SEARCHING AND EXPLORING DATA IN A SOFTWARE ARCHITECTURE FOR FILM-INDUCED TOURISM

Sandro Savino, Nicola Orio

Department of Cultural Heritage, University of Padua, Italy
{sandro.savino, nicola.orio }@unipd.it

Abstract. Film induced tourism is a recent phenomenon, rising increasing interest in tourism management and promotion. A research project on this topic has been recently investigated at the Department of Cultural Heritage of the University of Padua, resulting in the development of a software architecture for the promotion of film-induced tourism, capable of storing and providing rich information about the movies produced in a selected geographic area. This paper presents the design and implementation of the solutions developed to search and explore the data stored in the software architecture.

Keywords: film-induced tourism, film annotation, user interfaces, personalization.

1 Introduction

Film-induced tourism has been defined by Sue Beeton as "visitation to sites where movies and TV programs have been filmed as well as to tour to production studios, including film-related theme parks" [1]. Film-induced tourism is an interesting asset that can be exploited in destination management and destination marketing: it attracts new visitors and also tourists who have already seen an area; it is largely independent from seasonal trends; it conveys tourists from overcrowded sites to new and less explored ones; and eventually can be suitable for a substantial re-branding of a certain area. The increasing interest in it is testified by the growing number of film-induced tourism related initiatives undertaken at international level by public and private bodies, which developed movie maps and movie tours or exploited the success of a particular movie as a tool for destination branding [2].

The Department of Cultural Heritage of the University of Padua recently investigated the topic of film-induced tourism in a project that brought together the expertise of film scholars and computer scientists; the goal of the project was to develop an information system that combines the data about a geographical area and the movies produced in it with the purpose of fostering film-induced tourism. The developed system is composed of two main components: a knowledge base storing all the relevant data about the movies shot in a given territory together with touristic information of the film locations, and an interface to query the knowledge base and retrieve the information.

A novel characteristic of the system resides in the idea to relate the information not to a movie as a whole (as it is instead commonly done in current systems) but to excerpts of the movie itself (called clips), leading to a database that has a very precise and fine-grained description of what appears in the movie; furthermore all these data are stored and organized in the system with the precise purpose to serve not only the tourist, but all the actors involved in film-induced tourism, as film-makers and destination managers.

The design and implementation of the model for this system can be found in [6]; this paper describes how the user interface has been developed and the techniques used to query the knowledge base to retrieve the information; the paper is organized as follows: the first part illustrates in more detail the context, the aim and the requirements of both the information system in general and of the interface to be developed; Section 3 reviews of the existing literature on these topics; in Section 4 we describe how we modeled the user experience and Section 5 details how the actual system was implemented; conclusion and future work are left in the final part of the paper.

2 Analysis of Requirements

The analysis of the requirements done at the beginning of the project led us to aim to a system capable of storing and retrieving rich and fine-grained data about movies and locations and to provide these data to different users. The intended users of the system were modeled in three main groups, based on their different interests:

- Tourists, interested in information to plan or enrich their visit;
- Agents of the touristic industry, interested in information to exploit and promote locations;
- Representatives of the movie industry, interested in information about locations and film-making related services.

In a previous work [6] we described how the system was modeled in order to store data that is informative for each of the different user groups; in particular, as a novel characteristic of the system, we decided that it should be able to store information at a fine-grained level, allowing to annotate portions of the movie. The idea is to record not only the places seen during the movie, but also other information like the actors on screen, any interesting objects on the scene, the emotions evoked, and to relate them to the specific portion of the movie they belong to. All these information can be then used to search and explore the data stored in the system, allowing the users to access and organize them in a manner that best suits their own needs. Figure 1 depicts an excerpt of the ER diagram, where the main entities and their relationships are represented. Each MOVIE is referenced by number of excerpts (ESTRATTO) selected by the film scholars according to their artistic relevance and by a number of moments-of-interest (MOI) that are short sequences selected according to their touristic interest. Both ESTRATTO and MOI have a relationship with one (the former) or many (the latter) points-of-interest (POI), that is georeferenced locations that are related to the MOVIE (setting, filming location). Each MOI shares a CLIP with one ESTRATTO. Searchable information is stored in full text descriptions

(TESTO), information about involved persons (PERSONA) and keyword-based descriptions (TAG).

Once the model was finally implemented and we started to populate it, we had two big challenges to face: the first was how to enable the system to organize and relate the data and the second was how to present them to the different users. In particular we were interested in building a system that was attractive and informative to the tourists user group; true to our intent to promote tourism through cinema and also to promote cinema through tourism, we wanted our system to enable the user to explore both the geographic space and the cinematographic space, creating an engaging, culturally rich, experience.

In Section 5 we describe the solutions found to organize and explore the data; in particular we will describe how we implemented the search function, the user profiling and the recommendation system.

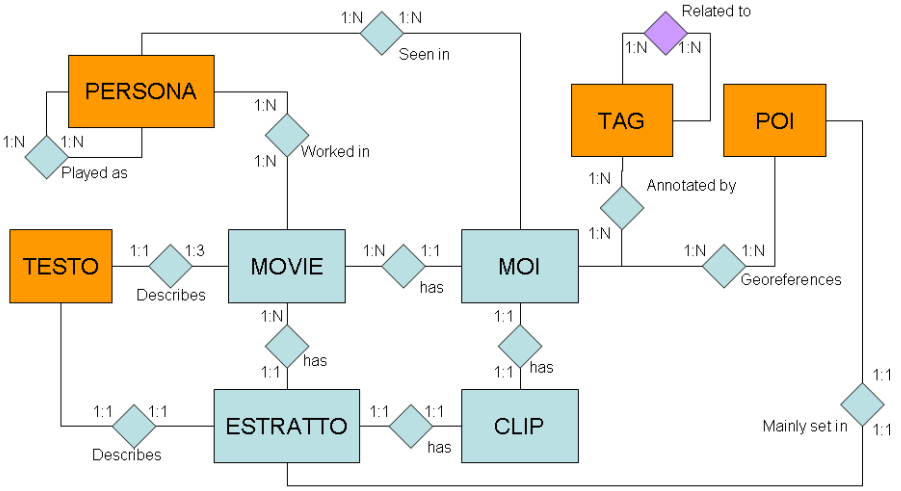

Figure 1:ER diagram of the main entities and relations of the schema; entities in orange are used in the free text search; the relation in purple is computed using a lexical database.

3 Related Work

While little or no works can be found in literature approaching the topic of film-induced tourism from the perspective of information systems and digital libraries, the proposed approach shares some common elements in the way user experience is modeled and in the way the search and recommendation functions have been developed.

The idea that users movements in a place, either the inner space of a museum or the open spaces of a city, can be classified according to user interest was first introduced in [7]. The user models (named after the behavior of four animals: ant, fish, grasshopper and butterfly) have been exploited in a number of approaches [8] [9]. More in general, the need to personalize multimedia access and to tailor it to

different categories of users has been described by [10] as a relevant challenge in human centered multimedia.

In our system, touristically relevant movie excerpts are described also by semantic tags. The effectiveness of social tags for search and retrieval in an interactive context has been investigated in [11], we experienced similar results also in the case of tags created by a pool of experts. To partially overcome this approach, following the approach initially proposed in [12] we use WordNet in order to expand tag descriptions. The next section illustrate how we used some of the techniques above to develop our own solution.

4 Modeling the user experience

As written above, our system has different intended users, divided in three user groups: tourists, agents of the tourism industry and professionals of the film industry. Each group has its own interests: tourist usually like to explore a place, both in a geographical, visual way and in a cultural sense; the second group is interested in materials for the touristic promotion of a location, such as texts, descriptions, pictures and video clips that can be used for instance to produce a brochure or an itinerary; the latter is more interested in finding information about filming, as a location for a new movie or a list of local professionals that can be hired for that purpose.

Table 1. The users of our system, with different interests and ways to interact with the system.

Group	User type	Interests	Interaction	Data presentation
1	Tourist	discover new places, learn new things both on movies and the territory	website, mobile app	maps, itineraries, recommendations, video clips
2	Tour operator, destination manager	touristic promotion, creation of touristic itinerary	DMS via RESTful API	descriptions, video clips, search results
3	Film maker, location manager	discover locations for new movies, see how the territory was read in previous movies	website	maps, video clips, data sheets, lists, search results

In order to provide data that could be useful for every user, we developed a system capable of storing a lot of information and we started to gather and produce these data, collecting them from archives, writing descriptions, geo-referencing and annotating movies.

To make this content accessible, we need also to develop an adequate interface, enabling all user groups to reach the information they wanted. In particular we took extra care in designing an interface that could appeal the tourist user group: they are usually the main target of film-induced tourism initiatives (and of our system

as well) but, as in most cases where the same content is provide to professional and casual user, they are also the more "volatile" because they are not driven by any real necessity (on the contrary of the professionals using our system for work-related purposes); hence, in order to attract them, we wanted our system able to provide an engaging and rewarding user experience.

To achieve this, we developed an interface that lets the user explore our database as he would explore the touristic space, using what we call a "path". A path is a sequence of pieces of information extracted by our database, e.g. the general description of a movie, the profile of a director, a geo-referenced video clip of a movie; each step of a path is related to the subsequent one by some means as for example spatial proximity or content similitude. While populating the database, the content is created in such a way that all the data in the database is geo-referenced or it is related to a geo-referenced element (e.g. a director has a 1:N relation to the locations used in all his movies).

The user can follow a predefined path, whose content has been selected by film or tourism experts, and thanks to the geo-reference embedded in each part of the path, if the user is equipped with a mobile device, he can actually follow the path, using our system as a guide to explore the geographic space around him. Conversely, the user can explore the area and let the system enrich his visit by providing geo-localized information about the surrounding context; this way to explore the data can also be simulated by virtually navigating an online map and does not require the user to actually walk around. A search function enables the user to query the database for a specific text (e.g. name, place, keyword); each search result is a possible starting point of a path that the user can begin to follow to explore the database.

Most of the content in the database has more than one relation with other contents (e.g. a place or an actor appearing in different movies); on the one hand, this lets the user branch off a given path at any moment and start a different one; on the other hand, this often requires the system to filter the amount of choices offered to the user, to avoid overwhelming him/her. The filtering module tries to minimize the number of choices and, at the same time, to maximize their diversification, to encourage the user to explore new content; with respect to this, offering a loosely related option is not considered an error as it adds a sort of serendipity to the exploration of the data.

To make the experience more customized, some personalization has been implemented in the filtering: users are profiled depending on their previous searches and choices and the content suggested to them through recommendations is influenced by the profile associated to the user. The users belonging to the second group are less likely to need all these functionalities; for them we opted to offer a software interface instead of a graphical interface: by providing a list of APIs, tour operators and destination managers can access the data stored in the database with their DMS or other software. The APIs give access to some data that are not visible from the user interface described above because of scarce interest to the tourist (e.g. copyrights holders).

The requirements of the users in the third group are a mix between those of the first two: they can use some of the functionalities provided by the GUI (e.g. the geographical exploration, the visualization of clips) but they are also interested in some data that is not provided to the tourist (e.g. the list of professionals who worked

in a production). To accommodate these needs, we decided to offer a signed-on access to the website: once logged in, this particular type of user is recognized and the GUI shows the extra information needed.

In the following section we will describe how we implemented the interfaces described above, with particular detail to the techniques used in the development of that for the first group of user.

5 Implementation of the solutions

As highlighted above, the key features to make the system appealing to our intended users are to have a lot of data, to have them accessible and, possibly, related to each other. To achieve this we exploited a number of information retrieval techniques.

First of all, to make the data easily accessible, we implemented a free text search function, backed by a search engine that indexes most of the texts of our database (see Figure 2 for the list of searchable fields). The search engine has been developed within the project and it handles single word searches and multi word searches separately. The first case is the simplest: the word is searched for an exact match in the indexed tables of the database, with the exception of the tables POI and PERSONA, where it is searched for an approximate match to account for the cases where the word is a part of the name of a place (e.g. “Piccadilly” in “Piccadilly Circus”) or part of the name of an actor (e.g. “Loren” in “Sofia Loren”); if the word has a match with the name of an actor, the search is extended including all the characters associated to that actor. In the case of a multi word search string, the whole string is first searched for an exact match in the tables TAG and PERSONA and as an approximate match in POI, then the string is tokenized and a single word search is started on each token.

The search is performed as case insensitive and a classical stop-list is used to remove parts of speech (pronouns, articles and prepositions) from the computation.
The output of a search is a list of clips to watch; if searching for a director’s name, a link to a description and filmography of the director is added to the results.

To order the results a score is assigned to each of them: the exact matches have full score; approximate matches are scored by a custom function that measures the similarity between the word found and the word searched for; for matches in tables other than PERSONA, POI and TAG, a value combining *tf* and *idf* is computed. Tokens have a lower score (computed as a fraction), as they represent a partial match; however if different tokens point to the same element, their score is increased.

To rank the clips in the search result, the system takes also into account the scores assigned to them by the film experts (stored in the table MOI) and the ratings in the social media. In this way retrieved items are sorted also according to their quality (scores by scholars) and their popularity (ratings by the public).

The search function uses query expansion to widen the number of results found; query expansion is used only to search in the TAG table and was implemented using MultiWordNet [13], a free Italian version of WordNet. The search uses an auxiliary table containing each element of the TAG table plus the synsets related to it found iterating recursively on the lexical database. This table is populated

automatically and it is updated every time the content of the database changes (e.g. a new movie is added); its contents however can be also manually checked and modified using a custom control panel that lets the administrator decide which synsets store in the table for a given tag.

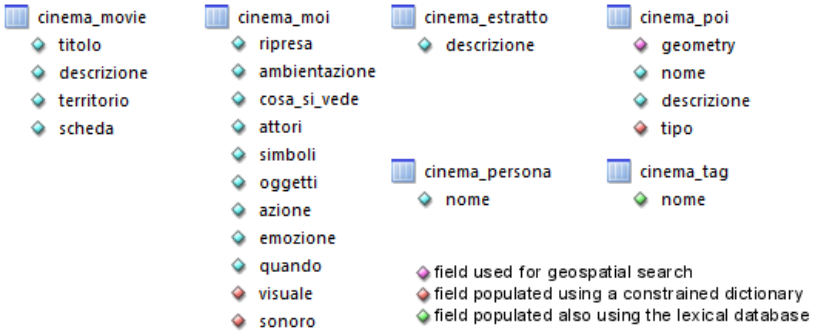

Figure 2: Attribute list used by the search engine.

To make the user experience more varied and to invite the user to explore our database and discover new things, we developed a recommendation system. Once the user has selected a first content object (e.g. clicking on a search result, or on one of the predefined path), the system generates a list of related content that is proposed to the user to continue his exploration. As with the search results, recommendations are mostly a link to another clip to watch. To produce this list the system searches which contents have the same tags of the present one; this relies on the same table used for the query expansion, containing all the tags and their related synsets. Given a clip currently being watched by the user, the recommendations might include clips shots in the same place or in the surroundings, clips with the same actor or from the same movie, clips conveying the same emotions or showing the same object. In a way similar to the search results, recommendations are ranked and ordered; however recommendations are also filtered and limited in numbers, with the filter trying to propose an heterogeneous list of different choices (e.g. different movies, different places).

Recommendations are ranked only on their scores and ratings; in the case the user is accessing the system from a mobile device and his location is known, they are also ordered by the geographical distance to the user.

To improve the recommendations, we also implemented some user customization. Customization is used to modify the list of recommendations in order to make them fit more the user's interests; as such it is triggered once the user has chosen a first content. Inspired by the paper [7] we decided to model our users in three profiles:

- Profile A: the user is more interested in the touristic aspect of his visit, i.e. he/she likes to explore the geographic space;
- Profile B: the user is more interested in the cinematographic aspect, i.e. he/she likes to explore a movie;

- Profile C: the user is in-between profile A and B, with no special preferences, he/she is simply attracted from what is good, interesting.

Whenever a user accesses the system, the software proposes a list of recommendations that is balanced, i.e. containing an equal number of clips that are set far from each other (profile A: exploring the territory), clips that belong to the same movie (profile B), clips that have high ratings (profile C); once the user starts to explore the recommendations, the system records which choices have been chosen and biases later recommendation lists to contain more options of the type the user seems more interested in. At present, the user cannot manually adjust his/her profile although this is an important aspect that will be taken into account in future development.

Table 2. The profiles used to improve the recommendation system.

Profile	Interested in moving around the territory	Interested in following the same movie	Interested in watching high rated contents
A	Yes	No	No
B	No	Yes	No
C	No	No	Yes

The system was developed and is currently deployed as a closed beta. The architecture consists of a database to store and query the data, a storage for the video contents, both the movies to annotate and the clips, a server with a backend interface to annotate the movies, insert data and manage the system, and a frontend interface to browse and explore the data, aimed mainly to the tourist user group.

The server was developed in PHP on a hosted Apache web server. The server contains an administration panel, to manage the system and a backend panel to populate the database; at the moment the database has been populated by a pool of experts in film studies using, where possible, controlled dictionaries for tags and places. The server also runs the code for searching and query expansion.

The web server exposes a list of RESTful APIs that enables other systems to interact with the system and query the database. This feature is aimed mainly to professionals (e.g. tour operator accessing the system using a Destination Management System -DMS) but resulted also very useful while developing the user interface.

The database was implemented in MySQL; this was almost a direct consequence of our choice to use Apache, as it can be hard to find a web host offering a different DBMS. Due to MySQL basic support of the SFA/SQL [SS12] standard, we developed a custom PHP library to complement the limited functionalities of the MySQL Spatial Extensions, implementing the missing spatial operations required to handle our geo-referenced data. The database was implemented following the model reported above (see Figure 1) and populated using the backend web interface; some auxiliary tables have been added to the DBMS to speed up the search of keywords and to cache the results of the matching functions. The DBMS also contains the lexical database used to perform query expansion.

To store the video content, we needed a solution providing safety (as the contents are copyright protected materials) and, in perspective, performance, scalability and low maintenance costs. Our choice was to store the data in the cloud using Amazon Web Services (AWS); in particular we chose the Simple Storage Service (Amazon S3) due to its easy use and cheap pricing, compared to ad hoc video hosting solutions (e.g. Vimeo PRO). We converted all the video in a format supported by HTML5 (H264 video codec and AAC audio codec), and stored them as MP4 files, that are then streamed via http; to keep the location of the files secure, a PHP proxy is used to relay the video streams to the client device. The clips are automatically generated from the annotated source files; each clip, resized at 480p resolution and compressed, is less than 2MB per 10 seconds, while a whole movie varies between 500 and 700 MB in size.

The user interface is implemented as a website developed to be accessible by computers, tablets and mobile devices: this choice looked favorable to the development of specialized interfaces for these different platforms; the website uses Bootstrap to be responsive and jQuery to load contents progressively to reduce the amount of data sent to the user; we use Google Maps as the backdrop maps for the geographic contents, but the system can easily switch to other map services; the video contents are provided using the capabilities of HTML5, thus leaving to the users' clients the choice of the video player.

As a design choice, the user interface is completely isolated from the server, exchanging data with it only through the APIs; the website is implemented in PHP executed on a different Apache web server; this server also runs the code for user profiling and recommendation generation.

The website can be used both by logged and not logged users; logged users can access some specialized content (if enabled by the administrator); all users' activity is tracked and the choices made (i.e., which of the proposed contents is read first) are used to associate the user to one of the different user profiles described above.

Another important feature of the GUI is the integration with the social media. Users can comment, "share" and "like" most of the contents of the website; user contribution increase the engagement of the users in the system and enriches it; the "like" are also used to create a user-based ranking of the content.

Inspired by the theory of gamification, the GUI offers also a mini-game that asks the users to recognize a clip or to position on the map the location of a clip; the contents of the mini-game is automatically generated from the data available. We tested the effectiveness of a gamified approach by making available the system at a booth during the European Researchers' Night 2015. Users enjoyed to play with movies and their city and reported that it was a novel and fun way to discover new places. Additionally, participating to the game raised interest towards the movies themselves, because most participants were not aware that they were located in Padua.

6 Conclusions and Future Work

The system developed is currently under test. The user interface, with the functionalities of search, recommendation and personalization offers an easy access to the information of the database and also a simple way to discover new contents. A screenshot of the interface is shown in Figure 3. Offering predefined paths, as well as the use of social media, are a good combination to invite the user of the group one to start exploring the system and all this together seems able to offer the engaging experience we were looking for.

Recommendations, that are mainly thought for the users of the group one, can also be useful for the users of the group three, as they can help them to discover similar places; furthermore, the use of controlled dictionaries in the annotation of some specific features of the movies, further improves the ability of our system to provide, through the search function, the data that this group of user may need.

The API developed allow to communicate with the system and they seem complete enough to cover the various needs to query the database that the user of the group two might have.

The use of a lexical database to perform query expansion proved to be more difficult than expected: the expansion (especially if iterated) may lead to unexpected results, and we had to develop a custom tool to let the administrator check which synsets are inserted in the database.

Future developments will regard a function to track the content already seen by the user, in order to avoid duplicate proposals, and a better integration with social media.

Acknowledgments. The work reported has been partially funded by the European Commission under the FSE (Fondo Sociale Europeo - European Social Fund) initiative. The authors thanks the colleagues in film study, Farah Polato and Giulia Lavarone, who collaborated to the system design and coordinated the population of the database

References

1. Beeton S.: Film-Induced Tourism, Channel View Publications, Clevedon (UK), 2005, p.11.
2. Di Cesare F., Rech G.: Le produzioni cinematografiche, il turismo, il territorio, Carocci, Roma, 2007, pp. 45-88.
3. Provenzano R.: Al cinema con la valigia. I film di viaggio e il cineturismo, FrancoAngeli, Milano, 2007, pp. 256-283.
4. Buhalis D., Law R.: Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research, Tourism Management, Volume 29, Issue 4, 2008, pp 609-623.
5. Joliveau T.: Connecting Real and Imaginary Places through Geospatial Technologies: Examples from Set-jetting and Art-oriented Tourism. The Cartographic Journal 46(1), 2009, pp. 36-45.
6. Lavarone G., Orio N., Polato F., Savino S.: Modeling the concept of movie in a software architecture for film-induced tourism. In Proc. of the 11th Italian Research Conference on Digital Libraries, 2015.
7. Veron E., Levasseur M.: Ethnographie de l'Exposition. Paris : Bibliothèque publique d'information : Centre Georges Pompidou, 1989.

8. Chittaro L., Ieronutti L.: A visual tool for tracing users' behavior in Virtual Environments. In Proc. of the working conference on Advanced Visual Interfaces, 2004, pp. 40-47.
9. Antoniou A., Lepouras G.: Modeling visitors' profiles: A study to investigate adaptation aspects for museum learning technologies. Journal of Computers for Cultural Heritage 3(2), 2010, electronic publication (19 pages).
10. Elgammal A.: Human-centered multimedia: Representations and challenges. In Proc. of ACM Workshop on Human-Centered Multimedia, 2006, pp. 11–18.
11. Kim Y.-M.: Social Tags in Text and Image Search. In Proc. of the Symposium on Information Interaction in Context, 2010, pp. 353-358.
12. M. Voorhees E.M.: Using WordNet to disambiguate word senses for text retrieval. In Proc. ACM SIGIR conference, 1993, pp. 171-180.
13. Pianta E., Bentivogli L., Girardi C.: MultiWordNet: developing an aligned multilingual database. In Proc. of the First International Conference on Global WordNet, 2002, pp. 293-302.

Figure 3: Screen capture of the user interface.