Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations

Iván Cantador[1], Martin Szomszor[2], Harith Alani[2],
Miriam Fernández[1], Pablo Castells[1]

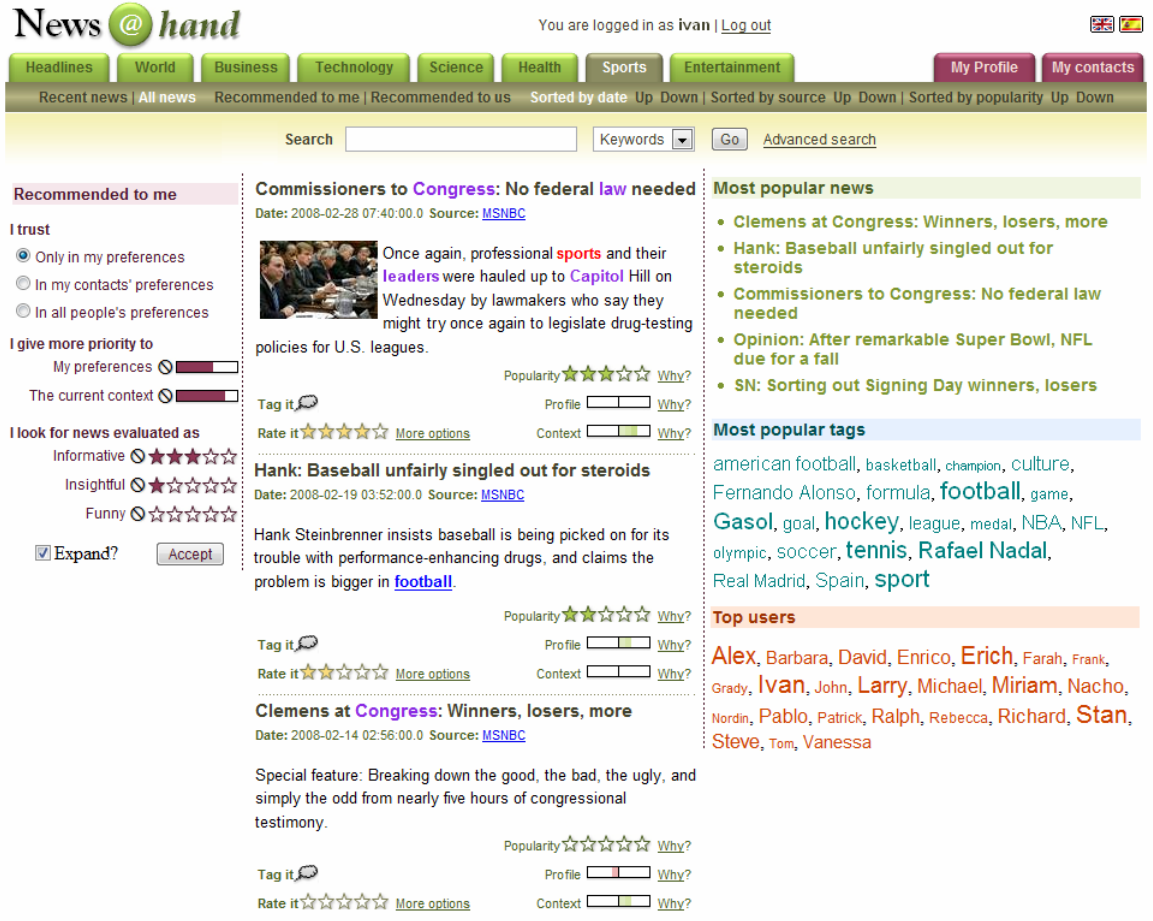
[1] *Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain*
{ivan.cantador, miriam.fernandez, pablo.castells}@uam.es

[2] *School of Electronics and Computer Science, University of Southampton, Southampton, UK*
{mns2, ha}@ecs.soton.ac.uk

1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008)
5th European Semantic Web Conference (ESWC 2008)
Tenerife, Spain, June 2nd, 2008

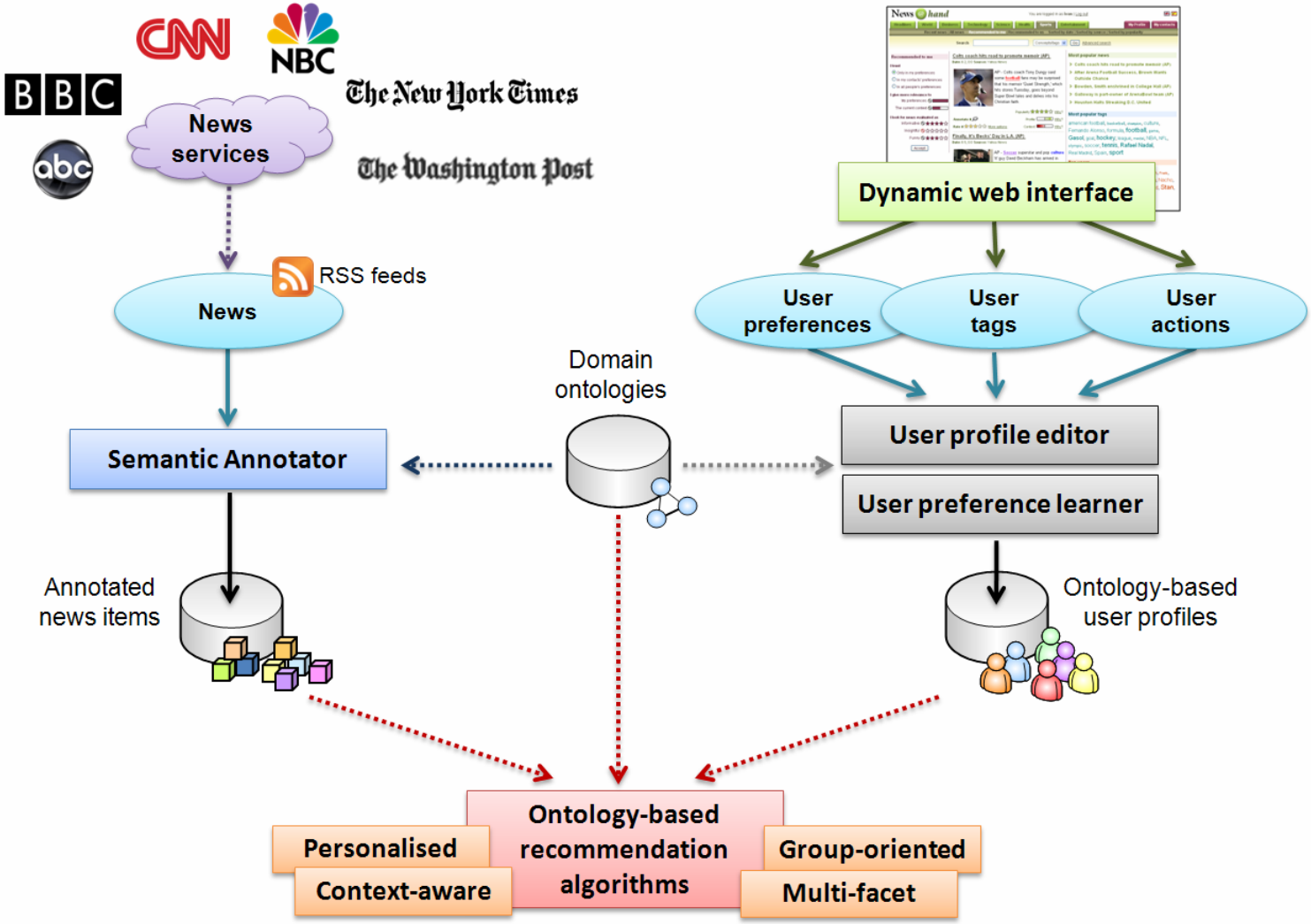
Motivation

- **News@hand**
  - An hybrid news recommender system which makes use of Semantic Web technologies to provide several on-line news recommendation services



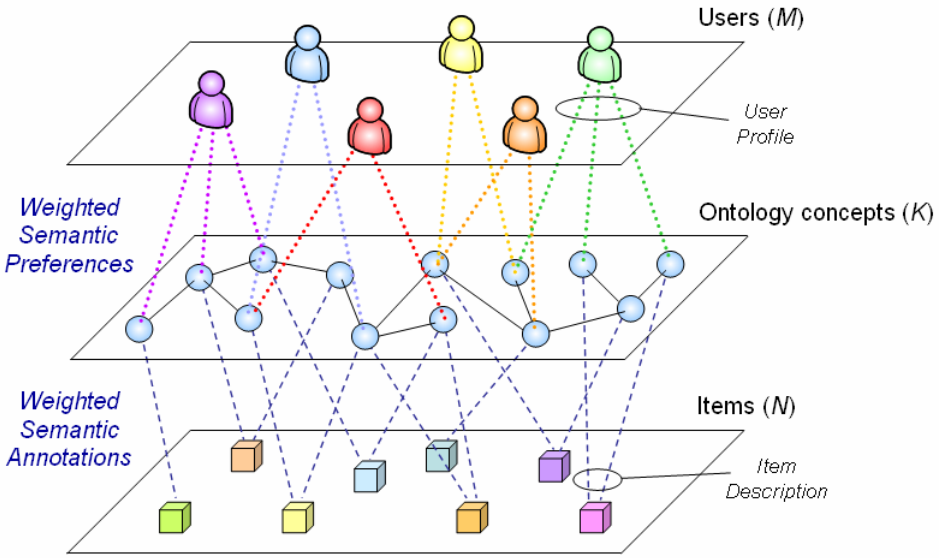
Motivation

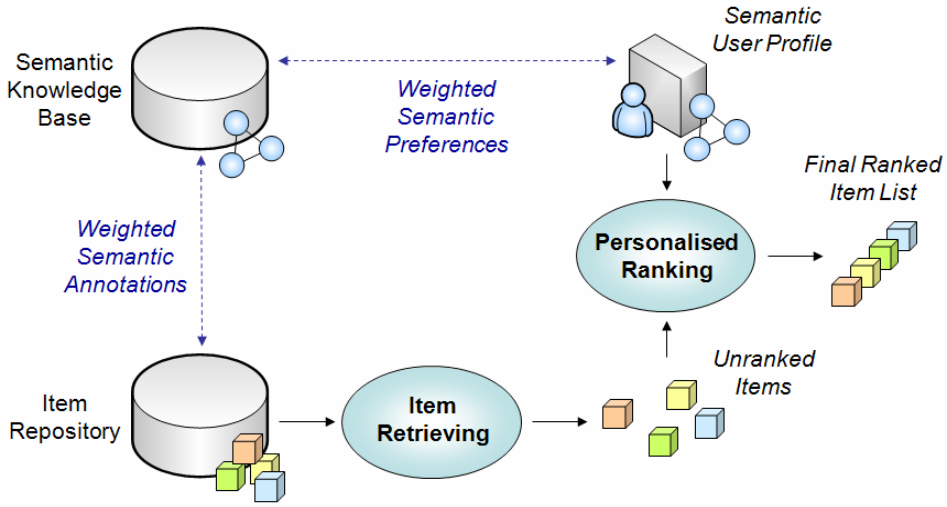
- **System architecture**



Motivation

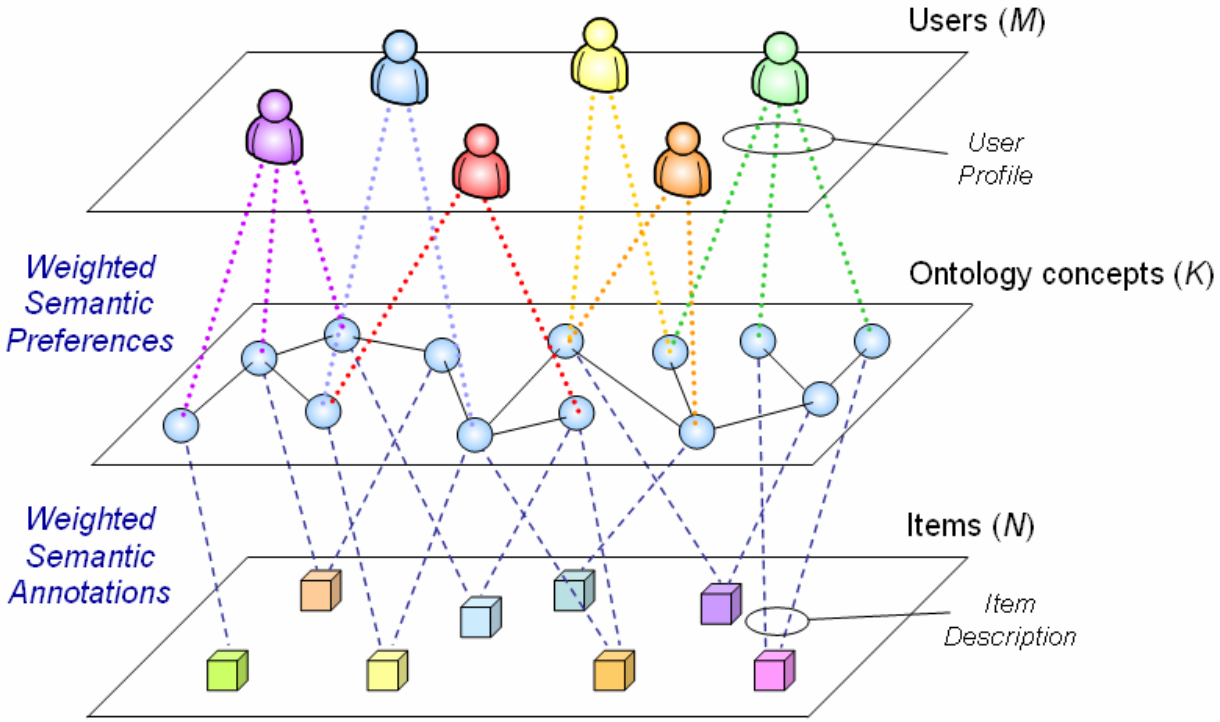
- **Knowledge representation**
  - **User profiles** and **item descriptions** are respectively represented as vectors $\mathbf{u_m}=(u_{m,1}, u_{m,2}, \ldots, u_{m,K})$ and $\mathbf{d_n}=(d_{n,1}, d_{n,2}, \ldots, d_{n,K})$, where $u_{m,k}, d_{n,k} \in [-1,1]$ are the weights that measure the relevance of concept c_k for user u_m and item d_n
  - **Recommendation models** are based on the definition of matching algorithms which make use of similarity measures such as $cos(\mathbf{u_m}, \mathbf{d_n})$





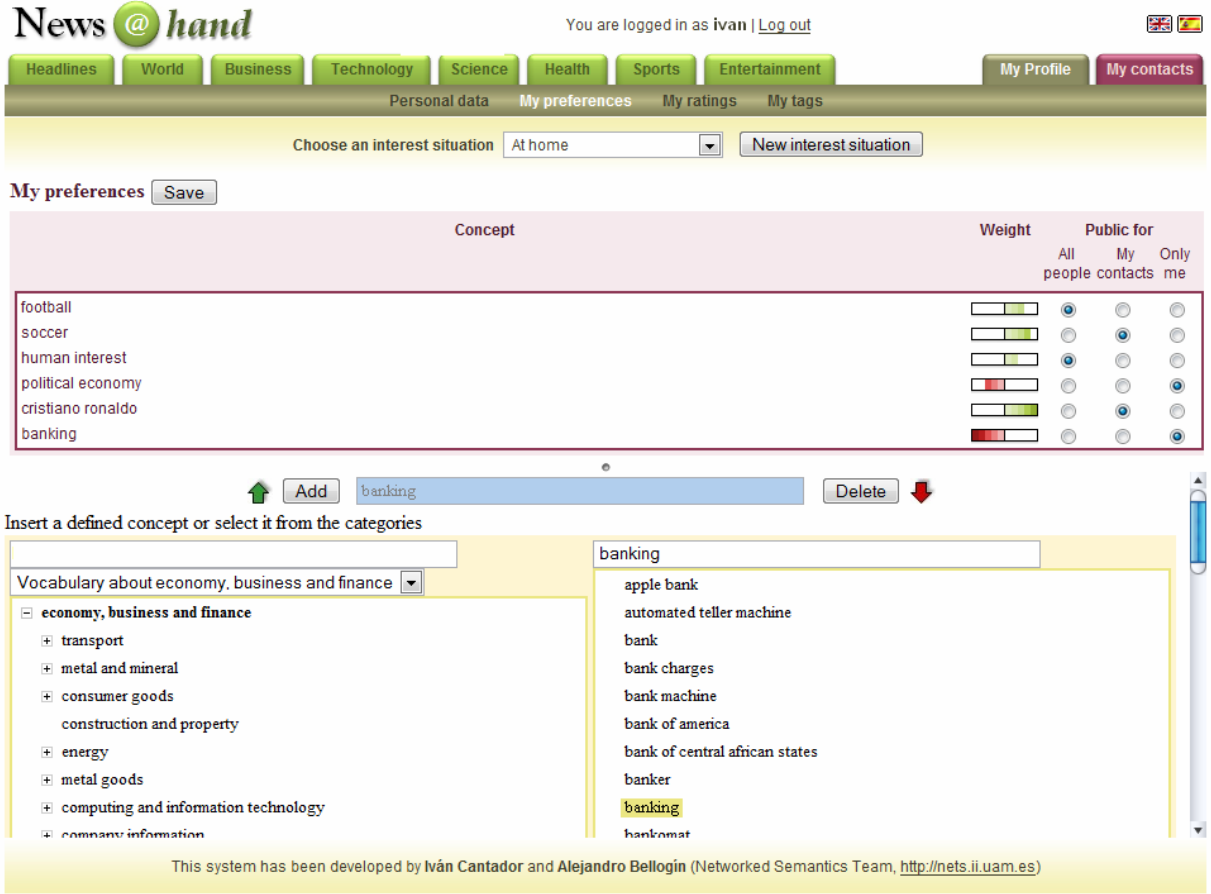
Research question

How do we obtain the semantic user preferences?



Research question

- **How do we obtain the semantic user preferences?**
  1. Manually defined by the user
     - News@hand already has a user profile editor ;-)



Research question

- **How do we obtain the semantic user preferences?**
  1. Manually defined by the user
     - News@hand already has a user profile editor ;-)
  2. (Semi-)automatically learned from user behaviour and/or relevance feedback
     - Work in progress

Research question

- **How do we obtain the semantic user preferences?**
  1. Manually defined by the user
     - News@hand already has a user profile editor ;-)
  2. (Semi-)automatically learned from user behaviour and/or relevance feedback
     - Work in progress
  3. Automatically learned from social tags!!!

The objective of this work is to automatically transform social tags into ontology concepts (classes or instances)

Relating social tags to ontological information

- **3 steps**
  1. Filtering social tags
  2. Obtaining semantic information about social tags
  3. Categorisation of social tags into ontology classes

Relating social tags to ontological information

- **3 steps**
  1. Filtering social tags
     - Lexical filtering
       - Discard those tags that are too long, too short, stop words, numbers*, unpopular tags
       - Convert special characters to a base form: zürich → zurich, mónaco → monaco
     - Compound nouns and misspellings (using Google "did you mean")
       - Split compound nouns: lordoftherings → lord of the rings
       - Correct misspellings: barclona → barcelona
     - Wikipedia correlation
       - Transform acronyms: nyc → new york city, diy → do it yourself
     - Morphologically similar terms (using Snowball stemming library)
       - Merge terms with the same stem: blog, blogs, blogging → blog
     - WordNet synonyms
       - Merge non-ambiguous synonyms: android = humanoid, thesis = dissertation

Relating social tags to ontological information

- **3 steps**
  2. Obtaining semantic information about social tags

Relating social tags to ontological information

- **3 steps**
  2. Obtaining semantic information about social tags

Field	Entry: java
wikipediaTerm	java
wikipediaFilter	none
isAbout	java island
equivalentTo	
ambiguousWith	java (programming language)
hasOtherMeanings	programming language | cigarettes | band | dance | board game
categories	java

Field	Entry: javanese language
wikipediaTerm	javanese language
wikipediaFilter	ambiguous
isAbout	indonesian spoken language
equivalentTo	
ambiguousWith	
hasOtherMeanings	
categories	

Field	Entry: java (programming language)
wikipediaTerm	java (programming language)
wikipediaFilter	ambiguous
isAbout	programming language
equivalentTo	
ambiguousWith	javanese language
hasOtherMeanings	
categories	java programming language | java platform | java specification requests | c programming language family | sun microsystems | concurrent programming languages | class-based programming languages | object-oriented programming languages | jvm programming languages

Relating social tags to ontological information

- **3 steps**
  2. Obtaining semantic information about social tags

Field	Entry: ny
wikipediaTerm	ny
wikipediaFilter	none
isAbout	state
equivalentTo	new york
ambiguousWith	new york city
hasOtherMeanings	magazine | album | typeface
categories	new york | new york | states of the united states | former british colonies

Field	Entry: new york
wikipediaTerm	new york
wikipediaFilter	none
isAbout	state
equivalentTo	
ambiguousWith	new york city
hasOtherMeanings	magazine | album | typeface
categories	new york | new york | states of the united states | former british colonies

Field	Entry: nyc
wikipediaTerm	nyc
wikipediaFilter	none
isAbout	
equivalentTo	new york city
ambiguousWith	
hasOtherMeanings	
categories	neighborhoods in new york city | new york | coastal cities in the united

Field	Entry: new york city
wikipediaTerm	new york city
wikipediaFilter	none
isAbout	
equivalentTo	
ambiguousWith	
hasOtherMeanings	
categories	neighborhoods in new york city | new york | coastal cities in the united

Relating social tags to ontological information

- **3 steps**

  3. Categorisation of social tags into ontology classes
     - Search for the term in Wikipedia and retrieve the corresponding wiki page
       - Brad Pitt → http://en.wikipedia.org/wiki/Brad_pitt
     - Extract the Wikipedia categories of the term from the retrieved page
       - action film actors, american film actors, …
     - Apply a morphologic matching algorithm between the term name and categories, and the names of the ontology classes
       - cinema, film, person, actor
     - Select the ontology class whose name is the “most similar” to the term name and/or categories
       - actor
     - Create an instance for the concept in the selected class
       - http://www.nets.ii.uam.es/iptc/entertainmet#actor_brad-pitt

Preliminary evaluations

- **Ontology population**
  - 20 users
  - Each user evaluated 25 instances of each ontology
  - Evaluation of an instance: (1) correct class, (2) correct ontology, (3) incorrect ontology

Table 1. Number of classes and instances of News@hand KB, and average population accuracy

Ontology	#classes	#instances	Avg. #instances/class	Avg. accuracy
arts, culture, entertainment	87	33,278	383	78.7 / 93.3
crime, law, justice	22	971	44	62.7 / 73.3
disasters, accidents	16	287	18	74.7 / 84.0
economy, business, finance	161	25,345	157	69.3 / 80.0
education	20	3,542	177	57.5 / 76.7
environmental issues	41	20,581	502	72.0 / 85.3
health	26	1,078	41	65.3 / 89.3
human interests	6	576	96	64.0 / 84.0
labour	6	133	22	70.7 / 78.7
lifestyle, leisure	29	4,895	169	72.0 / 90.7
politics	54	3,206	59	60.0 / 81.3
religion, belief	31	3,248	105	84.0 / 90.7
science, technology	50	7,869	157	68.0 / 86.7
social issues	39	8,673	222	70.7 / 85.3
sports	124	5,567	45	72.0 / 86.7
unrests, conflicts, wars	23	1,820	79	61.3 / 80.0
weather	9	66	7	69.7 / 89.5
	744	121,135	163 (avg.)	**69.9** / **84.4**

Preliminary evaluations

- **Item annotation**
  - 20 users
  - Each user evaluated 10 items from each section

Table 2. Average number of annotations per news item, and average annotation accuracies

	headlines	world	business	technology	science	health	sports	entertainment
#news items	2,660	2,200	1,739	303	346	803	603	1,044
#annotations	18,210	17,767	13,090	2,154	2,487	4,874	2,453	5,343
#annotations/item	7	8	8	7	7	6	4	5
Avg. accuracy	71.4	72.7	79.2	76.3	74.1	73.1	75.8	76.0

- **Recommendations**
  - 20 users
  - Each user evaluated recommendations for 10 profiles from del.icio.us and Flickr

Table 3. Average relevance values for the 5 top ranked news items recommended by News@hand

	headlines	world	business	technology	science	health	sports	entertainment
keyword-based	46.3	34.3	39.0	43.5	35.9	21.1	58.0	33.5
News@hand	57.0	53.2	72.8	94.0	60.9	40.6	98.2	60.4

Conclusions and Future Work

- **Conclusions**
  - Combination of folksonomy information and knowledge information available in the Semantic Web is a powerful and promising approach to provide flexible, multi-domain recommendations
    - Easy adaptation to new vocabularies
    - Supervised representation of semantic knowledge
- **Future work**
  - More detailed experimentation
  - Investigation of ambiguty issues
    - Correlation of tags (e.g., java ← programming, netbeans, sun microsystems) within and between user profiles

Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations

Iván Cantador[1], Martin Szomszor[2], Harith Alani[2],
Miriam Fernández[1], Pablo Castells[1]

[1] *Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain*
{ivan.cantador, miriam.fernandez, pablo.castells}@uam.es

[2] *School of Electronics and Computer Science, University of Southampton, Southampton, UK*
{mns2, ha}@ecs.soton.ac.uk

1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web (CISWeb 2008)
5th European Semantic Web Conference (ESWC 2008)
Tenerife, Spain, June 2nd, 2008