

# A Proposal for an Open Local Movie Recommender

Federica Cena, Elisa Chiabrando, Andrea Crevola\*, Martina Deplano, Cristina Gena, and Francesco Osborne

University of Torino, \*3Juice - Italy {cena, chiabrando, deplano, gena, osborne}@di.unito.it, {andrea.crevola}@3juice.com

**Abstract.** This position paper describes the initial research assumptions and work carried on in the context of the ReAL CODE, a project aimed at realizing an intelligent and tailored distribution of digital creative contents towards end users. Contents belong to the movie domain and particular attention will be given to local contents, that will be promoted through recommendations in order to support local cultural heritage. The main research goals of the project are open data and interoperability of domain and user data. Our initial efforts on these key directions will be described.

**Keywords:** interoperability, open data, recommender, movies

## 1 Introduction

Movies belong to cultural heritage domain of a given place. On the one side the notion of cultural heritage links two concepts: heritage and culture. The term "heritage refers to local landscape, architecture and artefacts that make a particular place unique. The term cultural puts less emphasis on the place and, more generally, regards knowledge, and local traditions of a territory. We term cultural heritage resources both the physical heritage and the intangible activities related to it, like movies filmed and related to these places.

On the other end, it is not unusual to consider movies as belonging to cultural heritage and history of a given place. Increasingly, historians have moved away from a history that chronicles battles, treaties, and presidential elections to one that tries to provide an image of the way daily life unfolded for the mass of people. Film has an important role to play in these histories. [...] But when we focus on social and cultural history, especially the important role of leisure in the lives of ordinary people, film not only provides evidence and records but takes on a key role<sup>1</sup>. Movies filmed in a given place and in a given time are not only a priceless statement of the daily life and a picture of those places, but, when their authors and producers are somehow linked to that place, they also witness the cultural life and the cultural ferment of territory. The link to

<sup>1</sup> <http://historymatters.gmu.edu/mse/film/socialhist.html>

the evolving local movie resources produced in our region<sup>2</sup> is one of the key point of ReAL CODE<sup>3</sup> (Recommendation Agent for Local Contents in an Open Data Environment), a project aimed at designing an intelligent and tailored distribution of digital creative contents towards end users. Contents belong to the movie domain and particular attention will be given to local contents - namely movies produced and realized in the Piedmont region - which will be promoted through recommendations.

More in detail, one of the main goal of the project is to develop an open, interconnected and semantic database of local creative contents, in line with current trends of *Web 3.0* and *Open Data*. On the one hand the idea is to retrieve content information from different open datasets (Section 3) and automatically map it to the ReAL CODE own knowledge and format. On the other hand ReAL CODE own information will be available in a semantic and open format, so that everyone will be free to use, reuse and redistribute it. Another important goal of the project is to achieve user model interoperability (Section 3) in order to obtain a big amount of information about the user and her interest and to overcome the cold start problem in recommendations.

As for the user experience, the challenge is to realize an innovative and engaging application, that differs from traditional movie recommendation sites. In the ReAL CODE system, movies will be considered as an important part of people life, related to specific moments, places and people (Section 2). Based on user activity on the system and on imported user model, ReAL CODE will suggest the user movies that she could like, focusing on local contents. Movie operators could also join the system in order to insert and promote their own contents; in this way, the project will support the local cultural industry.

## 2 A Scenario

Similarly to music, movies are often related to people's special moments and they recall feelings and emotions. Independently from the feelings they trigger, movies are related to particular moments, to other people and to places.

**Final User.** *Marta enters ReAL CODE through Facebook login. Using information stored in the Facebook account, the system shows her some movies she liked or added to her profile. Among these movies there is "Roman Holiday", a movie with her favourite actress, Audrey Hepburn, that she used to see during her teenage years with her friends. She selects it and adds a few information about her sight experience: where she saw it (in her hometown, Rome), when (she can select a specific date) and with whom (her friend Sofia, who is also a*

---

<sup>2</sup> The movie industry plays a key role in the Piedmont Region, where we live. Piemonte Film Commission is a very active entity in producing movies and organizing movie-related events. The city of Torino hosts different film festivals (e.g. The International Turin Film Festival) and the National Cinema Museum, visited every years by thousand of national and international visitors.

<sup>3</sup> <http://www.ReALCODE.it/>. ReAL-CODE is a research project funded in the context of POR FESR 2007/2013 of the Piedmont Region, Italy.

Facebook's friend). She decides that she will add the other optional information (tag, vote, comments...) later. Immediately Sofia receives a Facebook notification about Marta's activity because ReAL CODE is in-and-out integrated with Facebook. Since she is pleased to remember this event, she decides to join ReAL CODE herself and she adds a comment on Marta's activity. After having added a minimum number of movies, the system can show Marta's "movie history" in multiple way: on a timeline, on a map or on a sort of diary. The system also shows her some recommendations about movies she could like. Beyond traditional suggestions (blockbusters movies, top rated movies, etc) the system recommends Marta local movies release.

**Producer.** Francesco is a small movie producer that lives in a town near Turin. His company has a YouTube channel and a Facebook page. He decides to join ReAL CODE in order to increase its visibility, especially on niche and local audience. He imports into ReAL CODE information and videos from YouTube and Facebook and adds some other details about his company. Since next month he will participate with his last short movie to the Turin Film Festival, he accesses on ReAL CODE, adding specific information about his short movie, such as the movie categories, movie tags, etc. In this way, this movie could be recommended to potentially interested users.

### 3 Data interoperability for User and Domain Models

**User Model interoperability.** The emergence of social networks on different domains (Facebook, Twitter, aNobii, LinkedIn) has made available an enormous amount of user data, such as demographic information, current location, friendship network, job and interests. Aggregating these data may be useful to solve the cold-start and sparsity problems, but it requires a deep knowledge of the user profiles in the various social networks and how combining data to obtain a complete and effective user model.

Some recent works [1][2] analyze and create user models from popular social networks. In Abel et al.[1] the authors capture user information of Twitter, Facebook, LinkedIn and social tagging activities in Flickr, Delicious and StumbleUpon to evaluate the performance of a strategy based on several cross-system user modeling. They improve recommendation quality in a significantly manner. Likewise, Shapira et al.[2] integrate Facebook data with the recommendation process and compare traditional different collaborative filtering methods with their cross-domain recommendation.

Differently from these previous works, we propose to combine content-based and collaborative filtering-based methods to suggest interesting movies according to a user model created from data extracted from different social networks. For our aims, the most interesting social networks are Youtube and Facebook.

Through the YouTube API<sup>4</sup>, data we are considering to use are: *favorite videos* (videos flagged as favorite by the user), *watch history*(videos watched by

---

<sup>4</sup> Youtube API 2.0 <https://developers.google.com/youtube/>

the user), *uploads feed* (videos uploaded by the user), *video recommendations* (videos that may appeal to a user), *user subscription feed* (channels and people the user has subscribed to), *standard channel feeds* (channels that reflect user preferences). The two main problems we encountered with YouTube are: (1) most of the titles does not match exactly with official movies titles; (2) if the user's channel on Youtube has not been linked to her Google+ account, we can not find the corresponding user personal data.

On the other hand, personal data and information on users' interests can be collected through the Facebook Graph API and FQL interface<sup>5</sup>: user personal information (*i.e.* birthday, gender, geolocation); her friendship network; the lists of her favorite books, music and movies; the list of posts in her stream created by third application, such as Youtube. However, movies in Facebook are not categorized (*i.e.* comedy or horror), so it is necessary to retrieve this information from other external datasets, such as the ones described below. Moreover, recently, Facebook has introduced the notion of action on particular objects of the system (fitness, music, news, video and books), allowing users to tell sort of "stories" about what she did or what she would like to do. In particular, user may either tell that she watched a film and then evaluate it, or what she would like to see it in the future. Our idea is to import these data in the ReAL CODE, reason on them, and then use them both for the user timeline and for recommendations.

A recent innovation of Facebook allows users to **tell stories** on Facebook through a structured, strongly typed API<sup>6</sup> about what she did or what she would like to do. In particular, user may either tell that she watched a film and then evaluate it, or what she would like to see it in the future. Our idea is to import these data in the ReAL CODE reason on them, and then use them both for the user timeline and for recommendations.

**Domain Interoperability.** The Open Data initiative promotes the idea that the data should be freely available to everyone to use and re-publish. "Openness" of data enables the construction of a place in the web for global sharing, the "Web of Data". In this line, many information about the movie domain can be found in the Web and used in our project.

*LinkedMDB*<sup>7</sup> aims at being the first open semantic web database for movies, including a large number of interlinks to several datasets on the open data cloud. Data can be accessed using traditional Web browsers, Semantic Web browsers, SPARQL clients, but there are not APIs available.

*TheMovieDB*<sup>8</sup> provides a large movies database (118,000 titles) classified in 18 categories with other information (id, alternative title, cast, images, keywords, trailers, similar movies and so on) available through APIs.

---

<sup>5</sup> Facebook Graph API <http://developers.facebook.com/docs/reference/apis/>

<sup>6</sup> Recently, Facebook is introducing the notion of action on particular objects of the system (fitness, music, news, video and books) <http://developers.facebook.com/docs/reference/apis/>

<sup>7</sup> LinkedMDB <http://www.linkedmdb.org>

<sup>8</sup> TheMovieDB <http://www.themoviedb.org>

*YouTube* is an important data source, even if it contains only videos and clips, and not whole movies. However, information such as trailer and popular scenes can be very useful to enrich the domain model. Some of the available data from the API are: Videos related to a specified video; Title and author of the video (user ID, name, etc); Average rate and statistics (how many views, likes, bookmarks). There is not a semantic ontology describing categories of movies, but users can freely select some categories at the uploading time.

*DBpedia*<sup>9</sup> allows to extract structured data from Wikipedia and to link other data sets to it. 172,000 films instances are present, classified in a consistent ontology formed by a three-level taxonomy covering different general classes (genre, themes, year, location, nationality, etc) and 58 classes regarding genre of film, the most important category for us. These data are accessible via a SPARQL query endpoint. We chose DBpedia as a primary source for movie domain thanks to the advantages of its knowledge base: it covers many domains; it represents real community agreement; it automatically evolves as Wikipedia changes; and it is truly multilingual. Moreover, it is very informative in relation to movie domain, and this allow us to have a complete description of a film. For our purpose, it is particularly important the field “subjects”, since it contains the ontological categories the movie belongs to, according to Wikipedia ontology, and we can use this information for user modeling purpose, making inference on user’s interest in categories of films, starting from her interest in a specific movie. For the same reason, we will use also the other datasets when linked to DBpedia, since in this way it is possible to access Wikipedia categorization and integrate information with possibly missing ones.

**Open Movie Recommendation** Our idea is to create an open movie recommender, *i.e.* a recommender that uses open data from the Web as knowledge base and then combines content-based and collaborative filtering to suggest local interesting movies. Thus, after have retrieved the user personal information and the list of movies and friends in Facebook, ReAL CODE creates the user model of the specific user. Then, it finds meta information about movies (*i.e.* category, actors, register, and so on) from the external movies datasets seen above in order to create a domain model that can be used in the content-based recommendation. Moreover, ReAL CODE will use the information about user’s social network in Facebook and in Youtube to suggest movies that are interesting for similar users belonging to the user’s network.

## 4 Conclusion

Social Open Web (of Data) offers a lot of information we can get both about the user and the movie domain. Our next efforts will be concentrated on how to integrate, map, and reason on these data in order to have a knowledge base structured in such a way to provide new tailored user recommendation on local

---

<sup>9</sup> DBpedia <http://dbpedia.org/>

movies and to be open and available to other applications. Moreover, we are working in realizing an innovative user experience by means of game activities, that will be used also to infer new data about their preferences and thus enrich the user model.

Moreover, in order to improve the user model without just asking users to add explicit information and attract them to the ReAL CODE application, a set of Games With A Purpose [3] will be available. Through the gaming activity of the users and their mobile devices, the system will infer interesting data about their preferences and geolocalization. These data will be then used to provide customized recommendations. For example, one of the GWAPs provides a set of movies among which the player has to guess the correct ratings given by a random friend. Thanks to the confirmation by her friend, we can give points to the player as a reward, and implicitly update the user model of her friend. Another GWAP allows the player to add questions about movies and respond to ones posted by other users.

Thanks to this gaming activity, we can infer users' interests based on the selected movie with which they want to play. Furthermore, we can also take advantage of the human computation [4] approach for a community control of the correctness of the quizzes. Through the geolocalization of mobile devices, one typology of GWAP is connected to a treasure hunt approach: given a movie set in the current player city, she can collect points by checking-in in the places related to the plot of the movie. This game can be very useful also for the promotion of the tourism in a selected city, by incentivizing the users to visit specific places through gaming activities that can make the user experience more fun and appealing.

## 5 Acknowledgments

We would like to thank professor Giancarlo Ruffo for his helpful advices.

## References

1. F. Abel, E. Herder, G.-J. Houben, N. Henze, and D. Krause. Cross-system user modeling and personalization on the social web. *User Modeling and User-Adapted Interaction*, 23, 2013.
2. B. Shapira, L. Rokach, and S. Freilikhman. Facebook single and cross domain data for recommendation systems. *User Modeling and User-Adapted Interaction*, 23, 2013.
3. L. Von Ahn. Games with a purpose. *Computer*, 39(6):92–94, 2006.
4. L. Von Ahn. Human computation. In *Design Automation Conference, 2009. DAC'09. 46th ACM/IEEE*, pages 418–419. IEEE, 2009.