An Enhanced Semantic Layer for Hybrid Recommender Systems: Application to News Recommendation

Iván Cantador, Pablo Castells, Alejandro Bellogín
Departamento de Ingeniería Informática, Universidad Autónoma de Madrid, Spain

ABSTRACT

Recommender systems have achieved success in a variety of domains, as a means to help users in information overload scenarios by proactively finding items or services on their behalf, taking into account or predicting their tastes, priorities or goals. Challenging issues in their research agenda include the sparsity of user preference data, and the lack of flexibility to incorporate contextual factors in the recommendation methods. To a significant extent, these issues can be related to a limited description and exploitation of the semantics underlying both user and item representations. We propose a three-fold knowledge representation, in which an explicit, semantic-rich domain knowledge space is incorporated between user and item spaces. The enhanced semantics support the development of contextualisation capabilities, and enable performance improvements in recommendation methods. As a proof of concept and evaluation testbed, the approach is evaluated through its implementation in a news recommender system, in which it is tested with real users. In such scenario, semantic knowledge bases and item annotations are automatically produced from public sources.

Keywords: recommender systems, collaborative filtering, context modelling, semantics, domain knowledge, ontologies

1. INTRODUCTION

The growing volume and complexity of digital information (news, blogs, music, movies, etc.) of potential value to our daily activities challenge the limits of human processing capabilities in a wide array of information seeking and e-commerce tasks. Users need help to cope with a wealth of readily available information, in order to reach the most interesting items in online retrieval spaces. The problem is not just to find the needle in the haystack, but to select the best among thousands of needles. In such information overload scenarios, recommender systems become particularly appealing as a means to help users make choices, by proactively finding items or services on their behalf, taking into account or predicting their tastes, priorities or goals.

Recommender systems came forth by the early nineties as an emerging area at the confluence of Artificial Intelligence and Information Retrieval, to address the essential research problem of predicting or estimating the relevance of items that a user has not seen or searched for. The way in which the estimation is computed raises the distinction of two main recommendation strategies (Adomavicius & Tuzhilin, 2005): *content-based filtering* (CBF) strategies, which predict the relevance of an item for a user according to the relevance that other similar items seemed to have for him in the past, and *collaborative filtering* (CF) strategies, which predict the relevance of an item for a user by considering the relevance that other items had in the past for similar people.

It has been generally observed that combining CBF and CF methods, known as *hybrid recommendation*, is usually the best approach to mitigate the limitations CBF and CF approaches suffer separately (Adomavicius & Tuzhilin, 2005). Hybrid recommendation approaches are becoming an integral part of a large number of important e-commerce and leisure Web sites like *Amazon*, *Netflix*, and many online retailers, where recommendation models have proved successful. Nonetheless, ample room and need for further improvements remains in the current generation of recommender systems to achieve more effective algorithms, in a wider variety of applications. These improvements include, among others:

- Better coping with low data density situations in the available evidence of user preference, e.g. user input scarcity in initial situations (*cold-start problem*). When a new user enters a recommender system, no personal profile is yet available for him, and no proper recommendations can be made. Similarly, until a new item is rated by a substantial number of users, a CF recommender is not able to recommend it, as observed user preference for specific items is the data CF relies upon. This problem is particularly prevalent in *sparse* domains such as the News, where there is a constant stream of new items, and each user only rates a few.
- The consideration of *contextual information* in the recommendation strategies. Traditional recommenders operate on the two-dimensional Users×Items space, i.e., they make recommendations based solely on user and item information, and do not take into consideration additional contextual information that may be crucial in some applications. In many situations, the utility of a certain item to a user may largely depend on the circumstances under which the item would be utilised, or the temporary purpose and changing goals of users with respect to the items. Using multidimensional settings, the inclusion of knowledge about the user's task, goals, environment, etc. into the recommendation algorithm can lead to better recommendations.

Among other directions, the enhancement of the semantics representation of user preferences and item contents is being identified as a key outlook to achieve qualitative steps forward on the above problems (Anand & Mobasher, 2007; Mobasher & Burke, 2007). Classic techniques usually describe user and item profiles in terms of identifiers and numerical preference values, plain keywords, and/or attribute/value pairs with controlled vocabularies. The latent semantic meanings underneath the user and item spaces involved in recommendations, and the semantic relations between their elements, are largely underexploited when building recommendations.

Following this direction, and aiming to address the aforementioned limitations of current recommendation technologies, we propose a three-folded knowledge model, in which a space for interrelated semantic concepts is incorporated between the user and item spaces. The concepts are defined by ontology classes and instances, describing one or several domains. On top of this, user and item profiles are described by vectors consisting of weighted concepts from the ontology space. In this respect, our contribution is the definition of a formal knowledge representation of user preferences and item contents, which is not ambiguous, and takes into account arbitrary (i.e., not pre-established) semantic relations between concepts.

In order to address the *cold-start* and *sparsity* problems, we propose a strategy that spreads the weights of the ontological concepts available in user and item profiles towards other concepts that are connected through semantic relations of the domain ontologies. The semantic propagation strategy proposed herein is based on Constrained Spreading Activation techniques (Cohen & Kjeldsen, 1987; Crestani, 1997), considering the attenuation of weights as the expansion grows away from the initial set, with loop control in the propagation paths, and the possibility to bound the expansion distance. Our contribution is the design of a novel mechanism that extends the semantic descriptions of user preferences and item contents through the ontological relations of the involved concepts.

Based on the proposed ontology-based user and item profile models, we define a personalised recommendation approach based on an adaptation of the vector-space information retrieval model (Baeza-Yates & Ribeiro-Neto, 1999), which exploits the semantic extension technique mentioned above. Analogously, we also define the notion of *semantic context* as the set of ontological concepts present in the annotations of the items recently browsed or evaluated by the user. The context is described by a vector representation, so it can be easily combined with the basic personalised model.

Furthermore, alternatively to global comparisons of profiles made by state of the art recommenders, we propose a diversity-aware approach that leverages partial yet meaningful similarities between users even if they are not globally alike, but share strongly similar preferences in focused areas of interest. The hypothesis is that since user interests are not made of a single piece, an approach that deals with them as such would have inevitable limitations. We realize this principle by splitting user profiles according to meaningful groups/layers of preferences shared among users, and establishing user similarities based on the sub-profiles rather than the global profiles. Thus, coincidences of unusual preferences have further chances of being found when dealing with smaller profiles, focused on specific semantic areas of interest and taste. The semantic information plays a key role in this strategy, as a basis to relate preference data together, and determine meaningful interest layers. The relations between users at the different semantic layers represent different latent communities of interest, and are used to provide recommendations in more focused or specialised conceptual areas, even when the whole user profiles are globally fairly dissimilar. Our contribution can be expressed as building hybrid recommendation models that combine user profiles collaboratively at various semantic levels, in response to different groups of shared preferences.

In this paper, we present an evaluation of the above approaches in a controlled yet live and open setting, enabling a realistic evaluation of context-aware recommendation with interactive user feedback. Framing our scientific contributions within a consistent connection to the perspective of a feasible working application, the evaluation is conducted on a news recommender prototype, News@hand, in which all the proposed methods are integrated. As the implementation and evaluation of our methods require the availability of appropriate semantic resources, not readily available today from public sources, automatic knowledge base creation (i.e., ontology population) and semantic annotation techniques have been developed and evaluated as part of the work presented here.

The rest of the paper is organised as follows. Section 2 presents an overview and discussion of works that are related to our research. Section 3 describes the proposed ontology-based representation for recommender systems. Section 4 explains the recommendation methods we have built upon that enhanced semantic representation. Section 5 describes News@hand, the news recommender system that integrates the recommendation methods, and is been used as an evaluation platform. Section 6 presents the experiments performed with News@hand and results thereof. Finally, Section 7 provides conclusions and future work.

2. RELATED WORK: SEMANTICS IN RECOMMENDER SYSTEMS

One of the trends of work in the introduction of semantics for recommendation has been oriented towards the exploitation of social information in the recommendation methods. Approaches have been proposed that automatically collect explicit or implicit social network information from the Web and other sources, in order to apply Semantic Network Analysis methods for the study of online communities. Flink (Mika, 2005) is a system for the extraction, aggregation and visualisation of online social networks. It employs semantic technologies for reasoning with personal information extracted from a number of electronic information sources including Web pages, emails, publication archives, and FOAF profiles. Extending the traditional bipartite model of ontologies (classes and instances) with the social dimension leads to a tripartite model of the Semantic Web, namely the layer of communities and their relations (users), the layer of semantics (ontologies and their relations), and the layer of content items and their relations (the hypertext Web).

Aside from explicit social relations, further research works have focused their attention on finding implicit relations among people. Hence, for example, (Liu, Maes & Davenport, 2006) presents an implementation of "taste fabrics", a semantic mining approach to model personal tastes for different topics of interests. The taste fabric affords a flexible representation of a user in the taste-space, enabling a keyword-based profile to be relaxed by a spreading activation pattern. An evaluation of taste-based recommendation shows it compares favourably to classic CF methods, and whereas CF is an opaque mechanism, recommendation using taste fabrics can be effectively visualised, thus enhancing transparency and user trust.

In addition to the explicit and implicit definition of social relations (and the subsequent discovery of communities) to be exploited by recommender systems, other works have studied the incorporation of semantic-based knowledge representations to describe user and/or item profiles, and making enhanced, more understandable recommendations. An adaptation of the item-based CF method integrating semantic similarities for items with rating- or usage-based similarities is presented in (Mobasher, Jin & Zhou, 2004). The reported experimental results demonstrate the integrated approach yields significant advantages both in terms of improving accuracy, as in dealing with sparse datasets. An approach to ontological user profiling in recommender systems is presented in (Middleton, Roure & Shadbolt, 2004). Working on the problem of recommending on-line academic research papers, the authors present two systems, Quickstep and Foxtrot, which create user profiles monitoring the behaviour of the users, and gathering relevance feedback from them. The obtained profiles are represented in terms of a research topic ontology. Research papers are classified using ontological classes, and the proposed recommenders suggest documents seen by similar people on their current topics of interest. In this scenario, ontological inference is shown to ease user profiling, external ontological knowledge seems to successfully improve the recommendations, and the profile visualisation is used to enhance profiling accuracy.

More recently, (Anand & Mobasher, 2007) take up again the issue that most currently available recommender systems still tend to use very simplistic user models to generate recommendations. The authors contend for a fundamental shift in terms of how a user is modelled in a recommender. Specifically, they distinguish between a user's long term and short term memories, and propose a recommendation process that uses these two memories. Context-based retrieval cues are obtained to retrieve relevant preference information stored in the long term memory, and the identified relevant preferences are used in conjunction with the information stored in the short term memory to make recommendations. The paper introduces three types of contextual cues: collaborative, behavioural and semantic, and provides empirical evidence that the approach improves recommendation quality. An implementation of the semantic contextualisation proposed in the previous work is described in (Sieg, Mobasher & Burke, 2007). In this case, the authors present a strategy for personalised search that involves building models of user contexts as ontological profiles by assessing implicitly derived interest scores to concepts defined in a domain ontology. A spreading activation algorithm is used to maintain the interest scores based on the user's ongoing behaviour. The conducted experiments show that re-ranking search results based on the interest scores and the semantic evidence in an ontological user profile are effective in presenting the most relevant results to the user. Finally, (Shoval, Maidel & Shapira, 2008) proposes the incorporation of a common ontology that enables describing both the users' and items' profiles with concepts taken from the same vocabulary. Based on this representation approach, and utilising the ontology hierarchy, the authors present a content-based method for filtering items for a given user. The active user's profile is compared with the item profiles using a similarity measure that takes into account the occurrence of common concepts in both profiles, as well as the existence of "related" items according to their position in the ontology hierarchy. Based on the computed similarities, items are ranked for the user. At the time of this writing, the method is being implemented in ePaper, a personalised electronic newspaper, using an ontology that mirrors the two first levels of the IPTC[1] news taxonomy.

Our semantic-based knowledge representation and recommendation proposals, and their integrated implementation in a news recommender system, are related to the works outlined in this section.

- **Ontology-based knowledge representation**. Similarly to (Mika, 2005), we base and focus our research on a tripartite knowledge model, where user and item spaces are connected through a semantic one. As done in (Shoval, Maidel & Shapira, 2008), we propose to build this layer in terms of concepts available in domain ontologies. In our case, domain ontologies and items are respectively populated and annotated in an automatic way.
- **Spreading of semantic preferences**. The extension of ontology-based user profiles through the semantic relations of the domain ontologies (Mobasher, Jin & Zhou, 2004; Sieg, Mobasher & Burke, 2007) is also studied in this work. We show that this strategy is beneficial to mitigate the cold-start and sparsity problems in a user study.
- **Implicit communities of interest**. Like (Liu, Maes & Davenport, 2006), we discover implicit user relations (communities) from the similarities existing among semantic user preferences. Differently, in our approach, the identification of such communities is carried out at different semantic interest layers, laying the ground for building what we shall call multilayered Communities of Interest.
- **Personalised and context-aware recommendations**. Personalisation (Anand & Mobasher, 2007) and contextualisation (Sieg, Mobasher & Burke, 2007) of content retrieval exploiting an ontological knowledge representation are proposed and evaluated in this work.
- **Hybrid recommendations**. Explicit item-based collaborative recommendation from ontological user profiles was presented in (Middleton, Roure & Shadbolt, 2004). Here, we propose the exploitation of the underlying multilayered communities found by our approach,

[1] IPTC News Codes, http://www.iptc.org/NewsCodes

for making hybrid recommendations.

- **A prototype recommender system**. The integration and evaluation of our content-based and collaborative recommendation strategies in a news recommender system is reported herein. Similarly to ePaper system (Shoval, Maidel & Shapira, 2008), our prototype will make use of the IPTC news codes ontology to describe both user and item profiles, but will extend such taxonomy with concepts obtained from news contents.

3. KNOWLEDGE REPRESENTATION

Our general recommendation approaches make use of explicit user profiles (as opposed to for example sets of preferred items). User preferences are represented as vectors $\mathbf{u} = (u_1, u_2, \ldots, u_K) \in [-1,1]^K$ where the weight $u_k \in [-1,1]$ measures the intensity of the interest of user $u \in \mathrm{U}$ for concept $c_k \in \mathrm{O}$ (a class or an instance) in a domain ontology O, K being the total number of concepts in the ontology. A positive value indicates that the user is interested in the concept, while a negative one reflects a user's dislike for the concept. Similarly, items $i \in \mathrm{I}$ are assumed to be described (annotated) by vectors $\mathbf{i} = (i_1, i_2, \ldots, i_K) \in [0,1]^K$ of concept weights, in the same space as user preferences. The recommendation approaches we shall develop upon this basis (which are described later in Section 4) do not prescribe a particular granularity for the semantic representation, the domain ontology, or the user profiles. This is largely generic in the representation and recommendation approaches, and left for the application to instantiate it. Section 5 will illustrate the instantiation of the framework by the integration of a medium-sized domain ontology in a news recommender prototype.

An ontology-based representation is richer and less ambiguous than a keyword-based or item-based model, providing a number of benefits:

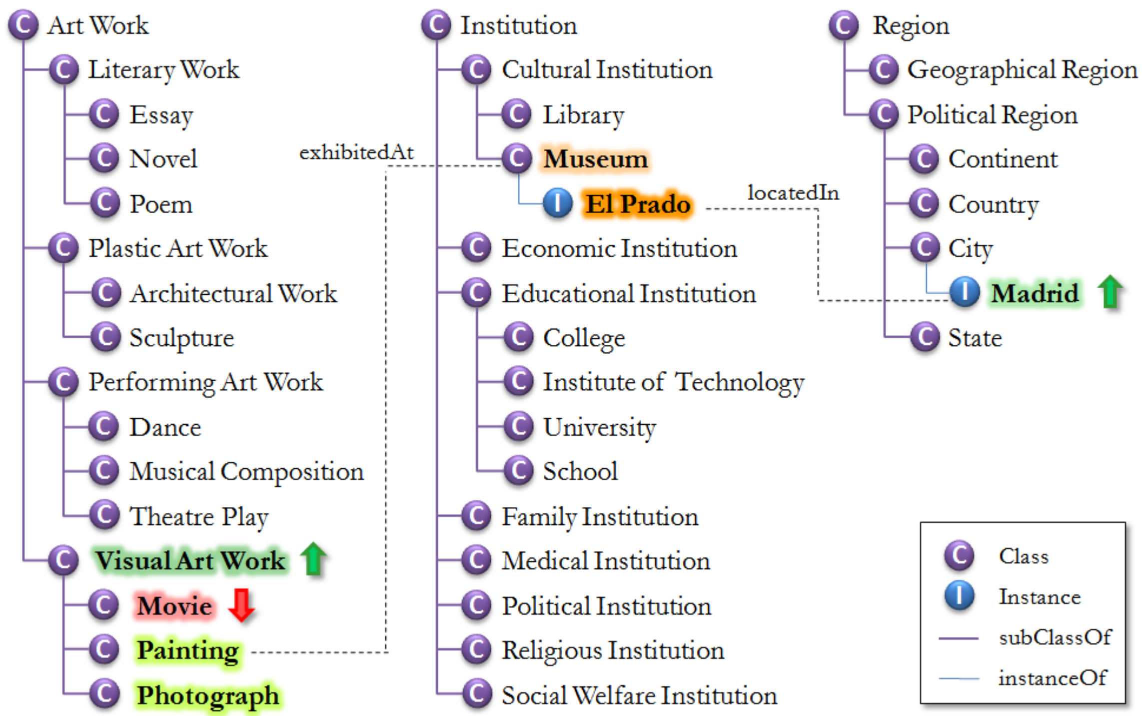
- **Semantic detail**. A concept-driven representation of user preferences is generally more meaningful for and useful for personalisation. It enables more precise specifications of users' tastes, with less ambiguity than is involved in plain keyword terms. For instance, if the system just knows about a user interest involving the keyword "java", without further information the system cannot tell whether the user is interested in the programming language or the Pacific island. Stating the preference in terms of a concept such as "ProgrammingLanguage:Java" (the instance *Java* of the *Programming Language* class) is not ambiguous in that sense, and further, allows the exploitation of any known semantics associated with the concept (e.g. Java is object-oriented), its class (e.g. related to computers), etc., which may be available in a domain knowledge base.

- **Hierarchical representation**. Ontology concepts are represented in a hierarchical way, through different hierarchy properties, such as *subClassOf*, *instanceOf* or *partOf*. Parents, ancestors, children and descendants of a concept give valuable information about the semantics of the concept. For instance, the concept *leisure* might be highly enriched by the semantics of each leisure activity, which would be described by the taxonomy that the concept could subsume.

- **Inference**. Ontology standards, such as RDF and OWL, support inference mechanisms that can be used to enhance the representation of user tastes and interests, so that, for instance, a user keen on *animals* (superclass of *dog*) could be also interested in items about dogs. Inversely, a user interested in *skiing*, *snowboarding* and *ice hockey* can be inferred with a certain confidence to be globally interested in *winter sports*. Also, a user keen on *Spain* could be assumed to like *Madrid*, through the *locatedIn* transitive relation, assuming that this relation had been seen as relevant for inferring previous underlying user's interests.

Figure 1 shows an example of conceptualised preferences. Circles indicate ontology classes and instances, solid lines represent hierarchical relations between pairs of classes (*subclassOf* property) or between classes and instances (*instanceOf* property), and dotted lines are associated to other arbitrary semantic properties. Having a set of three ontologies with information about art works, institutions and regions, let us suppose a user indicates an interest for the topic "visual art works", which is represented in the ontologies as a class *Visual Art Work* inheriting from the main class *Art Work*. The system is then able to infer preferences for *Visual Art Work* subtopics (trough the general property *subClassOf*), obtaining finer grain details about the user's preferences, such as potential interests in "paintings" and "photographs". Note that original and more specific preferences will prevail over the system's inference. In this case, as highlighted in the figure, the user is not interested in the concept "movie", whose negative weight prevails over the higher-level topic inference. Other arbitrary semantic relations can be exploited for preference extension, as shown in the figure (relations *exibitedAt*, *locatedIn*).

Figure 1: Representation of user preferences as concepts of domain ontologies



In addition to the above benefits, this kind of knowledge representation provides advantages in recommendation scenarios thanks to the use of semantic-based technologies.

- **Portability**. Based on XML standards, the domain knowledge, user profile, and item annotation information could be easily distributed, adapted and integrated in different recommender systems for different applications.
- **Domain independence**. Using a semantic-based knowledge representation, recommendation algorithms can be designed independently from the domain of discourse. Ontology hierarchies, concepts and relations are the elements to be taken into consideration for the definition of new recommenders. In principle, no domain-dependent restrictions would affect the implementation and reuse of such systems. This is not feasible for example

in model-based recommender systems, where probabilistic models are built from the available data, and cannot be used in different domains, unless the entire model is rebuilt with new data.

- **Multi-source annotation**. Assuming the existence of manual or automatic mechanisms to semantically annotate any type of content (text, images, audio, video, etc.), ontology-based recommenders could suggest items from multiple different sources without the need of changing their inner recommendation algorithms.

The use of ontology-based knowledge frameworks in recommender systems may also provide a series of advantages with respect other techniques, such as a better comprehension and explainability of the user's preferences and recommendations, an easy adaptation and extension of further personalization and recommendation models, and a potential integration of several sources of user information, which e.g. may be used to perform cross-domain recommendations. These represent promising research issues to address in the future.

3.1. Semantic extension of user preferences

To overcome cold-start and sparsity in user profiles, we propose a semantic preference spreading mechanism that expands the initial set of preferences stored in user profiles through explicit semantic relations with other concepts in the ontology. The activation of user preferences is based on an approximation to conditional probabilities. Let $u_k \in [-1,1]$ be the preference (dislike/interest) of the user u for the ontology concept $c_k \in$ O. The probability that c_x is relevant for the user can be expressed in terms of the probability that c_x and each concept $c_y \in$ O directly related to c_x in the ontology belong to the same interest topic, and the probability that c_y is relevant for the user. A similar formulation could be given for non-relevant concepts.

Let R be the set of all relations in O. The spreading strategy is based on weighting each relation $r \in$ R with a measure $w(r, c_x, c_y)$ that represents the probability that given the fact that $r(c_x, c_y)$ holds, c_x and c_y belong to the same topic. This is used for estimating the relevance of c_y when c_x is relevant for the user. With this measure, concepts are expanded through the relations of the ontology using a Constrained Spreading Activation (CSA) mechanism over the semantic network defined by these relations. As a result, the initial user profile $\mathbf{P}_u = \{c_x \in \mathrm{O} | u_{c_x} \neq 0\}$ is extended to a larger vector $\mathbf{EP}_u$, which is computed as:

$$\mathbf{EP}_u[c_y] = \begin{cases} \mathbf{P}_u[c_y] & \text{if } \mathbf{P}_u[c_y] > 0 \\ R\left(\{\mathbf{EP}_u[c_x] \cdot power(c_x)\}_{c_x \in \mathrm{O}, r(c_x, c_y)}\right) & \text{otherwise} \end{cases}$$

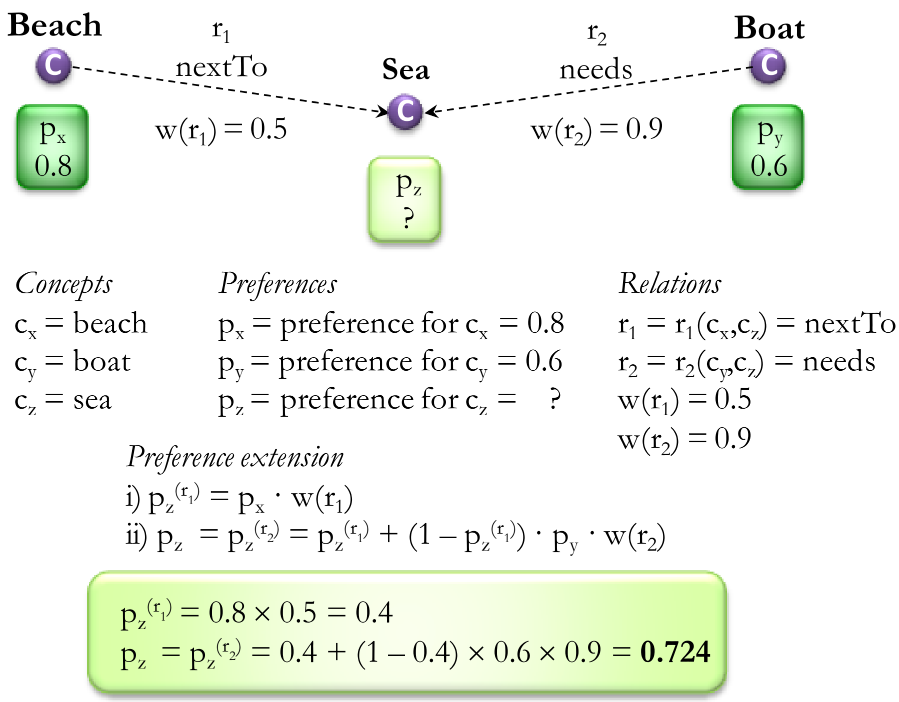
where $power(c_x) \in [0,1]$ is a propagation power assigned to each concept c_x (1 by default), and

$$R(\mathbf{X}) = \sum_{S \subset \mathbb{N}_n} \left\{(-1)^{|S|+1} \times \prod_{i \in S} x_i\right\},$$

having $\mathbf{X} = \{x_i\}_{i=0}^{n}, x_i \in [0,1]$.

The above formula is an adaptation of CSA (Cohen & Kjeldsen, 1987; Crestani, 1997) to an ontology-based personalized retrieval setting developed in (Vallet et al, 2007). Figure 2 shows a simple example of the preference extension process, where three concepts are involved. The user has preferences for two of these concepts, which are related to a third through two different ontology relations. The extension shows how a third preference is inferred, accumulating the evidence of relevance from the original two preferences.

Figure 2: Example of semantic preference extension computation



The relation weights in our experiments are different for each semantic relation type, and could in fact be different depending on the concepts being related, though in our experiments we used fixed values for each relation type, that is, $w(r, c_x, c_y) = w(r)$. Relation weighting is a modular problem in our approach, which can be addressed by several methods (co-occurrence analysis, etc.). In our experiments, we used manual values taken from prior work (Vallet et al, 2007).

Our constrained spreading activation technique can reach any concept which is connected through a semantic path (formed by semantic relations of the ontology) with the current concepts of the user's profile and context (see Sections 4.1 and 4.2). The psedudocode of the technique is shown in Figure 3. The spreading mechanism is characterised not only by the propagation decay factors (i.e., the multiplying values $w(r)$ in [0, 1]) of the semantic relations between concepts, but also by several constraints such as 1) the minimum threshold weight value ε a concept has to have in order to propagate its activation to related concepts, 2) the maximum distance (number of propagation steps) n_e to be reached by the spreading algorithm, and 3) the maximum fan-out (i.e., the number of outgoing links) n_f a concept can propagate its activation through, to reduce the "hub effect" in concepts with many relations to other concepts.

Figure 3: Pseudocode of the semantic spreading algorithm

```
function expand(P, EP, w) {
   // Init the expanded concept weights with the input ones
   for ( cx ∈ 𝒪 ) {
      EP[cx] = P[cx]
   }

   // Create a priority queue based on concept weights (initially null)
   Q ← buildPriorityQueue(𝒪×{prev=0,hierarchyLevel=0,expansionLevel=0})

   while ( Q.isEmpty() == false ) {
      // Extract the next concept to expand
      (cx, prevx, hierarchyLevel, expansionLevel) ← Q.pop()

      // Check the minimum concept weight constraint
      if ( EP[cx] < ε ) {
         exit         // The remaining concept weights are also below ε
      }

      // Check the maximum expansion constrain
      if(expansionLevel ≥ ne){
         goto while
      }

      // Expand the neighbourhood of the current concept
      for({r, cy} ∈ cx.getNeighbourhood()){
         prevy = EP[cy]

         // Check the hierarchical level expansion constrain
         if(EP[cy] = 1 OR (r.isHierarchical() AND hierarchyLevel ≥ nh)){
            goto for
         }

         // "Undo" the last update from cx
         EP[cy] ← (EP[cy] - w(r,cx,cy)*power(cx)*prevx) /
                  (1 - EP[cy]*w(r,cx,cy)*wf(cx,nf)*power(cx)*prevx

         // Do the propagation taking into account the fan-out factor
         EP[cy] ← EP[cy] + (1 - EP[cy])*w(r,cx,cy)*wf(cx,nf)*power(cx)*EP[cx]

         if(r.isHierarchical()){
            hierarchyLevel++;
         }

         Q.push(cy,prevy,hierarchyLevel,expansionLevel)
      }

      expansionLevel++
   }
}
```

4. RECOMMENDATION METHODS

Building upon the knowledge representation model presented in the previous section, we propose several recommendation methods, which have been integrated in the News@hand system. Our evaluation methodology foresees an experimental design supporting the analysis of the specific effect of each of these techniques in the overall system performance, as will be reported in Section 6. The first method, explained in Section 4.1, suggests items to a single user considering only the preferences described in his profile. The second, presented in Section 4.2, incorporates semantic contextual information into the above content-based recommendation method. Specifically, we define the notion of semantic context as the set of ontological concepts present in the annotations of items recently browsed or evaluated by the user. Finally, in Section 4.3, we extend our content-based recommendation method through a content-based collaborative (i.e., hybrid) strategy, which establishes and exploits user relations according to semantic similarities between user and item profiles.

4.1. Personalised recommendation

Our notion of personalised recommendation is based on the definition of a matching algorithm that provides a personal relevance measure $pref(u,i)$ of an item i for a user u. This measure is set according to the semantic preferences **u** of the user and the semantic annotations **i** of the item, and is based on the cosine function for vector similarity computation:

$$pref(u,i) = \cos(\mathbf{u},\mathbf{i}) = \frac{\mathbf{u} \cdot \mathbf{i}}{||\mathbf{u}|| \times ||\mathbf{i}||}.$$

The formula matches two weighted-concept vectors and produces a value in [0,1]. Values close to 0 are obtained when the two vectors are dissimilar, and indicate that user preferences negatively match the content metadata. On the other hand, values close to 1 indicate that user preferences significantly match the content metadata, which means a potential interest of the user for the item.

In Section 6, we empirically show that this and subsequent recommendation models perform better when they are used with user profiles extended through the semantic extension mechanism introduced in Section 3.1.

4.2. Context-aware recommendation

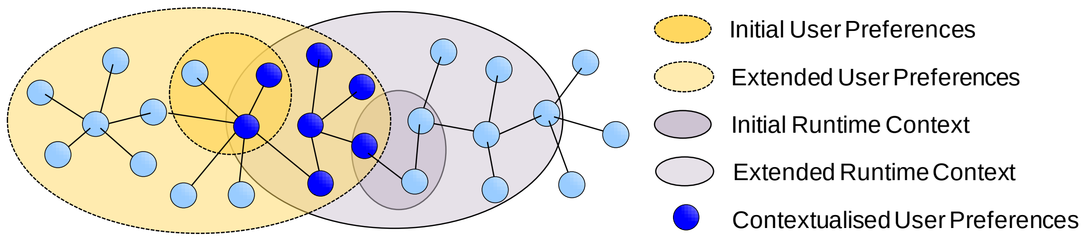
Context is a difficult notion to capture in a recommender system (Adomavicius & Tuzhilin, 2011; Dey, 2001), and the elements that can be considered under the notion of context are manifold: user tasks/goals, recently browsed/rated items, physical environment and location, time, social environment, etc. As representative examples, the reader is referred e.g. to (Ahn et al., 2007; Billsus & Pazzani, 2000; Räck, Arbanowski & Steglich, 2006; Sujiyama, Hatano & Yoshikawa, 2004). Complementarily to these, we propose a particular notion of context for semantic content retrieval: the *semantic runtime context*, which we define as the background topics $\mathbf{C}_u^t$ under which activities of a user u occur within a given unit of time t. A runtime context is represented in our approach as a set of weighted concepts from a domain ontology O. This set is obtained by collecting the concepts that have been involved in user’s actions (e.g., accessed items) during a session. Similarly to (Middleton, Roure & Shadbolt, 2004), the context is built in such a way that the importance of concepts $c_k \in$ O fades away with time (number of steps back when the concept occurred) by a decay factor $\xi \in [0,1]$:

$$\mathbf{C}_u^t[c_k] = \xi \cdot \mathbf{C}_u^{t-1}[c_k] + (1-\xi) \cdot \mathbf{Req}_u^t[c_k]$$

where $\mathbf{Req}_u^t \in [0,1]^{|\mathrm{O}|}$ is a vector whose components measure the degree in which the concepts c_k are involved in the user’s request at time t. This vector can be defined in multiple ways, depending on the application: a query concept-vector, a concept vector containing the most relevant concepts in an item, the average concept-vector corresponding to a set of items marked as relevant by the user, etc. It also can be defined in terms of semantic concepts captured from the external physical environment through sensors and ubiquitous devices (Haake et al., 2010). The decay factor ξ establishes the number of action units in which a concept is considered as in the current semantic context, i.e., how fast a concept is “forgotten” by the system when recommendations have to be made.

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 ontology, following the CSA technique explained 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 4). The perceived effect of contextualisation is that user interests that are out of focus, under a given context, are disregarded, and those that are in the semantic scope of the ongoing user activity are considered for recommendation.

Figure 4: Extension and contextualisation of user preferences



After the semantic user profile $\mathbf{P}_u^t = \mathbf{u}$ and context $\mathbf{C}_u^t$ are propagated through the ontology relations, a combination of their expanded versions $\mathbf{EP}_u^t$ and $\mathbf{EC}_u^t$ is exploited for making context-aware recommendations using the following expression:

$$\begin{aligned} pref_C(u,i) &= \lambda \cdot pref(\mathrm{EP}_u, i) + (1-\lambda) \cdot pref(\mathrm{EC}_u, i) \\ &= \lambda \cdot \cos(\mathbf{EP}_u, \mathbf{i}) + (1-\lambda) \cdot \cos(\mathbf{EC}_u, \mathbf{i}) \end{aligned}$$

where $\lambda \in [0,1]$ measures the strength of the personalisation component with respect to the current context. This parameter can be manually established by the user, or dynamically adapted by the system according to multiple factors, such as the current size of the context, and the automatic detection of a change in the user's search focus.

4.3. Hybrid recommendation

Complementarily to the two previous methods, which apply to single users, we explore the potential of the ontology-based representation to enhance collaborative recommendation techniques. Of all three approaches this is the one which more deeply explores the conjunction and joint processing of complex hybrid networks involving users, items and concepts. In the approach, links among concepts (semantic networks), between users and items (ratings), and between items and concepts (annotations) are explicit, where as relations between users are implicitly (collaboratively) derived from their links to concepts and items.

We propose to exploit the links between users and concepts to extract relations among users and derive semantic Communities of Interest (CoI) according to common preferences. Analysing the structure of the ontologies, and taking into account the preference weights of the user profiles, we cluster the domain concept space generating groups of interests shared by several users. Thus, those users who share interests of a specific concept cluster are connected in the community, and their preference weights measure their degree of membership to each cluster.

Specifically, a vector $\mathbf{c}_k = (c_{k,1}, c_{k,2}, \ldots, c_{k,M}) \in [-1,1]^M$ is assigned to each ontology concept c_k present in the preferences of at least one user, where $c_{k,m} = u_{m,k}$ is the weight of concept c_k in the profile of user u_m. Based on these vectors, a classic hierarchical clustering strategy is applied. The obtained clusters represent the groups of preferences (topics of interests) in the concept-user vector space shared by a significant number of users.

Once the concept clusters are created, each user can be assigned to a specific cluster. The similarity between a user's preferences $\mathbf{u} = (u_1, u_2, \dots, u_K) \in [-1,1]^K$ and a cluster C_q is computed by:

$$sim(u, C_q) = \frac{\sum_{c_k \in C_q} u_k}{|C_q|},$$

where c_k represents the concept that corresponds to the u_k component of the user preference vector, and $|C_q|$ is the number of concepts included in the cluster.

Taking into account the concept clusters, user profiles are partitioned into semantic segments. Each of these segments corresponds to a concept cluster, and represents a subset of user interests shared by the users who contributed to the clustering process. By thus introducing further structure in user profiles, it is now possible to define relations among users at different levels, obtaining a multilayered network of users.

Using our semantic multilayered CoI proposal, we present two recommendation models that generate ranked lists of items taking into account the obtained links between users. The first model, which is labelled as *UP* (i.e., user profile based), generates a unique ranked list based on the similarities between the items and all the existing semantic clusters. The second model, labelled *UP-q*, provides a ranking for each semantic cluster C_q. The strategies are formalised next. In the following, for a user profile $\mathbf{u}$, an information item vector $\mathbf{i}$, and a cluster C_q, we denote by $\mathbf{u}^q$ and $\mathbf{i}^q$ the projections of the corresponding concept vectors onto cluster C_q, i.e., the *k*-th components of $\mathbf{u}^q$ and $\mathbf{i}^q$ are u_k and i_k respectively if $c_k \in C_q$, and 0 otherwise.

Model UP

The semantic profile of a user u is used by the system to return a unique ranked list. The preference score of an item i is computed as a weighted sum of the indirect preference values based on similarities with other users in each cluster. The sum is weighted by the similarities with the clusters, as follows:

$$pref(u, i) = \sum_{\mathrm{q}} \mathrm{nsim}(i, C_q) \sum_{\mathrm{v} \neq u} \mathrm{nsim_q}(u, v) \cdot \mathrm{sim_q}(v, i),$$

where

$$\mathrm{sim}(i, C_q) = \frac{\sum_{c_k \in C_q} i_k}{\|\mathbf{i}\| \sqrt{|C_q|}}, \mathrm{nsim}(i, C_q) = \frac{\mathrm{sim}(i, C_q)}{\sum_{C \in \mathrm{C}} \mathrm{sim}(i, C)}$$

are the single and normalised similarities between the item i and the cluster C_q, C being the set of all clusters,

$$\mathrm{sim_q}(u, v) = \cos(\mathbf{u}^q, \mathbf{v}^q) = \frac{\mathbf{u}^q \cdot \mathbf{v}^q}{\|\mathbf{u}^q\| \times \|\mathbf{v}^q\|}, \mathrm{nsim_q}(u, v) = \frac{\mathrm{sim_q}(u, v)}{\sum_{w \neq u} \mathrm{sim_q}(u, w)}$$

are the single and normalised similarities at layer q between users u and v, and

$$\mathrm{sim_q}(u, i) = \cos(\mathbf{u}^q, \mathbf{i}^q) = \frac{\mathbf{u}^q \cdot \mathbf{i}^q}{\|\mathbf{u}^q\| \times \|\mathbf{i}^q\|}$$

is the similarity at layer q between item i and user u.

The idea behind this first model is to compare the current user's interests with those of the others users, and, taking into account the similarities among them, weight all their complacencies about the different items. The comparisons are done for each concept cluster measuring the similarities between the items and the clusters. We thus attempt to recommend an item in a double way. First, according to the item characteristics, and second, according to the connections among user interests, in both cases at different semantic layers.

Model UP-*q*
The preferences of the user are used by the system to return one ranked list per cluster, obtained from the similarities between users and items at each cluster layer. The ranking that corresponds to the cluster for which the user has the highest membership value is selected. The expression is analogous to *UP* model equation, but does not include the term that connects the item with each cluster C_q:

$$pref(u,i) = \sum_{v \neq u} \mathrm{nsim_q}(u,v) \cdot \mathrm{sim_q}(v,i),$$

where q maximises $\mathrm{sim}(i, C_q)$.

Analogously to the previous model, this one makes use of the relations among the user interests, and the user satisfactions with the items. The difference here is that recommendations are done separately for each layer. If the current semantic cluster is well identified for a specific item, we expect to achieve better precision/recall results than those obtained with the overall model.

5. NEWS@HAND

Specific aspects of the recommendation methods described in the previous section were evaluated in prior studies with artificial datasets created from external sources and standard collections (Cantador, Bellogín & Castells, 2008). The experiments with standard datasets are needed to isolate and focus on specific aspects of our approaches, and adhere to established methodologies and benchmarks. At the same time, these experiments require simplifications and adaptations of our approaches to the constraints and limitations of the collection data, which restricts the scope of techniques that can be evaluated. In particular, standard datasets do not properly support the evaluation of contextual methods involving dynamic interaction with users, in a fair enough approximation of the conditions of a real setting. Also, we aim to study the combined effect of the different methods, which in standard experiments would require different datasets.

To meet these aims, we have implemented News@hand, a news recommender system in which all the proposed recommendation strategies are integrated, and where textual contents of news are annotated with concepts belonging to a set of ontologies covering various domains. With this system, users can interact with the methods for longer periods of time, and further information can be captured as required to measure the effectiveness of the proposed techniques. Besides this integrative evaluation, the experience with News@hand has been useful to uncover the difficulties involved in the transition of the presented ontology-based models and strategies to a real application. While building the system, a number of research challenges emerged, for which additional solutions have been developed. Specifically, we needed to implement a technique to populate (i.e., create instances in) the domain ontologies, and an automatic mechanism to semantically annotate the news articles.

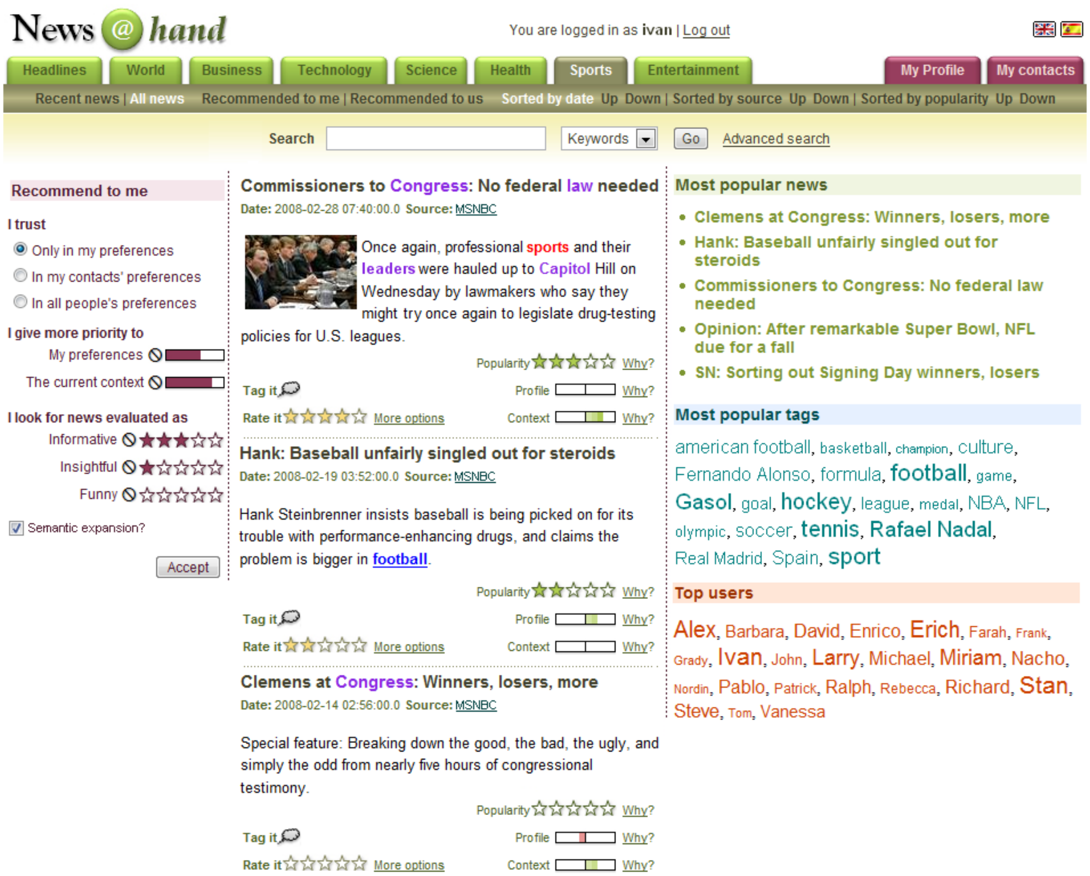
In the following, we describe News@hand, highlighting the important aspects of its architecture (Section 5.1), and explaining the followed ontology population (Section 5.2), and item annotation (Section 5.3) strategies.

5.1. Architecture

Like in other systems (Ahn et al., 2007; Jones, Quested & Thomson, 2000; Nadjarbashi-Noghani et al., 2005), in News@hand, news are automatically and periodically retrieved from several on-line news services via RSS feeds. Using Natural Language Processing (NLP) and indexing tools, the title and summary of the retrieved news are annotated with concepts (classes and instances) of the domain ontologies available to the system. Thus, for example, all the news about actors, actresses and similar terms may be annotated with the concept "actor". News@hand ontologies contain concepts of multiple domains such as education, politics, science, technology, business, entertainment, sports, etc. As done in (Ahn et al., 2007; Billsus & Pazzani, 1999), a TF-IDF technique is applied to assign weights to the annotated concepts, measuring their importance (informativeness) to the news contents in the document repository.

News@hand has a client/server architecture, where users interact with the system through a Web interface in which they receive on-line news recommendations, and update their semantic profiles. A dynamic graphical interface (Figure 5) allows the system to automatically store all the users' inputs, analyse their behaviour with the system, update their semantic preferences, and adjust the recommendations in real time. Similarly to (Claypool et al., 1999; Das et al., 2007; Chu & Park, 2009), implicit and explicit user preferences are taken into account, via manual ratings and tags, and via automatic learning from the users' actions. News@hand has a preference learning module that identifies concepts annotating recent consumed (browsed, rated, tagged) items which are susceptible of being included in the long term user profile. We do not explain News@hand preference learning mechanism since it is out of the scope of the paper. The reader is referenced to (Cantador et al. 2008, Picault et al., 2011) for further details. On the other hand, the system integrates specific facilities for manual user profile building, such as an ontology browser, and an ontology concept search engine for instant ontology class and instance suggestion, in the form of a query completion facility in the search box.

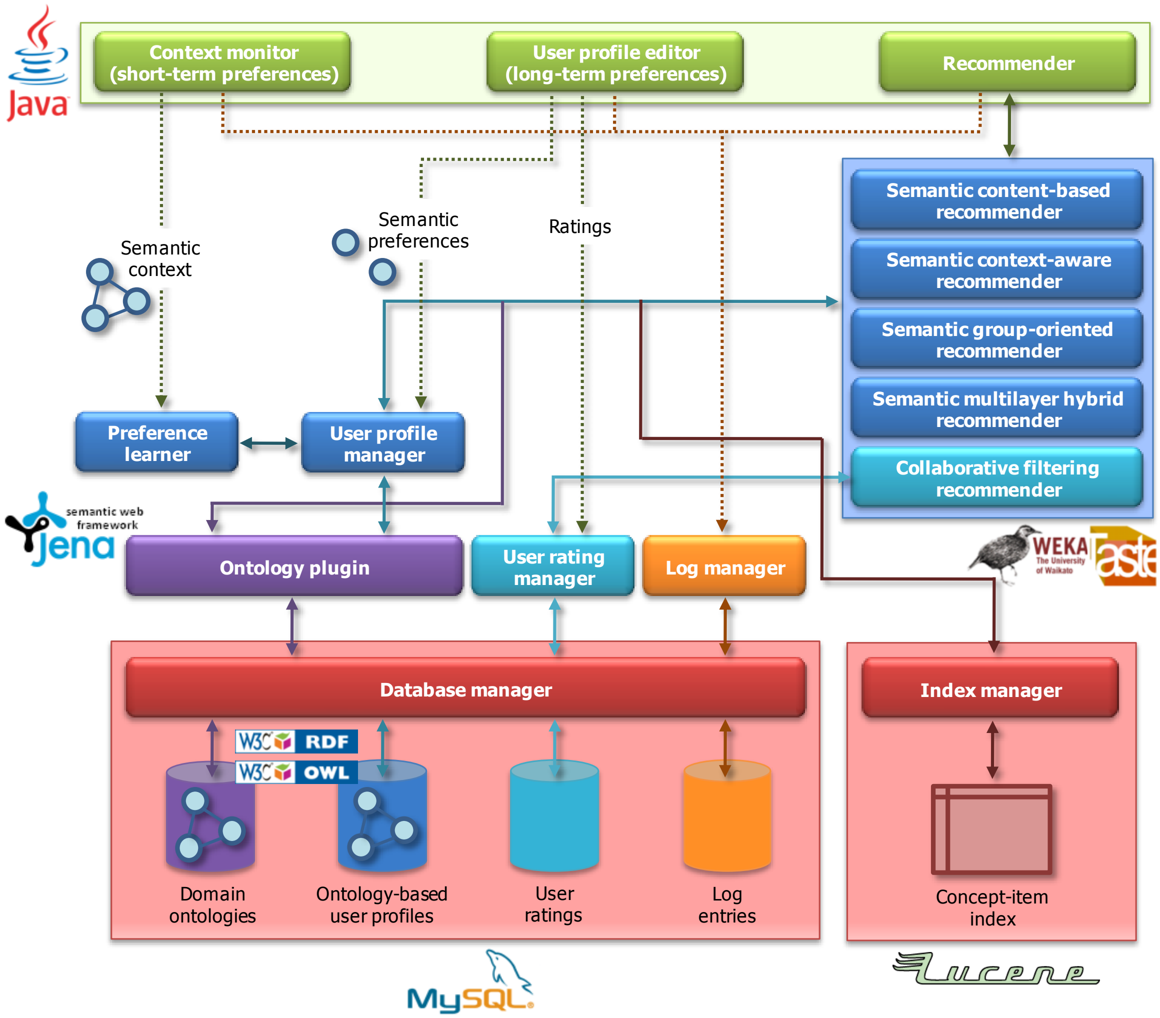
Figure 5: A typical news recommendation page in News@hand



Leveraging the semantically annotated news items, the defined ontology-based user profiles, and the knowledge represented by the domain ontologies, a set of recommendation algorithms is executed. Specifically, News@hand integrates all the recommendation models explained in Section 4, i.e., personalised, context-aware, and multilayer recommenders. Figure 6 shows a detailed schema of the system modules which are directly involved in the domain-independent semantic-based recommendation and user profiling processes. In the figure, the arrows indicate dependency relationships from a source to a target component. Three main layers of related modules can be distinguished:

- The **server-side access layer** (top part of the figure) is composed by those modules that receive requests from a client interface, and return the corresponding results: short- and long-term preference reads/updates, and recommendation responses.
- The **recommendation layer** (right part of the figure) contains and combines the proposed semantic-based personalised, context-aware and hybrid recommenders.
- The **data access layer** (bottom part of the figure) provides functionalities to manage the domain, user preference, user rating, log, and item annotation information exploited by the system using ontologies, databases and indices.

Figure 6: Architecture of News@hand



5.2. Knowledge base

In this section, we describe the Knowledge Base (KB) creation methods developed as part of the addressed research problem. A total of 17 ontologies have been used. They are adaptations of the IPTC ontology[2], which contains concepts of multiple domains. They have been populated with semantic information extracted from news contents and social tags, applying an automatic population mechanism that exploits Wikipedia, and is explained below. A total of 137,254 Wikipedia entries were used to populate 744 classes with 119,497 instances. Table 1 gathers the characteristics of the generated KB. In (Cantador, Bellogín & Castells, 2008b), we present a manual evaluation in which the ontology population process obtained 69.9% and 84.4% average accuracy values for class and ontology assignments respectively (see Table 1), according to 8,500 assessments provided by 20 subjects.

[2] IPTC ontology, http://ir.ii.uam.es/news-at-hand/iptc-ontology_v01.rdfs

Table 1: Number of classes and instances available in News@hand KB, and average accuracy of class/ontology assignments

Ontology	#classes	#instances	Avg. #instances/class	Memory (KB)	Avg. accuracy
Arts, culture, entertainment	87	33,278	383	5,347	78.7 / 93.3
Crime, law, justice	22	971	44	444	62.7 / 73.3
Disasters, accidents	16	287	18	358	74.7 / 84.0
Economy, business, finance	161	25,345	157	8,468	69.3 / 80.0
Education	20	3,542	177	649	57.5 / 76.7
Environmental issues	41	20,581	502	692	72.0 / 85.3
Health	26	1,078	41	967	65.3 / 89.3
Human interests	6	576	96	288	64.0 / 84.0
Labour	6	133	22	688	70.7 / 78.7
Lifestyle, leisure	29	4,895	169	820	72.0 / 90.7
Politics	54	3,206	59	2,989	60.0 / 81.3
Religion, belief	31	3,248	105	711	84.0 / 90.7
Science, technology	50	7,869	157	1,591	68.0 / 86.7
Social issues	39	8,673	222	2,649	70.7 / 85.3
Sports	124	5,567	45	6,454	72.0 / 86.7
Unrests, conflicts, wars	23	1,820	79	355	61.3 / 80.0
Weather	9	66	7	92	69.7 / 89.5
	744	**119,497**	**134.3** (avg.)	**33,562**	**69.9 / 84.4** (avg.)

The ontologies are populated with semantic concepts associated to noun terms extracted from the news contents to be annotated and recommended, and tags manually introduced by users (Szomszor et al., 2008). These terms are categorised as common nouns (e.g., *actor*) and proper nouns (e.g., *Brad Pitt*). The terms belonging to the first category are easily processable because their corresponding semantic concepts are the terms themselves. In this case, with simple morphological transformations, the concepts can be found in English dictionaries like WordNet (Miller, 1995). The terms of the second category may result in a complex processing. In order to infer their semantic concepts, general multi-domain semantic knowledge is needed. For News@hand, we propose to extract that information from Wikipedia. We have implemented an automatic mechanism that creates ontology instances using, among other things, the Wikipedia categories of the terms. We explain in detail the whole population process in the following.

5.2.1. Semantic information extraction

Many of the entities are ambiguous, having several meanings for different contexts. For instance, the same term "java" could be assigned to a picture of the Pacific island, or a Web page about the programming language. One approach to address disambiguation is by using the information available in Wikipedia.

A Wikipedia article is fairly structured: the title of the page is the entity name itself, the content is divided into well delimited sections, and a first paragraph is dedicated to possible disambiguation for the corresponding term. For example, the page of the entry "apple" starts with sentences such as “This article is about the fruit...”, “For the Beatles multimedia corporation, see...”, “For the technology company, see...”.

Apart from these elements, every article contains a set of collaboratively generated categories. Hence, for example, the categories created for the concept “Teide” are: world heritage sites in Spain, Tenerife, mountains of Spain, volcanoes of Spain, national parks of Spain, stratovolcanoes, hotspot volcanoes, and decade volcanoes. Processing the previous information, we could infer that “Teide” is a volcano located in Spain.

Disambiguation and categorisation information have been therefore extracted from Wikipedia for every concept appearing in our news item and social tag datasets.

5.2.2. Categorisation of terms into ontology classes

The assignment of an ontology class to a Wikipedia entry is based on a morphological matching measure between the name and the categories of the entry, and the “names” of the ontology classes. The ontology classes with most similar names to the name and categories of the entry are chosen as the classes whereof the corresponding individual (instance) is to be created. The created instances are assigned a URI containing the entry name, and RDFS labels with the Wikipedia category names.

To better explain the proposed matching method, let us consider the following example. Let “Brad Pitt” be the concept we want to instantiate. If we look for this concept in Wikipedia, a page with information about the actor is returned. At the end of the page, several categories are shown: “action film actors”, “American film actors”, “American television actors”, “best supporting actor Golden Globe (film)”, “living people”, “Missouri actors”, “Oklahoma (state) actors”, etc.

After retrieving that information, all the terms (tokens) appearing in the name and categories of the entry (which we will henceforth refer to as entry terms) are morphologically compared with the names of the ontology classes (by the name of a class we mean all the possible textual forms of the class, assuming a class-label mapping is available, as is usually the case). Applying singularisation and stemming, and computing the Levenshtein distance, only the entry terms that match some class name above a certain similarity threshold (set to 2 in the experiments) are kept. For instance, suppose that “action”, “actor”, “film”, “people”, and “television” are the entry terms sufficiently close to some ontology class names, e.g., “Actions”, “Actors”, “Films”, “People” and “Television”, respectively.

To select the most appropriate ontology class among the matching ones, we firstly create a vector whose components correspond to the filtered entry terms, taking as values the numbers of times each term appears in the entry and category names together. In the example, the vector might be as follows: {(action, 1), (actor, 6), (film, 3), (people, 1), (television, 1)}, assuming that “actor” appears in six categories of the Wikipedia entry “Brad Pitt”, and so forth. Once this vector has been created, one or more ontology classes are selected by the following heuristic:

- If a single component holds the maximum value in the vector, we select the ontology class that matches the corresponding term. Here, we are assuming that in Wikipedia, a concept belongs to several categories that are semantically related through hierarchical relations (i.e., there are subcategories or sibling categories between them), and are morphologically similar, sharing main terms, e.g., subcategories of “Actors” are “Female actors”, “Male actors”, “Ancient actors”, “Child actors”, “Fictional actors”, etc.
- In case of a tie between several components having the maximum value, a new vector is created, containing the matched classes plus their taxonomic ancestor classes in the

ontologies. Then, the weight of each component is computed as the number of times the corresponding class is found in this step. Finally, the original classes that have the highest valued ancestor in the new vector are selected.

In our example, the weight for the term “actor” is the highest, so we select its matching class as the category of the entry. Thus, assuming that the class matching this term was *Actor*, we finally define *Brad Pitt* as an instance of *Actor*.

Now suppose that, instead, the vector for *Brad Pitt* was {(actor, 1), (film, 1), (people, 1)}. In this case, there would be a tie in the matching classes, and we would apply the second case of the heuristic. We take the ancestor classes, which could be for example “cinema industry” for “actor”, “cinema industry” for “film”, and “mammal” for “person”, and create a weighted list with the original and ancestor classes. Then, we count the number of times each class appears in the previous list, and create the new vector: {(actor, 1), (film, 1), (person, 1), (cinema industry, 2), (mammal, 1)}. Since the class *Cinema industry* has the highest weight, we finally select its sub-classes *Actor* and *Film* as the classes of the instance *Brad Pitt*.

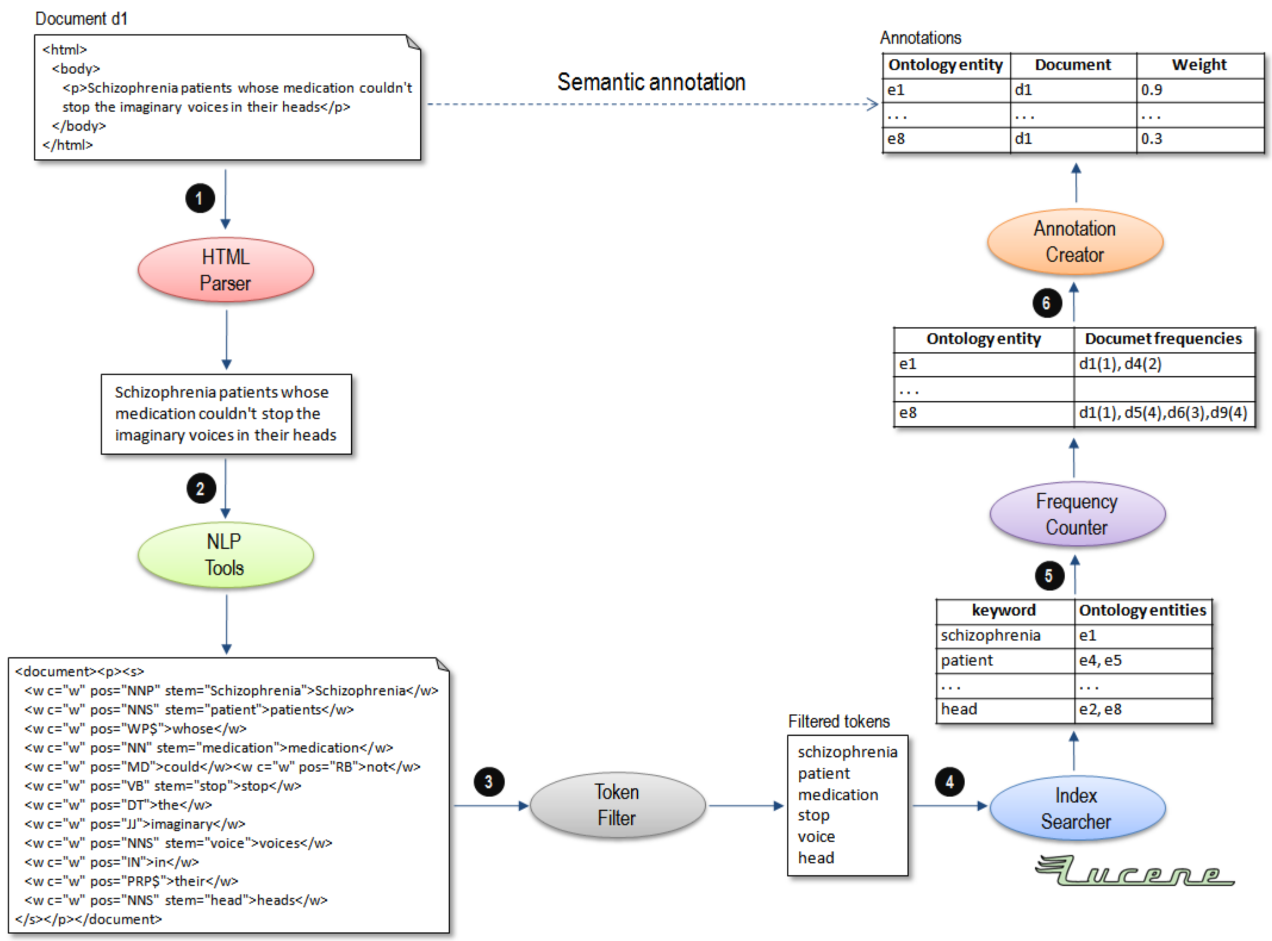
We must note that our ontology population mechanism does not necessarily generate individuals following an “is-a” schema, but a more relaxed, fuzzier semantic association principle. This is not a problem for our final purposes, since the annotation and recommendation methods are themselves rooted on models of inherently approximated nature, for example regarding the relationships between concepts and item contents. Nonetheless, there is a wide range of works on ontology population, and alternative techniques could be investigated in the future.

5.3. Item annotation

News@hand periodically retrieves news items from websites of well-known news and media sources. These items are obtained via RSS feeds, and contain information of published news articles: their title, summary of 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 with concepts (classes and instances) existing in the domain ontologies, and have been previously indexed.

Using Wraetlic NLP tools (Alfonseca et al., 2006), an annotation module removes stop words, and extracts relevant (simple and compound) terms, categorised according to their Part of Speech (PoS): nouns, verbs, adjectives, etc. Then, nouns are morphologically compared with the names of the classes and instances of the domain ontologies. The comparisons are done using an ontology index created with Lucene, and according to fuzzy metrics based on the Levenshtein distance. For each term, if similarities above a certain threshold are found, the most similar semantic concepts are chosen and added as annotations of the news items. After all the annotations are created, a TF-IDF technique computes and assigns weights to them. Figure 7 shows a more detailed view of the annotation mechanism, which takes as input the HTML document to annotate, and the system ontology indices, and returns as output new entries for the annotation database.

Figure 7: Semantic annotation mechanism



The steps illustrated in the figure are:

- A Web document is parsed removing HTML tags and meaningless textual parts (in terms of not having or being related to news contents).
- The remaining text is analysed by the Wraetlic tools to extract the PoS and the stem of each term.
- The information provided by the linguistic analysis is used to filter the less meaningful terms (determinants, prepositions, etc.), and to identify those sets of terms that can operate as individual information units.
- The filtered terms are searched in the ontology indices, obtaining the subset of semantic entities to annotate.
- The annotations are weighted according to the semantic entity frequencies within individual documents and the whole collection.
- The annotations are added to a relational database.

The next subsections explain in more detail the previous steps and provide information about the gathered and annotated news contents. We run our semantic annotator on a set of 9,698 news items daily retrieved during two months. The ontological KB from which we obtained the semantic concepts appearing in the annotations is the one explained in Section 5.3. A total of

66,378 annotations were created. Table 2 describes the information gathered and annotated for each news section. In (Cantador, Bellogín & Castells, 2008b), we present a manual evaluation in which the semantic annotation process of news items obtained an average accuracy of 74.8% (see Table 2), according to 800 assessments provided by 20 subjects.

Table 2: Average number of annotations per news item, and average accuracy of the annotation process for each news section

News section	#news items	#annotations	Avg. #annotations/item	Avg. accuracy
Headlines	2,660	18,210	7	71.4
World	2,200	17,767	8	72.7
Business	1,739	13,090	8	79.2
Technology	303	2,154	7	76.3
Science	346	2,487	7	74.1
Health	803	4,874	6	73.1
Sports	603	2,453	4	75.8
Entertainment	1,044	5,343	5	76.0
	9,638	**66,369**	**6.5** (avg.)	**74.8** (avg.)

5.3.1. Natural language processing of news contents

The NLP of news contents is performed by means of the Wraetlic linguistic-processing tools, an XML suite for processing texts which performs the following tasks:

- **Segmentation**: identification of lexical units in the texts. It is done by two components: a *tokeniser* which finds word boundaries, and a sentence splitter which locates the sentence boundaries.
- **Part-of-Speech (PoS) tagging**: assignment of a PoS to each token. A *PoS tagger* labels each token with its corresponding PoS. Wraetlic tools utilise the PoS tags of the Penn Treebank corpus[3], and take into consideration the grammatical context of a word (i.e., its surrounding terms) to infer its PoS.
- **Morphological analysis**: study of the inner structure of the words. For each token, a morphological analyser identifies the root (stem), which contains the basic meaning of the word, and the bound morphemes (prefixes and suffixes), which vary the basic meaning, e.g., by pluralizing a noun (e.g., “parent” and “parents”), or by changing an adjective into a noun (e.g., “wide” and “width”).

5.3.2. Semantic annotation

The semantic annotator identifies ontology entities (classes and instances) within the text documents, and generates links between the identified ontology entities and the documents using index structures.

Semantic annotations are assigned weights that reflect how well the ontology entities represent the meaning of the document. Weights are computed by an adaptation of the TF-IDF

[3] The Penn Treebank Project, http://www.cis.upenn.edu/~treebank

algorithm, and based on the frequency of the occurrences of each ontology entity within the document. Initially, the frequency of occurrences of an entity in a document was defined as the number of times any of its associated "mappings" appears in the document text. In preliminary experiments, however, we realised that quite a number of occurrences were missed, since we were not considering pronouns as entity occurrences. To slightly overcome this limitation, we included a modification in the algorithm to also count pronoun occurrences in a sentence if an entity was previously identified. This modification does not help to increase the annotation accuracy or incorporate new annotations, but enhances the preciseness of the annotation weights that will be later used during the recommendation processes.

The weight $w_{k,n}$ in the annotation of a document i_n with an ontology entity c_k is computed as:

$$w_{k,n} = TF\text{-}IDF_{k,n} = \frac{freq_{k,n}}{\max_j freq_{j,n}} \times log \frac{N}{N_k},$$

where $freq_{k,n}$ is the number of occurrences in i_n of the keywords attached to c_k, $\max_j freq_{j,n}$ is the frequency of the most repeated ontology entity in i_n, N_k is the number of documents annotated with c_k, and N is the total number of documents.

On the other hand, we exploit the PoS information provided by Wraetlic tools to identify and discard those words that typically do not provide significant semantic information, and to group sets of words that can operate as individual semantic information units. The following are some examples of the considered word group patterns.

- *Noun* + *noun*. E.g., "tea cup".
- *Proper noun* + *proper noun*. E.g., "San Francisco".
- *Proper noun* + *proper noun* + *proper noun*. E.g., "Federico García Lorca".
- *Abbreviation* + *proper noun* + *proper noun*. E.g., "F. García Lorca".
- *Abbreviation* + *abbreviation* + *proper noun*. E.g., "F. G. Lorca".
- *Participle* + *preposition*. E.g., "located in", "stored in".
- *Modal verb* + *participle* + *preposition*. E.g., "is composed by", "is generated with".

6. EXPERIMENTS

The News@hand system was implemented to support comparative evaluations of the proposed recommendation models with real users, without the inherent restrictions that previous isolated (off-line) experiments with public datasets imposed (Cantador, Bellogín & Castells, 2008). The integration of the recommendation models in a single evaluation framework enables A/B tests in which each particular recommendation functionality is activated/deactivated for comparison. Specifically, we aim to evaluate and empirically compare the effect of semantic preference expansion and contextualization on our personalized recommender, and assess the potential improvements resulting from the proposed hybrid recommendation approach. By alternately activating/deactivating these functionalities, we discriminate, observe and measure the effect of each other separate from the rest. In our experiments, we state the following hypotheses:

- Extending the semantic descriptions of user preferences through the ontological relations of the involved concepts will mitigate sparsity and cold-start user situations. By applying semantic preference expansion, user profiles become larger, covering more areas of the conceptual space, and resulting in a higher likelihood of finding user and item similarities and correlations.

- Adding semantic context into the personalised recommendation process allows casting the user's preferences into the scope of the ongoing user activity, and obtaining more accurate recommendations.
- Building hybrid models that combine user profiles collaboratively at various semantic levels, in response to different groups of shared preferences, will enhance personalised recommendations, and would help to address the grey-sheep effect.

In this section, we present the conducted experiments, but before that, we identify the evaluation cases we should investigate in order to cover the validation of the above recommendation functionalities. We also list general steps that should be followed in the identified evaluation cases.

Activation/deactivation of functionalities

The following are identified as the significant comparisons to be investigated in order to properly assess the performance of the personalisation and recommendation functionalities.

- **Test 1**. **Evaluation of the semantic preference extension mechanism**, by activating and deactivating it in the personalised recommendation model. When no semantic expansion is applied, the above model reduces to a keyword vector space model, which is analogous to our semantic-based approach but performing exact matching between keyword vectors.
- **Test 2**. **Evaluation of the effect of semantic contextualisation**, by activating and deactivating it in the personalised recommendation model. In this case, the semantic preference extension mechanism is activated to better find intersections between the extended versions of the user profile and context.
- **Test 3**. **Evaluation of the hybrid recommendation approach against the personalised recommender**. In this case, the semantic preference extension mechanism is also activated since it enhances the semantic clustering process (Cantador et al., 2008).
- **Test 4**. **Evaluation of the hybrid recommendation approach against a rating-based CF strategy** (Konstan et al., 1997) in order to measure the quality of provided recommendation. In this case, it is important to note that our aim is not no conduct an exhaustive empirical comparison of our hybrid recommender with state of the art CF strategies (Koren & Bell, 2011), but to show the feasibility of our approach in terms of obtaining performance values of comparable magnitude with respect to widely used standard recommendation approaches. In order to conduct more precise experiments, a significant amount of rating data would be needed, and issues such as the ontology population and semantic annotation processes would have to be improved.

Table 3 shows the functionalities in each of the four proposed testing cases. We consider the basic ontology-based approach without semantic preference extension as a form of simple keyword-based content retrieval technique. Since, in this case, we do not perform any semantic inference, concept vectors are treated as plain keyword vectors.

Table 3: Functionalities to be evaluated (activated) in each testing case

		Personalisation functionalities			Recommendation Functionalities	
		Keyword-based personalisation	***Ontology-based personalisation***	***Context-aware personalisation***	***Collaborative filtering***	***Hybrid recommendation***
Evaluation of personalisation	*1*	X	X			
	2		X	X		
Evaluation of recommendation	*3*		X			X
	4				X	X

Execution of evaluation tasks

We propose an experimental protocol where every subject performs several tasks. Each pair of tasks is aimed to evaluate a specific testing case. A user does not have to deal with all the testing cases, but only a subset evenly distributed (according to a *Latin square* design) so that users and tasks do not introduce any bias in the performance of the different configurations. An average result is finally obtained for each evaluation case from the corresponding tasks performed by the users.

The following is a general scheme about how the experimentation has to be conducted.

- T specific search tasks are defined. We set $T = 6$.
- Each user performs $2S$ tasks (with $S \leq T/2$). We set $S = 2$ (i.e., 4 search tasks per user).
- The tasks of each user will be used to evaluate S testing cases: a pair of tasks addresses a specific testing case activating or deactivating the involved functionalities.
- The first task for each user is used to make the user familiar with the system functionalities and experiment tasks. Thus, the evaluation of the results obtained in this task is omitted.
- Average precision/recall results are measured for each testing case and all users involved in the testing case.

The experiments described in the next subsections have been designed following the previous evaluation methodology. The definition of the tasks and the computation of the precision/recall values will be different depending on which functionality is tested.

6.1. Evaluating personalised and context-aware recommendations

We conducted an experiment to evaluate the precision of the personalisation and contextualisation functionalities. With this experiment we also wanted 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.

The experiment was done with sixteen subjects, recruited among members of our department. They were PhD students and academics staff members. The experiment consisted of two phases, each composed of two different tasks.

- In the first phase, only the personalisation module was active, and its tasks were different in having the semantic preference extension enabled or disabled. The baseline in this phase is the keyword-based recommender, and the goal is to evaluate the effect of semantic expansion of user preferences.

- In the second phase, the semantic preference extension was activated, and the contextualisation was alternately activated and deactivated. On its second task, we also enabled the personalised recommendation in order to compare the effect of combining personalisation and contextualisation.

Search tasks

A task was defined as finding out 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 manually defined (see below). Table 4 shows a summary of the involved tasks. To simplify the searching tasks, they are defined for pre-established sections and queries. 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 from the query "music".

Table 4: Summary of the search tasks performed in the experiment

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

The configuration and assignment of the tasks were set according to the following principles:

- A user should not repeat a query during the experiment.
- The domains should be equally covered by each experiment phase.
- A user has to manually define a user profile once in the experiment.

For each phase, the combination of personalised and context-aware recommendations was established as a linear combination of their results using two weights $w_p, w_c \in [0,1]$:

$$score(u,i) = w_p \cdot pref(u,i) + w_c \cdot pref_C(u,i).$$

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

User profiles

Static user profiles were used for each domain. Some of them were common predefined profiles, and others were created by the users during the experiment using the profile editor of

News@hand. In addition, some tasks were done with user profiles containing concepts belonging to all the three domains.

Table 5 lists those concepts included in the predefined domain-driven user profiles. Each domain was described with six semantic concepts, appearing in a significant number of item annotations. Note that each domain may be described by concepts belonging to different ontologies, and may be covered with news items of different news sections.

Table 5: Topics and concepts allowed for the predefined user profiles in the evaluation of personalised and context-aware recommenders

Domain	Concepts
Telecommunications	internet, network, satellite, technology, telecommunications, website
Banking	bank, banking, business, economy, euro, dollar
Social care	drug, health, immigration, safety, social abuses, terrorism

Analogously to the predefined user profiles, those manually created by the evaluators using the profile editor of News@hand contained semantic concepts of the above three domains. In this case, the evaluators were free to select their preferences from concepts available in the entire KB. No restriction was placed on the number, type (classes or instances) and ontology of the concepts. Table 6 shows the concepts included for each domain, and the average size of the sixteen profiles. For instance, in *Telecommunication* domain, 55 preferences were declared using 30 different semantic concepts, producing an average of 3.4 preferences per user. On average, each profile contained 3.2 preferences of each domain.

Table 6: Topics and concepts of the manually-defined user profiles in the evaluation of personalised and context-aware recommenders

Domain	Concepts	#pref.	Avg. #pref./user
Telecommunications (30 concepts)	blackberry, cell phone, computer programming, computer sciences, computing and information technology, digital voice, email, encryption, file sharing, free downloads, internet, internet history, mobile network operator, network theory, networks, router, search engine, signal processing, social search, software, technology, telecommunications, television, tfidf, video arcade, video call, video game, voice over internet, web crawler, web search	55	3.4
Banking (25 concepts)	bank, bank charges, bank machine, bank of america, banker, banking, business, cash, credit card, dollar, economy, euribor, euro, euro interbank offered rate, finance, foreign exchange market, funds, ibank, macroeconomics, microfinance, money, payment system, stock, stock broking, trade policy	46	2.9
Social care (26 concepts)	abstinence, abuse, adoption, charity, children, civil society, drug, drug trafficking, family, gay, health, homophobia, homosexuality, immigration, pornography, safety, sexuality, smoking, social abuses, social change, social development, social groups, teenagers, terrorism, victims, volunteerism	51	3.2

Steps for the evaluation of the personalised recommendation

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

- Launch the query with the personalisation module deactivated.
- Rate the top 15 news items. The allowed rating 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 are considered as our baseline case.
- Launch the query with the personalisation module activated (and the semantic extension enabled/disabled depending on the case).
- Rate the new top 15 news items as explained before.

Steps for the evaluation of the context-aware recommendation

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

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

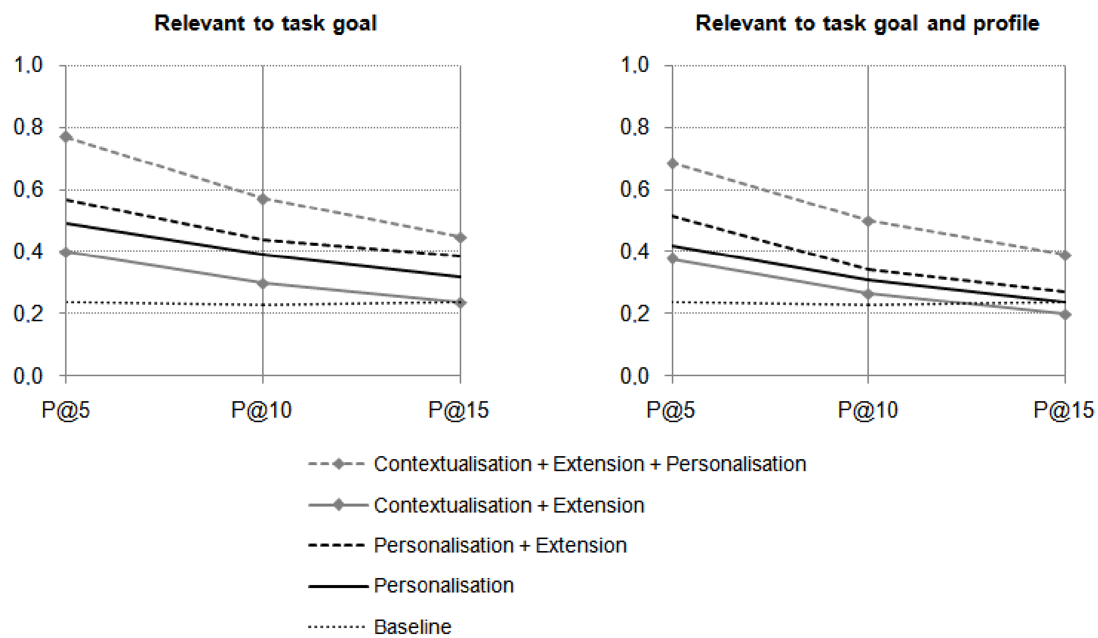
Results

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

$$P@N = \frac{\#\{relevant\ items\ in\ the\ top\ N\ news\ items\}}{N}.$$

Figure 8 shows the average results for the sixteen users, taking into account those items evaluated as relevant to the task goal, and also the user profile. In both cases, the recommendation models outperformed the baseline case (except some, for P@15), especially for the five 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 extension seemed to be an essential component. It accounts for a 10% improvement in the personalisation precision in this experiment. This is in line with results we obtained in previous studies (Cantador, Bellogín & Castells, 2008; Cantador, Castells & Bellogín 2007) for different domains and applications of our semantic extension approach, thus adding empiric evidence on the benefits of the semantic user preference expansion in our content-based recommender. Finally, the combination of personalised and context-aware recommendations (plus semantic extension) gave the best results, achieving a P@5 value of 80%. The results in P@10 showed a similar trend, slightly less far above the baseline, and only for P@15 two configurations perform the same as or very slightly below the baseline, namely semantic personalisation and contextualisation alone. We performed a two paired sample Wilcoxon signed rank test, comparing 1) the baseline against the rest of the algorithms, 2) the personalised recommender with and without semantic preference expansion, and 3) the effect of enabling and disabling semantic contextualisation, obtaining respectively p values $p \leq 0.002$, $p \leq 0.05$ ($p = 0.049$ for results marked as relevant to the task goal, $p = 0.040$ for results marked as relevant to the task goal and the user's profile), and $p \leq 0.006$ ($p = 0.006$ for results marked as relevant to the task goal, $p = 0.005$ for results marked as relevant to the task goal and the user's profile). For all the recommendation models, no significant differences were found in the recommendation performance when using pre-defined and manually created user profiles. We think that this was due to the fact that the average numbers of preferences in both types of profiles were quite similar (see Tables 5 and 6).

Figure 8: 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



Apart from the computation of the precision values, we also asked the evaluators to provide comments and suggestions about the system. The most remarkable feedback we obtained can be summarised in the following points:

- **The contextualisation of recommendations is a useful functionality**. The users noticed and positively assessed how news items relevant to the current search goal move up to the top positions of the ranked lists when the context-aware recommender is activated.
- **A disambiguation mechanism should be included within the annotation process**. The users found out annotations whose terms appeared in their profiles but having different meanings. This not only worsened the generated recommendations, but also the users' evaluations.
- **A content-based collaborative approach to enrich the semantic profiles may be beneficial**. Several users declared some preferences assuming that related ones (e.g., synonyms) were going to be implicitly taken into account. A mechanism to exploit co-occurrences among preferences of different users could be useful to automatically add

related semantic concepts into the profiles.

- **The incorporation of a user preference recommender would be helpful**. Despite the facilities offered by the ontology browser and the auto-complete concept search boxes of News@hand, several users missed the fact of having concept suggestions (e.g., in the form of "related preferences are…") when they had to create their profiles.

6.2. Evaluating content-based collaborative recommendations

A second experiment was conducted with News@hand to evaluate the multilayer hybrid recommenders. The objective of this experiment was to compare the recommendations provided by our hybrid models with those obtained using a classic CF approach. Again, an off-line execution of the recommendation strategies over a set of user profiles and ratings was performed in order to compute accuracy measures.

The sixteen members of our department who participated in the previous experiment were again requested to take part of the evaluation presented herein. Three phases were followed by each user, assessing news recommendations for three news sections: *Business*, *Sports* and *World* (see below why we selected these sections). For each phase, two tasks were defined:

- In the first task, the users had to rate a number of news items from a random list. The goal of this task was to obtain a significant amount of personal rating data, with which collaborative filtering approaches could be executed in an offline process.
- In the second task, the users had to rate several news items from a list generated with the personalisation functionality activated. The goal of this task was to ensure obtaining a set of user ratings with which content-based, collaborative filtering and hybrid approaches could be compared.

Search tasks

A task was defined as finding out and rating those news items that were "related to" a personal user profile. By "related to" we mean that a news item contains semantic annotations whose concepts appear in the user's profile. Note that a concept could be assigned negative or positive weights within a profile, so the evaluation of an item might have a low (close or equal to 1 star) or a high (close or equal to 5 stars) rating values.

User profiles

Similarly to the experiment described in Section 6.1, the evaluators were asked to choose their preferences. However, in this case, they could only select preferences from a given list of semantic concepts. They were provided a form with a list of 128 concepts, classified in 8 different domains. From this list, the users had to select a subset of concepts, and assign them negative/positive weights according to personal interests. Table 7 shows the concepts available for each domain, and the average number of preferences per user. On average, each profile was created with 7.8 preferences per domain, duplicating the preferences introduced by the users when they had to manually search the concepts in the ontology browser (see Section 5.2).

Table 7: Topics and concepts allowed for the user profiles in the evaluation of the hybrid recommenders

Domain	Concepts	#preferences	Avg. #pref./user
Computers Technology Telecommunications	computer, digital, ebay, google, ibm, internet, mass, media, microsoft, networking, online, satellite, software, technology, video, website	135	8.4
Wars Armed conflicts	al-qaeda, army, battle, combat, crime, kidnapping, kill, memorial, military, murder, peace, prison, strike, terrorism, war, weapons	104	6.5
Social issues	aids, assassination, babies, children, death sentence, divorce, drugs, family, health, hospital, immigration, love, obesity, smoking, suburb, suicide	115	7.2
Television Cinema Music	actor, bbc, cinema, cnn, film, grammy, hollywood, movie, music, musician, nbc, radio, rock, oscar, singer, television	129	8.1
Sports	baseball, cricket, football, lakers, nascar, nba, new england patriots, new york giants, nfl, olympics, premier league, running, sports, soccer, super bowl, tennis	168	10.5
Politics	george bush, condolezza rice, congress, democracy, elections, government, hillary clinton, john maccain, barack obama, parliament, politics, president, senate, senator, voting, white house	104	6.5
Banking Economy Finance	banking, business, cash, companies, earnings, economy, employment, finance, fraud, gas price, industry, marketing, markets, money, oil price, wall street	120	7.5
Climate Weather Natural disasters	air, climate, earth, earthquake, electricity, energy, fire, flood, forecast, fuel, gas, pollution, sea, storm, weather, woods	128	8.0

Once the user profiles were created, we identified which news sections contained news items annotated with the most popular (i.e., the most used) preferences. The goal was to define an item set from which the recommenders could provide a significant number of personalised recommendations. Finally, we selected the news sections mentioned previously: *Business*, *Sports* and *World*.

Steps for the evaluation of the collaborative filtering and hybrid recommendations

The users had to perform three tasks, each of them in one of the following news sections: *Business*, Sports and *World*. Successively, for each section, a user had to:

- Deactivate the personalisation functionality, and display the news items of the section. The goal is to present to all the users the same set of news items, in order to obtain a “shared” group of rated items.
- Rate 20 news items that are related (with negative or positive weights) to the user profile. Taking into account the similarities between item annotations with user preferences, assign a 1-5 start rating to the selected news items. No restriction is placed on which items have to be

rated.

- Activate the personalisation functionality, and display again the news items of the section. This time the order (ranking) of the news items is different to the one shown previously. The goal here is to present to each user a set of news items that might be related to his semantic profile. Thus, content-based similarities could be found among profiles of different users.
- Rate (as explained before) 50 news items not evaluated previously.

With this strategy, the sixteen users provided a total of 3,360 ratings for 859 different news items.

Results

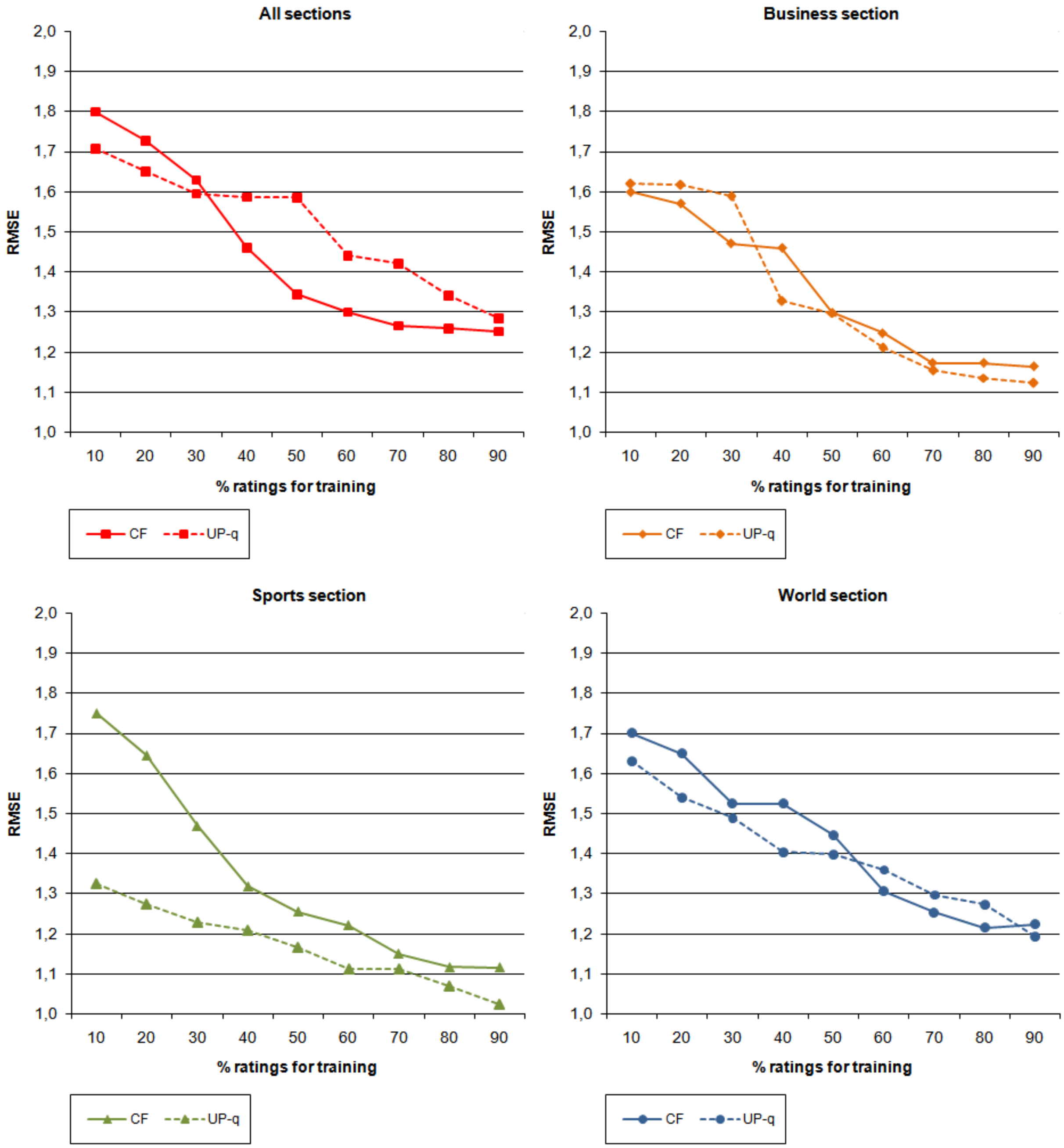
The purpose of the experiment was to compare the accuracy values obtained with our multilayer hybrid recommendation model UP-q, with those achieved by a classic item-based CF strategy (Konstan et al., 1997).

We computed the accuracy of the recommendations using different percentages of the user ratings to build (train) and evaluate (test) off-line the models. For comparative purposes, we computed the Root Mean Squared Error (RMSE) between the actual ratings $r_{u,i}$ introduced by the users u for items i belonging to the test set T, and the ratings $\hat{r}_{u,i}$ predicted by the above recommenders:

$$RMSE = \sqrt{\frac{1}{|T|}\sum_{(u,i)\in T}(r_{u,i} - \hat{r}_{u,i})^2}\,.$$

Figure 9 shows separately the average results for the items belonging to the three considered news sections. In *Business* and *World* sections, the accuracy values of both models seem to be very similar. For the *World* section, the UP-*q* strategy performs slightly better than CF when 10% to 50% of the ratings were used to build the recommenders. In the *Sports* section, the UP-*q* model provides better recommendations at all sparsity levels. The user profiles created with concepts of this domain were rich, enabling the discovery of similarities among the user interests. For the *Business* section, however, there is no significant difference. Checking the news items profiles, we noticed that there was a relative small number of annotations about banking, economy and finance. As one might expect, our semantic-based approach is sensitive to the availability of semantic information. In the general case where items of the three sections were taken into account, the hybrid model seems again to give more accurate recommendations when few ratings are available. Specifically, utilising 10%, 20% and 30% of the rating information, the UP-*q* error is lower than the error obtained with the CF strategy.

Figure 9: Average Mean Squared Error of item-based collaborative filtering (CF) and semantic multilayer hybrid (UP-q) recommendation strategies using 10%, 20%, …, 90% of the available ratings for building (training) the models, and the rest for testing



The observed results show that the semantic approach enhances the recommendation performance as soon as the semantic data provides a fair coverage. Further research might be worth on this point to more closely study the dependence between recommendation accuracy and semantic data sparsity. On the other hand, we see that the hybrid recommender improves the performance of the baseline in cold-start and sparsity situations. However, when enough user data is available, CF becomes better than our approach. This motivates further research to find an optimal point up to which our hybrid approach should be applied, after which CF should take over.

Apart from the computation of accuracy metrics, we gathered more subjective assessments of the system. We asked the evaluators to provide us comments about the recommendations obtained during the experiment. The most remarkable observations were the following:

- **Very similar news items were closely shown**. The non-diversity problem has not been addressed in this work. In the current version of the system, a certain news item can be retrieved from different RSS sources, and might be recommended to the user several times. Various users did not rate some items because they had already evaluated very similar ones.
- **A disambiguation mechanism should be included within the annotation process**. As noticed in the evaluation of the personalised and context-aware recommenders, the users found out semantic annotations with wrong meanings.
- **The contextualisation of recommendations is a desirable functionality even when collaborative item suggestions are provided**. Several users missed the activation of the context-aware recommender for this experiment. They also suggested us to consider additional sources of context, such as the semantics of news items linked through spatial (location) and temporal relations.
- **The rating of news items according to the user profile seemed to be difficult in some cases**. Several users found difficult to rate some news items because they could not easily distinguish between interesting and pleasant-reading articles.

7. CONCLUSIONS AND FUTURE WORK

Our research elaborates on the incorporation of a conceptual space describing and connecting user preferences and item contents, as a means to enhance recommendations. The specific results of our research are:

- The definition of a formal (ontology-based) knowledge model, supporting the expression of explicit semantic relations between concepts.
- The design of flexible semantic content-based recommenders, allowing the contextualisation of the recommendations.
- The design of semantic-layered hybrid recommenders, drawing further benefit from content-based and collaborative filtering approaches.
- The integration and joint evaluation of the proposed approaches with end users in a news recommender system. The semantic recommendation methods are executed in parallel in the system, and their outputs are combined in different configurations.

The results confirm previous observations from partial, separate evaluations of the methods, within the restrictions of standard datasets, and provide additional findings which cannot be obtained in the isolated experiments. The personalised recommendations help users find relevant news articles, and the semantic extension of user preferences supports a richer matching between user and item profiles, improving precision for the top suggested items, and mitigating the cold-start and sparsity problems. The incorporation of contextualisation in the personalisation mechanism got positive subjective feedback from the users, and supported faster discovery of items related to current search goals. Finally, layered hybrid recommendations enhance content-based collaborative approaches in the experiments when partial (interest-focused) comparisons of user profiles are computed. In this context, the detection of relations among users at multiple interest layers may be reducing the effect of the grey sheep problem, as shown in the reported results. This issue has to be investigated in depth in future work.

The implementation of a recommender system based on a semantic representation of user preferences and item features raised interesting additional challenges which we addressed as well. First, we had to build a knowledge base from scratch comprising different domains. For that purpose, we developed an automatic ontology population mechanism that extracts semantic information from several public information sources such as WordNet and Wikipedia. Next, we annotated news contents with classes and instances from the domain ontologies. To this end, we developed an automatic semantic annotator that makes use of NLP tools to analyse and process texts, retrieving their semantic concepts. Finally, we provided easy to use interactive facilities for users to edit their semantic profiles. The experimental work has produced valuable feedback from the users about system functionalities and outputs. Among other issues, user comments evidenced the need for a more elaborated semantic disambiguation step in the annotation process. Preliminary results in this direction are reported in (Cantador et al., 2011). The users also noticed the need of addressing the non-diversity problem, as very similar (or even the same) items were presented closely in the recommendation sets. Moreover, they suggested additional improvements in the personal profile editor, such as the integration of a real-time preference recommender which takes into account concepts related to the already introduced ones (synonyms, co-occurrences, etc.).

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, **17**(6), 734-749.

Adomavicius, G., & Tuzhilin, A. (2011). Context-Aware Recommender Systems. In Ricci, F., Rokach, L., Shapira, B., Kantor, P. B. (Eds.) *Recommender Systems Handbook*, 217-253.

Ahn, J., Brusilovsky, P., Grady, J., He, D., & Syn, S. Y. (2007). Open User Profiles for Adaptive News Systems: Help or Harm? In *Proceedings of the 16th International World Wide Web Conference*, 11-20.

Alfonseca, E., Moreno-Sandoval, A., Guirao, J. M., & Ruiz-Casado, M. (2006). The Wraetlic NLP Suite. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*.

Anand, S. S., & Mobasher, B. (2007). Contextual Recommendation. In *Berendt, B., Hotho, A., Mladenic, D., Semeraro, G., From Web to Social Web: Discovering and Deploying User and Content Profiles*, 142-160.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. Addison-Wesley.

Billsus, D., & Pazzani, M. J. (1999). A Personal News Agent that Talks Learns and Explains. In *Proceedings of the 3rd International Conference on Autonomous Agents*, 268-275.

Billsus, D., & Pazzani, M. J. (2000). User Modeling for Adaptive News Access. *User Modeling and User-Adapted Interaction*, **10**(2), 147-180.

Cantador, I., Bellogín, A., Fernández-Tobías, I., & López-Hernández, S. 2011. Semantic Contextualization of Social Tag-based Item Recommendations. In *Proceedings of the 12th International Conference on E-Commerce and Web Technologies*. To appear.

Cantador, I., Bellogín, A., & Castells, P. (2008). A Multilayer Ontology-based Hybrid Recommendation Model. *AI Communications*, **21**(2-3), 203-210.

Cantador, I., Bellogín, A., & Castells, P. (2008b) Ontology-based Personalised and Context-aware Recommendations of News Items. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, 562-565.

Cantador, I., Fernández, M., Vallet, D., Castells, P., Picault, J., and Ribière, M. (2008). A Multi-Purpose Ontology-Based Approach for Personalised Content Filtering and Retrieval. In Wallace, M., Angelides, M., and Mylonas, P. (Eds.), *Advances in Semantic Media Adaptation and Personalization*, 25-51.

Cantador, I., Szomszor, M., Alani, H., Fernández, M., & Castells, P. (2008). Enriching Ontological User Profiles with Tagging History for Multi-Domain Recommendations. In *Proceedings of the 1st International Workshop on Collective Semantics: Collective Intelligence and the Semantic Web*, 5-19.

Chu, W., & Park, S. T. (2009). Personalized Recommendation on Dynamic Content Using Predictive Bilinear Models. In *Proceedings of the 18th International World Wide Web Conference*, 691-700.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin, M. (1999). Combining Content-Based and Collaborative Filters in an Online Newspaper. In *Proceedings of the SIGIR 1999 Workshop on Recommender Systems: Algorithms and Evaluation*.

Cohen, P. R., & Kjeldsen, R. (1987). Information Retrieval by Constrained Spreading Activation in Semantic Networks. *Information Processing and Management*, **23**(4), 255-268.

Crestani, F. (1997). Application of Spreading Activation Techniques in Information Retrieval. *Artificial Intelligence Review*, **11**(6), 453-482.

Das, A., Datar, M., Garg, A., & Rajaram, S. (2007). Google News Personalisation: Scalable Online Collaborative Filtering. In *Proceedings of the 16th International World Wide Web Conference*, 271-280.

Dey, A. K. (2001). Understanding and Using Context. *Personal Ubiquitous Computing* **5**(1), 4-7.

Haake, J. M., Hussein, T., Joop, B., Lukosch, S., Veiel, D., & Ziegler, J. (2010). Modeling and Exploiting Context for Adaptive Collaboration. *International Journal of Cooperative Information Systems*, **19**(1-2), 71-120.

Jones, G. J., Quested, D. J., & Thomson, K. E. (2000). Personalised Delivery of News Articles from Multiple Sources. In *Proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries*, 340-343.

Konstan, J. A., Miller, B. N., Maltz, D., Herlocker, J. L., Gordon, L. R., & Riedl, J. (1997). GroupLens: Applying Collaborative Filtering to Usenet News. *Communications of the ACM*, **40**(3), 77-87.

Koren, Y., Bell, R. B. (2011) Advances in Collaborative Filtering. In Ricci, F., Rokach, L., Shapira, B., Kantor, P. B. (Eds.) *Recommender Systems Handbook*, 145-186.

Liu, H., Maes, P., & Davenport, G. (2006). Unravelling the Taste Fabric of Social Networks. *International Journal on Semantic Web and Information Systems*, **2**(1), 42-71.

Middleton, S. E., Roure, D. D., & Shadbolt, N. R. (2004). Ontology-based Recommender Systems. In *Staab, S., Studer, R., Handbook on Ontologies*, 477-498. Springer-Verlag.

Mika, P. (2005). Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. *Journal of Web Semantics*, **3**(2-3), 211-223.

Miller, G. (1995). WordNet: A Lexical Database for English George. *Communications of the ACM*, **38**(11), 39-41.

Mobasher, B., Jin, X., & Zhou, Y. (2004). Semantically Enhanced Content-based Collaborative Filtering on the Web. In *Berendt, B., Hotho, A., Mladenic, D., Van Someren, M., Spiliopoulou, M.: Web Mining: From Web to Semantic Web*, 57-76.

Nadjarbashi-Noghani, M., Zhang, J., Sadat, H., & Ghorbani, A. A. (2005). PENS: A Personalised Electronic News System. In *Proceedings of the 3rd Annual Communication Networks and Services Research Conference*, 31-38.

Picault, J., Ribière, M., Bonnefoy, D., & Mercer, K. (2011) How to Get the Recommender Out of the Lab? In Ricci, F., Rokach, L., Shapira, B., Kantor, P. B. (Eds.) *Recommender Systems Handbook*, 333-365.

Räck, C., Arbanowski, S., & Steglich, S. (2006). Context-aware, Ontology-based Recommendations. In *Proceedings of the 2006 International Symposium on Applications and the Internet Workshops*, 98-104.

Shoval, P., Maidel, V., & Shapira, B. (2008). An Ontology- Content-based Filtering Method. *International Journal of Information Theories and Applications*, **15**, 303-318.

Sieg, A., Mobasher, B., & Burke, R. (2007). Ontological User Profiles for Personalized Web Search. In *Proceedings of the AAAI 2007 Workshop on Intelligent Techniques for Web Personalization*, 84-91.

Sujiyama, K., Hatano, K., & Yoshikawa, M. (2004). Adaptive Web Search Based On User Profile Constructed Without Any Effort From Users. In *Proceedings of the 13th International World Wide Web Conference*, 675-684.

Szomszor, M., Alani, H., Cantador, I., OHara, K., & Shadbolt, N. R. (2008). Semantic Modelling of User Interests based on Cross-Folksonomy Analysis. In *Proceedings of the 7th International Semantic Web Conference*, 632-648.

Vallet, D., Castells, P., Fernández, M., Mylonas, P., & Avrithis, Y. (2007). Personalized Content Retrieval in Context Using Ontological Knowledge. *IEEE Transactions on Circuits and Systems for Video Technology* **17**(3), 336-346.