SCIRES-IT
SCIentific RESearch and Information Technology
Ricerca Scientifica e Tecnologie dell'Informazione
Vol 1, Issue 2 (2011), 129-158
e-ISSN 2239-4303, DOI 10.2423/i22394303v1n2p129

A SEMANTIC APPROACH FOR RECOMMENDATIONS GENERATION: SOME CULTURAL HERITAGE APPLICATIONS

Maurizio De Tommasi[*], *Gianluca Lorenzo*[*], *Gianluca Solazzo*[*]

[*]Centro Cultura Innovativa d'Impresa, Università del Salento, Italy (maurizio.detommasi@unisalento.it, gianluca.lorenzo@unisalento.it, gianluca.solazzo@unisalento.it)

Abstract

The growing availability of data in the information systems has raised the challenging problem of distinguishing between the resources that belong to the same information context. Starting from the hypothesis that the information system is based on Semantic Web technologies, is it possible to use these technologies to make an information system more adaptive to user requirements in order to enable personalization and differentiation mechanisms in the information delivery process?

This paper proposes an approach to building recommendations by using Semantic Web technologies, in order to give the users a different access to the information. The outcome is a semantic recommender engine, capable of retrieving and ranking semantically annotated resources, by using a set of domain ontologies and a semantic matching algorithm. We are showing some applications of this model in the Cultural Heritage domain in which the presented approach seems to be particularly effective, due to the richness of semantic structures and models existing for such domain.

1. Introduction

The World Wide Web technology has radically changed the way people access knowledge sources, enabling new search and retrieval processes on unstructured, heterogeneous and distributed sources of information. The web success story is based on a tremendous growth in terms of number of nodes, users and available knowledge. The number of web users has in fact moved from 16 million in 1996 to 1billion of 2006, while the number of web sites is changed from the few of the 1994 to the millions of today, transforming the web into a plethora of knowledge in which highly diverse information is linked in extremely complex and arbitrary fashion. The simple usage of the web access technologies has supported this growth and at the same time is preventing the web maturity. Several estimates of the total number of web pages indicate that, due to a rapid growth of the web, most search engines are only finding a fraction of all the available sites [22]. If we are able to exclude topological reasons, due to the asymmetric behavior of the web link, that prevent automatic search mechanism to reach certain web regions, the main challenge is currently related to the creation of smarter and powerful indexing systems, capable of selecting and indexing web pages with at a faster rate than they appear on the web. In the last years different search engines are competing for web users share. This competition, that currently sees Google as winner, is based on the challenge for new and more efficient algorithms for searching and indexing web pages. These search mechanisms, although extremely complex and more and more efficient, are mainly based on syntactic analysis. The need for a semantic layer, allowing an explicit representation of the semantics of data, is fundamental if we take into account that:

– the main part of the current web content is thought to be used by human beings and it cannot be easily manipulated by automatic systems;
– the web, because of its effectiveness as a business tool, is becoming more and more the web of services.

The next generation of the web will be the Semantic Web. The semantic web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation[5]. According to Tim Berners-Lee, who first introduced it, the Semantic Web is a Web of data that can be processed directly or indirectly by machines [4], capable of allowing automatic access to information through a computable semantic and a set of meta-data. The explicit representation of the semantics of data in relation with the creation of domain theory could enable a new web, with improved capability and a higher level of service. The transformation from web to web of knowledge is mainly based on the acquisition and the redesign of information through an innovative way of representing information which can increase the level of description with an explicit definition of its own semantics. The introduction of a format that is understandable by both humans and machines will allow, on one side, humans to represent semantics and, on the other side, extended and specialized reasoning systems to support humans to access the right knowledge and service. From this perspective, semantic web is a direct implication of Artificial Intelligence technologies that, instead of reproducing some of the human being capabilities, will try to complement human capabilities in interacting with a widest and complex environment of knowledge and services. According to Sowa, the three main components in the knowledge representation are logic, ontology and computing power [11]. Logic is related to a formal representation language. Ontology takes care about meaning and about taxonomy of classes. Computing power enables the capability to implement automatic manipulation and reasoning. The Berners-Lee Semantic Web architecture is mainly organized as a stack of layers:

– the level of data, mainly represented through XML, describes data and defines how data should be formatted;

– the level of schema defines dictionaries and grants the mechanical rules for semantic interoperability;

– the ontological level provides common comprehension about terms and supports domain vocabularies definition;

– the logical level grants the rules and conventional semantics that enables knowledge representation [23].

Although the Semantic Web represents the reference framework for the web evolution, it is currently facing the challenge of becoming the wide adopted solution for the web infrastructure. Limiting factors in the Semantic Web adoption are mainly related to production of ontologies, an activity that is far fromt he common user capabilities. According to the multiple layers Semantic Web structure, in order to enable a semantic web scenario in a specific domain, it is necessary to develop ontologies describing the knowledge item and the service characterizing that domain. Although much research effort and investment has been done to simplify ontology design and to develop domain ontologies, many domains still remain uncovered and, therefore, not suitable for adoption of Semantic Web technologies. On the other hand, syntactic search and retrieval mechanisms are becoming more and more effective, reducing the need for Semantic Web based system. An excellent driver for the Semantic Web diffusion has been the business adoption of these technologies. Many companies and industrial sectors, looking for the interoperability of knowledge and services invested money on research and industrial applications devoted to Semantic Web adoption in order to achieve interoperability among different systems. The Semantic Web approach promises to achieve semantic interoperability using a Reference Ontology and a Semantic Annotation Language based on the former. Among typical business applications of Semantic Web approach there are [24]:

– information and knowledge management systems(for the organization and the retrieval of enterprise knowledge);

– e-business and inter firms coordination and orchestration systems(by annotating local resources - such as information and processes -they support business cooperation among enterprise software applications).
– e-commerce and customer relationship technologies (semantic matchmaking of requested and offered services);

From this perspective, the adoption of Semantic Web technologies is particularly relevant in those business domains in which specific needs have supported the creation of a set of the domain ontologies, or where the use of annotation mechanisms could support a better application of Semantic Web paradigm to the Service Oriented Architecture adoption. With the increasing growth in the popularity of Web services, discovery of relevant Web services becomes a significant challenge. The syntactic approach, extremely effective in the document management approach and extremely powerful on the current web, is absolutely useless in the Web Service search and the retrieval process since the business aspects related to the web services are poorly described in its WSDL. To provide an effective solution to this problem it is necessary to shift toward a semantic level of interaction making explicit the semantics hidden in an application interface. Tourism, and more in details Cultural Heritage driven tourism, is extremely interesting domain of application for Semantic Web technologies for both the reasons that we mentioned above, since:

– in tourism many initiatives and projects (Fetish[1], Teschet[2],Mais[3]) developed ontologies for domain interoperability;
– in tourism, due to its intrinsic fragmentation, Semantic Web Service approach is particularly requested as a solution for inter-firms coordination and orchestration.

In this technological context, how is it possible to take advantage of such paradigms in order to handle the large amount of information that currently Information Systems supply? Moreover, how is it possible to fulfill the user expectations and needs in terms of adaptation to user's requirements? The key relies on the integration between Information Systems and tools that perform the delivery of personalized informative resources [12].

[1] Federated European Tourism Information Systems Harmonisation, http://www.fetish.t-6.iton
[2]Technology System for Cultural Heritage in Tourism, http://www.teschet.org
[3] Multichannel Adaptive Information Systems, http://www.mais-project.it

In this paper, we present our approach for a generation of the semantic recommendation, based on the domain ontologies processing and on the semantic annotation of resources in Cultural Heritage applications. Moreover we introduce a new metric, the Power Accuracy Measure (PAM), in order to measure the Recommender Engine's accuracy. While there is still much work to be done to fulfill our long-term goals, we believe that the work we present here can serve as a road map to build similar systems.

2. Recommender Systems: a Literature Review

Recommender Systems (RS) became very popular in the 90's, by offering a solution to the problem of information overload in the World Wide Web. In a few years, many approaches have been developed and used, and each of them presents some benefits and some disadvantages. Recommender Systems are able to learn over time user preferences and, through their analysis, they are able to automatically identify and propose relevant products or services to the user. Recommender Systems are also able to dynamically track how the user interests change by building a user profile from his preferences. They can "observe" the behavior of the user during his interaction with the information system, building and updating his profile preferences. The acquisition of the user knowledge is very important because it is necessary to collect a huge amount of information in order to grant the correctness of the profiles.

The term *personalization* usually refers to the process of providing updated information in the most suitable way with respect to the user's needs. *Personalization* is also referred to as a one-to-one marketing techniques used in the management of user-related information in order to tailor a business to a specific user rather than to a broad group of customers with different characteristics.

Personalization should be intended as the activity that allows the establishment of longterm relationship between the user and an information system, in which the platform learns more and more information about the users. As a consequence it will better satisfy their specific needs and establish a trusted relationship with the users. If an information system contains a component that is able to monitor the user's behavior during their work sessions, and collect and organize user data(age, profession, interests, preferences, etc.), updating the online selection (history), and reasoning in order to deliver personalized content, the quality of theuser's experience in the interaction with the system can be improved. Currently, the most diffused systems for the real-time personalization are able to recognize the user, either explicitly at each login or implicitly through some cookies stored by a web browser tracking the users' navigation history, and to adapt the presentation of contents and services. In the e-Business scenario, the main ways to get real-time personalization are:

– identification: the user is recognized and welcomed each time he logs on tothe system;
– customization: the user can decide to customize the supplied services to his needs;
– narrow-casting: the user can choose to be advised for the same kind of events by email, sms, etc.;
– recommendation: the system itself proposes products/services that match the user's needs, both explicit and implicit (elaborated by the system).

Systems that improve customer loyalty, thanks to stored memory about the user, can contribute to create an added value relationship between the user and the system.

2.1 Main Technological Characteristics of Recommender Systems

Actual Recommender Systems can be divided in three categories [2]:

– Content-based Recommender Systems: in the content-based approach, the system tries to recommend items similar to those in which a given user has indicated interest in the past. The content-based approach to Recommender System has its root in Modeling and Information Filtering. A pure content-based system has several shortcomings. Generally, only a very shallow analysis of certain kinds of content can be supplied. In some domains, useful feature extraction methods do not exist for certain items (such as movies, music, restaurants). A second problem, which has been studied extensively in several domains, is "over-specialization". When the system can only recommend items scoring highly against a user's profile, the user is restricted to see items similar to those already rated. Finally, there is a problem common to most recommendation systems that is to elicit user feedback. Rating items is an onerous task for users, so with the pure content-based approach, a user's own ratings are the only factor influencing future performance, and there is actually no way to reduce the quantity without also reducing performance[1],[10],[8],[7],[25].

– Collaborative Recommender Systems: in the collaborative approach, the system identifies users whose preferences are similar to those of the given user and recommends items they have liked. Thus, a pure collaborative recommendation system is one which does not require any analysis of the items. Recommendations for a user are made on the basis of similarities to other users. Pure collaborative recommendation solves all of the short comings given for pure content-based systems. By using other users' recommendations, we can deal with any kind of content and recommends items with dissimilar content to those seen in the past. Since other users' feedback influences what is recommended, there is the potential to maintain effective performance due to fewer ratings individual users must give. However, this approach does introduce some problems. For example, if a new item appears in the database there is no way it can be recommended to a user until more information about it is obtained through other user ratings or by specifying other items it is similar to. Thus, if the number of users is small in comparison to the volume of information in the system, there is the risk the coverage ratings can become very sparse, making the collection of recommendable items thin. A second problem is related to unusual clients, for which there are other particularly similar users, leading to poor recommendations [30],[20],[27].

– Hybrid Recommender Systems: this approach tries to leverage the positive aspects of both content-based and collaborative-filtering systems, while avoiding their drawbacks. Generally, in order to determine recommendations a hybrid Recommender System implements algorithms that use both content and item's attributes as well as user's opinions [9].

The main research questions related to the Recommender Systems are the following:

1. Knowledge acquisition techniques: this problem consists in deciding which techniques should be used to collect the user related information. The user knowledge can be obtained in both explicit and implicit ways. The implicit knowledge acquisition is the preferred way to collect the information due to its low impact on the user interaction with the system. The transparent monitoring of user activity is useful to discover the behavioral data. It needs although a certain degree of interpretation in

order to understand the reasons behind the user's behavior. It is therefore, a process prone to errors. The explicit knowledge acquisition requires that the user periodically interacts with the system in order to provide a feedback. This kind of knowledge has a high degree of confidence because it is directly provided by the user, and it is not obtained after an interpretative process. The explicit feedback can consist of interests, item preferences or priorities. It is possible to provide explicit feedback by defining the rules for the selection and filtering of the information [27].

2. Information sharing: this problem consists in deciding how the user knowledge can be processed to create the user profiles. For example, it is very useful to share users' feedback in order to improve future recommendations. It is also very useful to share a set of the most preferred items in order to increase the number of the training set and to improve the classification accuracy.

3. User profile representation: this problem consists in how to represent a profile in a suitable form. For example, a vector space model can be used to represent user profiles as a vector of characteristics. It is easy to apply machine-learning techniques to this kind of representation for elaborating recommendations [21].

4. Recommendation techniques [30], [13]: one of the most important requirements for a Recommender System is to use a recommendation technique capable of producing suitable suggestions for every user. There is a large number of recommendation techniques, but most of them can be associated to one of the following categories:

– Machine learning techniques that use similarity as the parameter to classify interesting items [19],[25];

– Filtering rules, that use heuristics to classify items according to a possible interest [3];

– Collaborative filtering techniques, with which it is possible to recommend items that users, with a similar profile, have chosen in the past. Statistic functions are used to calculate the recommendations, by discovering the similarity of the profiles [17], [33].

3. Semantic recommendation generation

As we said in the previous sections, due to the huge amount of the information available on an information system, providing the users with personalized information is becoming a hard task. The complexity of retrieving personalized information is also actually related with the growing user expectations.

Starting from the hypothesis that a given information system is based on Semantic Web technologies, is it possible to use them to make an information system adaptive to user requirements in order to enable personalization and differentiation mechanisms in the information delivery process?

To answer to this question we developed an approach for the generation of a semantic recommendation using a semantic annotation of resources and a semantic processing of resources in order to compute a degree of affinity between a target resource and a set of the available ones. The whole approach is based on the existence of a set of domain ontologies that represent the layer in which the semantics associated to the resources resides.

Our approach consists of the following steps:

– setup of domain ontologies;
– semantic annotation of resources;
– setup of the resource processing and recommendation.

With the ontology setup activity, the semantic layer composition has to be chosen. An ontology is a data model that represents a domain and it is used to reason about the objects in that domain and the relations between them. Usually, information can be modeled within a single domain, but with this approach we take into account also the opportunity to make use of more than one domain ontology.

In the second step, semantics is associated to resources within an information system. This is a delicate issue within the approach because it can be achieved in many ways and each one may differently affect future choices for system implementation. Semantic annotation aims at providing a richer and more formal service description and creating an agreement on a vocabulary (a set of terms)and on a semantic structure for information exchange about a specific domain. As emerged from the literature and from other related works [18], [31], semantic annotation can be done in a semi-automatic or manual way. It can be also considered embedded or not embedded in the technological description of the available resources. Our approach takes in consideration the manual way to add a semantic layer to the resource descriptions, while we use both the embedded and the non-embedded fashion in the applications of the approach. The last step in this approach is the setup of the resource processing and recommendation, which consists in the tuning of the semantic matching algorithm, whose model is described in the following section.

3.1 Semantic matching algorithm model

The semantic matching algorithm represents the way the recommender engine calculates a similarity between a resource's semantic description and the description of a target resource. It takes into account the semantics associated to them. In our discussion, we will present the algorithm based on the evaluation of the *DegreeOfMatch()* function, that is an example of feature-based similarity approach developed by Paolucci et al. [28]. This approach is extended in order to evaluate the semantic similarity on *n* different ontologies. More in particular, we assume that for each resource belonging to an Information System, there is a Semantic Description File associated.

In this paragraph a semantic matching function will be defined. This function will calculate the degree of semantic similarity among each resource description and the target resource, allowing a ranking of available resources.

Given the Ontology Ω and, given RD that is the generic Resource Description,subset of Ωand given TD that is the generic Target Resource Description, subsetof Ω and hold by the Information System, we can define the function SemSimilarity():

$$SemSimilarity_{\Omega}(RD \ \in \ Y', TD) \rightarrow r \ \in [0,1]$$

Thus, SemSimilarity() is a function that returns a real number value representing the similarity degree between its arguments by measuring the semantic similarity on all the semantics contained in the semantic descriptions. We should observe that the Resource Description RD, can be defined as: $RD = \{RD_1, RD_2, RD_3, \dots, RD_n\}$ where each $RD_i$is a concept of the ontology. In the same way the generic Target Resource Description TDcan be defined as: $TD = \{TD_1, TD_2, TD_3, \dots, TD_n\}$where each TD_i is a concept of the Ω ontology. In order to define the SemSimilarity() function it is necessary to identify a function whose role is to measure the degree of match among two different concepts of the same ontology. This function will be applied in order to measure the relation between the concepts in the target resources and in the available ones.

$$DegreeOfMatch_{\Omega}(RD_i \in RD,\ TD_i \in RD) \rightarrow r\ \in [0,1]$$

In particular, the *DegreeOfMatch* is given by the minimum distance between the concepts in the ontology view. It is possible to distinguish four different kinds of matches [28]:

1. an *exact match* can occur in two cases: in the simplest situation when two concepts coincide, and also when the concept specified in the target resource description is a direct specialization (first level specialization) of the concept specified in the generic resource description and contained in the ontology;
2. *plugIn* occurs when the concept specified in the generic resource description is a direct specialization of the concept specified in the target resource. This kind of relation is weaker than the previous;
3. *subsumes* occurs when the concept specified in the target resource is a specialization of the concept specified in the generic resource description;
4. *fail* occurs when no transitive relation exists between two specific concepts.

The four degree states that the *DegreeOfMatch* can assume will be associatedto discrete values, considering that the preferable degree is the ”exact” and theless preferable is the “subsumes”. For example, it is possible to assign discretevalues to the four states as follows:

DegreeOfMatch	*Value assigned*
exact	1
lugIn	0.7
subsumes	0.35
fail	0

In order to define *SemSimilarity()* it might be necessary to introduce somerules in order to be able to choose the best match and to reduce the computational complexity of the algorithm. If one of the additional attributes thatis contained in Target Resource Description matches with more than one ofthe Resource Descriptions attributes, we must choose the best combination. Ingeneral, we can say that given a user profile with *n* attributes and a ResourceDescription with *m* attributes, there will be $n \times m$ pairings: we must choosethe n distinct pairings with the highest value. With this assumption, given$TD = \{TD_1, TD_2, TD_3, \dots, TD_n\}$and $RD = \{RD_1, RD_2, RD_3, \dots, RD_m\}$:

$$NSemSimilarity_{\Omega}(RD, TD) = \Sigma_{i=1}^{n}[max_{j=1}^{m}(DegreeOfMatch_{\Omega}(RD_j, TD_j))]$$

The value returned by the *NSemSimilarity* is normalized with the number ofpreferences contained in the user profile:

$$SemSimilarity_{\Omega}(RD, TD) = \frac{NSemSimilarity_{\Omega}(RD, TD)}{n}$$

We can generalize this result to a complex condition. We could have to dealwith an information system which needs different reference ontologies in order tobe effective. In this specific condition it is quite reasonable that both the ResourceDescription and the Target Resource Description will contain a concept from thedifferent Ontologies.
Let Ω be the set of Ontologies used in the system and Ω_α the generic ontologyof Ω such that $\Omega = \{\Omega_1, \dots, \Omega_n\}$. A specific Resource Description is defined as:

$$RD = \{RD_1, RD_2, RD_3, \dots, RD_m\}$$

where each $RD_i$is a concept of a given ontology Ω_j:

$$RD = \left\{RD_1, \dots, RD_g \in \Omega_1, RD_{g+1}, \dots, RD_l \in \Omega_{2,}, \dots \dots, RD_{q+1}, \dots, RD_m \in \Omega_\eta\right\}$$

In the same way a specific Target Resource Description is defined as:

$$TD = \{TD_1, TD_2, TD_3, \dots, TD_t\}$$

where each TD_i is a concept of a given ontology Ω_j:

$$TD = \{TD_1, \dots, TD_p \in \Omega_1, TD_{p+1}, \dots, TD_r \in \Omega_{2,}, \dots \dots, TD_{s+1}, \dots, TD_t \in \Omega_\eta\}$$

we can define:

$$NSemSimilarity_{\Omega_\alpha}(RD, TD) = \Sigma_{i=1}^{g}[max_{j=1}^{p}(DegreeOfMatch_{\Omega_\alpha}(RD_j, TD_i))]$$

Taking in to account that, according to the definition of *DegreeOfMatch*, the *DegreeOfMatch* among concepts of different ontologies is always zero, wecan conclude that:

$$NSemSimilarity_{\Omega}(RD, TD) = \Sigma_{\rho=1}^{\eta} NSemSimilarity_{\Omega_\rho}(RD, TD)$$
$$= \Sigma_{i=1}^{g}\left[max_{j=1}^{p}\left(DegreeOfMatch_{\Omega_1}(RD_j, TD_i)\right)\right] + \cdots$$
$$\dots + \Sigma_{i=q+1}^{m}\left[max_{j=s+1}^{t}\left(DegreeOfMatch_{\Omega_\eta}(RD_j, TD_i)\right)\right]$$

This result is important since it allows us to improve the effectiveness of theselection adding a specific weight ω_ρ to each ontology, according to its perceivedvalue for the customer.

$$NSemSimilarity_{\Omega}(RD, TD) = \Sigma_{\rho=1}^{\eta}\omega_\rho NSemSimilarity_{\Omega_\rho}(RD, TD)$$
$$\text{with } \Sigma_{\rho=1}^{\eta}\omega_\rho = 1$$

Finally, the value returned by the *NSemSimilarity* is normalized with the number of preferences contained in the target resource.

$$SemSimilarity_{\Omega}(RD, TD) = \frac{NSemSimilarity_{\Omega}(RD, TD)}{n}$$

4. An accuracy metric for the semantic recommender system

In order to evaluate the quality of our system we introduce an evaluation metricwith the aim of measuring the recommendation's *accuracy*. An accuracy metricempirically measures how close a recommender system's predicted ranking ofitems for a user differs from the user's true preference ranking [15].

Basic assumption for our accuracy metrics measurements is that if the usercould examine all the items available, he or she would place them in an ordering ofpreference. Moreover, since our recommender system returns a ranked list wherehighest ranked items are predicted to be the most preferred, we assume thatthe user will generally view the recommended items starting at the top of thelist and descending the list with a gradually decreasing interest until the neededinformation is met or a certain number of results are examined (or a certain timelimit is reached). This premise has been already used by[6], who introduced a newevaluation metric for recommender systems designed for tasks where the user ispresented with a ranked result, and is unlikely to browse very deep into the list.Therefore, our focus is on the measurement of the ability of our recommendationalgorithm to produce a recommended ordering of items that matches how theuser would have ordered the same items.

4.1 The Power Accuracy Measure (PAM)

Several accuracy metrics exist for evaluating algorithms to present ranked recommendation lists to the user. Predictive accuracy metrics, which measure howclose the recommender's predicted ratings are to the user's ratings, (as for instance the *Mean Absolute Error* and related metrics) are less appropriate fortasks where a ranked result is returned to the user who only views the top of thelist and may care about errors in items that are ranked high. For these tasks,rank accuracy metrics are more appropriate. Among these metrics, the mostknown and applied are Pearson' product-moment, Spearman's ρ and Kendall's Tau correlation measures. They are described in [15]. Nevertheless, such metricssuffer from the weakness that interchanges between two recommendations at thetop of the ranking have the same weight as interchanges at the bottom of theranking. In order to overcome this problem, Breese et al. [6] presented a new rankaccuracy metric, called "half-life utility metric", which attempts to evaluate theutility of a ranked list to the user by describing the likelihood that a user will vieweach successive item with an exponential decay function. The strength of thedecay is described by a parameter called “half-life” which is the rank of the itemon the list such that there is a 50-50 chance that the user will review that item.The utility metric applies most of the weight to early items, with every successive rank having exponentially less weight in the measure. Another descriptionof this metric can be found in [14]. Nevertheless, the half-life utility metric isused in collaborative recommender systems where it is possible to determine adefault rating for all items.

Therefore, we propose a new ad hoc accuracy metric based on the distancebetween the user’s ranked list and the predicted recommendation list, whichtakes into account the item’s relative position in the list. In other terms, thedistances between the items that are top ranked have more substantial negativeimpact on the outcome of the metric. So that, we introduce a weighted distancefunction in which the weights are calculated trying to approximate as much aspossible the actual user’s behavior and degree of interest when browsing the results. In our accuracy metric we adopt the Stevens’ power law [32], which modelsthe relationship between the magnitude of a physical stimulus and its perceivedintensity. In our case, this law models the relationship between the degree ofimportance of items predicted with a recommendation for a particular user andhis or her perceived importance. In other terms, even if the user’s behavior when browsing the recommendation list is strongly influenced by the degree ofsatisfaction gained by the items reviewed, we can suppose that, when a user ispresented with a ranked list, he or she would have a predetermined attitude toreview items with a gradually decreasing interest. It is evident that the importance the user gives to the top item in the list is more than the importance givento the bottom item of the list.

The general form of the Stevens’ power law is:

$$S = kI^{a} \tag{1}$$

where S is the intensity of sensation, k is a constant, I is the magnitude of the physical stimulus, and a is an exponent. The value of a is dependent on the type of stimulation. We use such law to determine the weights in our accuracy metric.

Regarding the distance between the predicted ranked list and the user’sranked list, we use the deviation distance measure. The deviation of the item *k*between the recommendation list *P*and the user’s list *T*is the absolute difference between the position *i* of the item in the list *P* and its position *j* in *T*. Thedeviation distance is the sum of the deviations of all *n* items:

$$d_{dev}(P,T) = \sum_{j=1}^{n} |i - j| \quad (2)$$

where *T(j)* = *P(i)* and $i \in \{1..n\}$ is the position of the item *T(j)* in the list *P*.
Therefore, the normalized weighted accuracy metric that we propose, named*Power Accuracy Measure*, is given by:

$$PAM = \frac{1}{n-1} \sum_{j=1}^{n} \omega_j |i - j| \quad (3)$$

where *T(j)* = *P(i)* and $\frac{1}{n-1}$ is the normalization factor.
Weight ω_j is given by

$$\omega_j = k(n + 1 - j)^a \quad (4)$$

where $k = \frac{1}{\sum_{j=1}^{n} k(n+1-j)^a}$ is a constant that allows to have$0 \le \omega_j \le 1$ and $\sum_{j=1}^{n} \omega_j = 1$

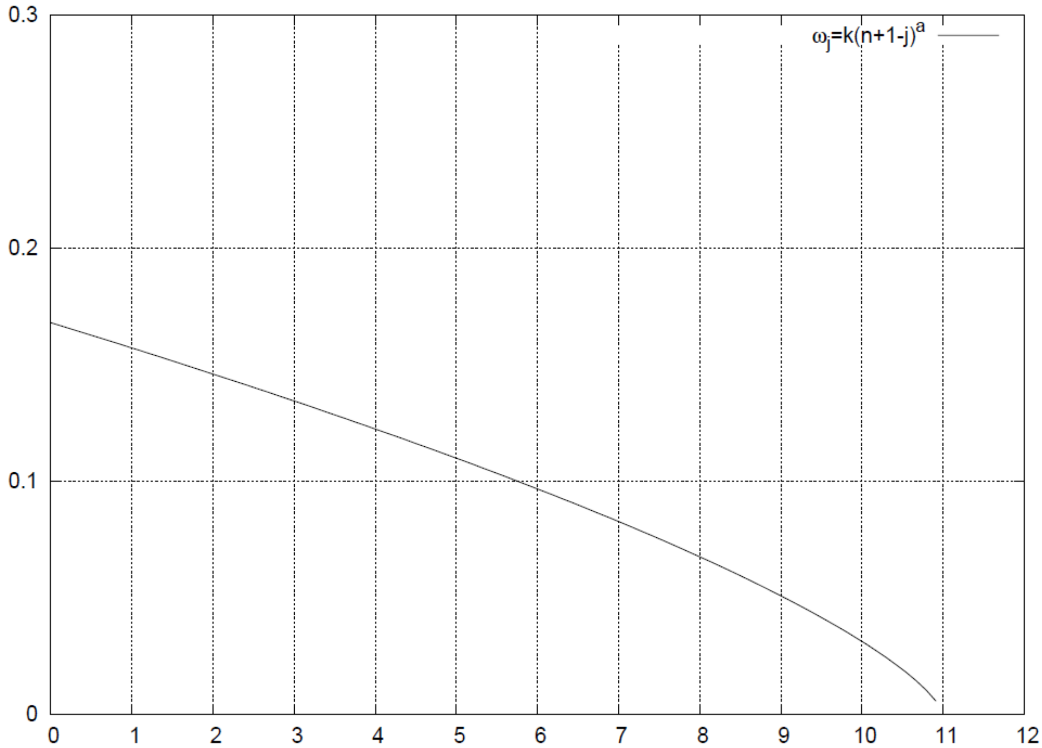

Fig. 1: Representation of $\omega_j (n = 10$ and $a = 0.7)$

Fig. 1 shows the decay of the weight function for a = 0, 7 when n = 10. The value of the exponent a has to be empirically determined, starting from different sets of weights collected directly from the users involved in the experiments.

We calculate the PAM for each user involved in the experiments. Then, in order to study how the PAM varies, we calculate the mean ($\overline{PAM}$) and the standard deviation (σ_{PAM})which show the recommender system's average behavior and the range in which the PAM values lie.

Moreover, in order to verify the behavior of our recommender system when the number of the items changes we calculate the mean and the standard deviation for different values of n.

5. Applying the approach in the context of Cultural Heritage in Tourism

In this section we present the Recommendation Engine we designed and developed for the Teschet platform.The Teschet platform is characterized by the Semantic Web paradigm implementation, and it creates a distributed thematic network to publish services, semantically described and accessible via web, through the aggregationof several actors (SMEs, government bodies, etc.). Thus, the Teschet technological platform is a tool for distributed services integration.

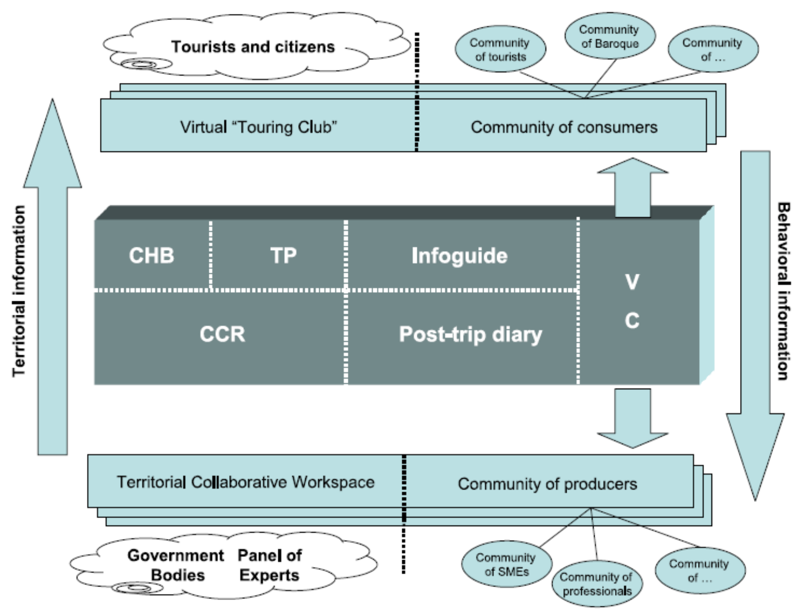

Fig. 2: Teschet platform

Services in the Teschet platform represent just some of all the possible implementable services. They actually comprehend:

– tourist-cultural heritage fruition and valorization services, such as:
  • Cultural Holiday Builder (CHB)
  • Travel Planner
  • Infoguide
  • Post-Trip diary
  • Virtual Community for tourists
– tourist-territorial analysis and marketing services, such as:
  • Territorial analysis
  • Virtual Community for government bodies

All these services are accessible via web, and they provide relevant informationfor a tourist when organizing, consuming and remembering the touristic-culturalexperience. Specifically, we will focus on the Cultural Holiday Builder that realizes two main functionalities:

– choice of the travel destination with respect to the cultural thematic of specific interest;
– choice and reservation/booking of the services related to the cultural offeringwithin the desired destination.

With the CHB, the user can access a virtual space, such as virtual travellingbag, in which he can hold the information about sites and legs associated toeach tourist route created. The functionalities offered by the CHB allow to highlight the characteristics of a territory for a better holiday planning and abetter fruition of the heritage within the territory, increasing the tourist finalsatisfaction.

5.1 Teschet Recommender Engine

The Teschet Recommender Engine is a tool that supports the functionality offered by the Cultural Holiday Builder (CHB), by providing a user, who is planning to travel, with a wider set of touristic/cultural resources according to twodimensions: a ”typological affinity” and a

”geographic proximity”. Thus, theTeschet Recommender Engine recognizes and proposes, among all the touristic/cultural resources in a region, those that are the most similar to a specificselected resource (target resource). A recommendation is generated through:

– a selection of the resources that are in the same region as the target resource;
– a subsequent resource rank estimated with:
 • a semantic matching algorithm on the concepts that describe the target resource typology, by using an ontological representation of the Teschet domain;
 • the resource localization within a county or a city;

The Teschet Recommender Engine utilizes a technique for the similarity evaluation between two different resources, based on the algorithm presented in Section 3.1. More specifically, when a user is browsing the available resources of aterritory within the CHB, as showed in the Fig. 3, and asks for similar resources,the Recommender Engine extracts typology and localization information fromthe target resource. Then it starts to filter the resources available in the TeschetPlatform according to the localization information previously obtained, and theresult set of the resource gets ranked according to the semantic matching onthe typology information. As previously said, resources belonging to the Teschetplatform are the tourist and cultural heritage sites from different part of Italy,whose characteristics are contained in a description file semantically enrichedand accessible by the recommender engine. The effectiveness of the semantic layer is granted by the Teschet domain ontology of the cultural heritage and tourist resources, made by about 100 concepts.

Fig. 3: The Cultural Holiday Builder front-end

Each resource is classified as an instance of a concept belonging to this ontology and enriched by other descriptions related to the geographic position within a territory (county, region, nation). Here is an example of a semantic resource description file made with DAML[4] which is being designed as an XML-based semantic language that ties the information on a page to the machine-readable semantics (ontology):

```
<rdf:RDF >
<teschetOntology:VillaMonumentale rdf:about="http://.../Teschet" >
        <teschetOntology:descrizione>...
        </teschetOntology:descrizione>
        <teschetOntology:sito turistico rdf:resource="http://.../Teschet"/>
        <teschetOntology:sito in>
                <teschetOntology:Comune rdf:about="http://.../Teschet">
                        <teschetOntology:sito in provincia>
                                <teschetOntology:Provincia rdf:about="http://.../Teschet" >
                                        <teschetOntology:sito in regione>
                                                <teschetOntology:Regione rdf:about="http://.../Teschet">
                                                        <teschetOntology:sito in nazione>
                                                                <rdf:Description rdf:about="http://.../Teschet">
                                                                        <rdfs:label xml:lang="it">
                                                                                THE NATION
                                                                        </rdfs:label>
                                                                </rdf:Description>
                                                        </teschetOntology:sito in nazione>
                                                        <rdfs:label xml:lang="it">THE REGION </rdfs:label>
                                                </teschetOntology:Regione>
                                        </teschetOntology:sito in regione>
                                        <rdfs:label xml:lang="it">THE COUNTY </rdfs:label>
                                </teschetOntology:Provincia>
                        </teschetOntology:sito in provincia>
                        <rdfs:label xml:lang="it">THE CITY </rdfs:label>
                </teschetOntology:Comune>
        </teschetOntology:sito in>
        <rdfs:label xml:lang="it">Villa Sticchi </rdfs:label>
</teschetOntology:VillaMonumentale>
</rdf:RDF>
```

[4] http://www.daml.org

Each *sito_in_* attribute expresses the geographic localization of a resource. The *sito_in_regione* is processed to filter the set of the available resource. The semantic matching is computed by comparing the target resource typology with the available resource typologies. This matching produces a rank as described in section 3.1. The ranking is finally updated if the available resource belongs to the same city and/or county as the target resource as showed in 4. In this way it is possible to obtain a list of the most similar resources that will be displayed to the user.

Fig. 4: The rank of the resources produced by the Recommender Engine

5.2 Experimental results

In order to evaluate the quality of our system, some offline experiments[5] havebeen conducted within the Teschet projects with the purpose of measuring therecommendation's *accuracy*. Therefore, the focus of our experiments is to measure the ability of how our recommendation algorithm will produce a recommended ordering of items that matches the order of the same items that theuser would have produced. The experiment took place over one month periodwhere 10 computer science researchers participated. The collected dataset consists of 10 users who ranked a total of 30 items. Both the ontology and the servicedescriptions (of about 10 concepts each) were developed in the previous phaseof the project. User profiles have been generated with a semi-automatic processwhich integrated data about a touristic/cultural resources type and a geographicposition. In particular for our experiments,

[5] Evaluations can be completed using offline analysis, live user experiments, or a combination of the two. In offline experiments, the recommendation algorithm is used to predict certain values from a dataset, and then the results are analyzed using some metric.

the user profiles were built by mapping the resource types with the concepts in the given domain ontology andcompleting this info with the geographic position given by the Teschet platform.For our experiments, we used an exponent a = 0, 7.[6] Fig. 1 shows the decayof the weight function for a = 0, 7 when n = 10. In this section we discuss theresults of the conducted experiments.

Table 1: Comparing PAM with different number of items.

User	PAM (n=10)	PAM (n=20)	PAM (n=30)
user1	0,21549812	0,197992865	0,21939058
user2	0,204115938	0,126544523	0,106790511
user3	0,009149882	0,164890019	0,204560224
user4	0,114580733	0,132352153	0,166332684
user5	0,238471112	0,12819527	0,10298586
user6	0,159133378	0,133454888	0,147318458
user7	0,186973182	0,195032762	0,208014077
user8	0,249546441	0,132421164	0,141531927
user9	0,241791043	0,172805543	0,137779375
user10	0,19584843	0,166919798	0,155387774
$\overline{PAM}$	0,181510826	0,155060898	0,159009147
σ_{PAM}	0,073103608	0,027934509	0,040753014

Table 1 shows the different values of PAM calculated for all users and withdifferent values of n. It shows also the mean value $\overline{PAM}$and the standarddeviation σ_{PAM}.

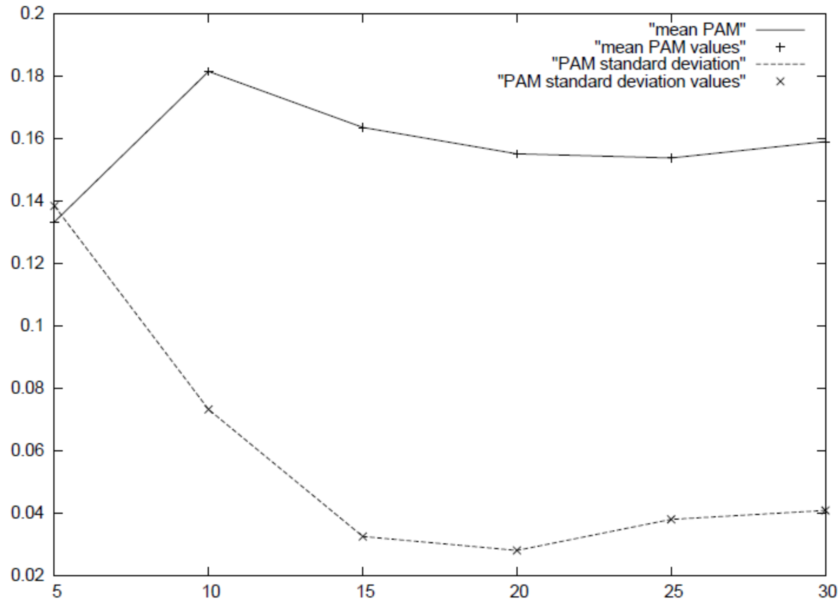

Fig. 5: Values of $\overline{PAM}$and $\sigma_{PAM}$for various sizes of n.

[6] The value of exponent a has been determined empirically, starting from different sets of weights collected directly from the users involved in the experiments.

A careful observation of Tab. 1 leads us to the conclusion that the recommendation system has a good level of accuracy, in fact the highest value of PAMobtained is about 0,249. Looking at the first column of Tab. 1 (i.e. when the valueof an item is equal to 10), we can see that the $\overline{PAM}$is about 0,181 which meansthat the number of the relevant items ranked erroneously is very low. Moreover,the σ_{PAM} is about 0,073, which means that the recommender system's accuracyvaries very slightly around the mean value. Looking at the other columns of Tab.1 we notice that the variations in the size of *n* don't affect the quality of therecommendation. In fact, the PAM mean value does not change substantially when*n* increases. This can be clearly seen in Fig. 5 which shows the $\overline{PAM}$and theσ_{PAM} for various sizes of *n*:when n increases, after an initial transition, the recommendation's behavior tends to be stable without losing accuracy, with the σ_{PAM} that holds low values.

Fig. 6 shows the Gaussian distribution of PAM values with n = 10, n = 20and n = 30. It provides an estimation of recommender engine's precision aroundthe mean PAM.

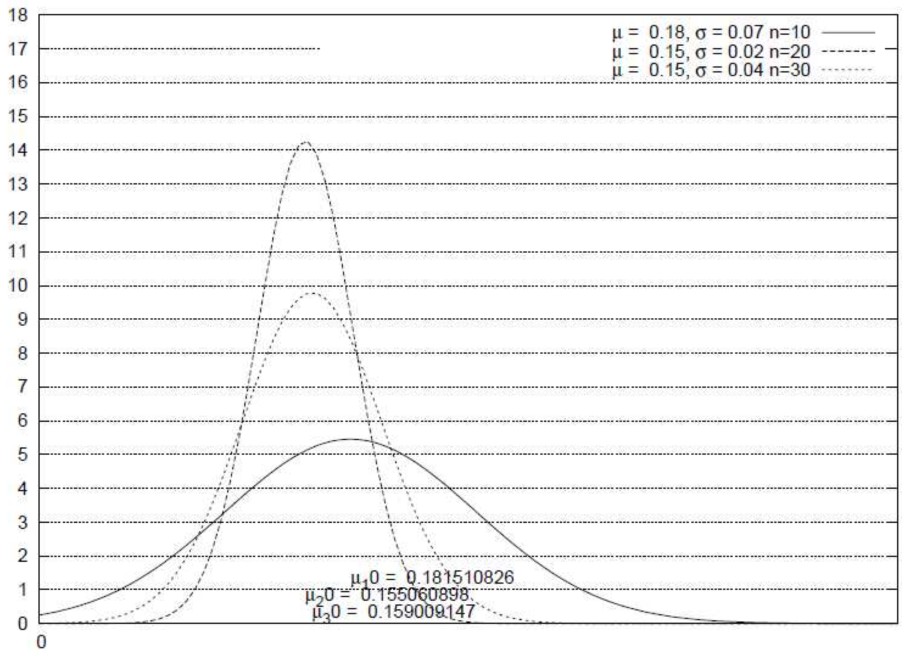

Fig. 6: Distribution of PAM values with *n* = 10, *n* = 20 and *n* = 30.

Concluding, the experiments run reveal an interesting behavior of our system, presenting a good accuracy which also seems to be independent from the size of items available for the recommendation.

5.3 Semantic approach's advantages and weakness

Recommender Systems best practices presented in the literature states that statistical approach to the recommendation generation is the most used approachthrough several contexts and application domains. This is particularly true whenthe number of items or user affinities, that has to be considered, increases overtime. So it's clear that, for applications with huge load of data available andfor which it's crucial to provide a quick and precise (in terms of exactness)recommendation, the statistical approach to recommendation (hybrid techniques more than others) is the successful one. As a matter of facts, the semanticapproach to recommendation generation can be successful in more limited application contexts, providing advantages that can be split in two specific categories:

1. technological advantages descending from the use of ontologies that avoidsspecific problems, including[29]:
   – to guarantee the inter-operability of system resources and the homogeneity of the representation of information
   – to allow the dynamic contextualization of user preferences in specificdomains.
   – to improve communication processes between agents and between agentsand users.
   – the ability to semantically extend descriptions of user contextual factors.
   – to improve the representation and description of different system elements.
   – to improve the description of system's logic by admitting the inclusionof a set of rules.
   – to provide the necessary means to generate descriptions enriched by webservices and facilitate their discovery by software agents.
2. Context application advantages descending from specific application domains. In the domain of Cultural Heritage and tourism services, the domain modeling is carried out in a great detail, as it is delegated to domain experts. It is worth to mention that there are many defined international standards which aim to describe Cultural Heritage domain such as the CIDOC[7]Conceptual Reference Model (CRM) that provides definitions and a formalstructure for describing the implicit and explicit concepts and the relationships used in the cultural heritage documentation. In Italy, the referencestandard for the Cultural Heritage cataloguing is given by ICCD

[7]http://cidoc.mediahost.org

(IstitutoCentrale per il Catalogo e la Documentazione)[8] who is in charge for the definition of cataloguing standards for archaeological, architectural, environmental, historical and artistic heritage. Another interesting work is an Italian project called ”Ontologie archivistiche”[9], based on the Semantic Web technologies thatare creating, through a shared conceptual base, a collaborative system forthe analysis and the ontological description of the national archivist system.International and national standards are the real drivers for the domain modeling developing initiative which can affect the accuracy of the semantic recommender engine output that makes this approach preferable over thestatistical one, as it also results from the experimental data presented in thispaper.

6. Conclusions and the future research agenda

In this paper, we tried to answer to the question of how it is possible to use the semantic web technologies to support an information system in order to enable personalization in the information delivery process. We presented our approach using semantic recommendation generation, based on the semantic annotation of resources and a semantic matching algorithm to evaluate similarity and generate recommendations. We also showed how this approach was applied to the Cultural Heritage domain. The ongoing test and the experimental results made on the Recommender Engine, demonstrated also that the system is quite good accurate, with a low number of relevant items ranked erroneously. These results were supported by a metric proposed within this work, the Power Accuracy Measure (PAM), that is based on the distance between the user’s ranked list and the predicted recommendation list, taking into account the item’s relative position.

Following our research agenda we are ready to test the proposed approach in X-Net.Lab[10] project. X-Net.Lab will contribute to the modernization of thesouthern touristic system through the development of architectures, organizational and technological systems integrating cultural heritage and Agrifood assets in the dynamics of growth and development of the touristic sector. Theidea of the project borrows the biological metaphor of Ecosystem[26] to create environments, software architectures and tools enabling the

[8]http://www.iccd.beniculturali.it/index.php?it/115/standard-catalografici
[9]http://www.archivi.beniculturali.it/servizioII/progetti/ontologie.html
[10]The X-Net.Lab project was presented in response to the call for public-private laboratories published by MIUR - Ministry of Education, University and Research

transformation to the Digital Business Ecosystem. One of the technical characteristics of the platform designed for the X-Net.Lab project is to use a CBD[16] approach basedon richly described components named XBC (eXtended Business Component)which implement services provided by the SMEs (Small Medium Enterprise).The system should be able to suggest possible partnerships to the firms withother SMEs of the Digital Business Ecosystem, in order to increase the revenuefor all partners. The selection of these collaborations is a possible applicationof the recommendation algorithm appropriately tailored on the functional andbusiness description of the components as well as on the services offered. Inparticular the reference standards are:

– ICCD for the cultural heritage resources classification;
– GS1[11] for the classification of goods and services of the agrifood firms;
– UNSPSC (United Standard Products and Services Code)[12] for the classification of goods and services traded in the tourism sector.

The use of standards adopted by the project with the components descriptiondeveloped and deployed on the X-Net.Lab platform provides the informationthat the recommender engine uses for the generation of the possible partnershiplist. The application of the approach in this context is extremely challenging, because it aims at generating recommendation even in the cases in whichthere is no formal ontology, but only different domain models as backbone for the engine. We are still evaluating the feasibility of our semantic approach insuch new context. Any success with good accuracy values measured with PAM,would prove that the generalized approach can be extended to any applicationenvironment with a formal domain modeling.

[11]GS1 is the main international classification system used to identify the goods or services provided by companies. The GS1 System of standards is the most widely-used supply-chain standards system in the world - www.gs1.org
[12]UNSPSC is a taxonomy of products and services for use in eCommerce. UNPPSC ismanaged by GS1

REFERENCES

[1] B. Arslan and F. Ricci, editors. *Workshop on Recommendation and Personalization in eCommerce: Case-Based Session Modeling and Personalization in a Travel Advisory System (RPEC'02)*, Malaga, Spain, May 28th 2002.

[2] M. Balabanovi and Y. Shoham. Fab, content-based, collaborative recommendation.*Communications of the ACM*, 40(3):66–72, 1997.

[3] C. Basu, H. Hirsh, and W. Cohen. Recommendation as classification: usingsocial and content-based information in recommendation. In *AAAI '98/IAAI'98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence*, pages 714–720, Menlo Park,CA, USA, 1998. American Association for Artificial Intelligence.

[4] Tim Berners-Lee and Mark Fischetti. *Weaving the Web; The Original Design andUltimate Destiny of the World Wide Web by Its Inventor (2 Cassettes)*. HarperAudio, 1999.

[5] Tim Berners-Lee, James Hendler, and Ora Lassila. The Semantic Web. *ScientificAmerican*, May 2001.

[6] John S. Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictivealgorithms for collaborative filtering. pages 43–52.

[7] R. Burke. The wasabi personal shopper: a case-based recommender system. In*AAAI '99/IAAI '99: Proceedings of the sixteenth national conference on Artificial intelligence and the eleventh Innovative applications of artificial intelligenceconference innovative applications of artificial intelligence*, pages 844–849, MenloPark, CA, USA, 1999. American Association for Artificial Intelligence.

[8] R. Burke. Knowledge-based recommender systems. *Encyclopedia of Library andInformation Systems*, 69, Supplement 32, 2000.

[9] R. Burke. Hybrid recommender systems: Survey and experiments. *User Modelingand User-Adapted Interaction*, 12(4):331–370, 2002.

[10] R. Coster, A. Gustavsson, T. Olsson, and A. Rudstrom, editors. *Workshop on Recommendation and Personalization in e-Commerce: Enhancing Web-Based Configuration with Recommendations and Cluster-Based Help (RPEC'02)*, Malaga, Spain,May 28th 2002.

[11] Sowa J. F. *Knowledge Representation: Logical, philosophical, and computationalfoundations*. Pacific Grove, 2000.

[12] Peter W. Foltz and Susan T. Dumais. Personalized information delivery: an analysis of information filtering methods. *Commun. ACM*, 35(12):51–60, 1992.

[13] X. Fu, J. Budzik, and K. J. Hammond, editors. *Proceedings of the 2000 International Confonference on Intelligent User Interfaces: Mining Navigation History forRecommendation (IUI'00)*, New Orleans, Louisiana, 2000.
[14] David Heckerman, David Maxwell Chickering, Christopher Meek, Robert Rounthwaite, and Carl Myers Kadie. Dependency networks for inference, collaborativefiltering, and data visualization. *Journal of Machine Learning Research*, 1:49–75,2000.

[15] Jonathan L. Herlocker, Joseph A. Konstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*,22(1):5–53, 2004.

[16] P. Herzum and O. Sims, editors. *Business Component Factory*. John Wiley &Sons, 1999.

[17] T. Hofmann. Latent semantic models for collaborative filtering. *ACM Trans. Inf.Syst.*, 22(1):89–115, 2004.

[18] A. Kiryakov, B. Popov, I. Terziev, D. Manov, and D. Ognyanoff. Semantic annotation, indexing, and retrieval. *Journal of Web Semantics*, 2, 2005.

[19] J. Kolodner. *Case-based reasoning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1993.

[20] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying Collaborative Filtering to Usenet News. *Communications ofthe ACM*, 40(3):77–87, 1997.

[21] B. Krulwich. Lifestyle finder: Intelligent user profiling using large-scale demographic data. *AI Magazine*, 18(2):37–45, 1997.

[22] S. Lawrence and L. Giles. Searching the World Wide Web. *Science*, 280:91–94,1998.

[23]Buranarach M. *The Foundation for Semantic Interoperability on the World WideWeb*. sadsad, 2001.

[24] M. Missikoff, F. Schiappelli, and F. Taglino. A controlled language for semanticannotation and interoperability in e-business applications.

[25] R. J. Mooney and L. Roy. Content-based book recommending using learning fortext categorization. In *DL '00: Proceedings of the fifth ACM conference on Digitallibraries*, pages 195–204, New York, NY, USA, 2000. ACM Press.

[26] F. Nachira. Towards a network of digital business ecosystems fostering the local development. Technical report, European Commission DG INFSO, 2002.http://www.digital-ecosystems.org/doc/discussionpaper.pdf.

[27] D. Nichols. Implicit rating and filtering. In *Proceedings of 5th DELOS Workshopon Filtering and Collaborative Filtering*, pages 31–36. ERCIM, 1998.

[28] M. Paolucci, T. Kawamura, T. R. Payne, and K. P. Sycara. Semantic matchingof web services capabilities. In *ISWC '02: Proceedings of First International Semantic Web Conference on The Semantic Web*, pages 333–347, London, UK, 2002.Springer-Verlag.
[29] E. Peis, J. M. Morales del Castillo, and J. A. Delgado-Lopez. Analysis of the stateof the topic. *Hipertext.net*, (6), 2008.

[30] J. B. Schafer, J. Konstan, and J. Riedi. Recommender systems in e-commerce.In *EC '99: Proceedings of the 1st ACM conference on Electronic commerce*, pages158–166, New York, NY, USA, 1999. ACM Press.

[31] K. Sivashanmugam, K. Verma, A. Sheth, and J. Miller. Adding semantics to webservices standards. In *ICWS'03: Proceedings of the 1st International Conferenceon Web Services, Las Vegas, Nevada*, pages 395–401, London, UK, 2003. Springer-Verlag.

[32] Stanley S. Stevens. On the psychophysical law. *Psychol Rev*, 64(3):153–181, 1957.

[33] C. Zeng and L. Zhou C. Xing. Similarity measure and instance selection for collaborative filtering. In *WWW '03: Proceedings of 12th international conference onWorld Wide Web*, pages 652–658, New York, NY, USA, 2003. ACM Press.