

A Generic Semantic-based Framework for Cross-domain Recommendation

Ignacio Fernández-Tobías
Escuela Politécnica Superior
Universidad Autónoma de Madrid
28049 Madrid, Spain

ign.fernandez01@estudiante.uam.es

Iván Cantador
Escuela Politécnica Superior
Universidad Autónoma de Madrid
28049 Madrid, Spain

ivan.cantador@uam.es

Marius Kaminskas
Faculty of Computer Science
Free University of Bozen-Bolzano
39100 Bolzano, Italy

mkaminskas@unibz.it

Francesco Ricci
Faculty of Computer Science
Free University of Bozen-Bolzano
39100 Bolzano, Italy

fricci@unibz.it

ABSTRACT

In this paper, we present an ongoing research work on the design and development of a generic knowledge-based description framework built upon semantic networks. It aims at integrating and exploiting knowledge on several domains to provide cross-domain item recommendations. More specifically, we propose an approach that automatically extracts information about two different domains, such as architecture and music, which are available in Linked Data repositories. This enables to link concepts in the two domains by means of a weighted directed acyclic graph, and to perform weight spreading on such graph to identify items in the target domain (music artists) that are related to items of the source domain (places of interest).

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *information filtering, retrieval models*. I.2.4 [Artificial Intelligence] Knowledge Representation Formalisms and Methods – *semantic networks*

General Terms

Algorithms, Design, Human Factors.

Keywords

Recommender systems, cross-domain recommendation, knowledge extraction, semantic networks, Linked Data, DBpedia.

1. INTRODUCTION

The vast and ever increasing amount, complexity and heterogeneity of digital information – news, images, videos, music tracks – overwhelm human processing capabilities in a wide array of information seeking and e-commerce tasks. In order to tame such information overload, recommender systems can help users to make choices, by proactively finding relevant items or services, taking into account or predicting the users' tastes, priorities and goals [1].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
HetRec '11, October 27, 2011, Chicago, IL, USA

Copyright © 2011 ACM 978-1-4503-1027-7/11/10... \$10.00

Recommender systems are becoming an integral part of a large number of important e-commerce and leisure Web sites like Amazon¹, Netflix², Last.fm³, and in many online retailers, recommendation models have proved to be successful. Nonetheless, ample room and need for further improvements remain in the current generation of recommender systems to achieve a more effective human decision support, in a wide variety of applications and scenarios. Among these improvements, new research works have addressed the functionality of providing recommendations of items in one domain using the preferences expressed on items in a second domain or to build recommendations on a domain that are adapted to items in another domain, such as a book that suits a recommended travel [8][10][12][15].

In fact, the vast majority of the currently available recommender systems predict the user's interest for items in a specific and limited domain without considering information that can be extracted from other, possibly similar or related domains. For instance, Netflix suggests movies and TV series, and Last.fm makes personal recommendations of music artists and tracks using the feedback of the user only on the items in the target recommendation domain. However, in some e-commerce sites, such as Amazon, it could be useful to exploit user ratings for items in several domains simultaneously, and offer the user joint personalized recommendations of items belonging to multiple domains, e.g. by suggesting the user not only a particular movie, but also music CDs, books or videogames that are somehow related to that movie. Or in a travel application, it would be interesting to suggest a cultural event to a user that has booked a particular and recommended hotel.

To address such challenges, which are produced by cross-domain reasoning for recommendation building [6], a number of specific issues must be investigated [15]. Firstly, in the target application domain, we must verify the hypothesis that at community level, cross-domain user preferences do really exist, i.e., there are correlations between user preferences for items belonging to different domains. Next, at individual level, we should be able to develop recommendation models where user preferences for items in certain domains are used to predict user preferences for items in

¹ Amazon online shopping, <http://www.amazon.com>

² Netflix on-demand video streaming, <http://www.netflix.com>

³ Last.fm online radio, <http://www.last.fm>

other domains. Finally, we have to evaluate how effective cross-domain item recommendations are. One can conjecture that in this scenario the generated recommendations will be less accurate, but more diverse than those produced in single-domain systems.

In this paper, we focus on a particular scenario where we provide recommendations for music artists and tracks adapted to a place of interest (POI) in a city that the user may be visiting or browsing information on. Previous works [9][10] have proved that there exist latent similarities between items in the two domains – POIs and music. Therefore, a match between these two types of items can be established. In the quoted papers, the matching is computed using social tag-based profiles of the items and specific similarity measures for tagged resources. From that result, and aiming to finally complement the social recommendation models proposed in [10], we present here an ongoing research work on the design and development of a generic framework built upon semantic networks, which integrates and exploits knowledge on several domains to provide cross-domain adapted item recommendations. More specifically, we propose an approach that automatically extracts information about the above two domains (architecture attractions and music tracks) available in Linked Data⁴ repositories, links items in the two domains by means of a weighted directed acyclic graph, and performs weight spreading mechanisms on such graph to identify matching items in a target domain (music artists) from items of a source domain (places of interest identified by architectural attractions).

The rest of the paper is organized as follows. In Section 2, we briefly describe related works on cross-domain recommender systems and Linked Data. In Section 3, we present the particular cross-domain scenario in which we focus our attention: suggesting music artists and tracks for given places of interest. In Section 4, we describe our approach to extract and integrate knowledge about several domains online available in Linked Data repositories, and exploit it to provide cross-domain item recommendations. Finally, in Sections 5 and 6, we provide preliminary results, discussion and future work.

2. RELATED WORK

In this section, we summarize recent works on cross-domain item recommendation, and briefly describe the Linked Data initiative, from which our generic semantic-based framework for cross-domain recommendation is built.

2.1 Cross-domain Recommendation

Most current recommender systems make preference predictions for items belonging to individual domains. In [16], Yu observes that many recommender systems not only focus on a single area of interest – an application domain (e.g. travelling) –, but they also typically provide recommendations for only one item class or type (e.g. tourist attractions, touristic guides, books and websites, accommodation, and transport). Chung et al. [6] propose a classification for recommendation approaches distinguishing three levels of integration – single item type, cross-item type, and cross-domain systems –, and claim that integrated cross-domain recommender systems are particularly useful nowadays with the increasing diffusion of personalized, networked mobile devices.

In addition to mobile environments, the application of cross-domain item recommendation approaches has a special interest in many e-commerce and retailers websites, in which vendors offer a wide variety of items to increase company profits and strengthen the customers' loyalty. It is true that real online commercial

portals, such as Amazon, offer the user recommendations for items belonging to different domains, but exploiting only information about the user's preferences for items in the target recommendation domain. Moreover, in many cases these recommendations rely on the statistical analysis of popular items, neglecting the adoption of personalization strategies [15].

Despite its potential benefits, cross-domain recommendation has barely been investigated in the research literature. It is difficult to obtain datasets with user preferences crossing different domains, and the evaluation of that type of recommendation has to be carefully conducted. As shown by Winoto and Tang in [15], joint recommendations of items belonging to multiple domains may be less accurate, but more diverse than recommendations of items in a single domain.

Seminal work on cross-domain recommendation can be attributed to adaptive systems that make use of generic user models, and are able to mediate through different systems and application domains. In [11], Kobsa provides an extensive review of these systems, analyzing their purposes, services and design requirements, and categorizing them into shell systems – in which the user models form part of the applications –, central server systems – in which the user models are maintained by centralized components that communicate with the applications –, and agent-based systems – in which the user models are composed by heterogeneous decentralized components. Decentralized user modeling is indeed an emerging research topic in the field. Berkovsky et al. [4] present a generic framework to mediate the integration of data collected by several recommender systems, and discuss four major types of mediation: cross-user, cross-item, cross-context, and cross-representation. Some evaluations have shown that in certain conditions, user modeling data mediation improves the quality of recommendations, especially in the early stage of a recommender system life, i.e., in the cold start of the system. Szomszor et al. [13][14], on the other hand, propose an approach that aggregates personal social tagging information from different folksonomy-based user profiles, showing an increment on the coverage of user preferences.

In addition to cross-domain user modeling approaches, in the literature, there are recent research works that focus on the design of specific recommendation algorithms. In [6], Chung et al. present a cross-domain information filtering system that allows the user to set multiple criteria on attributes of items she is interested in. Berkovsky et al. [3] investigate a collaborative filtering approach in which user ratings (for movies) are partitioned according to domain-related attributes (e.g. film genres). A related approach is proposed by Li et al. [12], in which the bridge between domains (movies and books) is established by clustering rating matrices, and finding user-item patterns at the obtained cluster level. Finally, aiming to provide contextualized recommendations of music compositions when the user is visiting touristic attractions, Kaminskas and Ricci [9][10] investigate recommendation strategies that compute similarities between tag-based place of interest and music track representations.

2.2 Linked Data

Within the Semantic Web initiative, the Linked Data⁵ project aims at publishing structured datasets – usually described by standard metadata models such as RDF⁶ – on the Web, and setting (RDF)

⁴ Linked Data project, <http://www.linkeddata.org>

⁵ Linked Data, <http://linkeddata.org>

⁶ Resource Description Framework (RDF), <http://www.w3.org/TR/rdf-primer>

links between data items – usually called semantic entities – from different data sources. The adoption of Linked Data has thus led to the extension of the Web with a global data space connecting data from diverse domains such as people, companies, books, films, television, music, statistical and scientific data, and reviews [5]. This enables new types of applications. For instance, there are search engines that crawl Linked Data by following the links between data sources, and provide expressive query capabilities (see e.g. SPARQL⁷ RDF query language) over aggregated data, similar to how a relational database is queried today.

From the data sources available in the Linked Data cloud, DBpedia [2] can be considered as a core ontology, which is connected to many other data repositories. DBpedia is the Linked Data version of Wikipedia⁸, describing and linking more than 3.5 million concepts from a large variety of domains of human knowledge. Figure 1 shows a fragment of the Linked Data cloud that shows how DBpedia is linked with other data sources, some of them belonging to music and geography domains.

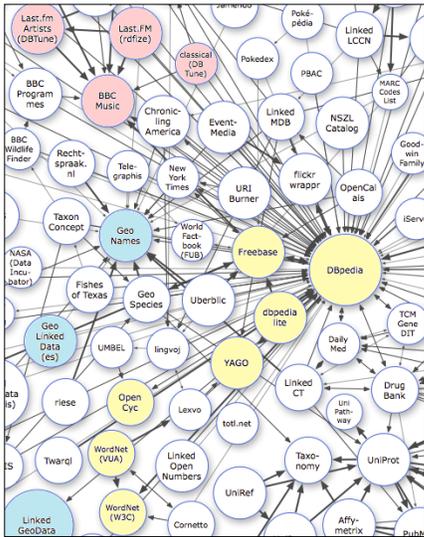


Figure 1. Fragment of the Linked Data cloud (as of September 2010). Colored circles represent examples of linked general knowledge, music, and geography ontologies.

3. CASE STUDY: LINKING PLACES OF INTEREST WITH MUSIC

Recommending music tracks that suit places of interest can be exploited in a number of engaging information services including music delivery ones. For instance, a mobile city guide may play music that suits the places visited by a tourist, thus providing a soundtrack to the sightseeing tour and enhancing the user’s experience. Such music recommendation can also be used in car entertainment and navigation systems, adapting music to places the car is passing by, or for enhancing the presentation of information in tourism websites. In these examples, it is clear that sometime the main focus of the recommendation process is the place of interest, and the music is an add on; in the other case, the recommendation is aimed at suggesting music tracks, and the place contributes to the customization of the music to the user’s needs and context.

⁷ Protocol and RDF Query Language (SPARQL), <http://www.w3.org/TR/rdf-sparql-query>

⁸ Wikipedia online encyclopedia, <http://www.wikipedia.org>

We must stress that in these applications “personalization”, intended as the adaptation of the recommendations to the user preferences, plays a secondary role. The first goal is to match music or artists to a POI such that a generic user will agree that these two items go well together: the music or artist have relationships with the POI that any user will tend to recognize as important and meaningful. In general, it is clearly challenging to match music to a place so that the user could appreciate such relation, and would prefer it to other music not explicitly matching the place. The main challenge that one must face when addressing this goal is related to the fact that music and POIs belong to different domains and there is no obvious way to match one with the other. In a previous work [9], as we mentioned earlier, we have used a repertoire of tags, which are describing emotional and physical characteristics, for annotating places and music (e.g. both a monument and a music track could be described as *strong*, *triumphant*, *heavy*). Such annotations allowed us to find matching pairs of POIs and music tracks. In this work, we explore a complementary research direction – finding semantic relations between places and music using Linked Data.

In order to motivate our approach, we present an example of a well-matched POI and music pair. Consider a given place of interest – the State Opera of Vienna, Austria. It is one of the most famous opera houses dating back to the 19th century and a prominent attraction for tourists visiting Vienna. A selection of well-fitting music for this place could consist of classical music pieces by composers who lived and worked in Vienna at a similar historical period, or were otherwise related to the venue itself. Among the best suiting selections would be operas by Austrian composers, which are frequently performed in this opera house, such as “Don Giovanni” by Mozart. Or, an even better match could be a composition by A. Shoenberg, who was active in Vienna in that period and revolutionized the classical music with the introduction of the atonal and dodecaphonic music. Such music selection, although rather straightforward to be done manually, can be a challenging task for a machine to perform. Performing this task automatically, for any given place of interest, requires a way to identify the meaningful relations between a POI and music. We are searching for these relations in the Linked Data cloud.

4. SEMANTIC-BASED KNOWLEDGE REPRESENTATION AND CROSS DOMAIN RECOMMENDATION

Our generic framework for cross-domain recommendation is built upon an ontology-based knowledge representation. This representation can be defined as a graph/network of semantic entities (concepts) of different domains interlinked by semantic relations (properties). Entities can be roughly categorized as classes and instances. Classes are types or categories of concepts, such as ‘city’, while instances are particular individuals of classes, such as ‘Madrid’. Moreover, semantic relations can link classes (e.g. a city ‘belongs to’ a country), instances (e.g. Madrid ‘is the capital of’ Spain), or both types of entities (e.g. Madrid ‘is’ a city). Links can express hierarchical relationships, e.g. ‘subclass of’ and ‘instance of’, or have an arbitrary meaning.

In the above semantic graph-based representation, our final goal is to automatically find paths between an entity from a source domain (e.g. places of interest) to entities from a target domain (e.g. music artists), and select (recommend) some of the reached target entities according to certain criteria⁹.

⁹ Without loss of generality, our analysis is restricted to two domains, but it remains valid for more domains.

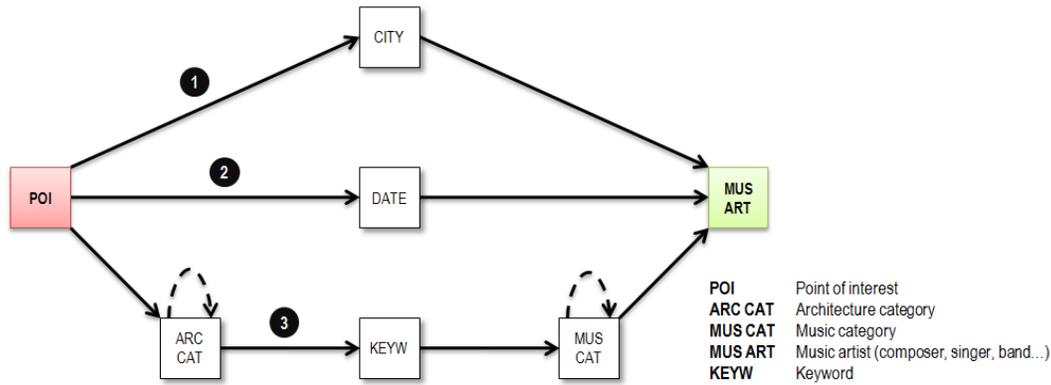


Figure 2. Semantic knowledge framework linking POIs and music artists.

To address that problem we exploit DBpedia since it currently represents the major multi-domain structured knowledge base in the Web, and provides an online query endpoint which we can easily use to access and gather the required information.

Our approach consists of three main stages. In the first stage, we restrict the subspace of DBpedia that is explored by identifying the DBpedia classes that belong to the two domains of interest, and the relations existing between instances of such classes. Next, we create a directed acyclic graph in which nodes represent the identified classes, and edges represent the selected relations (Figure 2). In this graph, there is a target node that does not have output relations, and corresponds to the entity class to recommend (‘Music artist’ in Figure 2). This stage is detailed in Section 4.1.

In the second stage, from the built graph, we generate one semantic network for a particular source instance (e.g. ‘Vienna State Opera’ in Figure 4). In practice, we query DBpedia to obtain instances related to the source instance according to the classes and relations defined in the above-mentioned graph. The retrieved instances are then incorporated into the network, and are used to query DBpedia for additional related instances. This process is repeated iteratively until reaching instances of the target class (Figure 4). This stage is detailed in Section 4.2.

Finally, we assign weights to the nodes and edges of the semantic network that was created in this process, and perform a weight spreading mechanism to filter (and recommend) the target instances with the highest spread relevance values. This stage is explained in Section 4.3.

4.1 Cross-domain Semantic Knowledge Framework

The first stage of our approach consists of defining a directed acyclic graph, which describes how the two domains of interest – source and target – are linked by means of semantic concepts and properties available in the used knowledge repositories (i.e., DBpedia in the implementation presented herein). In this graph, nodes are associated to classes, and edges correspond to relations between classes and/or instances.

The selection of classes and relations is guided by experts on the domains of interest and knowledge repositories. Figure 2 shows a cross-domain graph defined for the case study presented in Section 3 that aims at providing music artist recommendations for points of interest. The source and target domains are respectively characterized by POIs and music artists, so ‘POI’ and ‘Music artist’ classes represent the starting and ending nodes in the knowledge framework. Analyzing DBpedia, we identify three potential semantic paths from POI to Music artist entities. The

first path links POIs and music artists through the entity ‘City’, in the sense that there may exist music artists who were born, died or lived in the city of a particular POI. Similarly, the second path links POIs and music artists through the entity ‘Date’ since there may exist music artists who were born, died or lived in the year (decade, century, etc.) in which a POI was built. Finally, the third path links POIs and music artists in a more complex way. It utilizes the entity ‘Keyword’ to relate architecture and music categories, which are directly linked to POIs and music artists, respectively. Hierarchical relations between architecture/music categories are also taken into account (dashed edges in the figure).

Our framework also allows assigning relevance values for the considered semantic entities and relations, which may be used in the recommendation stage. These values could be assigned by the domain experts, or could come from the user’s profile. For instance, a domain expert may assign higher relevance to the class ‘City’ than to the class ‘Keyword’, since the former can be considered more informative to link a particular POI with related music artists. Similarly, specific instances like ‘Opera composers’ and ‘Classical music’ may receive high relevance if the user has a clear preference for them, hence producing personalized associations. Moreover, we consider the case in which relations also receive relevance values. For instance, to measure (e.g. by using TF-IDF weighting schemas) the relative strength with which the keyword ‘modern’ describes the music genre ‘Rock’, with respect to other more/less informative keywords.

To take into account the above issues, we propose the generic model shown in Figure 3. Formally, let $\mathcal{E} = C \cup I$ be the set of class and instance entities. We define a function $rel_e: \mathcal{E} \rightarrow \mathbb{R}_+$, with $\mathbb{R}_+ = \{x \in \mathbb{R} : x > 0\}$, which represents the relevance value assigned to entities $E \in \mathcal{E}$. We also define a function $rel_r: \mathcal{E} \times \mathcal{E} \rightarrow \mathbb{R}_+$ that represents the relevance value assigned to relations $R \in \mathcal{E} \times \mathcal{E}$ between pairs of entities (classes or instances).

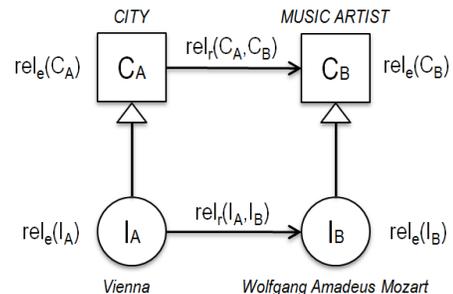


Figure 3. Relevance values assigned to semantic entities (classes and instances) and relations between them.

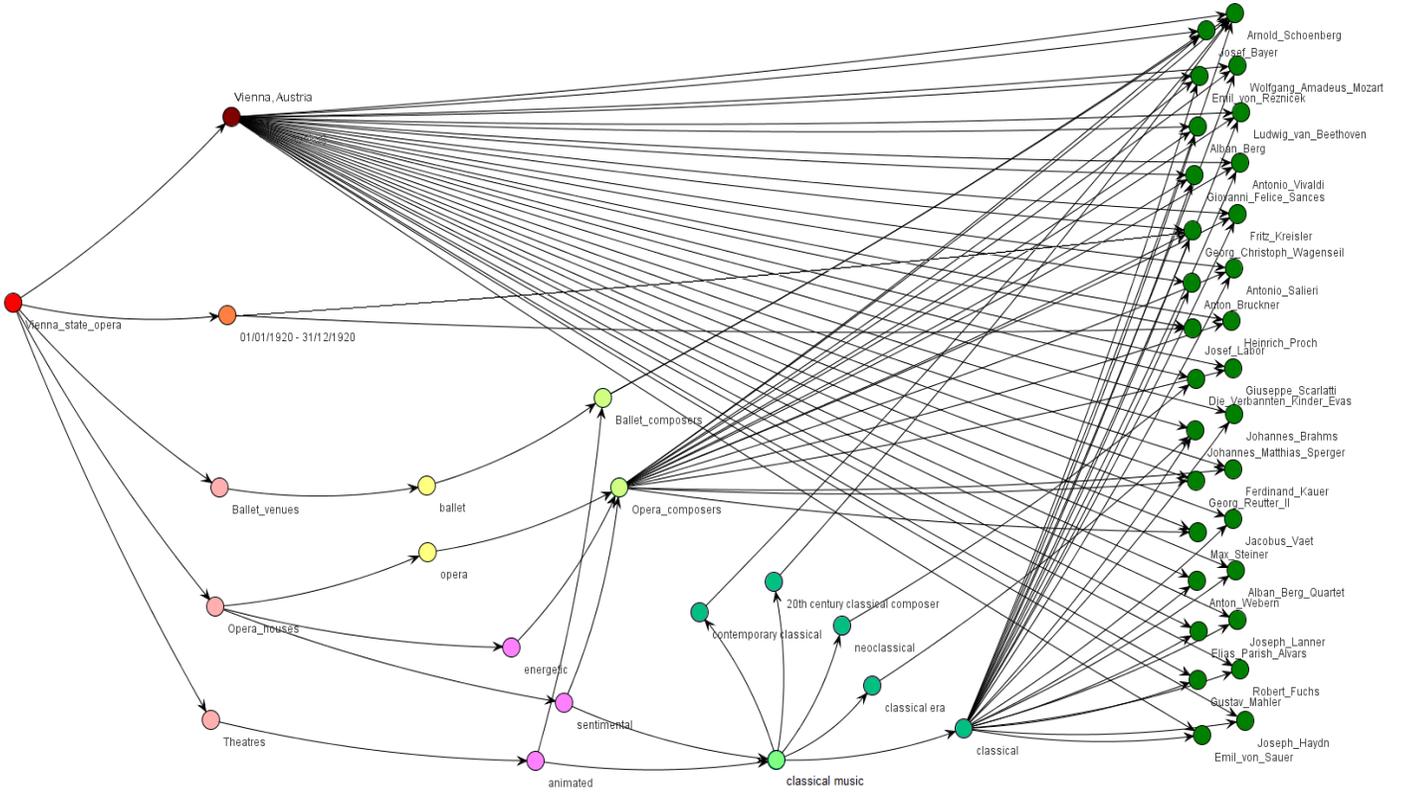


Figure 4. Semantic-based network obtained for the POI ‘Vienna State Opera’. Node colors represent the different types of semantic entities: POIs, cities, dates, architecture categories, tags, keywords, music categories, music genres, and music artists.

4.2 Cross-domain Semantic Network

The second stage of our approach consists of building a semantic network that explicitly links a particular instance in the source domain with its related instances in the target domain. This network is created by exploring the structured knowledge through the classes and relations established in the semantic framework (Section 4.1). Figure 4 shows a semantic network obtained for the POI ‘Vienna State Opera’. On the right side of the figure, we plot the retrieved music artists. In the center, we plot the instances and relations that link the above POI and music artists. The colors of the nodes represent the different instance classes: POIs, cities, dates, architecture and music categories, and keywords. Note that as input data, we have also considered emotional tags used in [10], and music genres obtained from Last.fm. The motivation of that will be clarified later on in this paper.

Although not illustrated in the figure, the semantic network does contain weights assigned to instances and relations. These weights are obtained from the corresponding relevance values, and would be finally exploited by a graph-based recommendation algorithm.

In general, we define the weight of a relation as a function $V: I \times I \rightarrow \mathbb{R}_+$ that depends on the relevance values of the connections between each pair of instances $I, I' \in I$, and between their classes $C_I, C_{I'} \in C$:

$$V(I, I') = f(\text{rel}_r(I, I'), \text{rel}_r(C_I, C_{I'}))$$

where $V(I, I') = 0$ if there is no link from I to I' .

Similarly, we define the weight of an instance I_k as a function $W: I \rightarrow \mathbb{R}_+$ that not only depends on the relevance values of the instance and its class C_{I_k} , but also on the weights of predecessor connected instances I_1, \dots, I_{k-1} and the weights of their relations

$$V(I_1, I_k), \dots, V(I_{k-1}, I_k):$$

$$W(I_k) = g(\text{rel}_e(I_k), \text{rel}_e(C_{I_k}); W(I_1), \dots, W(I_{k-1}); V(I_1, I_k), \dots, V(I_{k-1}, I_k))$$

In the next section, we provide particular definitions of these weight functions for a simple recommendation algorithm.

4.3 Cross-domain Recommendation

The third and last stage of our approach is the recommendation algorithm that exploits the weighted connections between instances in the semantic network. To preliminarily test the feasibility of our approach, we have implemented a simple retrieval algorithm that spreads the entity relevance weights through the network by iteratively adding them based on the relation weights. The algorithm is as follows.

First, the weight V of the relation between each pair of connected instances $I, I' \in I$ is computed by a linear combination of the relevance values assigned to the semantic connections (I, I') and $(C_I, C_{I'})$, where $C_I, C_{I'} \in C$ are the classes of instances I, I' , respectively.

$$V(I, I') = \alpha \cdot \text{rel}_r(I, I') + (1 - \alpha) \cdot \text{rel}_r(C_I, C_{I'}), \alpha \in [0, 1]$$

Next, the entities are sorted by graph topological ordering (I_1, \dots, I_k) . Following this ordering, each instance $I_k \in I$ is assigned a weight based on the weights of predecessor connected instances in $\{I_1, \dots, I_{k-1}\}$, the weights of the connections $V(I_1, I_k), \dots, V(I_{k-1}, I_k)$, and the relevance value of the instance class C_{I_k} .

$$W(I_k) = \beta \cdot \sum_{p=1}^{k-1} W(I_p) \cdot V(I_p, I_k) + (1 - \beta) \cdot \text{rel}_e(C_{I_k}), \beta \in [0, 1]$$

For simplicity purposes, in the previous formula, we do not include the instance relevance value $\text{rel}_e(I_k)$.

Finally, the algorithm returns a ranked list with the highest weighted target instances.

As future work, we plan to investigate more sophisticated algorithms. We are interested in developing constrained spreading activation mechanisms, taking into account factors such as path lengths, node in/out-degrees, weight propagation thresholds, and non-positive relevance values. We also want to explore alternative approaches such as Ford-Fulkerson’s algorithm [7] for flow networks.

5. PRELIMINARY RESULTS

We have developed a computational architecture that implements the proposed semantic-based framework for cross-domain recommendation. As a proof of concept, our system has been specialized to address the case study in which music artists are recommended for a particular place of interest. The system is, however, modular and flexible, and can be easily adapted to other domains of interest. The implementation of the algorithms involved in the three stages presented in Section 4 is generic, so special instantiations of the proposed algorithms are possible.

The system operates in both offline and online modes. In offline mode, the system iteratively queries DBpedia to obtain all the entities (and relations) defined in the knowledge framework for the source and target domains. The acquired data is then stored into a relational database. In online mode (i.e., at execution time), and for a particular input entity, the system has access to the database to retrieve instances and relations related to the input entity (a POI in our case), and builds with all of them the corresponding cross-domain semantic network. Over this semantic network, the system finally performs the graph-based retrieval algorithm to return a ranked list of target instances.

Focusing on the POI-music artist case study, in Section 5.1, we briefly describe the knowledge acquisition process, and present some statistics about the dataset generated for POIs of 21 European cities. In Section 5.2, we provide and explain music artist recommendations made for two example POIs.

5.1 Knowledge Acquisition

In the first stage of our approach, a domain expert has to identify the semantic entities and relations available in DBpedia that can be used to describe and link the domains of interest – architecture and music –, and more specifically the source and target entities – architectural POIs and music artists. Based on the defined framework, the system acquires the data querying DBpedia.

To preliminary test the feasibility of the approach, we executed it for POIs in 21 European cities, which are shown in Table 1. The knowledge acquisition process is summarized next where we distinguish whether the retrieved information belongs to the architecture or music domain.

Table 1. Cities currently available in our system database.

Cities
Amsterdam, Brussels, Copenhagen Barcelona, Madrid, Seville Berlin, Hamburg, Munich Dublin, Edinburgh, London Florence, Milan, Rome Lyon, Paris, Bern Prague, Vienna, Warsaw

5.1.1 Architecture Knowledge Acquisition

Regarding the architecture domain, we first obtained from DBpedia architecture taxonomies whose categories (classes)

would be directly linked to POIs. Specifically, we retrieved the taxonomies derived from the root (Wikipedia) categories `Architectural_styles`, `Visitor_attractions`, `Architecture_by_country`, `Years_in_architecture`, and `Architectural_history`. For instance, to obtain the direct subcategories of `Architectural_styles` we launched the following RDQL query to DBpedia:

```
SELECT ?x WHERE {
  ?x
  http://www.w3.org/2004/02/skos/core#broader
  <http://DBpedia.org/resource/Category:Architectural_styles>
  . }
```

As shown in Table 2, we retrieved a total of 697 architecture categories, which were integrated into two main taxonomies: architecture structures and architectural styles.

Table 2. Architecture categories.

	Architecture structures	Architectural styles	Total (distinct)
#Categories	438	298	697
#Categories linked to POIs	166	64	229

Once the architecture taxonomies were built, we were able to retrieve all the POIs of a particular city. We queried DBpedia for those entities that are linked to the city category (through the `dcterms:subject` property) and to any of the stored architecture categories. As shown in Table 3, we obtained an average of 147.5 POIs per city and 1.4 architecture categories per POI, and a total of 229 architecture categories linked with POIs.

Having collected and linked city, POI and architecture category entities, we proceeded to get more metadata about them. Specifically, we queried DBpedia for the entities’ English abstracts, which were processed to extract date information. We also created a limited set of keywords from the architecture category nouns. Table 3 shows some statistics about this metadata.

Table 3. POI and architecture database characteristics.

Architecture	
#POIs	3098
#POIs with date information	2005
Avg. #POIs/City	147.524 (185.712)
#Architecture categories related to POIs	229
Avg. #Architecture categories / POI	1.402 (0.769)
#Architecture categories with keywords	181 (with 109 keywords)
Avg. #Keywords / Architecture category	1.094 (0.310)

5.1.2 Music Knowledge Acquisition

For the music domain, we followed the same procedure as for the architecture domain. We first queried DBpedia for building music taxonomies. In this case, the root (Wikipedia) categories were `Musical_subgenres_by_genre`, `Musical_genres_by_region`, `Centuries_in_music`, `Musical_eras`, `Musicians_by_genre`, `Composers_by_genre`, and `Singers_by_genre`. Table 4 shows that we collected a total of 1116 distinct music categories, associated to three taxonomies: music artists, composers, and genres.

Table 4. Music categories.

	Artists	Composers	Genres	Total (distinct)
#Categories	467	126	525	1116
#Categories linked to POIs	110	48	153	309

Analogously to the architecture domain, once the music categories were built, we could access DBpedia to collect relevant music artists (i.e., musicians, composers, singers, and bands). Specifically, we queried DBpedia to retrieve those entities linked to the cities and to the collected music categories. In this case, the semantic relations between music artists and cities were ‘birth place’ (dbpprop:placeOfBirth, dbpedia-owl:birthPlace), ‘death place’ (dbpprop:placeOfDeath, dbpedia-owl:deathPlace), ‘origin place’ (dbpprop:origin, dbpedia-owl:origin), and ‘residence place’ (dbpprop:residence, dbpprop:hometown, dbpedia-owl:residence, dbpedia-owl:hometown). On the other hand, the relations between music artists and music genres were given by dbpprop:genre and dbpedia-owl:genre properties.

Table 5 shows some statistics about the collected data. We obtained a total of 1568 music artists: an average of 76 artists per city, and 131 artists per genre. We also obtained additional metadata for music categories and artists, such as keywords, Last.fm tags and genres, and emotional tags used in [10]. This metadata is not detailed here because it is out of the scope of the paper.

Table 5. Music database characteristics.

Music	
#Music artists	1568
Avg. #Music artists / City	76.048 (151.508)
Avg. #Music artists / Genre	131.846 (130.888)
#Music categories related to artists	309
Avg. #Music categories / Music artist	1.719 (0.982)
#Music categories with keywords	511 (with 109 keywords)
Avg. #Keywords / Music category	1.235 (0.459)

5.2 Recommendation Examples

In Tables 6 and 7, we show the lists of top 10 recommended music artists for two different POIs: the State Opera of Vienna, Austria, and the Wembley Stadium of London, UK. It can be seen that, in general, our approach suggested 18th-19th century composers for the Austrian opera house, and British modern rock bands for the British sport arena.

Table 6. Top 10 recommended music entities for POI ‘Vienna State Opera’ (Vienna, Austria).

Music artist	Top music genres	Born/Death Countries	Date
Wolfgang Amadeus Mozart	Classical Instrumental	Austria Austria	18th century
Arnold Schoenberg	Classical Avant-garde	Austria USA	20th century
Emil von Reznicek	Classical Opera	Austria Germany	20th century
Alban Berg	Classical Contemporary	Hungary Austria	20th century
Ludwig van Beethoven	Classical Instrumental	Germany Austria	19th century
Antonio Vivaldi	Classical Baroque	Italy Austria	18th century
Giovanni Felice Sances	Classical Baroque	Italy Austria	17th century
Fritz Kreisler	Classical Violin	Austria USA	20th century
Georg Christoph Wagenseil	Classical Baroque	Austria Austria	18th century
Antonio Salieri	Classical Italian	Italy Austria	19th century

Table 7. Top 10 recommended music entities for POI ‘Wembley Stadium’ (London, UK).

Music artist	Top music genres	Born/Death Countries	Date
Beady Eye (Oasis band members)	Rock British	UK (origin)	2009
Operahouse	Indie Rock British	UK (origin)	2006
The Woe Betides	Rock Grunge	UK (origin)	2008
Skunk Anansie	Rock Female vocalist	UK (origin)	1994
The Fallen Leaves	Garage Acoustic	UK (origin)	2004
Ivryrise	Rock Alternative	UK (origin)	2007
Plastic Ono Band (John Lennon & Yoko Ono)	Experimental Avant-garde	UK (origin)	1969
We Are Balboa	Indie Rock Female vocalist	Spain-UK (origin)	2003
Goldhawks	Rock British	UK (origin)	2009
Teddy Thompson	Folk British	UK USA	1976

One of the benefits of our approach is its capability to provide explanations of its recommendations. These explanations may be based on the discovered semantic paths between the input POI and the suggested music artists, in the associated semantic network (Figure 4). From a particular music artist, we can go back through semantic paths until reaching the POI. In the process, we can generate an explanation sentence for each explored link (Figures 5 and 6). For instance, we may recommend ‘Mozart’ for ‘Vienna State Opera’ because Mozart was an Opera composer, Opera composers are related to Opera houses (through the keyword opera), and Opera house is the architecture category of Vienna State Opera.

<p>PLACE OF INTEREST: Vienna State Opera CITY: Vienna, Austria MUSIC ARTIST: Wolfgang Amadeus Mozart ARCHITECTURE CATEGORY: Opera houses KEYWORD: opera MUSIC CATEGORY: Opera composers MUSIC ARTIST: Wolfgang Amadeus Mozart TAG: energetic MUSIC CATEGORY: Opera composers MUSIC ARTIST: Wolfgang Amadeus Mozart TAG: sentimental MUSIC CATEGORY: Opera composers MUSIC ARTIST: Wolfgang Amadeus Mozart MUSIC GENRE: classical MUSIC ARTIST: Wolfgang Amadeus Mozart ARCHITECTURE CATEGORY: Theatres TAG: animated MUSIC GENRE: classical MUSIC ARTIST: Wolfgang Amadeus Mozart</p>
--



Figure 5. Semantic paths found between the POI ‘Vienna State Opera’ and the Austrian classical music composer ‘Wolfgang Amadeus Mozart’. Sentimental tags from [10] and Last.fm genres are incorporated into the semantic network.

PLACE OF INTEREST: Wembley Stadium

CITY: London, United Kingdom

MUSIC ARTIST: Beady Eye

TIME: 01/01/2007 - 31/12/2007

MUSIC ARTIST: Beady Eye

ARCHITECTURE CATEGORY: Music venues

ARCHITECTURE CATEGORY: Rock music venues

KEYWORD: rock

MUSIC CATEGORY: Indie rock

MUSIC ARTIST: Beady Eye

MUSIC CATEGORY: Rock music

MUSIC ARTIST: Beady Eye

TAG: strong

MUSIC CATEGORY: Rock music

MUSIC ARTIST: Beady Eye



Figure 6. Semantic paths found between the POI ‘Wembley Stadium’ (London, UK) and the British rock band ‘Beady Eye’ (formed by former members of Oasis band). Sentimental tags from [10] and Last.fm genres are incorporated into the semantic network.

6. DISCUSSION AND FUTURE WORK

We have presented an ongoing research work on the design and development of a generic semantic-based framework for cross-domain recommendation. Our approach aims at integrating open structured knowledge sources from multiple domains into a common semantic network representation. Over such network, the proposed approach can find semantic paths between entities in different domains of interest, and can return a ranked list of entities in the target domain.

In this preliminary study, as a proof of concept of our approach, we have investigated the possibilities of using Linked Data to discover semantic paths between places of interest and music artists.

As future work, we intend to combine the described approach with the tag-based matching of POIs and music tracks [9]. This would allow us to match places with individual music tracks. The two approaches can be combined using a cascade method – first, musicians that are semantically related to a given POI can be identified in the Linked Data cloud, and afterwards the individual music tracks by these musicians can be selected to suit the POI based on the emotional characteristics of the items.

Furthermore, it is important to investigate how user preferences for particular music can be taken into account, as in the current experiments we have not evaluated how personalization can be managed in the proposed framework.

Finally, in order to assess how the users evaluate the recommendations provided using the described approach, we intend to build a working system that would allow users to browse a repository of POIs with detailed descriptions for each place, and would automatically retrieve music suited for the POIs. A user study has to be carefully designed in order to understand whether the users perceive our music recommendations as meaningful, and prefer them over non-adapted music tracks.

7. ACKNOWLEDGMENTS

This work was supported by the Spanish Ministry of Science and Innovation (TIN2008-06566-C04-02), and the Community of Madrid (CCG10-UAM/TIC-5877).

8. REFERENCES

- [1] 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.
- [2] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z. 2007. DBpedia: A Nucleus for a Web of Open Data. *6th International Semantic Web Conference*, 722-735.
- [3] Berkovsky, S., Kuflik, T., Ricci, F. 2007. Distributed Collaborative Filtering with Domain Specialization. *1st ACM Conference on Recommender systems*, 33-40.
- [4] Berkovsky, S., Kuflik, T., Ricci, F. 2008. Mediation of User Models for Enhanced Personalization in Recommender Systems. *User Modeling and User-Adapted Interaction* 18(3), 245-286.
- [5] Bizer, C., Heath, T., Berners-Lee, T. 2009. Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems* 5(3), 1-22.
- [6] Chung, R., Sundaram, D., Srinivasan, A. 2007. Integrated Personal Recommender Systems. *9th International Conference on Electronic Commerce*, 65-74.
- [7] Ford, L. R., Fulkerson, D. R. 1956. Maximal Flow through a Network. *Canadian Journal of Mathematics* 8, 399-404.
- [8] González, G., López, B. and de la Rosa, J. LL. 2005. A Multi-agent Smart User Model for Cross-domain Recommender Systems. *IUI 2005 Workshop Beyond Personalization*, 93-94.
- [9] Kaminskas, M., Ricci, F. 2009. Matching Points of Interest with Music. *ECDL 2009 Workshop on Exploring Musical Information Spaces*, 68-73.
- [10] Kaminskas, M., Ricci, F. 2011. Location-Adapted Music Recommendation Using Tags. *19th International Conference on User Modeling, Adaptation and Personalization*, 183-194.
- [11] Kobsa, A. 2001. Generic User Modeling Systems. *User Modeling and User-Adapted Interaction* 11(1-2), 49-63.
- [12] Li, B., Yang, Q., Xue, X. 2009. Can Movies and Books Collaborate? Cross-domain Collaborative Filtering for Sparsity Reduction. *21st International Joint Conference on Artificial intelligence*, 2052-2057.
- [13] Szomszor, M., Alani, H., Cantador, I., O’Hara, K., Shadbolt, N. R. 2008. Semantic Modelling of User Interests based on Cross-Folksonomy Analysis. *7th International Semantic Web Conference*, 632-648.
- [14] Szomszor, M., Cantador, I., Alani, H. 2008. Correlating User Profiles from Multiple Folksonomies. *19th ACM Conference on Hypertext and Hypermedia*, 33-42.
- [15] Winoto, P., Tang, T. 2008. If You Like the Devil Wears Prada the Book, Will You also Enjoy the Devil Wears Prada the Movie? A Study of Cross-Domain Recommendations. *New Generation Computing* 26(3), 209-225.
- [16] Yu, C. C. 2004. A Web-based Consumer-oriented Intelligent Decision Support System for Personalized E-services. *6th International Conference on Electronic Commerce*, 429-437.