, abs/0804.2354*, 2008.

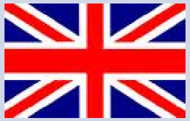
[Semeraro09a] G. Semeraro, P. Lops, P. Basile, and M. de Gemmis. Knowledge Infusion into Content-based Recommender Systems. In *Proceedings of the 2009 ACM Conference on Recommender Systems*, *RecSys 2009*, pages 301-304, New York, USA, October 22-25, 2009.

# Putting Intelligence into CBRS: Challenges & Research Directions

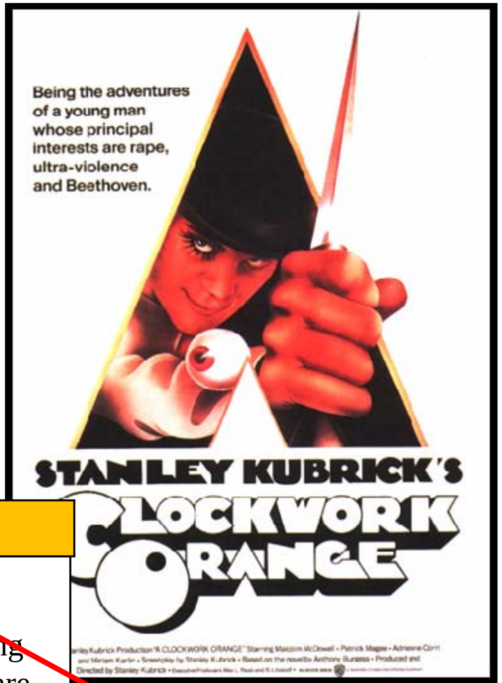
| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ "computational" serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# MARS (MultilAnguage Recommender System)
## cross-language user profiles

- WSD for building language-independent user profiles
- MultiWordNet as sense repository
  - ✓ Multilingual lexical database that supports English, Italian, Spanish, Portuguese, Hebrew, Romanian, Latin
  - ✓ Alignment between synsets in the different languages
    - – Semantic relations imported and preserved

| Language | Synset | Gloss |
|---|---|---|
|  | *world, human race, humanity, humankind, human beings, humans, mankind, man* | all of the inhabitants of the earth |
|  | *mondo, umanità, uomo, genere umano, terra* | insieme degli abitanti della terra, il complesso di tutti gli esseri umani |

# MARS (MultilAnguage Recommender System)
## cross-language user profiles





**ENGLISH description**

CLOCKWORK ORANGE

Being the adventures of a young man whose principal interests are rape, ultra-violence and Beethoven

**Bag of Synset**

“a12889641” “**n5477412**”
“**n3652872**” “**a2584413**”
“n3255687” “a3225896”
“**n32256325**” “**n225784**”
“**n255632**” “**Beethoven**”

MULTI WordNet

**ITALIAN description**

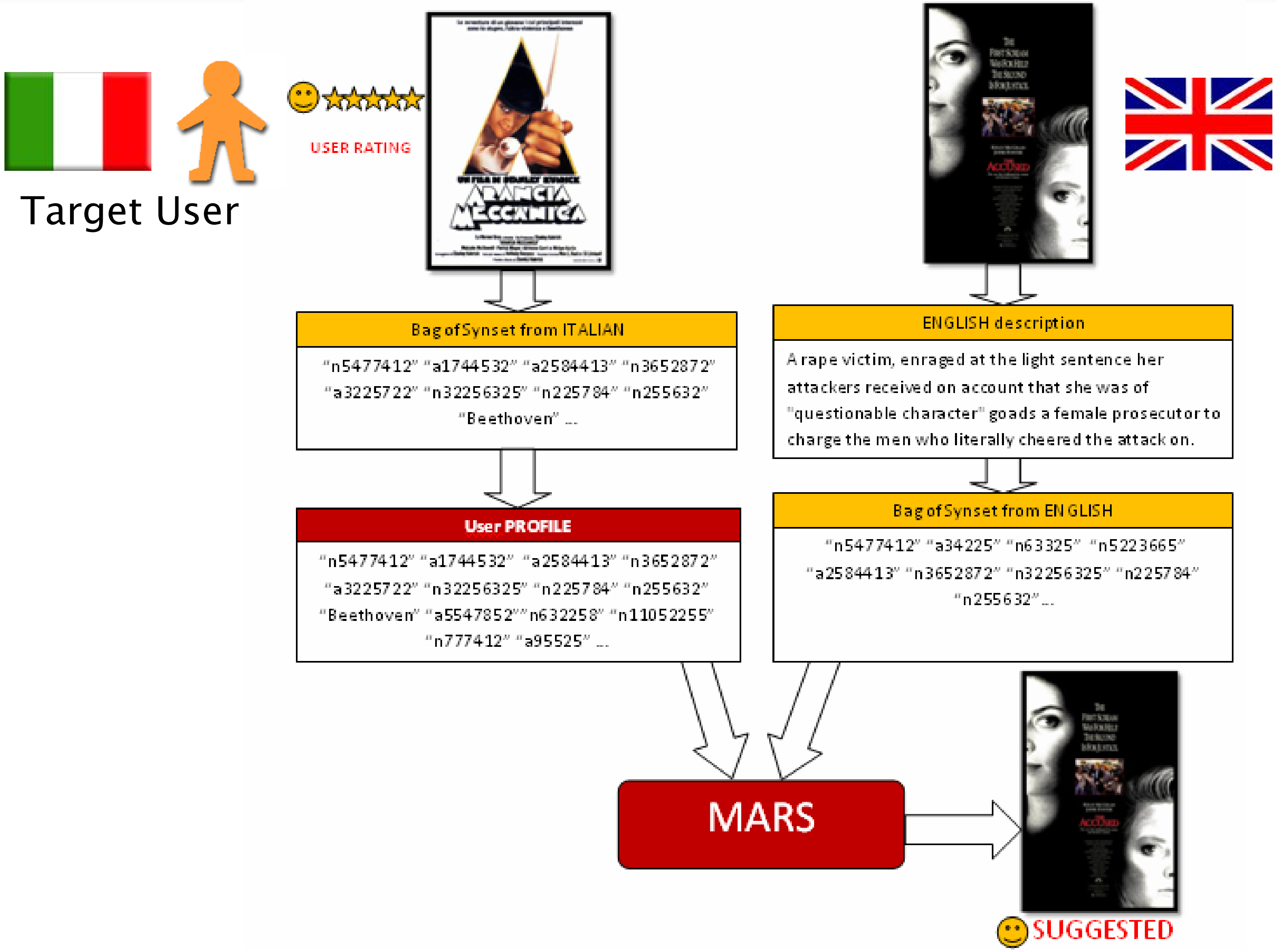
ARANCIA MECCANICA

Le avventure di un giovane i cui principali interessi sono lo stupro, l’ultra-violenza e Beethoven

**Bag of Synset**

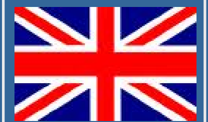
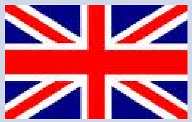
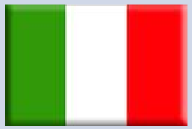
“**n5477412**” “a1744532”
“**a2584413**” “**n3652872**”
“a3225722” “**n32256325**”
“**n225784**” “**n255632**”
“**Beethoven**”

# MARS (MultilAnguage Recommender System)
## cross-language user profiles



# MARS (MultilAnguage Recommender System) preliminary results

- MovieLens 100k ratings dataset
- 613 users with ≥ 20 ratings selected from 943 different users
  - ✓ 520 movies and 40,717 ratings
  - ✓ movie content crawled from Wikipedia (English and Italian)
  - ✓ same movie - different descriptions in English and Italian
- Results in terms of $F_{ß=0.5}$ measure
  - ✓ no statistically significant difference wrt the baselines
- Neither content translations nor profile translations achieve the same effectiveness (they cannot avoid the negative impact of polysemy and lack of context)

| Profiles \ Recommendations |  |  |
|---|---|---|
|  | **64.91** | 63.98 |
|  | 63.70 | **63.71** |

# Putting Intelligence into CBRS: Challenges & Research Directions

| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ “computational” serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# Web 2.0 & User-Generated Content (UGC)

flickr BETA

myspace a place for friends

YouTube Broadcast Yourself

del.icio.us your bookmarks

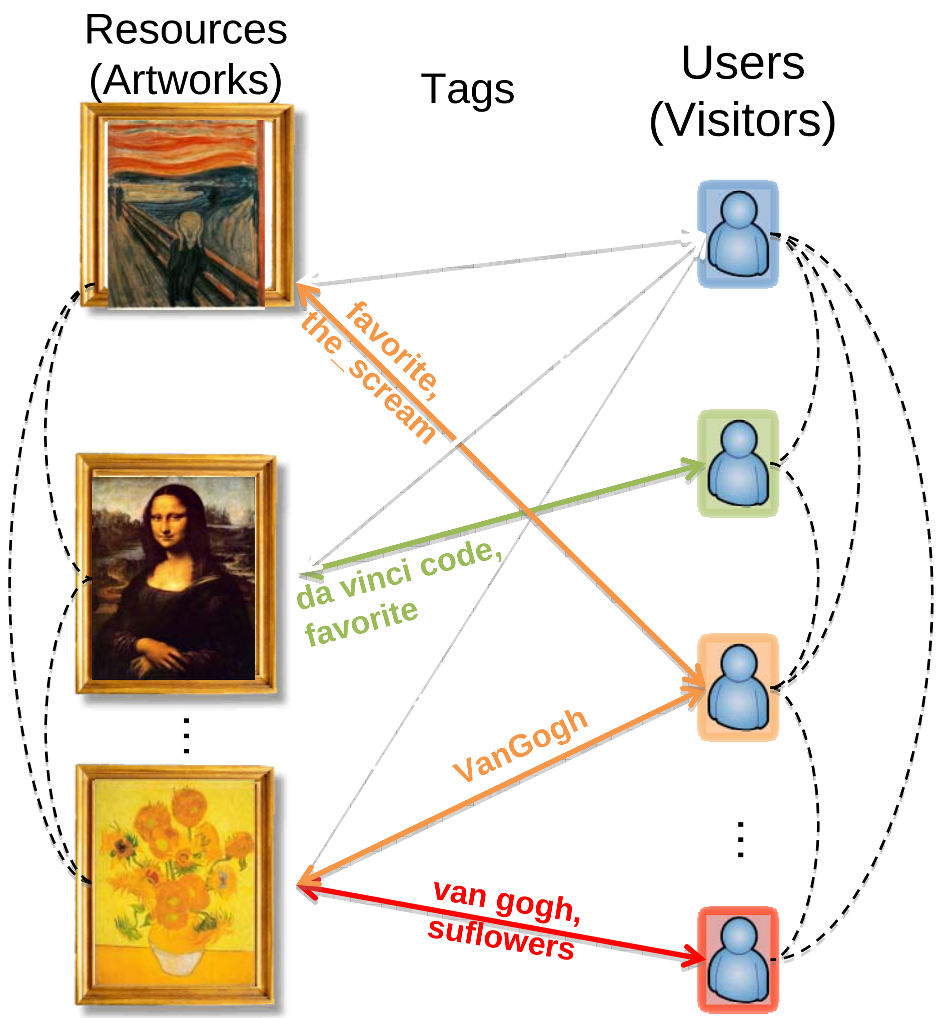
facebook

WordPress

LinkedIn

# Social Tagging & Folksonomies

- Users annotate resources of interests with free keywords, called ***tags***
- Social tagging activity builds a bottom-up classification schema, called a ***folksonomy***
  - Folksonomy: "Folks" + "Taxonomy"
- How to exploit folksonomies for advanced user profiling in CBRS?



CHAT

*C*ultural *H*eritage fruition & e-learning applications of new *A*dvanced (multimodal) *T*echnologies

ENGINEERING

Dipartimento di Informatica
Università degli Studi di Bari

*In the context of cultural heritage personalization, does the integration of UGC and textual description of artwork collections cause an increase of the prediction accuracy in the process of recommending artifacts to users?*

# FIRSt:
# Folksonomy-based Item Recommender syStem


- Artwork representation
  - ✓ Artist
  - ✓ Title
  - ✓ Description
  - ✓ Tags
- Semantic Indexing
  - ✓ Change of text representation from vectors of words (BOW) into vectors of WordNet synsets (BOS)
  - ✓ From tags to semantic tags
- Supervised Learning
  - ✓ Bayesian Classifier learned from artworks labeled with user ratings and tags

# FIRSt (Folksonomy-based Item Recommender syStem)
## Learning from Ratings & Tags

**27) Caravaggio - Deposition from the Cross**

**Textual description of items (static content)**

Descrizione dell'opera

The Deposition, considered one of Caravaggio's greatest masterpieces, was commissioned by Girolamo Vittrice for his famil chapel in S. Maria in Vallicella (Chiesa Nuova) in Rome. In 1797 it was included in the group of works transferred to Paris i execution of the Treaty of Tolentino. After its return in 1817 it became part of Pius VII's Pinacoteca. Caravaggio did not real portray the Burial or the Deposition in the traditional way, inasmuch as Christ is not shown at the moment when he is laid in t tomb, but rather when, in the presence of the holy women, he is laid by Nicodemus and John on the Anointing Stone, that is stone with which the sepulchre will be closed. Around the body of Christ are the Virgin, Mary Magdalene, John, Nicodemus and Mary of Cleophas, who raises her arms and eyes to heaven in a gesture of high dramatic tension. Caravaggio, who arriv in Rome towards 1592-93, was the protagonist of a real artistic revolution as regards the way of treating subjects and the us of colour and light, and was certainly the most important personage of the "realist" trend of seventeenth century painting.

**Social Tags**

Social Tags (from other users): caravaggio, deposition, christ, cross, suffering, religion

Inserisci il tuo voto e dei tag descrittivi (separati da una VIRGOLA, senza spazi)

**5-point rating scale**

1 ○ 2 ○ 3 ○ 4 ○ 5 ○

**Personal Tags**

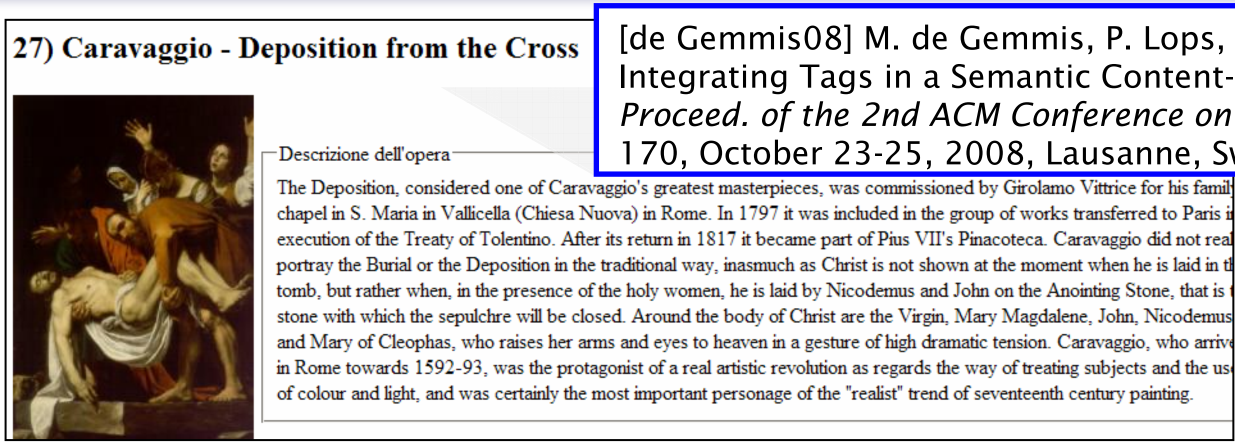
passion

Inserisci i voti e prosegui

# FIRSt (Folksonomy-based Item Recommender syStem)

## Tags within User Profiles

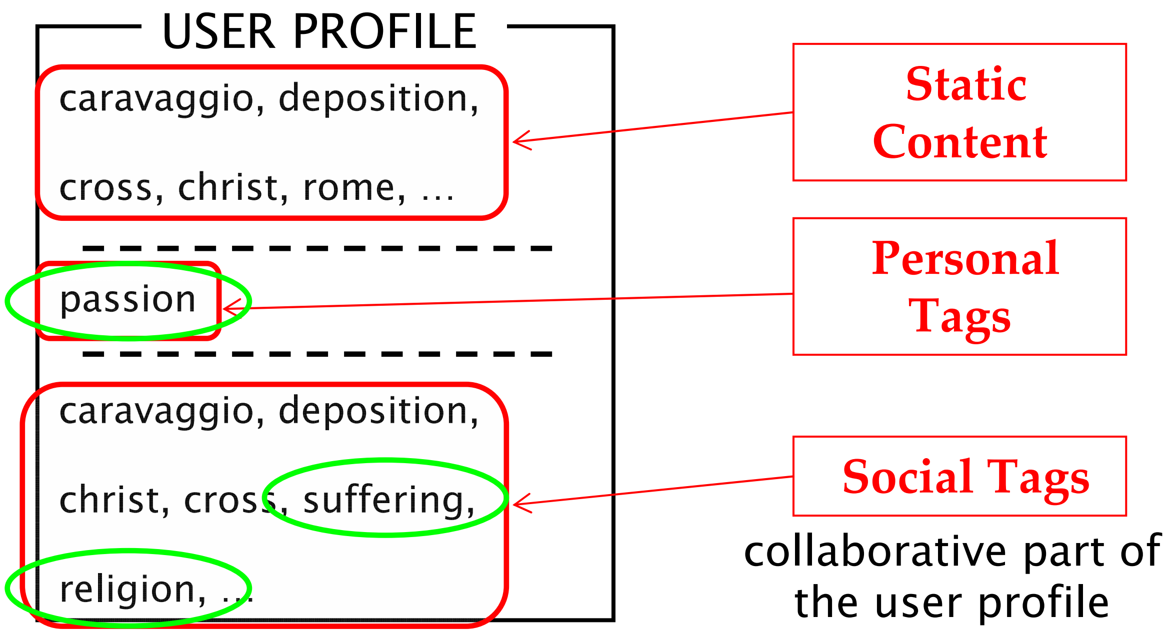
[de Gemmis08] M. de Gemmis, P. Lops, G. Semeraro, and P. Basile. Integrating Tags in a Semantic Content-based Recommender. In *RecSys '08, Proceed. of the 2nd ACM Conference on Recommender Systems*, pages 163–170, October 23-25, 2008, Lausanne, Switzerland, ACM, 2008.



**27) Caravaggio - Deposition from the Cross**

Descrizione dell'opera

The Deposition, considered one of Caravaggio's greatest masterpieces, was commissioned by Girolamo Vittrice for his famil chapel in S. Maria in Vallicella (Chiesa Nuova) in Rome. In 1797 it was included in the group of works transferred to Paris i execution of the Treaty of Tolentino. After its return in 1817 it became part of Pius VII's Pinacoteca. Caravaggio did not real portray the Burial or the Deposition in the traditional way, inasmuch as Christ is not shown at the moment when he is laid in t tomb, but rather when, in the presence of the holy women, he is laid by Nicodemus and John on the Anointing Stone, that is stone with which the sepulchre will be closed. Around the body of Christ are the Virgin, Mary Magdalene, John, Nicodemus and Mary of Cleophas, who raises her arms and eyes to heaven in a gesture of high dramatic tension. Caravaggio, who arriv in Rome towards 1592-93, was the protagonist of a real artistic revolution as regards the way of treating subjects and the us of colour and light, and was certainly the most important personage of the "realist" trend of seventeenth century painting.



# Experimental Evaluation

- Goal: Compare predictive accuracy of FIRSt when user profiles are learned from:
  - ✓ Static content only, i.e., textual descriptions of artifacts (content-based profiles)
  - ✓ both Static and Dynamic UGC (tag-based profiles). UGC can be:
    - Personal Tags, entered by a user for an artifact, i.e., the user's contribution to the whole folksonomy
    - Social Tags, i.e., the whole folksonomy of tags added by all visitors

# Experimental Setup

**Dataset**

1. 45 paintings from the Vatican picture-gallery
2. Static content (i.e., title, artist and description) captured using screenscraping bots

**Subjects**

1. 30 volunteers
2. average age ≈ 25
3. none reported to be an art expert

# Experimental Design

- 5 experiments designed
  - ✓ EXP#1: **Static Content**
  - ✓ EXP#2: **Personal Tags**
  - ✓ EXP#3: **Social Tags**
  - ✓ EXP#4: **Static Content + Personal Tags**
  - ✓ EXP#5: **Static Content + Social Tags**
- 5-fold cross validation
- Evaluation Metrics: Precision (Pr), Recall (Re), F1 measure

- One run for each user:
  1. Select the appropriate content depending on the experiment
  2. Split the selected data into a training set *Tr* and a test set *Ts*
  3. Use *Tr* for learning the corresponding user profile
  4. Evaluate the predictive accuracy of the induced profile on *Ts*

# Analysis of Precision

Content-based Profiles

Tag-based Profiles

Augmented Profiles

| Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|
| EXP#1: Static Content | 75.86 | 94.27 | 84.07 |
| EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| EXP#4: Static Content + Personal Tags | **78.04** | 93.60 | 85.11 |
| EXP#5: Static Content + Social Tags | **78.01** | 93.19 | 84.93 |

* Results averaged over the 30 study subjects

# Analysis of Precision

| | Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|---|
| Content-based Profiles | EXP#1: Static Content | 75.86 | 94.27 | 84.07 |
| Tag-based Profiles | EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| | EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| Augmented Profiles | EXP#4: Static Content + Personal Tags | **78.04** | 93.60 | 85.11 |
| | EXP#5: Static Content + Social Tags | **78.01** | 93.19 | 84.93 |

* Results averaged over the 30 study subjects

**Tag vs CB Precision not improved**

# Analysis of Precision

Content-based Profiles

Tag-based Profiles

Augmented Profiles

| Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|
| EXP#1: Static Content | 75.86 | 94.27 | 84.07 |
| EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| EXP#4: Static Content + Personal Tags | **78.04** | 93.60 | 85.11 |
| EXP#5: Static Content + Social Tags | **78.01** | 93.19 | 84.93 |

* Results averaged over the 30 study subjects

**Augmented vs CB Precision Improvement ≈ 2%**

# Analysis of Recall

| | Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|---|
| Content-based Profiles | EXP#1: Static Content | 75.86 | **94.27** | 84.07 |
| Tag-based Profiles | EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| | EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| Augmented Profiles | EXP#4: Static Content + Personal Tags | 78.04 | **93.60** | 85.11 |
| | EXP#5: Static Content + Social Tags | 78.01 | **93.19** | 84.93 |

* Results averaged over the 30 study subjects

**Tag vs CB**
**Recall decrease**
**1.62% – 3.77%**

# Analysis of Recall

| | Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|---|
| Content-based Profiles | EXP#1: Static Content | 75.86 | **94.27** | 84.07 |
| Tag-based Profiles | EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| | EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| Augmented Profiles | EXP#4: Static Content + Personal Tags | 78.04 | **93.60** | 85.11 |
| | EXP#5: Static Content + Social Tags | 78.01 | **93.19** | 84.93 |

* Results averaged over the 30 study subjects

**Augmented vs CB**
**Recall decrease:**
**0.67% – 1.08%**

# Analysis of F1

| | Type of Content | Precision* | Recall* | F1* |
|---|---|---|---|---|
| Content-based Profiles | EXP#1: Static Content | 75.86 | 94.27 | 84.07 |
| Tag-based Profiles | EXP#2: Personal Tags | 75.96 | 92.65 | 83.48 |
| | EXP#3: Social Tags | 75.59 | 90.50 | 82.37 |
| Augmented Profiles | EXP#4: Static Content + Personal Tags | 78.04 | 93.60 | **85.11** |
| | EXP#5: Static Content + Social Tags | 78.01 | 93.19 | **84.93** |

* Results averaged over the 30 study subjects

**Overall accuracy F1 ≈ 85%**

# Putting Intelligence into CBRS: Challenges & Research Directions

| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ "computational" serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# Serendipity: Definitions

❶ Serendipity

- ✓ Making discoveries, by accidents and sagacity, of things which one were not in quest of (Horace Walpole, 1754)
- ✓ The art of making an unsought finding (Pek van Andel, 1994) [vanAndel94]

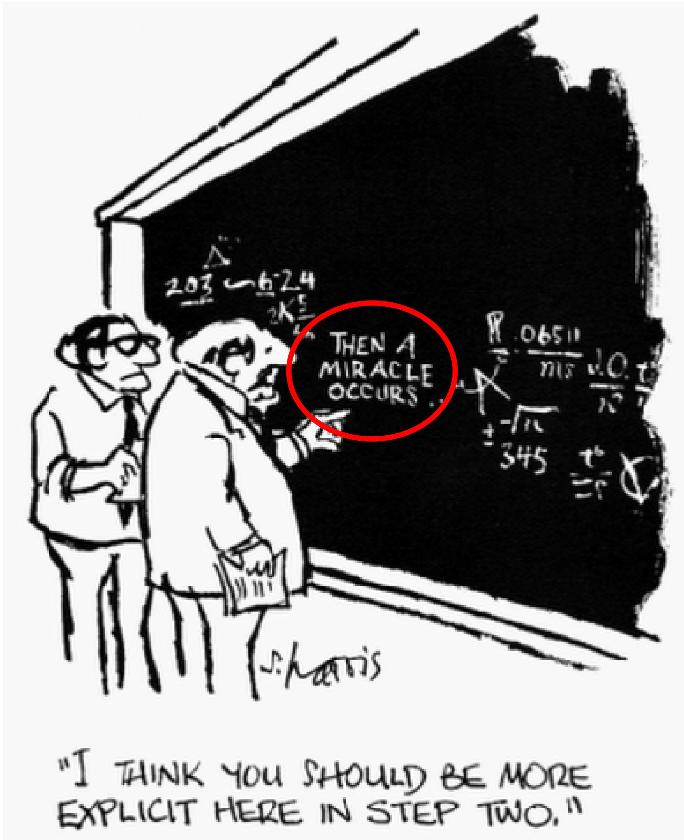
❷ Serendipitous ideas and findings

- ✓ Gelignite by Alfred Nobel, when he accidentally mixed collodium (gun cotton) with nitroglycerin
- ✓ Penicillin by Alexander Fleming
- ✓ The psychedelic effects of LSD by Albert Hofmann
- ✓ Cellophane by Jacques Brandenberger
- ✓ The structure of benzene by Friedric August Kekulé

[vanAndel94] van Andel, P. Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability. *The British Journal for the Philosophy of Science*, 45(2): 631-648, 994.

# The challenge

1. Serendipity in RSs is the experience of receiving an unexpected and fortuitous, but useful advice
   - ✓ it is a way to diversify recommendations
2. The challenge is programming for serendipity
   - ✓ to find a manner to introduce serendipity into the recommendation process in an operational way



# Strategies for *computational* serendipity [Toms00]

1. "Blind Luck": random recommendations
2. "Prepared Mind": Pasteur principle ("chance favors the prepared mind") - deep user modeling
3. "Anomalies and Exceptions": searching for dissimilarity [Iaquinta10]
4. "Reasoning by Analogy"


[Iaquinta10] L. Iaquinta, M. de Gemmis, P. Lops, G. Semeraro, P. Molino (2010). Can a Recommender System Induce Serendipitous Encounters? In: KYEONG KANG. *E-Commerce*, 229-246, VIENNA: IN-TECH, 2010.

[Toms00] Toms, E. Serendipitous Information Retrieval. In *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, Zurich, Switzerland: European Research Consortium for Informatics and Mathematics, 2000.

# Programming for Serendipity into CBRS: "Anomalies and Exceptions"

❶ Basic recommendation list defined by the best N items ranked according to the user profile

❷ Idea for inducing serendipity

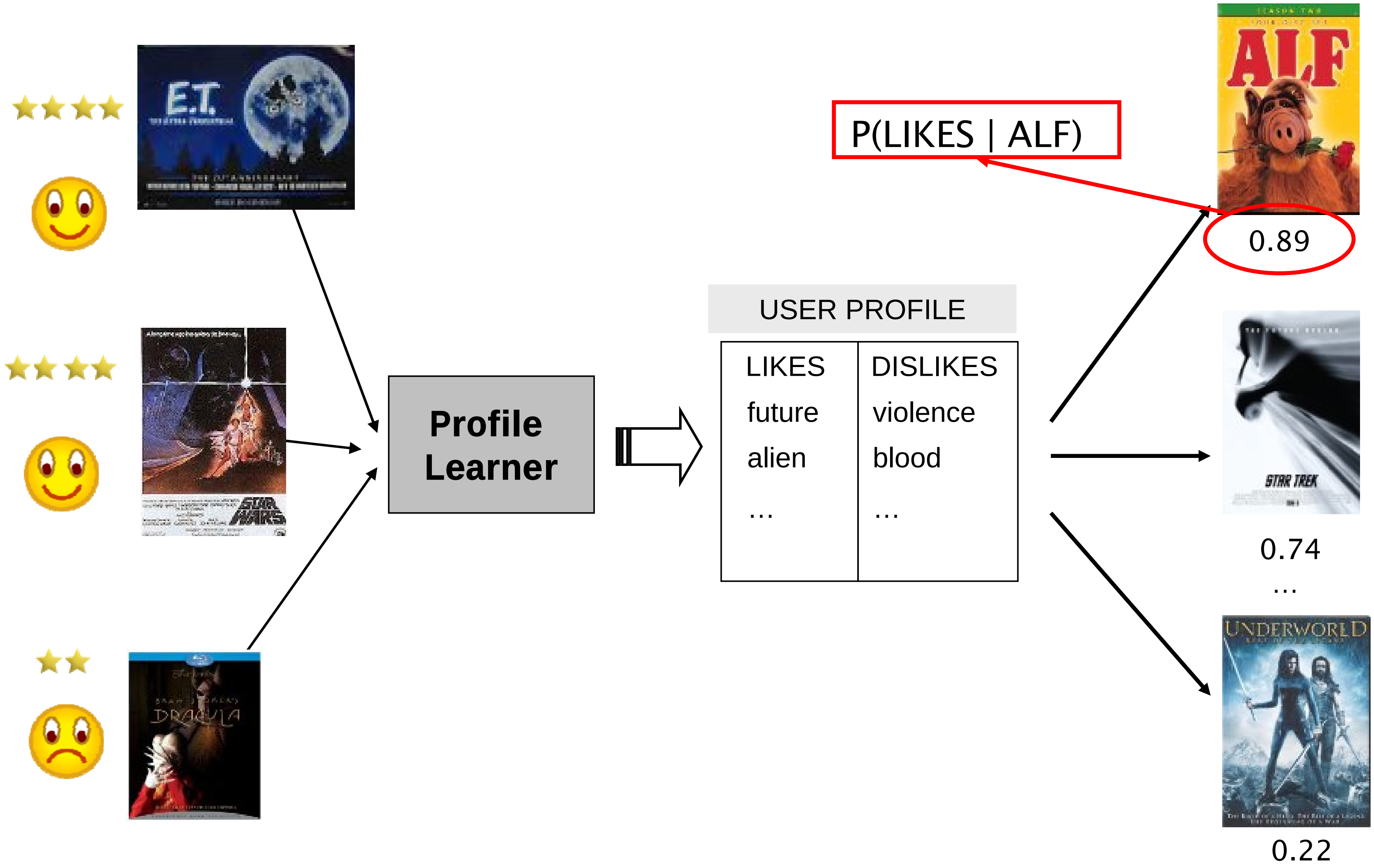
- ✓ extending the basic list with items programmatically supposed to be serendipitous for the active user

# ITem Recommender (ITR)

- Content-based recommender developed at Univ. of Bari [Semeraro07]
  - ✓ learns a probabilistic model of the interests of the user from textual descriptions of items
  - ✓ user profile = binary text classifier able to categorize items as interesting (LIKES) or not (DISLIKES)
  - ✓ a-posteriori probabilities as classification scores for LIKES and DISLIKES

[Semeraro07] G. Semeraro, M. Degemmis, P. Lops, and P. Basile. Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. In M. M. Veloso, editor, *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 2856–2861, Morgan Kaufmann, 2007.

# Recommendation process: Ranked list approach



# Programming for Serendipity into ITR: strategy

- Potentially serendipitous items selected on the ground of categorization scores for LIKES and DISLIKES
  - ✓ difference of classification scores tends to zero → uncertain classification

    $$| P(LIKES \mid ITEM) - P(DISLIKES \mid ITEM) | \approx 0$$

  - ✓ assumption:

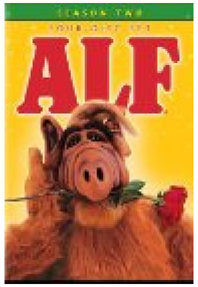
uncertain classification ≡ items not known by the user

# Programming for Serendipity into ITR: example

1. Basic recommendation list = *N* most interesting items
2. Ranked list of "unpredictable" items obtained from ITR
3. Basic recommendation list augmented with some serendipitous items

USER PROFILE

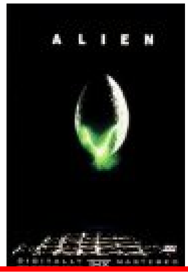
| LIKES | DISLIKES |
|---|---|
| future | violence |
| alien | blood |
| … | … |

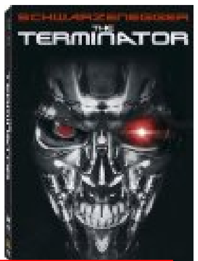





0.89 0.76 0.72

P(LIKES | ITEM)




0.01 0.02

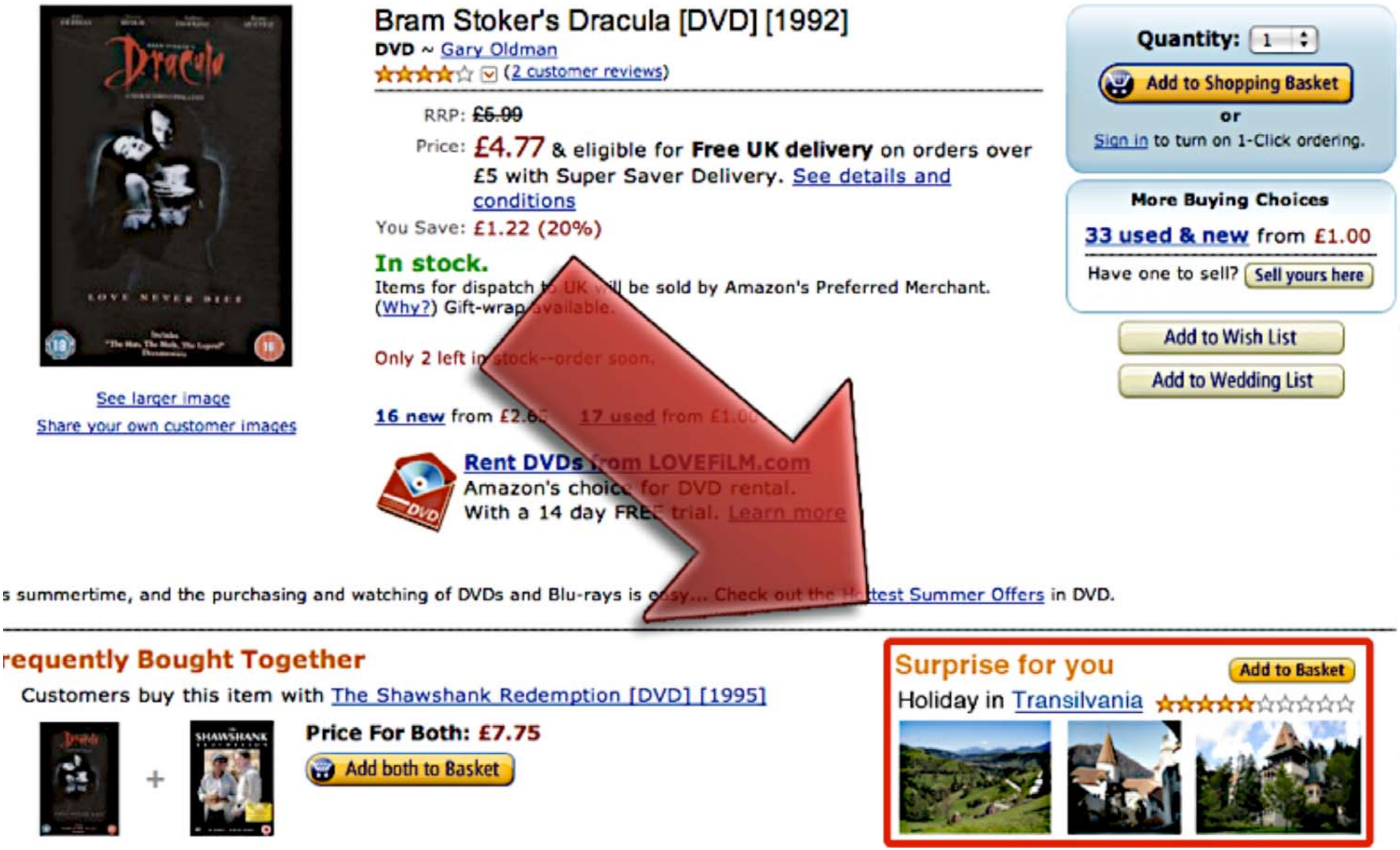
| P(LIKES | ITEM) – P(DISLIKES | ITEM) |

# What about evaluation?

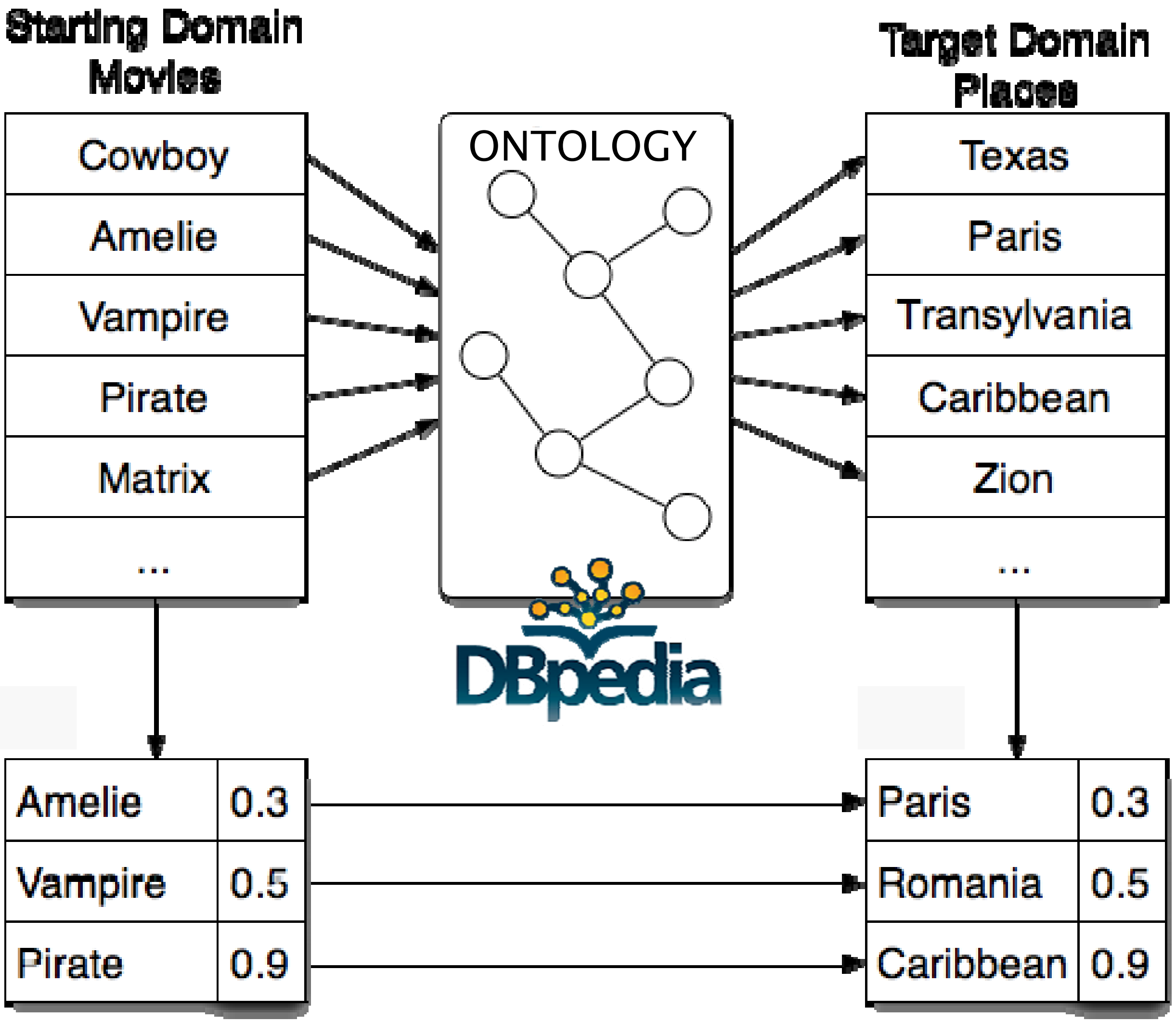
1. Classic evaluation metrics (Precision, Recall, F, MAE,…) don’t take into account obviousness, novelty and serendipity
   - ✓ Accurate recommendation $\neq$ Useful recommendation
   - ✓ emotional response associated with serendipity difficult to capture by conventional accuracy metrics
   - ✓ serendipity degree impossible to evaluate without considering user feedback

2. Novel metrics required
   - ✓ planned as a future work

# Programming for Serendipity: cross-domain recommendations



# "Reasoning by Analogy": a serendipity strategy for cross-domain recommendations




user profile for Movies

"parallel" user profile for Travels

# Ongoing work: DEVIUS

1. Analogy engine for computing “parallel” user profiles
   - ✓ Spreading activation on DBpedia for mapping between domains
2. Open source code of DEVIUS available in September
3. Experimental evaluation
   - ✓ books / movies

# Putting Intelligence into CBRS: Challenges & Research Directions

| PROBLEMS | CHALLENGES | RESEARCH DIRECTIONS |
|---|---|---|
| Limited Content Analysis | Beyond keywords: novel strategies for the representation of items and profiles | ▪ Semantic analysis of content by means of external knowledge sources<br>▪ Language-independent CBRS |
| | Taking advantage of Web 2.0 for collecting User Generated Content | Folksonomy-based CBRS |
| Overspecialization | Defeating homophily: recommendation diversification | ▪ "computational" serendipity → programming for serendipity<br>▪ Knowledge Infusion |

# Knowledge Infusion (KI)

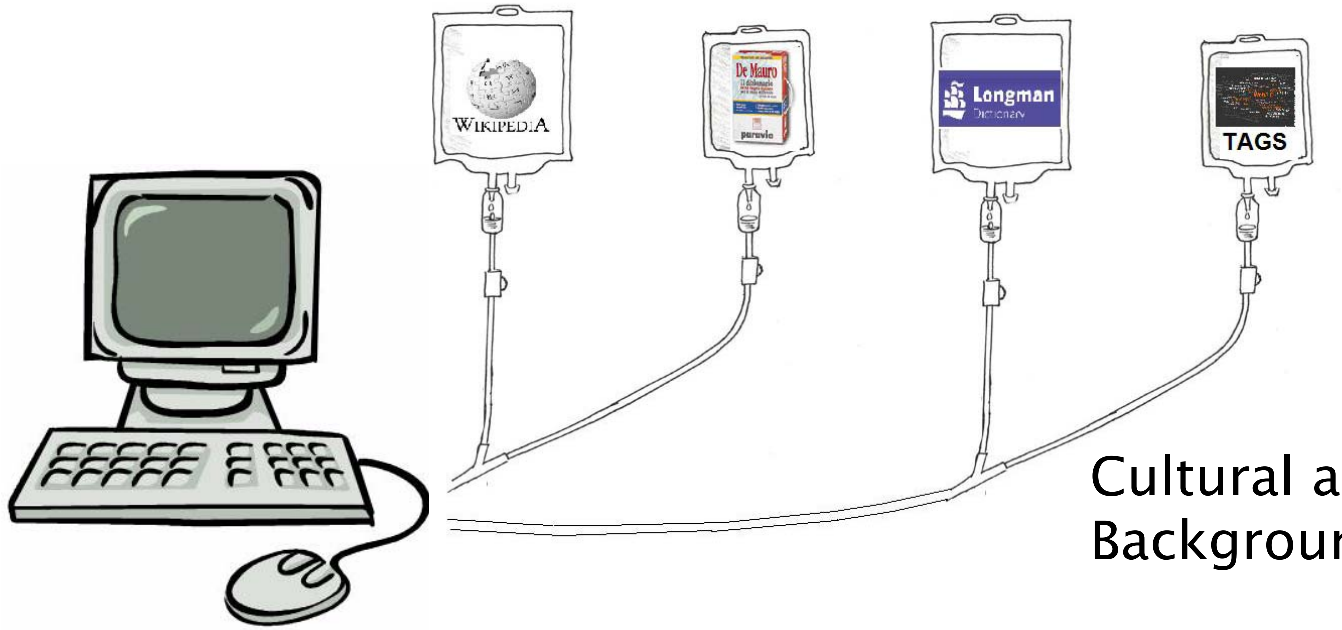
1. Humans typically have the *linguistic* and *cultural* experience to comprehend the meaning of a text
   - ✓ How to realize this *capability* into machines?
2. In NLP tasks, computers require access to vast amounts of common-sense and domain-specific world knowledge
   - ✓ Infusing lexical knowledge → Dictionaries (e.g. WordNet)
   - ✓ Infusing cultural knowledge→ Wikipedia
   - ✓ …

# Enhancing CBRS by KI

❶ Modeling the unstructured information stored in several (open) knowledge sources

❷ Exploiting the acquired knowledge in order to better understand the item descriptions and extract more meaningful features

❸ Inspired by a language game: The Guillotine [Semeraro09b]




Cultural and Linguistic Background Knowledge

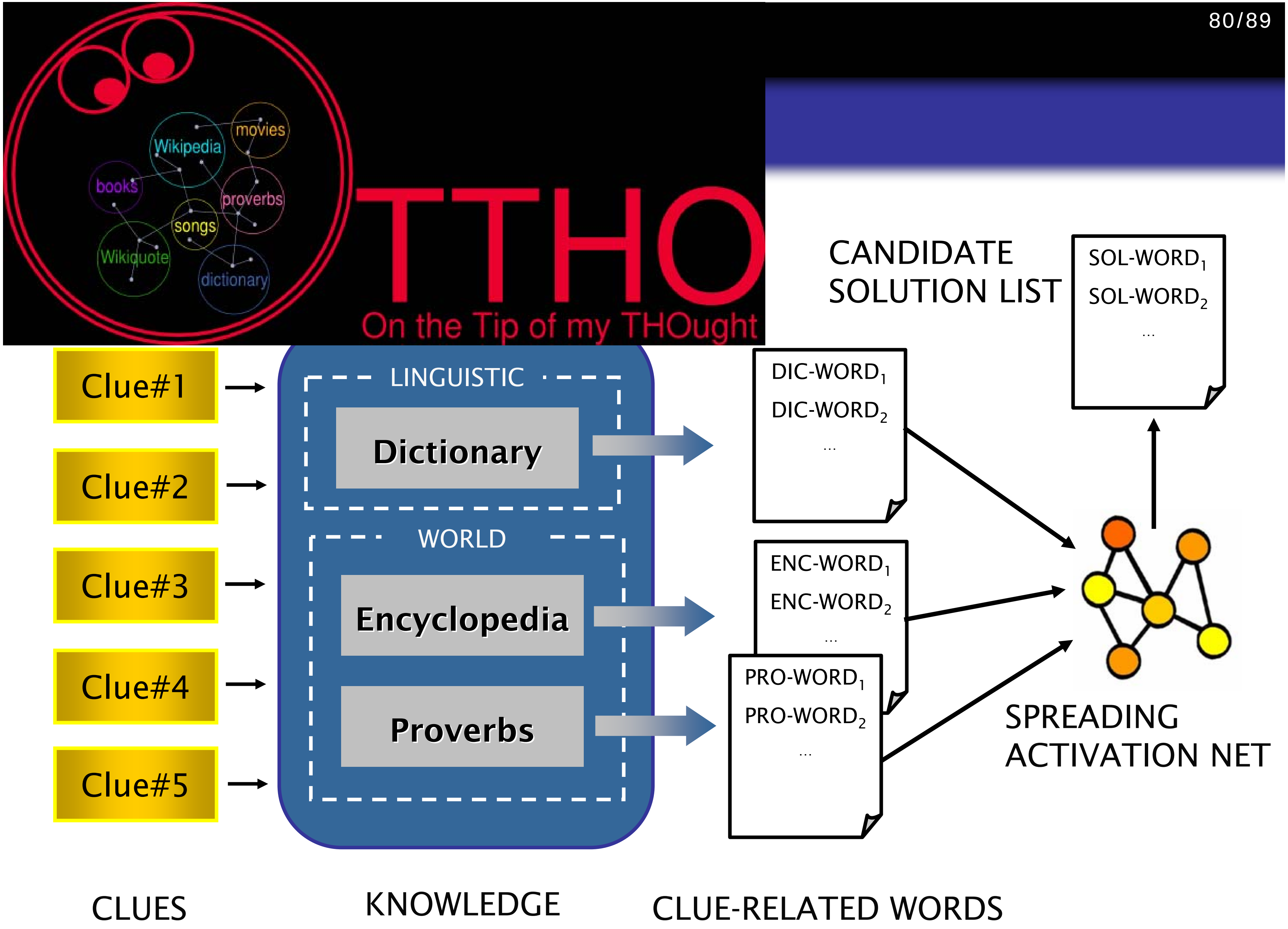
[Semeraro09b] G. Semeraro, P. Lops, P. Basile, and M. de Gemmis. On the Tip of my Thought: Playing the Guillotine Game. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, 1543-1548, Morgan Kaufmann, 2009.

# The Guillotine: the game


[Lops09] P. Lops, P. Basile, M. de Gemmis and G. Semeraro. "Language Is the Skin of My Thought": Integrating Wikipedia and AI to Support a Guillotine Player. In: R. Serra, R. Cucchiara (Eds.), *AI*IA 2009: Emergent Perspectives in Artificial Intelligence*, XIth International Conference of the Italian Association for Artificial Intelligence, Reggio Emilia, Italy, December 9-12, 2009. LNCS 5883, 324-333, Springer 2009.

# Let’s try to play the game

| | |
|---|---|
| APPLE | “An apple a day takes the doctor away” |
| JUDGMENT | Day of Judgment |
| SUNRISE | Beginning of the day |
| INDEPENDENCE | Independence day |
| SLEEPER | Daysleeper, a famous song by R.E.M. |


OTTHO
On the Tip of my THOught
Wikipedia
movies
books
proverbs
songs
Wikiquote
dictionary
CANDIDATE SOLUTION LIST
SOL-WORD$_1$
SOL-WORD$_2$
...
Clue#1
Clue#2
Clue#3
Clue#4
Clue#5
LINGUISTIC
Dictionary
WORLD
Encyclopedia
Proverbs
DIC-WORD$_1$
DIC-WORD$_2$
...
ENC-WORD$_1$
ENC-WORD$_2$
...
PRO-WORD$_1$
PRO-WORD$_2$
...
SPREADING ACTIVATION NET
CLUES
KNOWLEDGE
CLUE-RELATED WORDS

# What does OTTHO know about 'stars'?

STAR

KNOWLEDGE

LINGUISTIC
Dictionary

WORLD
Encyclopedia

User Generated Content
SOCIAL

Lemmas

| | LIGHT | ... | SKY |
|---|---|---|---|
| ... | | | |
| STAR | 0.55 | | 1.45 |
| ... | | | |

DICTIONARY MATRIX

SKY 1.45
LIGHT 0.55
...


**Lemma:** Definitions | *Compound Forms*

**Star:** any one of the distant bodies appearing as a point of light in the sky at night | *Fixed star,* i.e. one which is not a planet

Tags in items' tag cloud

| | ALIEN | ... | SPACE |
|---|---|---|---|
| ... | | | |
| STAR | 0.27 | | 1.41 |
| ... | | | |

TAG MATRIX

SPACE 1.41
ALIEN 0.27
...


"STAR, SPACE, ALIEN"

# KI@work for recommendation diversification


STAR

ROBOT

ALIEN

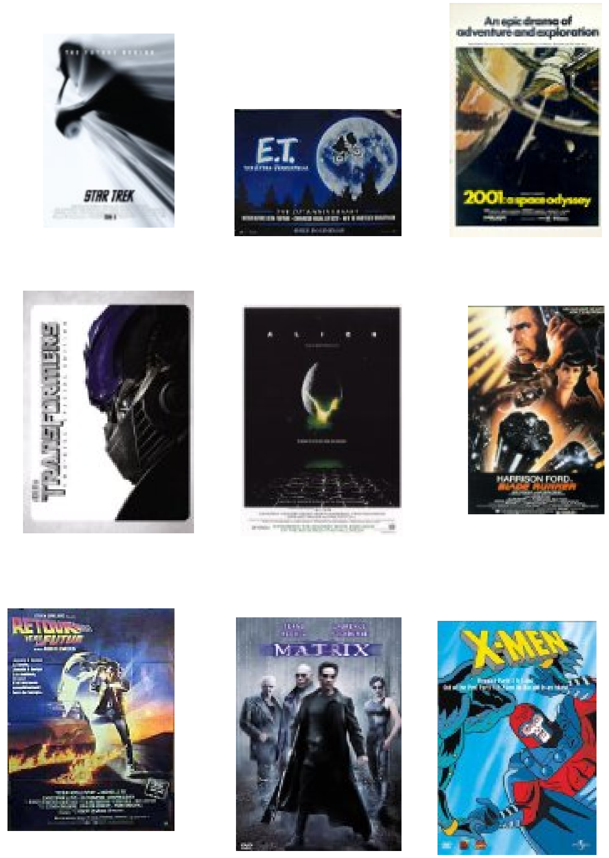
WAR

BATTLE

Plot Keywords

| | |
|---|---|
| SPACE | 0.36 |
| FUTURE | 0.10 |
| EXTRATERRESTRIAL | 0.08 |
| CYBORG | 0.07 |
| FIGHT | 0.02 |
| JUSTICE | 0.01 |
| ... | |

KI-LIST




Search Results

IMDb
The Internet Movie Database

# Concluding Remarks

❶ Research directions for overcoming some CBRS drawbacks

- ✓ main strategies adopted to introduce some semantics in the recommendation process
- ✓ main strategies for diversifying recommendations

❷ Research agenda: glean meaning and user thought from the precious boxes (brain, Web, social networks,...) they are hidden into:

- ✓ fMRI & Eye/Head-tracking technologies for a new generation of evaluation metrics
- ✓ Linked Open Data: interlinking user profiles with Semantic Web data and LOD
- ✓ Semantic Cross-system Personalization: semantic matching of user profiles coming from heterogeneous systems

# Thanks...

...for your attention...


...Questions?

**S**emantic
**W**eb
**A**ccess and
**P**ersonalization
research group
http://www.di.uniba.it/~swap

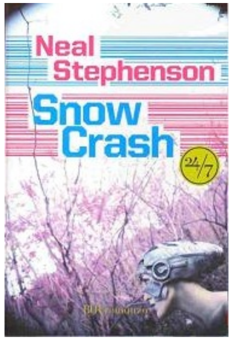
Pierpaolo Basile
Marco de Gemmis
Leo Iaquinta
Piero Molino
Fedelucio Narducci
Eufemia Tinelli
Annalina Caputo
Michele Filannino
Pasquale Lops
Cataldo Musto
Giovanni Semeraro

# Credits

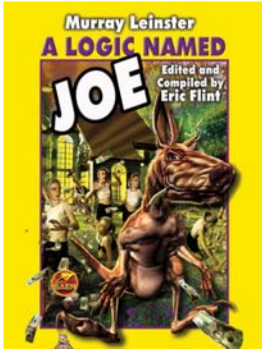
+ Arundhati Roy

+ Tullio De Mauro

+ Umberto Eco




+ The librarian




+ “A Logic Named Joe”


+ Ivonne Bordelois


+ Stefano Bartezzaghi “Accavallavacca”

+ Milena Jole Gabanelli

- Gaetano Bassolino & Emanuele Vizzini

# References 1/4

[Aciar07] S. Aciar, D. Zhang, S. Simoff, and J. Debenham. Informed Recommender: Basing Recommendations on Consumer Product Reviews. *IEEE Intelligent Systems*, 22(3):39–47, 2007.

[Basile07] P. Basile, M. Degemmis, A. Gentile, P. Lops, and G. Semeraro. UNIBA: JIGSAW algorithm for Word Sense Disambiguation. In *Proc.4th ACL 2007 International Workshop on Semantic Evaluations (SemEval-2007), Prague, Czech Republic*, 398–401, Association for Computational Linguistics, June 23-24, 2007.

[BlancoFernandez08] Y. Blanco-Fernandez, A. Gil-Solla, J. J. Pazos-Arias, M. Ramos-Cabrer, and M. Lopez-Nores. Providing Entertainment by Content-based Filtering and Semantic Reasoning in Intelligent Recommender Systems. *IEEE Trans. on Consumer Electronics*, 54(2):727–735, 2008.

[Cantador08] I. Cantador, A. Bellog´ın, and P. Castells. News@hand: A Semantic Web Approach to Recommending News. In Wolfgang Nejdl, Judy Kay, Pearl Pu, and Eelco Herderm (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems*, LNCS 5149, pages 279–283, Springer, 2008.

[Degemmis07] M. Degemmis, P. Lops, and G. Semeraro. A Content-collaborative Recommender that Exploits WordNet-based User Profiles for Neighborhood Formation. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research (UMUAI)*, 17(3):217–255, Springer Science + Business Media B.V., 2007.

[de Gemmis08] M. de Gemmis, P. Lops, G. Semeraro, and P. Basile. Integrating Tags in a Semantic Content-based Recommender. In *RecSys '08, Proc. of the 2nd ACM Conference on Recommender Systems,* pages 163–170, October 23-25, 2008, Lausanne, Switzerland, ACM, 2008.

[Eco07] U. Eco, *Sator arepo eccetera*. Bompiani, 2007 (in Italian).

# References 2/4

[Egozi09] O. Egozi. *Concept-Based Information Retrieval using Explicit Semantic Analysis*. M.Sc. Thesis, CS Department, Technion, 2009.

[Eirinaki03] LM. Eirinaki, M. Vazirgiannis, and I. Varlamis. SEWeP: Using Site Semantics and a Taxonomy to Enhance the Web Personalization Process. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 99–108, ACM, 2003.

[Gabri06] E. Gabrilovich and S. Markovitch. Overcoming the Brittleness Bottleneck using Wikipedia: Enhancing Text Categorization with Encyclopedic Knowledge. In *Proceed. of the 21th National Conf. on Artificial Intelligence and the 18th Innovative Applications of Artificial Intelligence Conf.*, pages 1301–1306, AAAI Press, 2006.

[Gabri07] E. Gabrilovich and S. Markovitch. Computing Semantic Relatedness Using Wikipedia-based Explicit Semantic Analysis. In Manuela M. Veloso, editor, *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, pages 1606–1611, 2007.

[Giles05] J. Giles. Internet Encyclopaedias Go Head to Head. *Nature*, 438:900–901, 2005.

[Herlocker04] Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems*, 22(1): 39-49, 2004.

[Iaquinta10] L. Iaquinta, M. de Gemmis, P. Lops, G. Semeraro, P. Molino (2010). Can a Recommender System Induce Serendipitous Encounters? In: KYEONG KANG. *E-Commerce*, 229-246, VIENNA: IN-TECH, 2010.

[Lees08] J. Lees-Miller, F. Anderson, B. Hoehn, and R. Greiner. Does Wikipedia Information Help Netflix Predictions? *Proceedings of the Seventh International Conference on Machine Learning and Applications* (*ICMLA*), pages 337–343, IEEE Computer Society, 2008.

# References 3/4

[Lops09] P. Lops, P. Basile, M. de Gemmis and G. Semeraro. "Language Is the Skin of My Thought": Integrating Wikipedia and AI to Support a Guillotine Player. In: R. Serra, R. Cucchiara (Eds.), *AI*IA 2009: Emergent Perspectives in Artificial Intelligence*, XIth International Conference of the Italian Association for Artificial Intelligence, Reggio Emilia, Italy, December 9-12, 2009. LNCS 5883, 324-333, Springer 2009.

[Lops10] P. Lops, M. de Gemmis, G. Semeraro. Content-based Recommender Systems: State of the Art and Trends. In: P. Kantor, F. Ricci, L. Rokach and B. Shapira, editors, *Recommender Systems Handbook: A Complete Guide for Research Scientists & Practitioners*, Chapter 3, pages 73-105, BERLIN: Springer, 2010.

[McNee06] S.M. McNee, J. Riedl, and J. Konstan. Accurate is not always good: How accuracy metrics have hurt recommender systems. In *Extended Abstracts of the 2006 ACM Conference on Human Factors in Computing Systems*, pages 1-5, Canada, 2006.

[Middleton04] S. E. Middleton, N. R. Shadbolt, and D. C. De Roure. Ontological User Profiling in Recommender Systems. *ACM Transactions on Information Systems*, 22(1):54–88, 2004.

[Pazzani07] Pazzani, M. J., & Billsus, D. Content-Based Recommendation Systems. *The Adaptive Web*. Lecture Notes in Computer Science vol. 4321, 325-341, 2007.

[Pedersen04] Pedersen, Ted and Patwardhan, Siddharth, and Michelizzi, Jason. WordNet::Similarity - Measuring the Relatedness of Concepts. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence* (AAAI-2004), pp. 1024-1025, San Jose, CA, July, 2004.

[Semeraro07] G. Semeraro, M. Degemmis, P. Lops, and P. Basile. Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. In M. M. Veloso, editor, *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, pages 2856–2861, Morgan Kaufmann, 2007.

# References 4/4

[Semeraro09a] G. Semeraro, P. Lops, P. Basile, and M. de Gemmis. Knowledge Infusion into Content-based Recommender Systems. In *Proceedings of the 2009 ACM Conf. on Recommender Systems*, *RecSys 2009*, pages 301-304, New York, USA, October 22-25, 2009.

[Semeraro09b] G. Semeraro, P. Lops, P. Basile, and M. de Gemmis. On the Tip of my Thought: Playing the Guillotine Game. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, 1543-1548, Morgan Kaufmann, 2009.

[Smirnov08] A. V. Smirnov and A. Krizhanovsky. Information Filtering based on Wiki Index Database. *CoRR, abs/0804.2354*, 2008.

[Toms00] Toms, E. Serendipitous Information Retrieval. In *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, Zurich, Switzerland: European Research Consortium for Informatics and Mathematics, 2000.

[vanAndel94] van Andel, P. Anatomy of the Unsought Finding. Serendipity: Origin, History, Domains, Traditions, Appearances, Patterns and Programmability. *The British Journal for the Philosophy of Science*, 45(2), pp. 631-648, 1994.

[Zuckerman08] E. Zuckerman. Homophily, serendipity, xenophilia. April 25, 2008. www.ethanzuckerman.com/blog/2008/04/25/homophily-serendipity-xenophilia/