The retrieval of moving images at Spanish film archives: the oversight of content analysis

Prof. Dr. Rubén Domínguez-Delgado
University of Seville
Avenida Américo Vespucio, s/n,
41092 Seville (Spain)
rdd@us.es

Prof. Dr. María-Ángeles López Hernández
University of Seville
Avenida Américo Vespucio, s/n,
41092 Seville (Spain)
alhernan@us.es

ABSTRACT

In the field of library and information science, content analysis is a crucial task for an effective retrieval of information by users at an archive. In this research, we analyse the current state of this task and the possibilities of retrieval of film information at six Spanish important film archives, interviewing their responsible librarians for film cataloguing and comparing the content fields inside their film cataloguing records used by each one of these six organizations.

Keywords

Moving-images librarianship / film librarianship / film archives / retrieval of film information / film content analysis / selective retrieval of information

INTRODUCTION

It is evident that doing an appropriate content crucial importance of doing an appropriate content analysis is of crucial importance for enabling effective search and retrieval of information that can satisfy the information needs of users at any moving-image archive. In the current century, significant advances have been produced in Spain about this task, inside the television branch of moving image librarianship, with bibliographic production about it (López-Hernández, 2003; Hidalgo-Goyanes, 2005; Agirreazaldegi, 2007; López-de-Quintana, 2007; De-Bustos-Pérez-de-Salcedo, 2007; Peñafiel & López, 2009; Caldera-Serrano & Sánchez-Jiménez, 2009; Aguilar-Gutiérrez & López-de-Solís, 2010; Caridad-Sebastián *et al*, 2011) and with revolutionary analytical methods, as a result of the deep transformation which supposed, from 2004, the application of the new media digital asset management in the main television newsrooms (Agirreazaldegi, 2007, p.

434).

In spite of the differences in the databases used by each television archive, we can affirm the selective retrieval of information is possible today in the main Spanish television archives. According to Caldera-Serrano & Sánchez-Jiménez (2009, p. 299), this type of retrieval of information "has been always one of the most wanted purposes by engineers and librarians who produced software for moving-image archives, making shorter the response time and making more pertinent the retrieval of information". And this means evident advantages for the users who visit these archives.

But drawing a parallelism between the television and the film branches of moving-image librarianship, we wonder if this worrying about an effective retrieval of moving images in television archives to provide the best possible access to information – in consonance with 1980 UNESCO recommendations (Unesco, 1980, p. 169) –, matches the current reality in the other important moving-image archives: film archives.

OBJECTIVES AND METHODOLOGY

Due to the scientific hole about this topic, we decided to do this research with the main purpose of knowing the current state of the possibilities of film retrieval of information by users in six important Spanish film archives, as well as the current state of the film content analysis they perform. These six film archives are: *Filmoteca Española*, *Filmoteca de Andalucía*, *Filmoteca de Catalunya*, *Filmoteca Vasca*, *Filmoteca de la Generalitat Valenciana* and *Centro Galego de Artes da Imaxe*.

To reach our goal, we have developed two methodological ways. Firstly, we have done a comparative study of the content fields inside the most exhaustive film cataloguing index cards – or film cataloguing records – which librarians of these six film archives use on their databases (taking into consideration that each archive use different kinds of records for different levels of analytical exhaustiveness, depending on the film collections). The purpose of this is to know if it exists or not a general worry for the film content during the design of these cataloguing index cards, as well

as if there is a consensus or heterogeneity in the film content analysis in Spain.

Due to the different names received in each film archive by the content fields inside the cataloguing records – for the same aspects of the films –, we have grouped all the content categories of the index cards in two big groups, which coincide with the two tasks that a content analysis implies: synopsis and indexing (Pinto-Molina, 1992, p. 89; Fox, 2005, p. 23). And inside the second task, we have grouped the content fields of the cataloguing records in four indexing categories: thematic descriptors, names descriptors, geographic descriptors and time descriptors (Caldera-Serrano y Sánchez-Jiménez, 2009, p. 293).

Secondly, we have interviewed to the librarians responsible for film analysis in the six selected film archives. To homogenize the getting of information, we have designed a survey with 5 questions (Table 1).

Question 1.- What are the fields which users use the most to search for film information on the database of your archive?
Question 2.- Do you analyse the content of films in your archive?
Question 3.- If you perform a film content analysis, what type of methodology do you apply (shot by shot, scene by scene, sequence by sequence…) and what type of films do you analyse using this methodology (fiction, non-fiction, newsreels…)?
Question 4.- Have you ever considered to apply a film content analysis methodology which let the users to search shots, scenes or sequences about a concrete content, event or epoch? ¿Do you think it would be interesting and viable to apply it currently in your archive?
Question 5.- Do you have a thesaurus or controlled vocabulary to describe the contents of films in your archive?.

Table 1. Survey for film archives.

RESULTS

The attention to film content in cataloguing index cards

After the study of the six cataloguing index cards from film databases, we observe the presence of the field synopsis in all of them, as we can see on Table 2.

Regarding the second important task about content analysis – indexing –, we have to emphasize that, except in the case of Filmoteca de Catalunya, it does not exist an specific area on the cataloguing index cards where to include all the assigned descriptors. Therefore we find throughout the cataloguing records, in a dispersed way, key concepts to search and retrieve the film information, grouped, besides, in fields with different names.

On the other hand, and as we can see on table 2, theme is the most important aspect in the task of indexing for these six film archives, considering that all of them include a specific field for the thematic descriptors which represent the content of the moving images. However, the rest of indexing categories – names descriptors, geographic or places descriptors and time descriptors – are omitted by some of the archives on their film cataloguing records.

Film Archive	Synopsis	Thematic indexing	Names indexing	Geographic indexing	Time indexing
Filmoteca Española	x	x			
Filmoteca de Andalucía	x	x		x	x
Filmoteca de Catalunya	x	x	x	x	
Filmoteca de la Generalitat Valenciana	x	x	x	x	
Filmoteca Vasca	x	x			
Centro Galego de Artes da Imaxe	x	x			

Table 2. Presence of aspects about film content on cataloguing index cards.

Concerning the names descriptors, it is striking that we can only find specific fields where to include people represented in films in two film archives – Filmoteca de Catalunya and Filmoteca de la Generalitat Valenciana –, although the first one assigns also on this field the names of organizations and not only people. In the rest of cards, we only find a field where to include the names of actors and actresses, but not the names of the characters they perform in the films. With these techniques of analysis, it is evident that a user of these film archives cannot search and retrieve concrete shots, scenes or even whole movies about any film character.

Concerning to the time indexing, there is not either a specific field for the dates or years represented in movies. In most cases, the cataloguing records only include the year of production of the film and, only occasionally, on the

synopsis, the year represented on the film. Only Filmoteca de Andalucía includes this specific content field on its cataloguing records. Under the circumstances, something so basic as searching and retrieval of information based on times represented on films is very difficult for users.

Finally, concerning to geographic indexing, only two film archives – Filmoteca de Catalunya and Filmoteca de la Generalitat Valenciana – assign geographic descriptors on their film cataloguing records. The other four film archives only include a formal field for the places where the films were shot, but not for the places which are represented on the films.

We must also say that none of the six archives distinguish on their film cataloguing records between viewed descriptors and heard descriptors, as Caldera-Serrano & Sánchez-Jiménez (2009, p. 296) recommend for an appropriate moving-image content analysis. Under the circumstances, the users who go to these archives do not know – after their searching and retrieval of film information and without viewing the results provided by the database – if the concept they are looking for can be viewed on the moving images or it is only heard.

In conclusion, there are many differences and deficiencies in the film content analysis performed by these six important Spanish film archives and also in the possibilities of retrieval of moving images by their users.

Film content analysis and retrieval of information according to film archives

After the interviews to librarians of the six film archives, all of them claim, as an answer to question 2 (Table 1), that they perform a content analysis of their movies.

However, as an answer to question 3, about the application of content analysis methodologies on shots, scenes or sequences, we can observe some differences. Alghough none of the six film archives perform this type of accurate or detailed analysis on all their films in a systematic way, four of them – Filmoteca Española, Filmoteca de Catalunya, Filmoteca de la Generalitat Valenciana and Centro Galego de Artes da Imaxe – claim that they perform this type of accurate analysis, sequence by sequence or scene by scene, on specific film collections throughout the whole film – the three last archives even shot by shot. However, except in the cases of Filmoteca Española and Filmoteca de Catalunya, they could not provide us any example of these content analysis. That is why we think that even the concept of accurate film content analysis is not very clear for the Spanish film archives.

Filmoteca de Andalucía is the only archive which always analyse the content of films considering each whole movie as the only unit, without describing the content of shots, scenes or sequences.

As there is not an unanimous application of accurate film content analysis methodologies on unities inferiors to the whole movie in the six selected film archives, we have asked the two film archives which do not apply it if they have ever considered this application. As an answer, Filmoteca de Andalucía has considered it and Filmoteca Vasca – which perform this type of analysis only on some specific shots of some film collections, but not throughout whole films – has not considered it.

We have also asked the six film archives – question 4 – if they consider interesting and useful for their users, on the one hand, and viable, on the other hand, the application of a scientific methodology of accurate film content analysis. All of them answered yes to the first question and only two of them do not consider viable to do it, because of the lack of human, economic and technical resources at the film archives.

Concerning to the way users who visit these six film archives search for moving images – and as an answer to question 1 – they mostly search using the field of the database *date* or *year* – option indicated by four of the six film archives –, followed by the fields *title*, *director*, *producer* or *actor/actress* – options indicated each one by three film archives. It is obvious that if the content of films is not systematically analysed, the options to search and retrieve film information are reduced or limited to the formal fields.

Other ways of searching and retrieval indicated by archives were by means of the fields *location of shot* – option indicated by two film archives –, *gender*, *theme* and *collection* – options indicated each one by one film archive.

Finally, concerning to the use of a thesaurus at the organizations, which would let assign homogeneous descriptors during the content analysis and search for information using a controlled vocabulary – question 5 –, only four of the six film archives have this important tool in librarianship – Filmoteca de Catalunya, Filmoteca Española, Filmoteca de Andalucía and Filmoteca de la Generalitat Valenciana –, although the thesaurus used by the last archive includes only formal or technical terms, but not terms about film content. Therefore, fifty per cent of the interviewed film archives do not have a thesaurus with terms about moving images content – Filmoteca Vasca, Filmoteca de la Generalitat Valenciana and Centro Galego de Artes da Imaxe –, what supposes some problems or disadvantages – synonymy or polysemy, for example – to search and retrieve film information by users.

CONCLUSIONS

Although it is certain that in the last decades we have significantly progressed in the preservationist aspect of film archives, we have not achieved the same in film content analysis and in the possibilities of retrieval of film

information at these archives by users. This field actually shows in Spain many deficiencies and it is characterized by the heterogeneity of criterions and models, as well as by the non systematical application of accurate film content analysis methodologies which let users a selective retrieval of information.

But we think it is fair and necessary, in the middle of the technological and digital era, that users of film archives have the same privileges as users of television archives in the searching and selective retrieval of moving images. This would mean a great satisfaction, time-saving and work-saving to find what they look for, something very important in a world dominated by the informative saturation and new technologies. If this does not happen, we take the risk this difficulty of finding film contents distance users of moving images from film archives, putting at risk the interest of citizens and researchers in this valuable cultural heritage.

That is why we require researchers open new lines of research about this topic, maintaining a continuous contact among researchers in the field of moving image librarianship, librarians who work in film archives and users who visit these archives, so we can make film heritage more accessible for citizens and we can also promote the use of film moving images in research.

REFERENCES

Agirreazaldegi, T. (2007). Claves y retos de la documentación digital en televisión". In *El profesional de la información*, vol. 16, n. 5 (pp. 433-442).

Aguilar-Gutiérrez, M. & López-De-Solís, I. (2010). Nuevos modos de trabajo de una redacción digital integrada: el caso de los servicios informativos de TVE. In *El profesional de la información*, v. 19, n. 4 (pp. 395-403).

Caldera-Serrano, J. & Sánchez-Jiménez, R. (2009). Recuperación de secuencias de información audiovisual con rdf y smil. In *El profesional de la información*, v. 18, n. 3 (pp. 291-299).

Caridad-Sebastián, M.; Hernández-Pérez, T.; Rodríguez-Mateos, D. & Pérez-Lorenzo, B. (2011). *Documentación audiovisual. Nuevas tendencias en el entorno digital*. Madrid: Síntesis.

De-Bustos-Pérez-de-Salcedo, P. (2007). Sistemas integrados y gestión documental. La experiencia en Telecinco. In *El profesional de la información*, v. 16, n. 5 (pp. 450-455).

Fox, V. (2005). *Análisis documental de contenido. Principios y prácticas*. Ciudad Autónoma de Buenos Aires: Alfagrama Ediciones.

Hidalgo-Goyanes, P. (2005). La documentación audiovisual en las televisiones. La problemática actual y el reto de la digitalización. In *Documentación de las ciencias de la información*, v. 28 (pp. 159-171). http://revistas.ucm.es/index.php/DCIN/article/view/DCIN0505110159A/19216

Lauwers, M. (2004). *Changing sceneries, changing roles: media management in the digital era*. Hilversum: nederlands instituut voor beeld en geluid.

López-de-Quintana, E. (2007). Transición y tendencias de la documentación en televisión: digitalización y nuevo mercado audiovisual. In *El profesional de la información*, v. 16, n. 5 (pp. 397-408).

López-Hernández, A. (2003). *Introducción a la documentación audiovisual*. Carmona: S&C ediciones.

Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (1980). Recomendaciones para la Salvaguardia y Conservación de Imágenes en Movimiento. In *Actas de la Conferencia General. 21ª Reunión. 23 de septiembre – 28 de octubre de 1980. Volumen 1. Resoluciones*. París: Unesco (pp. 167-172). http://unesdoc.unesco.org/images/0011/001140/114029s.pdf

Pinto-Molina, M. (1992). *El resumen documental. Principios y métodos*. Madrid: Fundación Germán Sánchez Ruipérez y Ediciones Pirámide.

Pinto-Molina, M.; García-Marco, F.J. y Agustín-Lacruz, M.C. (2002). *Indización y resumen de documentos digitales y multimedia. Técnicas y procedimientos*. Gijón: Ediciones Trea.

Rodríguez-Mateos, D. (2011). Nuevas tecnologías para la gestión de la documentación audiovisual. In Caridad Sebastián, M.; Hernández Pérez, T.; Rodríguez Mateos, D. y Pérez Lorenzo, B.. *Documentación audiovisual. Nuevas tendencias en el entorno digital*. Madrid: Síntesis (pp. 67-86).

Rodríguez-Mateos, D. & Pérez-Lorenzo, B. (2011). Análisis documental de contenidos audiovisuales. In Caridad Sebastián, M.; Hernández Pérez, T.; Rodríguez Mateos, D. y Pérez Lorenzo, B. *Documentación audiovisual. Nuevas tendencias en el entorno digital*. Madrid: Síntesis (pp. 111-134).