Intelligent Information Filtering

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

Semantic Web Access and Personalization research group
Dipartimento di Informatica
Università degli Studi di Bari "Aldo Moro"

April 27, 2009
Sala delle Statue, Palazzo Rospigliosi
Roma

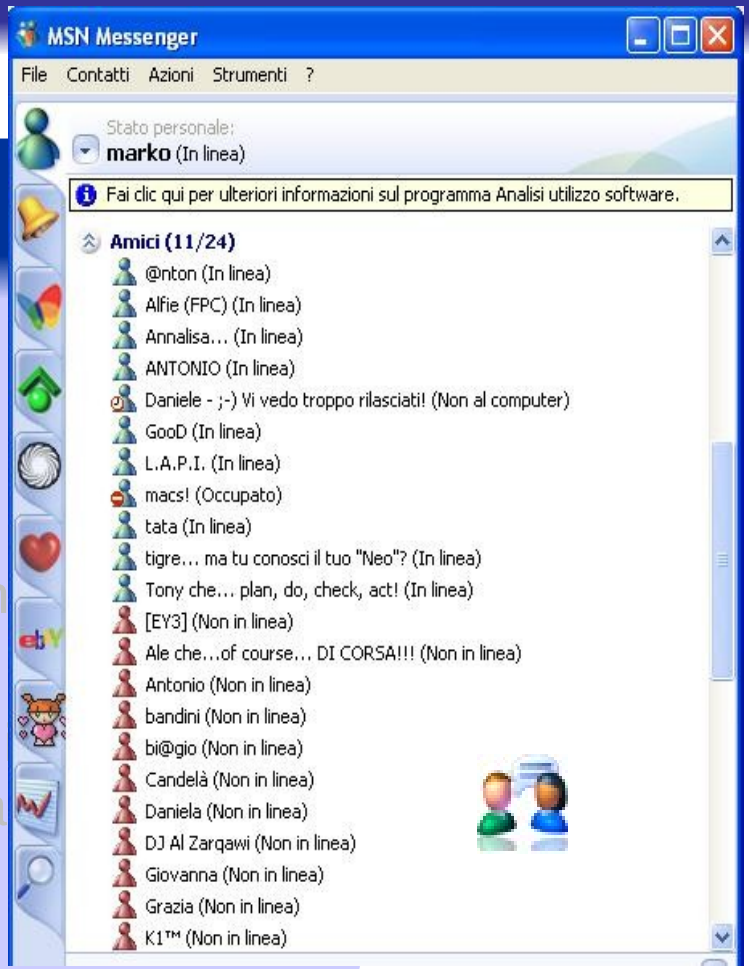
Outline

1. INTRODUCTION
   - ✓ The Information Overloading Problem
   - ✓ Personalization on the Web
2. INFORMATION SEEKING STRATEGIES
3. INFORMATION FILTERING
   - ✓ Collaborative Filtering
   - ✓ Content-based Filtering & User Profiling
4. IDEAS FOR INTELLIGENT INFORMATION FILTERING
5. CONCLUDING REMARKS
6. LIVE DEMO

Today's Information Society

People across the world…

- Chat
- Exchange e-mail, sms, pictures (mms)
- Buy products and services online
- Use search engines to find information their work and day-to-day life
- Exploit the Web for obtaining informat conventional sources like books, maga libraries



Today's Information Society

People across the world…

- Chat
- Exchange e-mail, sms, pictures
- Buy products and services online
- Use search engines to find information useful in their work and day-to-day life
- Exploit the Web for obtaining information instead of conventional sources like books, magazines and libraries

Today's Information Society

People across the world…

- Chat
- Exchange e-mail, sms, pictures
- Buy products and services online
- Use search engines to fi… their work and day-to-da…
- Exploit the Web for obta… conventional sources lik… libraries

Today's Information Society

People across the world…

- Chat
- Exchange e-mail, sms, pictures
- Buy products and services online
- Use search engines to find information useful in their work and day-to-day life
- Exploit the Web for obtaining information instead of co
  lib

Today's Information Society

People across the world…

- Chat
- Exchange e-mail, sms, pictures
- Buy products and services onlin
- Use search engines to find infor their work and day-to-day life
- Exploit the Web (instead of conventional sources like books, magazines and libraries) for obtaining information

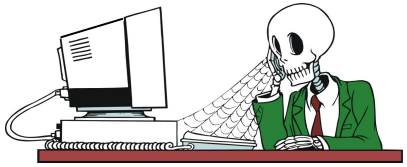
Information Overload

Problems…

- Explosion of irrelevant, unclear, inaccurate information
- Users overloaded with a large amount of information impossible to absorb

…and consequences

- Searching is time consuming
- Need for intelligent solutions able to support users in finding documents according to their interests



How Much Information?

1. Print, film, magnetic and optical storage media produced between 3.4 and 5.4 exabytes of unique information in 2002
   - ✓ 92% stored on magnetic media, mostly hard disks
   - ✓ 500-800 MB per person each year
2. Information flow through electronic channels - telephone, radio, TV, and the Internet – contained almost 18 exabytes of new information in 2002 (3 ½ more than recorded in storage media), 281 exabytes in 2007 (45 GB per person)

Medium	Terabytes
Radio	3,488
Television	68,955
Telephone	17,300,000
Internet	532,897
TOTAL	17,905,340

Medium	Terabytes
Surface Web	167
Hidden Web	91,850
E-mail	440,606
Inst. Messaging	274
TOTAL	532,897

Lyman, Peter and Hal R. Varian, *How Much Information*, School of Information Management and Systems, University of California at Berkeley, 2003. URL: http://www.sims.berkeley.edu/how-much-info-2003. Last access: May 23rd, 2007.

How to defend oneself from Information Overloading

Se si vuole trovare una metafora del rapporto fra l'uomo e i mezzi di comunicazione, Umberto Eco suggerisce quella dell'automobilista:

la tecnologia ha messo a disposizione vetture sempre più sofisticate, potenti e veloci; che vengano usate per portare una persona all'ospedale o per fare le gare di velocità sulle strade, dipende da chi è seduto al posto di guida.

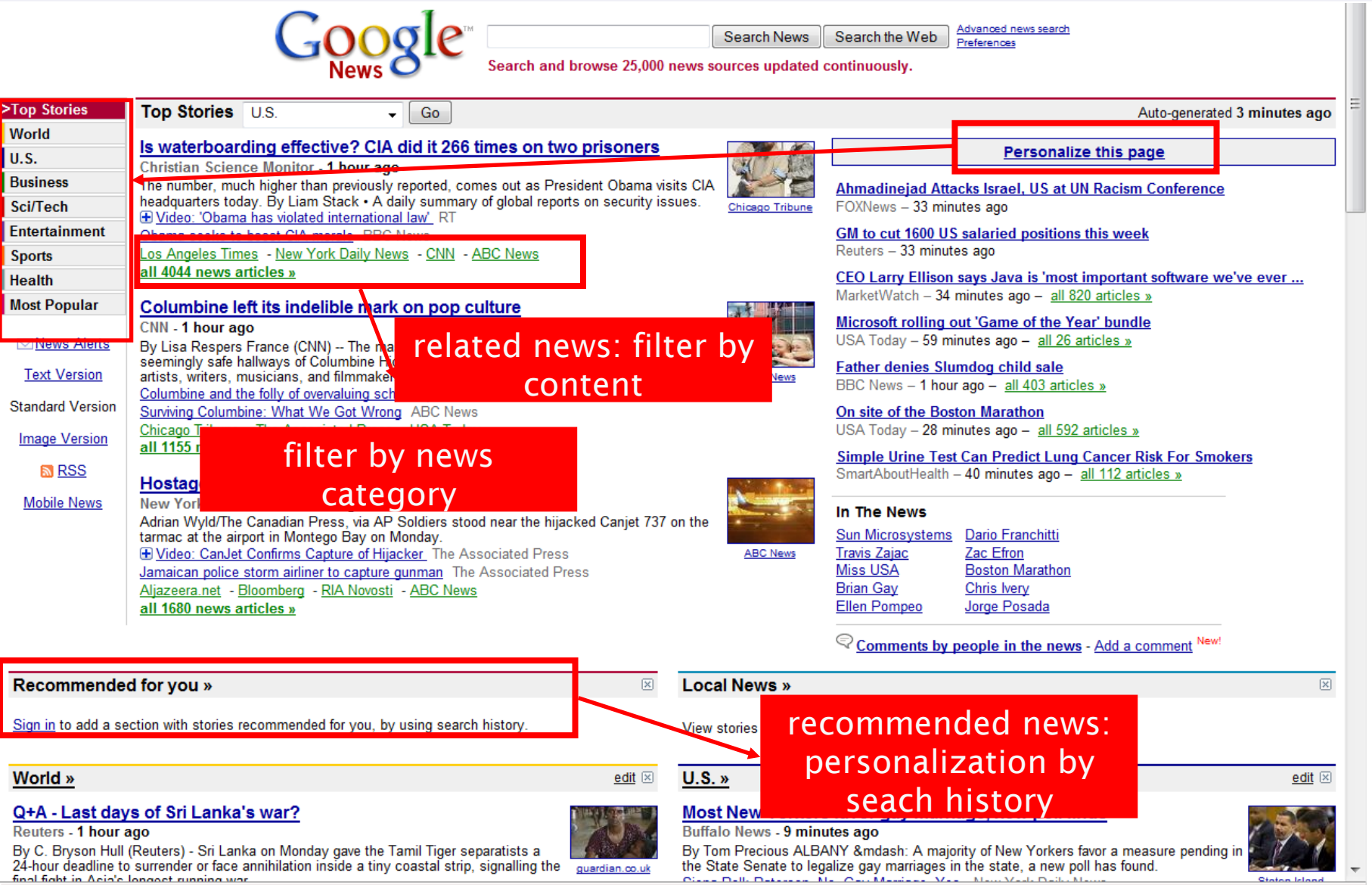
Lo stesso si può dire di quella che ormai è diventata una delle relazioni fondamentali della nostra vita quotidiana, cioè il nostro modo di interagire con i mass media, dalla televisione al telefonino, da Internet alla radio, dai libri ai cd e dvd (ebbene sì, anch'essi sono media), alla posta elettronica: dipende dalla cultura e dalla volontà di ciascuno di noi, educato soprattutto dalla scuola e dalla famiglia, mettere a punto una "dieta mediatica"
– suggerisce Gianfranco Bettetini – che non provochi né obesità né anoressia.

da: "Mettete a dieta i mass-media"

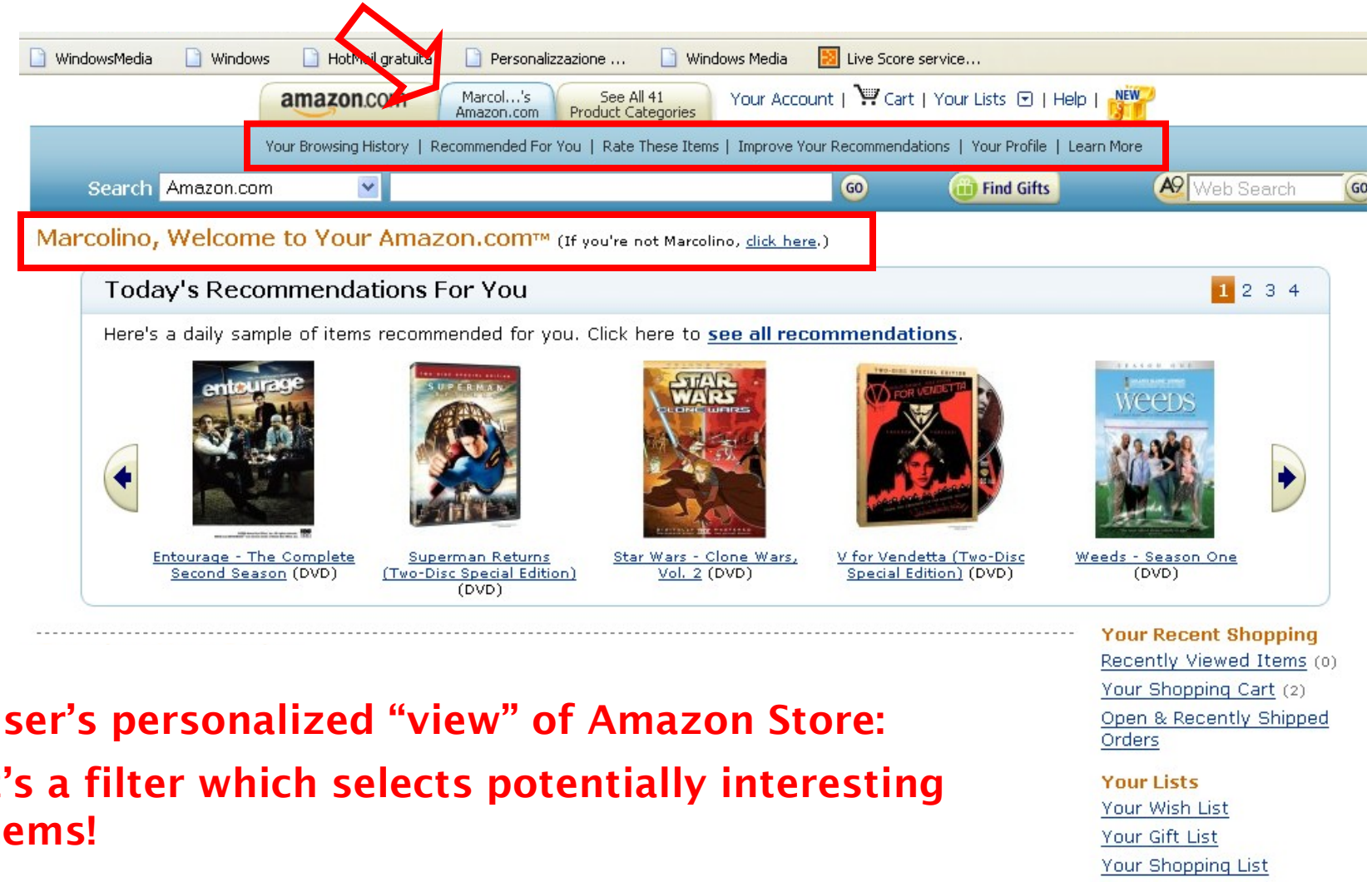
INTERVISTA A DUE VOCI CON UMBERTO ECO E GIANFRANCO BETTETINI

di Paolo Perazzolo, Famiglia Cristiana n.20 del 20-05-2007
(http://www.sanpaolo.org/fc/0720fc/0720fc54.htm)

How to defend oneself from Information Overload: My... Web - Google News



How to defend oneself from Information Overload: My… Web - Personalized Stores



User's personalized "view" of Amazon Store:

It's a filter which selects potentially interesting items!

Outline

1. INTRODUCTION
   - ✓ The Information Overloading Problem
   - ✓ Personalization on the Web
2. INFORMATION SEEKING STRATEGIES
3. INFORMATION FILTERING
   - ✓ Collaborative Filtering
   - ✓ Content-based Filtering & User Profiling
4. IDEAS FOR INTELLIGENT INFORMATION FILTERING
5. CONCLUDING REMARKS
6. LIVE DEMO

Popular Information Seeking strategies

Information Retrieval [Baeza-Yates and Ribeiro-Neto 1999]

- "Information Retrieval (IR) deals with the representation, storage, organization of, and access to information items"
- "…the user must first translate this information need into a *query* which can be processed by a search engine (or IR system)".
- "Given the user query, the key goal of an IR system is to retrieve information which might be useful or relevant to the user. The emphasis is on the retrieval of *information* as opposed to the retrieval of *data*".

Information Filtering [Hanani et al. 2001]

- "The aim of Information Filtering (IF) is to expose users to only the information that is relevant to them. Some examples of filtering applications are: filters for search results on the internet,… e-mail filters based on *personal profiles*, … filters for e-commerce applications that address products and promotions to potential customers only…"

Comparing IR and IF

Parameters	Information Retrieval	Information Filtering
representation of Information Needs	Queries	User Profiles
Goal	selection of items relevant for query	filtering out irrelevant items or collecting items
Frequency of use	ad hoc use one time users	repetitive use long term users
Type of Users	Not known to the system	"Profiled"
Database	(relatively) static	very large dynamic

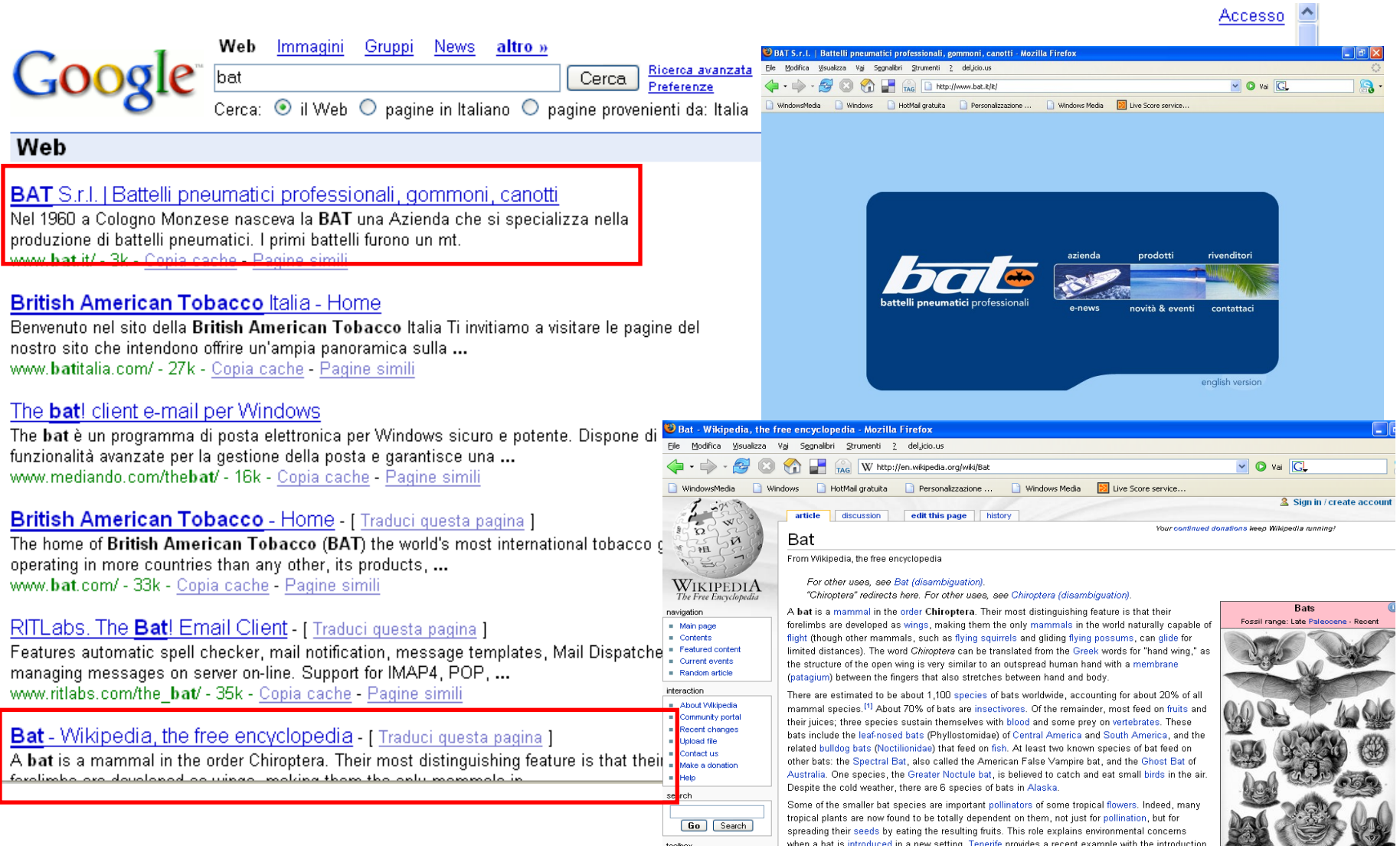
Table adapted from [Hanani et al. 2001]

Comparing IR and IF

Common Mechanisms

- **Representation:** Both the user's information need – query or profile – and the document set must be represented for *comparison*
- **Comparison:** String matching? Concept matching?
- **Feedback:** To improve the performance of the IR/IF system, a feedback mechanism is usually incorporated.

Some Problems in IR systems... Polysemy



Some Problems in IR systems... Synonymy

Google

Web Immagini Gruppi News altro »

chiropteran Cerca Ricerca avan Preferenze

Cerca: il Web pagine in Italiano pagine provenienti da: It

Web

Suggerimento: Cerca risultati solo in **Italiano**. Puoi specificare la lingua di ricerca su Preferenz

Chiropteran Flight - [Traduci questa pagina]
CHIROPTERAN FLIGHT. Bat. The Chiroptera, or bats, are the second most diverse group of mammals, and are the only mammals ever to evolve true powered flight. ...
www.ucmp.berkeley.edu/vertebrates/flight/bats.html - 7k - Copia cache - Pagine simili

chiropteran - definition of **chiropteran** by the Free Online ... - [Traduci
Definition of **chiropteran** in the Online Dictionary. Meaning of **chiropteran**. What d
chiropteran mean? **chiropteran** synonyms, **chiropteran** antonyms.
www.thefreedictionary.com/**chiropteran** - 30k - Copia cache - Pagine simili

Definition of **chiropteran** - Merriam-Webster Online Dictionary - [Tradu
Definition of **chiropteran** from the Merriam-Webster Online Dictionary with audio pronunciations, thesaurus, Word of the Day, and word games.
www.m-w.com/cgi-bin/dictionary?va=**chiropteran**s - 23k - Copia cache - Pagine sim

chiropteran - English-Italian Dictionary - WordReference.com
chiropteran - Dizionario Inglese Italiano online. ... Discussioni nei forum nel cui tito
presente la parola '**chiropteran**': ...
www.wordreference.com/enit/**chiropteran** - 9k - Copia cache - Pagine simili

chiropteran: Definition and Much More from Answers.com - [Traduci qu
chiropteran also chiropter n. A mammal, such as the bat, that is a member of the Chiroptera and has forelimbs modified as wings.
www.answers.com/topic/**chiropteran** - 35k - Copia cache - Pagine simili

Chiropteran Flight - Mozilla Firefox

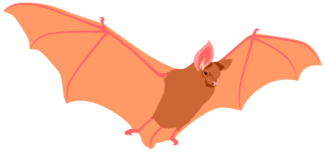
File Modifica Visualizza Vai Segnalibri Strumenti ? del.icio.us

http://www.ucmp.berkeley.edu/vertebrates/flight/bats.html

WindowsMedia Windows HotMail gratuita Personalizzazione ... Windows Media Live Score service...

Vertebrate Flight

CHIROPTERAN FLIGHT



The Chiroptera, or bats, are the second most diverse group of mammals, and are the only mammals ever to evolve true powered flight. The evolutionary origin of chiropterans is still somewhat of a mystery, because the fossil record of bats is scant. Some cladistic analyses indicate that bats are most closely related to the dermopterans, such as *Cynocephalus*, the colugo or "flying lemur" (which does not fly and is not a lemur). But others suggest that they are not so closely related. However, their ancestors may have been similar in some

Bat - Wikipedia, the free encyclopedia - Mozilla Firefox

File Modifica Visualizza Vai Segnalibri Strumenti ? del.icio.us

http://en.wikipedia.org/wiki/Bat

WindowsMedia Windows HotMail gratuita Personalizzazione ... Windows Media Live Score service...

Sign in / create account

article discussion edit this page history

Your continued donations keep Wikipedia running!

WIKIPEDIA
The Free Encyclopedia

navigation
- Main page
- Contents
- Featured content
- Current events
- Random article

interaction
- About Wikipedia
- Community portal
- Recent changes
- Upload file
- Contact us
- Make a donation
- Help

search
Go Search

Bat

From Wikipedia, the free encyclopedia

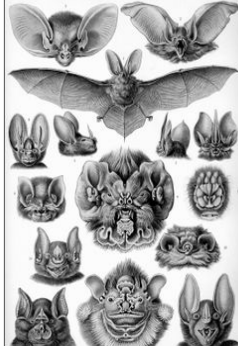
For other uses, see Bat (disambiguation).
"Chiroptera" redirects here. For other uses, see Chiroptera (disambiguation).

A **bat** is a mammal in the order **Chiroptera**. Their most distinguishing feature is that their forelimbs are developed as wings, making them the only mammal in the world naturally capable of flight (though other mammals, such as flying squirrels and gliding flying possums, can glide for limited distances). The word *Chiroptera* can be translated from Greek words for "hand wing," as the structure of the open wing is very similar to that of a human hand with a membrane (patagium) between the fingers and the arm stretching between hand and body.

There are estimated [illegible] species of bats worldwide, accounting for about 20% of all mammal species.[1] About 70% of bats are insectivores. Of the remainder, most feed on fruits and their juices; three species sustain themselves with blood and some prey on vertebrates. These bats include the leaf-nosed bats (Phyllostomidae) of Central America and South America, and the related bulldog bats (Noctilionidae) that feed on fish. At least two known species of bat feed on other bats: the Spectral Bat, also called the American False Vampire bat, and the Ghost Bat of Australia. One species, the Greater Noctule bat, is believed to catch and eat small birds in the air. Despite the cold weather, there are 6 species of bats in Alaska.

Some of the smaller bat species are important pollinators of some tropical flowers. Indeed, many tropical plants are now found to be totally dependent on them, not just for pollination, but for spreading their seeds by eating the resulting fruits. This role explains environmental concerns when a bat is introduced in a new setting. Tenerife provides a recent example with the introduction

Bats
Fossil range: Late Paleocene - Recent

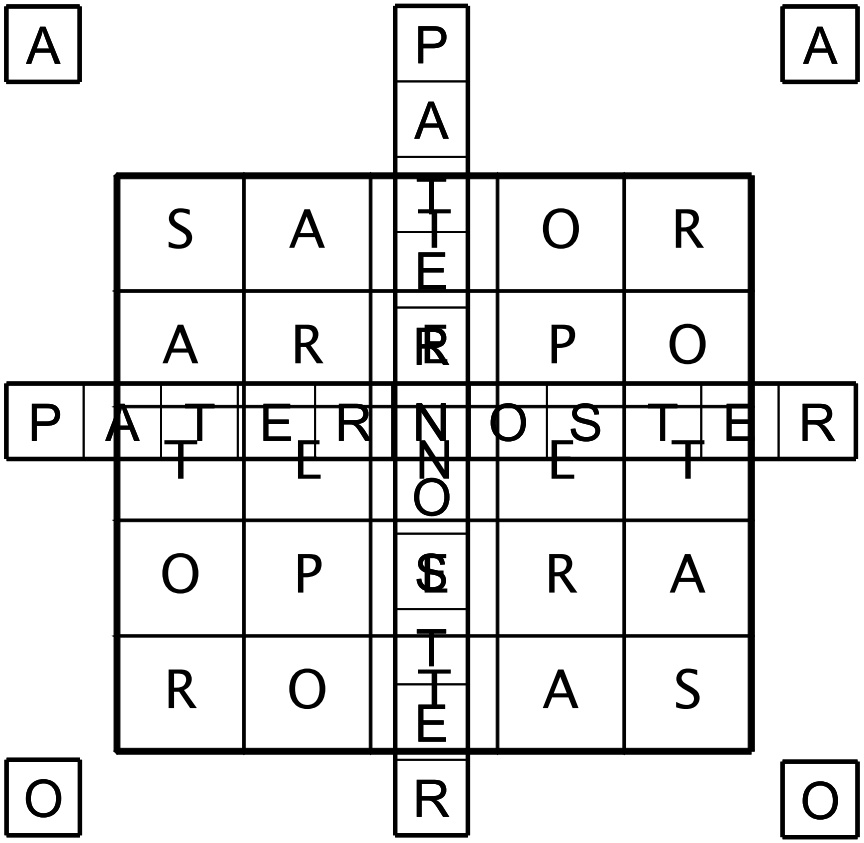


Missed!

Some Problems in IF systems...

1. IF systems perform the filtering task on the basis of user profiles
   - ✓ Structured model of the user interests
   - ✓ User profiles compared against item descriptions to provide recommendations
2. Problems: keywords not appropriate for representing content, due to polysemy, synonymy, multi-word concepts (*homography*, *homophony*) – "Sator arepo eccetera" (Eco, 2007)

Some Problems in IF systems... (cont'd)



Keyword-based Profiles

doc1

AI is a branch of computer science

doc2

the 2007 International Joint Conference on **Artificial Intelligence** will be held in India

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 2007 International Joint Conference on **Artificial Intelligence** will be held in India

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 2007 International Joint Conference on **Artificial Intelligence** will be held in India

doc3
apple launches a new product…

USER PROFILE

artificial	0.02
intelligence	0.01
apple	0.13
AI	0.15
…	

POLYSEMY

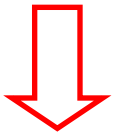
Tackling IR/IF problems

Putting Intelligence into Search/Filtering Tasks =

1. Semantics: concept identification in documents through advanced NLP techniques → "*beyond keywords*"

+

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



Information Seeking Support Systems (ISSSs):
tools that aid people in managing, analyzing, and sharing sets of retrieved information.
ISSSs provide search solutions that empower users to go beyond single-session lookup tasks.

G. Marchionini and R.W. White. Information-Seeking Support Systems. *IEEE Computer* 42(3):30-32, March 2009.

Meno's Paradox: Search is not so simple as it might seem

Meno: And how will you enquire, Socrates, into that which you do not know? What will you put forth as the subject of enquiry? And if you find what you want, how will you ever know that this is the thing which you did not know?

Socrates: I know, Meno, what you mean; but just see what a tiresome dispute you are introducing. You argue that man cannot enquire either about that which he knows, or about that which he does not know; for if he knows, he has no need to enquire; and if not, he cannot; for he does not know the very subject about which he is to enquire.

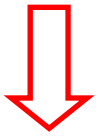
Plato Meno 80d-81a
`http://www.gutenberg.org/etext/1643`

Meno's question in our time

with modern ISSSs:

"How to discover the *magic words* that will connect you to the information you seek?"

The focus of many studies has been to point out methods for text processing to enable a system to make connections between the terminology of the user request (implicit/explicit) and related terminology in the information needed



Need for semantics for text interpretation

Beyond (Keyword) Search: Semantics by way of Knowledge Infusion

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

Outline

❶ INTRODUCTION
- ✓ The Information Overloading Problem
- ✓ Personalization on the Web

❷ INFORMATION SEEKING STRATEGIES

❸ INFORMATION FILTERING
- ✓ Collaborative Filtering
- ✓ Content-based Filtering & User Profiling

❹ IDEAS FOR INTELLIGENT INFORMATION FILTERING

❺ CONCLUDING REMARKS

❻ LIVE DEMO

Collaborative / Social Filtering

1. *Word of Mouth / "Wisdom of the Masses"*
2. Makes use of a database of user preferences in order to:
   - ✓ find users with *similar* interests
   - ✓ predict whether an *unseen* information item is likely to be of interest for a user based on how *other* users rated that item
3. Widely adopted in recommender systems [Resnick and Varian 1997, Linden et al. 2003]

Recommender Systems have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [Burke, 2002]

[Burke 2002] R. Burke. Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* 12(4):331-370, 2002.

[Resnick and Varian 1997] Resnick, P. and H. Varian. Recommender Systems. *Communications of the ACM* 40(3):56-58, 1997.

[Linden et al. 2003] Linden, G., B. Smith, and J. York. Amazon.com Recommendations: Item-to-Item Collaborative Filtering. *IEEE Internet Computing* 7(1), 76-80, 2003.

Collaborative / Social Filtering

1. *Word of Mouth / "Wisdom of the Masses"*
2. Makes use of a database of user preferences in order to:
   - ✓ find users with *similar* interests
   - ✓ predict whether an *unseen* information item is likely to be of interest for a user based on how *other* users rated that item
3. Widely adopted in recommender systems [Resnick and Varian 1997, Linden et al. 2003]

Recommender Systems provide personalized suggestions about items that the user might find interesting, by matching items to user profiles or groups [Kangas, 2001]

[Kangas 2001] Sonja Kangas. *Collaborative Filtering and Recommendation Systems*. Technical Report TTE4-2001-35. VTT Information Technology, Espoo, Finland, 2001.

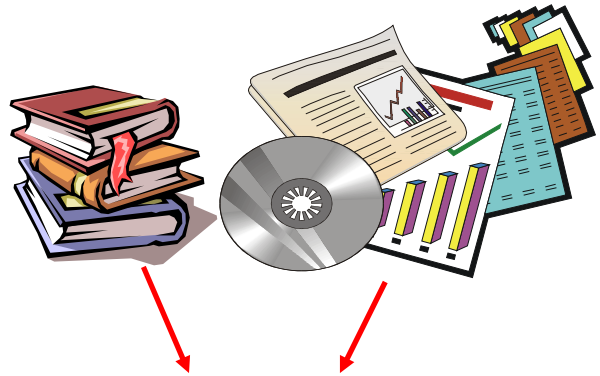
Collaborative / Social Filtering

1. *Word of Mouth / "Wisdom of the Masses"*
2. Makes use of a database of user preferences in order to:
   - ✓ find users with *similar* interests
   - ✓ predict whether an *unseen* information item is likely to be of interest for a user based on how *other* users rated that item
3. Widely adopted in recommender systems [Resnick and Varian 1997, Linden et al. 2003]

Recommendation process

Given a large set of items and a description of the user's needs, the goal is to present a small set of the items that are suited to the user needs

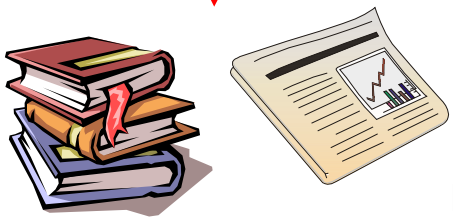
Content-based Filtering



Information Source

User Profile

User profile compared against documents for *relevance* computation



Documents recommended to the user

User

Content-based Filtering

1. Each user is assumed to operate independently
2. Items are represented by some features
   - ✓ Movies: actors, directors, plot, …
3. Machine Learning for automated inference of user profile from user's behavior
   - ✓ Relevance judgment on items, e.g. ratings
   - ✓ Training on rated items → user profile
4. Filtering based on the comparison between the *content* of the items and the user preferences as defined in the user profile
   - ✓ Keyword-based representation for content and profiles → string matching

Outline

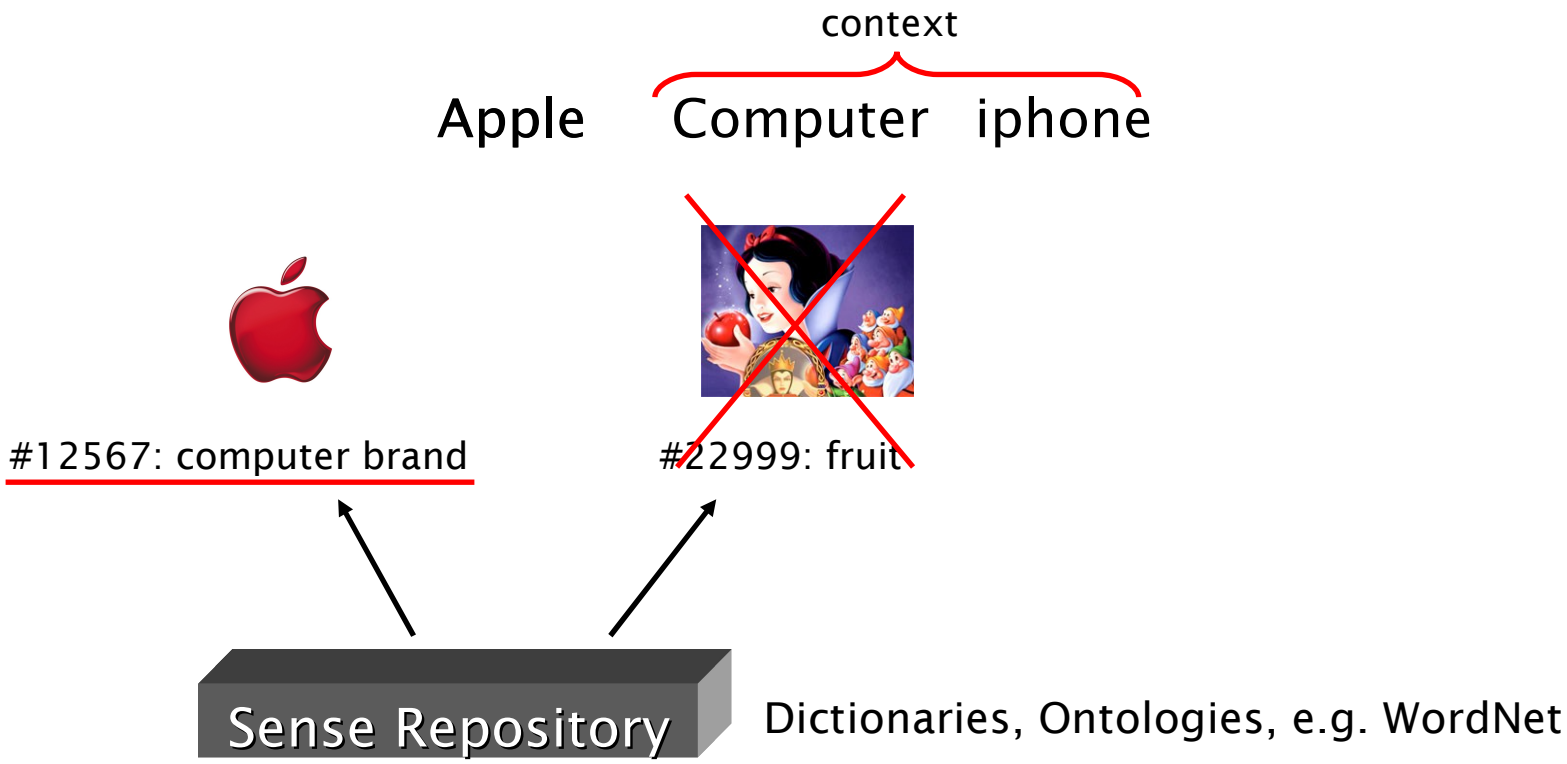
1. INTRODUCTION
   - ✓ The Information Overloading Problem
   - ✓ Personalization on the Web
2. INFORMATION SEEKING STRATEGIES
3. INFORMATION FILTERING
   - ✓ Collaborative Filtering
   - ✓ Content-based Filtering & User Profiling
4. IDEAS FOR INTELLIGENT INFORMATION FILTERING
5. CONCLUDING REMARKS
6. LIVE DEMO

Intelligent Content-based Filtering

1. Semantics
   - ✓ Novel strategies to represent items and user profiles
     e.g. based on knowledge infusion and NLP methods
2. Personalization
   - ✓ Novel strategies to build user profiles
     e.g. taking into account that information seeking can be seen no longer as a *solitary* activity
   - ✓ recent interest in collaborative search and social bookmarking

Word Sense Disambiguation (WSD): from words to meanings

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



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 2007 International Joint Conference on **Artificial Intelligence** will be held in India

doc3

apple launches a new product…

USER PROFILE

#12387	0.02
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 2007 International Joint Conference on **Artificial Intelligence** will be held in India

doc3

apple launches a new product…

USER PROFILE

#12387	0.17
apple	0.13
#12387	0.15
…	

SYNONYMY

ITR (ITem Recommender)

Sense-based Profiles

doc1
AI is a branch of computer science

doc2
the 2007 International Joint Conference on **Artificial Intelligence** will be held in India

doc3
apple launches a new product…

SEMANTIC USER PROFILE
sense identifiers rather than keywords

USER PROFILE

#12387	0.17
#12567	0.13
…	

POLYSEMY

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.

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.

Intelligent Content-based Filtering

❶ Semantics

- ✓ Novel strategies to represent items and user profiles
  e.g. based on knowledge infusion and NLP methods

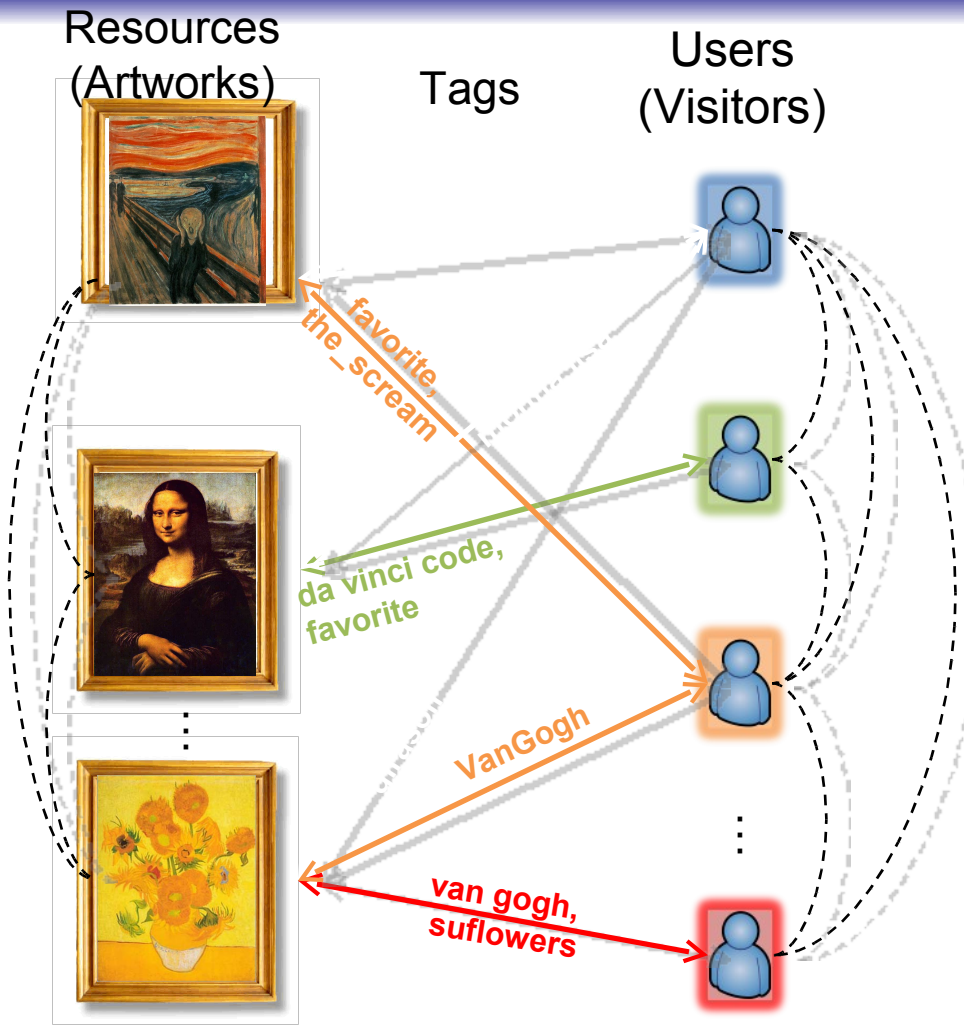
❷ Personalization

- ✓ Novel strategies to build user profiles
  e.g. taking into account that information seeking can be seen no longer as a *solitary* activity
- ✓ recent interest in collaborative search and social bookmarking

Web 2.0 & User-Generated Content (UGC)

Social Tagging & Folksonomies

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



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 arrive 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)

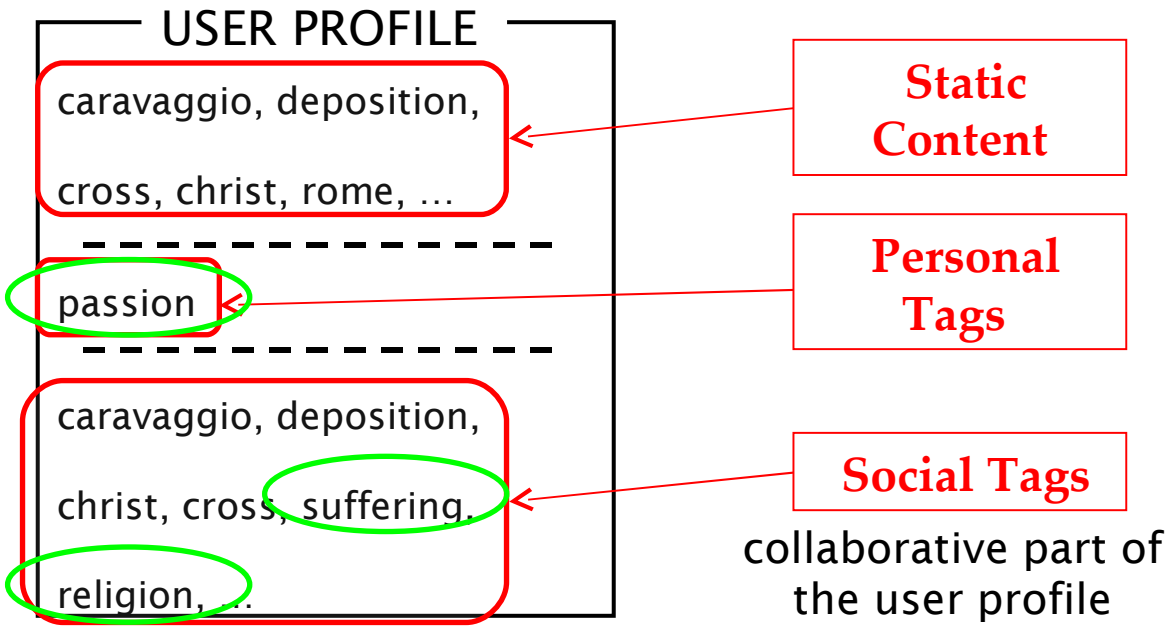
Tags within User Profiles

27) Caravaggio - Deposition from the Cross

Descrizione dell'opera

The Deposition, considered one of Carav
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 th
tomb, but rather when, in the presence of the holy women, he is laid by Nicodemus and John on the Anointing Stone, that is t
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 arrive
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.

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, October 23-25, 2008, Lausanne, Switzerland*, pages 163–170. Association for Computing Machinery, ACM, 2008.



Outline

1. INTRODUCTION
   - ✓ The Information Overloading Problem
   - ✓ Personalization on the Web
2. INFORMATION SEEKING STRATEGIES
3. INFORMATION FILTERING
   - ✓ Collaborative Filtering
   - ✓ Content-based Filtering & User Profiling
4. IDEAS FOR INTELLIGENT INFORMATION FILTERING
5. CONCLUDING REMARKS
6. LIVE DEMO

Concluding Remarks and Future Work

1. Need for *intelligent* solutions and tools for Information Access in the information overload era
2. New strategies for Information Filtering & Retrieval
   - ✓ Semantics: to capture the meaning of content and user needs
   - ✓ Personalization & User Profiling for adapting results to user information needs

Future Work

- ✓ Serendipity within Recommender System (avoiding the *homophily* trap [Zuckerman 2008])
- ✓ How to make serendipity operational? An initial attempt in [Iaquinta et al. 2008]

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

L. Iaquinta, M. de Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino. Introducing Serendipity in a Content-based Recommender System. In F. Xhafa, F. Herrera, A. Abraham, M. Koppen & J. M. Benitez, editors, *Proceed.of the 8th Int. Conf. on Hybrid Intelligent Systems HIS-2008* , 168–173. IEEE Computer Society Press, Los Alamitos, California, 2008.

Outline

❶ INTRODUCTION
- ✓ The Information Overloading Problem
- ✓ Personalization on the Web

❷ INFORMATION SEEKING STRATEGIES

❸ INFORMATION FILTERING
- ✓ Collaborative Filtering
- ✓ Content-based Filtering & User Profiling

❹ IDEAS FOR INTELLIGENT INFORMATION FILTERING

❺ CONCLUDING REMARKS

❻ LIVE DEMO

Intelligent Information Filtering & Retrieval @ SWAP group: the SENSEware suite

METAL	Multilanguage Text Analyzer
JIGSAW	Word Sense Disambiguator
ITR	Item Recommender
FIRSt	Folksonomy-based Item Recommender system
SENSE	SEmantic N-level Search Engine
OTHO	On the Tip of my THOught

Sensemaking

1. Many information search tasks are part of a broader class of tasks called *sensemaking*
2. *Sensemaking* involves finding and collecting information from large collections, organizing and understanding that information and producing some product
   - ✓ understanding a health problem to make a medical decision
   - ✓ deciding which laptop to buy

Peter Pirolli. Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales. *IEEE Computer* 42(3):33-40, March 2009.

Sensemaking

1. Searching and filtering are the basic steps of the *sensemaking* overall process
2. Reading and extracting information from retrieved documents are essential steps for decision making and other complex mental activities
   - ✓ Retrieval/Filtering are necessary but not sufficient

Peter Pirolli. Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales. *IEEE Computer* 42(3):33-40, March 2009.

Examples of challenging sensemaking processes

1. Just for fun:

   OTTHO (On the Tip of my THOught) [Semeraro et al. 2009]

   - ✓ La Ghigliottina: very popular linguistic game
   - ✓ Umberto Eco's article "La nostra ghigliottina quotidiana" on "L'Espresso"

2. More serious: Discovering "hidden" correlations among documents

U. Eco. La nostra ghigliottina quotidiana. April 24, 2008. http://espresso.repubblica.it/dettaglio/La-nostra-ghigliottina-quotidiana/2011914/18

G. Semeraro, P. Lops, P. Basile, and M. Degemmis. On the Tip of my THOught: Playing the Guillotine Game. In C. Boutilier, editor, *IJCAI 2009, Proceedings of the 21th International Joint Conference on Artificial Intelligence, Pasadena, California, July 11-17, 2009*, Morgan Kaufmann, 2009 (in press).

References

[Baeza-Yates and Ribeiro-Neto 1999] Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval. New York: Addison-Wesley, 1999.

[Basile et al. 2007] 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.

[Belkin and Croft 1992] Belkin, N.J. and W.B. Croft. Information Filtering and Information Retrieval: Two Sides of the Same Coin? Communications of the ACM, 35(12): 29-38, 1992.

[Burke 2002] R. Burke. Hybrid Recommender Systems: Survey and Experiments. User Modeling and User-Adapted Interaction. 12(4):331-370, 2002.

[Degemmis et al. 2007] 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.

[Hanani et al. 2001] Hanani, U., B. Shapira, and P. Shoval. Information Filtering: Overview of Issues, Research and Systems. User Modeling and User-Adapted Interaction, 11(3): 203-259, 2001.

[Herlocker et al. 1999] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl: 1999, An Algorithmic Framework for Performing Collaborative Filtering. In: Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 230-237.

[Iaquinta et al. 2008] L. Iaquinta, M. de Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino. Introducing Serendipity in a Content-based Recommender System. In F. Xhafa, F. Herrera, A. Abraham, M. Koppen & J. M. Benitez, editors, *Proceed.of the 8th Int. Conf. on Hybrid Intelligent Systems HIS-2008* , 168–173. IEEE Computer Society Press, Los Alamitos, California, 2008.

[Kangas 2001] Sonja Kangas. Collaborative Filtering and Recommendations systems. Technical Report TTE4-2001-35. VTT Information Technology, Espoo, Finland.

[Linden et al. 2003] Linden, G., B. Smith, and J. York. Amazon.com Recommendations: Item-to-

References

[Liu et al. 2004] Liu, F., Yu, C., and Meng, W. 2004. Personalized Web Search For Improving Retrieval Effectiveness. IEEE Transactions on Knowledge and Data Engineering 16, 1 (Jan. 2004), 28-40.

[Magnini et al. 01] Magnini, B. and C. Strapparava. Improving User Modelling with Content-based Techniques. In Proceedings of the Eighth International Conference on User Modeling. Sonthofen, Germany, pp. 74-83, Springer, 2001.

[Marchionini and White 2009] G. Marchionini and R.W. White. Information-Seeking Support Systems. *IEEE Computer* 42(3):30-32, March 2009.

[Miller 1990] G. Miller. WordNet: An On-Line Lexical Database. International Journal of lexicography, 3(4), 1990.

[Pazzani et al. 1997] Pazzani M. and D. Billsus. Learning and revising user profiles: The identification of interesting web sites. Machine Learning, 27(3):313–331, 1997.

[Pirolli 2009] Peter Pirolli. Powers of 10: Modeling Complex Information-Seeking Systems at Multiple Scales. IEEE Computer 42(3):33-40, March 2009.

[Resnick and Varian 1997] Resnick, P. and H. Varian. Recommender Systems. Communications of the ACM, 40(3), 56-58, 1997.

[Sarwar et.] B.M. Sarwar, G. Karypis, J.A. Konstan, J. Riedl. Analysis of Recommendation Algorithms for E-Commerce, ACM Conf. Electronic Commerce, ACM Press, 2000, pp.158-167.

[Semeraro 2007] Personalized Searching by Learning WordNet-based User Profiles. Journal of Digital Information Management, 5(5):309–322, 2007. ISSN 0972-7272. Digital Information Research Foundation, Chennai, India.

[Semeraro et al. 2007] G. Semeraro, M. Degemmis, P. Lops, and P. Basile. Combining Learning and Word Sense Disambiguation for Intelligent User Profiling. In Proc. of the 20th Int. Joint Conf. on Artificial Intelligence, 2007, Hyderabad, India, pages 2856-2861, 2007.

[Semeraro et al. 2009] G. Semeraro, P. Lops, P. Basile, and M. Degemmis. On the Tip of my THOught: Playing the Guillotine Game. In C. Boutilier, editor, *IJCAI 2009, Proceedings of the 21th International Joint Conference on Artificial Intelligence, Pasadena, California, July 11-17, 2009*, Morgan Kaufmann, 2009 (in press).

Thanks...

...for your attention...

...Questions?