Esta es la **versión de autor** de la comunicación de congreso publicada en:
This is an **author produced version** of a paper published in:

WI-IAT '08. IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, 2008. IEEE, 2008. 562-565

DOI: http://dx.doi.org/10.1109/WIIAT.2008.204

El acceso a la versión del editor puede requerir la suscripción del recurso
Access to the published version may require subscription

Ontology-based Personalised and Context-aware Recommendations of News Items

Iván Cantador, Alejandro Bellogín, Pablo Castells
Escuela Politécnica Superior, Universidad Autónoma de Madrid, Madrid, Spain
{ivan.cantador, alejandro.bellogin, pablo.castells}@uam.es

Abstract

News@hand is a news recommender system that makes use of semantic technologies to provide several on-line news recommendation services. News contents and user preferences are described in terms of concepts appearing in a set of domain ontologies. Based on the similarities between item descriptions and user profiles, and the semantic relations between concepts, content-based and collaborative recommendation models are supported by the system. In this paper, we evaluate a model that personalises the order in which news articles are shown to the user according to his long-term interest profile, and other model that reorders the news items lists taking into account the current semantic context of interest of the user. The combination of those models is investigated showing significant improvements on the experimental tasks performed.

1. Introduction

Thanks to the Web, people not only have access to more worldwide news information than ever before, but also they can obtain it in a more timely manner. On-line newspapers present breaking news on their websites in real time, and users can receive automatic notifications of them via RSS feeds. However, the increasing volume, growth rate and ubiquity of access to the contents challenge the limits of human processing capabilities. It is in such scenario where recommender systems can do their most, by scanning the space of choices, and predicting the potential usefulness of news items for each particular user, without explicitly specifying needs or querying for items whose existence is unknown beforehand.

Content-based and collaborative filtering techniques suffer from own limitations, and hybrid recommendation models have been proposed to address them by means of joined mechanisms. However, general common problems have not been fully solved yet, and further investigation is needed. For example, typical approaches are domain dependant. Their models are generated from information gathered within a specific domain, and cannot be easily extended and/or incorporated to other systems. Moreover, the need for further flexibility with the consideration of contextual features during the recommendation processes is also an unfulfilled requirement in most systems.

In this work, we present News@hand, a system which makes use of semantic technologies to recommend news.

The system supports different recommendation models for single and multiple users. The exploitation of meta-information in the form of ontologies that describe the recommended items and user profiles in a general, portable way, along with the capability of inferring knowledge from the semantic relations defined in the ontologies, are the key aspects of the proposals.

The recommendation models of News@hand have already been evaluated in different works. However, previous experiments were conducted for each model in an isolated way. Here, we present our first attempt to jointly evaluate some of those models, integrated and combined within the system. Specifically, we examine two models centred on a single user: one that personalise the ranking of news contents according to the user's profile, and other that takes into consideration the current context of interest of the user.

2. News@hand

As a hybrid recommender system, News@hand combines content features and collaborative information to make news suggestions. However, as opposite to previous systems, it uses a controlled and structured vocabulary to describe the news texts and user preferences. Item and user profiles are represented in terms of concepts appearing in domain ontologies, and semantic relations among those concepts are exploited to enrich the above representations, and are incorporated within the recommendation processes.

An ontology-based knowledge representation is less ambiguous than a keyword-based or item-based model. Furthermore, ontology standards, such as RDF and OWL, support inference mechanisms that can be used to enhance content retrieval [5]. A user interested in *natural disasters* (superclass of *hurricane*) is also recommended items about *hurricanes*. Similarly, a user fascinated about the life of *actors* can be recommended items in which the name of *Mel Gibson* appears, due to that person could be an instance of the class *actor*. Also, a user keen on *Australia* can be assumed to like *Sydney*, through the *locatedIn* transitive relation between these concepts.

Deriving benefit from the semantically annotated news items, the defined ontology-based user profiles, and the knowledge represented by the domain ontologies, a set of recommendation algorithms are executed. News@hand offers personalised, context-aware [6], group-oriented [3], and multi-facet recommendations [2].

2.1. Ontology population

As mentioned before, News@hand automatically annotates news items with semantic concepts that appear in the textual contents, and exist in the domain ontologies. These concepts can be ontology classes or instances. The ontology population (i.e., the transformation of unstructured, semi-structured and structured data into ontology instances) is performed when new contents are retrieved from the Web. In News@hand, a novel and simple mechanism to automatically populate ontologies using semantic information extracted from Wikipedia (http://www.wikipedia.org/) has been implemented [4].

2.2. Item annotation

News@hand periodically retrieves news items from the websites of well-known news media, such as BBC, CNN, NBC, etc. These items are obtained via RSS feeds, and contain information of published news articles: their title, summary of the contents, publication date, hyperlinks to the full texts and related on-line images. The system analyses and automatically annotates the textual information (title and summary) of the RSS feeds [5]. Using a set of NLP tools [1], an annotation module removes stop words and extracts relevant (simple and compound) terms, categorised according to their part of speech: nouns, verbs, adjectives or adverbs. Then, nouns are morphologically compared with the names of the classes and instances of the domain ontologies. The comparisons are done using an index created with Lucene (http://lucene.apache.org/), and according to fuzzy metrics. For each term, the most similar semantic concept is chosen and added as an annotation of the item. After all annotations are created, a TF-IDF technique computes and assigns weights to them.
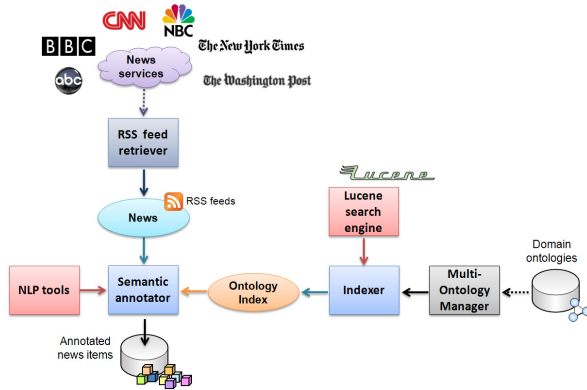

Figure 1. Item annotation mechanism of News@hand

3. Personalised recommendation models

3.1. Semantic content-based recommendations

Our notion of content retrieval is based on a matching algorithm that provides a personal relevance measure $pref(i_n, u_m)$ of an item i_n for a user u_m. This measure is set according to the semantic preferences of the user and the semantic annotations of the item, and is based on a cosine-based vector similarity $\cos(\mathbf{i}_n, \mathbf{u}_m)$. The measure can be combined with query-based scores $sim(i_n, q)$ and semantic context, to produce combined rankings [5][6].

We also propose a preference spreading mechanism, which expands the initial set of preferences stored in user profiles through explicit semantic relations with other concepts in the ontology (Figure 2). Our approach is based on Constrained Spreading Activation (CSA), and is self-controlled by applying a decay factor to the intensity of preference each time a relation is traversed [6].
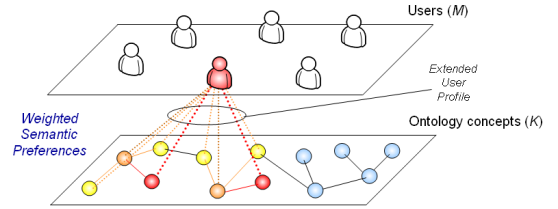

Figure 2. Semantic expansion of user preferences

3.2. Semantic context-aware recommendations

Context is a difficult notion to capture in a software system, and the elements considered in the literature under the concept of context are manifold: user tasks/goals, computing platform, network conditions, social and physical environment, location, time, text around a word, visual content of a graphic region, etc. Complementarily to the ones mentioned, we propose a particular notion useful in semantic content retrieval: that of semantic runtime context, which we define as the background topics under which user activities occur within a given unit of time. A runtime context is represented in our approach [6] as a set of weighted concepts from the domain ontologies. This set is obtained by collecting the concepts that have been involved in the interaction of the user (e.g. accessed items) during a session. The context is built in such a way that the importance of concepts fades away with time by a decay factor.

Once the context is built, a contextual activation of preferences is achieved by finding semantic paths linking preferences to context. These paths are made of existing relations between concepts in the ontologies, following the CSA technique introduced in section 3.1. This process can be understood as finding an intersection between user preferences and the semantic context, where the final computed weight of each concept represents the degree to which it belongs to each set (Figure 3). The perceived effect of contextualisation is that user interests that are out of focus, under a given context, are disregarded.
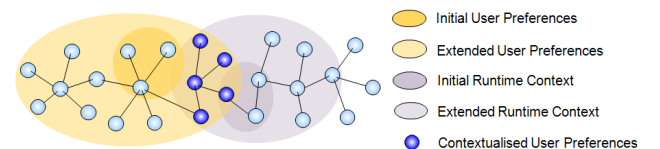

Figure 3. Semantic preference contextualisation

4. Experiments

We present an experiment conducted to evaluate the precision of the personalisation and context-aware recommendation functionalities of News@hand, and to investigate the influence of each mechanism in the integrated system, measuring the precision of the recommendations when a combination of both models is used.

4.2. Knowledge base

A set of 17 ontologies are used by the current version of the system. They are adaptations of the IPTC ontology, (http://nets.ii.uam.es/iptc-kb_v01.zip) which contains concepts of multiple domains such as education, culture, politics, religion, science, technology, business, health, entertainment, sports, weather, etc. They have been populated with concepts appearing in the gathered news items using semantic information from Wikipedia, and applying the population mechanism mentioned in section 2.1. A total of 137,254 Wikipedia entries were used to populate 744 ontology classes with 121,135 instances. The average accuracy of class assignments was 69.9%, and the average accuracy of ontology assignments arose 84.4% [4].

4.1. News item database

For two months we were daily gathering RSS feeds. A total of 9,698 news items were stored. With this dataset, we run our semantic annotation mechanism introduced in section 2.2, and a total of 66,378 annotations were obtained. We asked 20 graduate students of our department to evaluate 80 different news items from each of the 8 topic sections of News@hand, specifying whether the annotations of each item were correct or not. The annotation accuracies for each topic are presented in the table. An average accuracy of 74.8% was obtained [4].

4.3. Experimental setup

Sixteen members of our department were requested to participate. They were undergraduate/graduate students, and lecturers. The experiment consisted of two phases, and each phase was composed of two different tasks. In the first phase only personalisation was active, and its tasks were different in having the semantic expansion enabled or disabled. In the second phase, contextualisation and semantic expansion were active. On its second task we also enabled the personalised recommendations.

Search tasks

In the experiment, a task was defined as finding and evaluating those news items that were relevant to a given goal. Each goal was framed in a specific domain. We considered three domains: telecommunications, banking and social care issues. For each domain a user profile and two search goals were defined. Table 1 shows a summary of the involved tasks. To simplify the searching tasks (limiting the number of news items to be read by the users), they were defined for a pre-established section and query. Thus, for

example, the task goal of finding news items about software piracy, illegal downloads and file sharing, $Q_{1,2}$, was reduced to evaluate those articles existing in Entertainment section that were retrieved with the query "music".

Profile	Section	Query		Task goal
1 Telecom	*World*	$Q_{1,1}$	pakistan	News about media: TV, radio, Internet
	Entertainment	$Q_{1,2}$	music	News about software piracy, illegal downloads, file sharing
2 Banking	*Business*	$Q_{2,1}$	dollar	News about oil prices
	Headlines	$Q_{2,2}$	fraud	News about money losses
3 Social care	*Science*	$Q_{3,1}$	food	News about cloning
	Headlines	$Q_{3,2}$	internet	News about children, young people, child safety, child abuse

Table 1. Summary of the search tasks in the experiment

Table 2 shows the tasks performed by the 16 users.

	Personalised recommendations		Context-aware recommendations	
	Without expansion	With expansion	With expansion	
User	$w_p=1$, $w_c=0$	$w_p=1$, $w_c=0$	$w_p=0$, $w_c=1$	$w_p=0.5$, $w_c=1$
1	*$Q_{1,1}$	$Q_{2,1}$	$Q_{3,1}$	$^AQ_{1,2}$
2	$Q_{2,2}$	*$Q_{3,2}$	$^AQ_{2,1}$	$Q_{1,2}$
3	$Q_{3,1}$	$^AQ_{3,2}$	*$Q_{1,1}$	$Q_{2,1}$
4	$^AQ_{1,1}$	$Q_{1,2}$	$Q_{2,2}$	*$Q_{3,2}$
5	$Q_{1,2}$	*$Q_{2,2}$	$Q_{3,2}$	$^AQ_{2,1}$
6	$Q_{2,1}$	$Q_{3,1}$	*$^AQ_{3,2}$	$Q_{1,1}$
7	$Q_{3,2}$	$^AQ_{1,1}$	$Q_{1,2}$	*$Q_{2,2}$
8	*$^AQ_{2,2}$	$Q_{1,1}$	$Q_{2,1}$	$Q_{3,1}$
9	$Q_{1,1}$	$Q_{2,1}$	*$Q_{3,1}$	$^AQ_{3,2}$
10	$Q_{2,2}$	$Q_{3,2}$	$^AQ_{1,1}$	*$Q_{1,2}$
11	*$Q_{3,1}$	$^AQ_{2,2}$	$Q_{1,1}$	$Q_{2,1}$
12	$^AQ_{3,1}$	*$Q_{1,2}$	$Q_{2,2}$	$Q_{3,2}$
13	$Q_{1,2}$	$Q_{2,2}$	$Q_{3,2}$	*$^AQ_{1,1}$
14	*$Q_{2,1}$	$Q_{3,1}$	$^AQ_{2,2}$	$Q_{1,1}$
15	$Q_{3,2}$	*$^AQ_{3,1}$	$Q_{1,2}$	$Q_{2,2}$
16	$^AQ_{1,2}$	$Q_{1,1}$	*$Q_{2,1}$	$Q_{3,1}$

Table 2. Experiment tasks configurations

For each phase the combination of personalised and context-aware recommendations was established as a linear combination of their results using two weighs w_p, w_c:

$$score(i_n, u_m) = w_p \cdot pref(i_n, u_m) + w_c \cdot pref(i_n, u_m, context).$$

In the personalisation phase, the contextualisation was disabled (i.e., $w_c=0$). Its first tasks were performed without semantic expansion, and its second tasks had the semantic expansion activated. In the contextualisation phase, w_c was set to 1 and the expansion was enabled. Its first tasks were done without personalisation ($w_p=0$), and its second tasks were a bit influenced by the user profiles ($w_p=0.5$).

As mentioned before, a fixed user profile was used for each domain. Some of them were predefined profiles, and some of them where created by the users (those marked

with '*' in the table) using the profile editor of News@hand. In addition, some tasks were done with user profiles containing concepts belonging to all the three domains. They are marked with an 'A' in the table.

Evaluation of personalised recommendations

The objective of the two tasks performed in the first phase was to assess the importance of activating the semantic expansion in our recommendation models. The following are the steps the users had to do in these tasks:

- Launch the query with personalisation deactivated.
- Rate the top 15 news items. The allowed ratings values were: 1 if the item was not relevant to the task goal, 2 if the item was relevant to the task goal, and 3 if the item was relevant to the task goal and the user profile. These ratings, obtained without personalisation and semantic expansion, are considered as our *baseline case*.
- Launch the query with personalisation activated (and expansion enabled / disabled depending on the case).
- Rate again the top 15 items as explained before.

Evaluation of context-aware recommendations

The objective of the two tasks performed for the second phase was to assess the quality of the results when the contextualisation functionality is activated and combined with personalisation. The steps done in this case are:

- Launch the query with contextualisation deactivated.
- Rate the top 15 news items as explained before, and evaluate as relevant (clicking the title) the first two items which where related to the task goal. Doing this the current semantic context is updated.
- Launch the query with the contextualisation activated (expansion enabled, and personalisation enabled / disabled depending on the case).
- Rate again the top 15 items as explained before.

5.4. Results

Once the two evaluation phases were finished, we computed the precision values for the top N = 5, 10, 15 news items as:

$$P@N = 1/N \cdot \#\{\text{relevant items in the top } N \text{ items}\}.$$

Figure 4 shows the average results for the 16 users, taking into account those items evaluated as relevant to the task goal, and also to the user profile. In both cases, the recommendation models outperformed the baseline case, especially for the first top items. The P@5 values increased from 20% of the baseline case to almost 40% and 50% when contextualisation and personalisation functionalities were enabled. The semantic expansion seemed to be an essential component within the recommendation processes. It provided an improvement of 10% in the personalisation precision. Finally, the combination of personalised and context-aware recommendations (plus semantic expansion) gave the best results, achieving a P@5 value of 80%.

The obtained results reinforce conclusions previously

observed. Personalised recommendations help the users to find relevant news articles, and semantic expansion of user profiles ease the matching between user preferences and item annotations, improving precision values for the top suggested items.
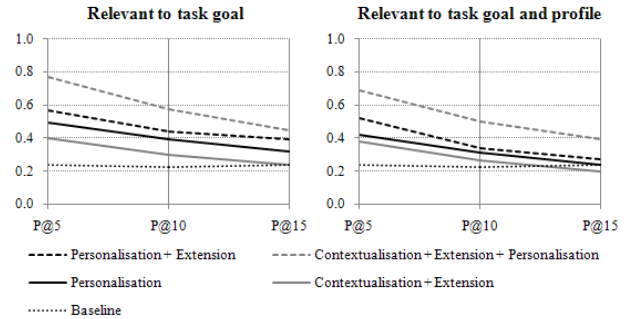
Figure 4. Average precision values for the top 5, 10 and 15 news items, taking into account those items evaluated as relevant to the task goal and the user profile

5. Summary and future work

We have presented News@hand, a news recommender system which makes use of semantic technologies to provide personalised and context-aware recommendations. In this paper, we describe experiments to preliminary assess their performance once they have been integrated and combined within the system.

We are gathering more news contents and user rating information in order to better evaluate the previous and the collaborative recommenders. We are also integrating and testing automatic mechanisms that learn user preferences from semantic contexts and freely-defined social tags.

Acknowledgements

References

[1] Alfonseca, E. Moreno-Sandoval, A., Guirao, J. M., Ruiz-Casado, M. (2006). The Wraetlic NLP Suite. *5th International Conference on Language Resources and Evaluation (LREC 2006)*.

[2] Cantador, I., Castells, P. (2006). Multilayered Semantic Social Networks Modelling by Ontology-based User Profiles Clustering: Application to Collaborative Filtering. *15th Intl. Conference on Knowledge Engineering and Knowledge Management (EKAW 2006)*, pp. 334-349.

[3] Cantador, I., Castells, P., Vallet, D. (2006). Enriching Group Profiles with Ontologies for Knowledge-Driven Collaborative Content Retrieval. *1st International Workshop on Semantic Technologies in Collaborative Applications (STICA 2006)*, pp. 358-363.

[4] Cantador, I., Szomszor, M., Alani, H., Fernández, M., Castells, P. (2008). Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations. *1st International Workshop on Collective Intelligence and the Semantic Web (CISWeb 2008)*.

[5] Castells, P., Fernández, M., Vallet, D. (2007). An Adaptation of the Vector-Space Model for Ontology-based Information Retrieval. *IEEE TKDE*, 19 (2), pp. 261-272.

[6] Vallet, D., Castells, P., Fernández, M., Mylonas, P., Avrithis, Y. (2007). Personalised Content Retrieval in Context Using Ontological Knowledge. *IEEE TCSVT*, 17 (3), pp. 336-346.