# User Requirements for Multimedia Indexing and Retrieval of Unedited Audio-Visual Footage - RUSHES

O. Schreer[1], L. Fuentes Ardeo[2], D. Sotiriou[3], A. Sadka[4], E. Izquierdo[5]

[1]*Fraunhofer Institute for Telecommunications/Heinrich-Hertz-Institut, Germany,* [2]*Euskal Telebista-Televisión Vasca S.A., Spain,* [3]*Athens Technology Centre, Greece,* [4]*Brunel University, UK,* [5]*Queen Mary, University of London, UK*
*Oliver.Schreer@hhi.fraunhofer.de, fuentes_leticia@eitb.com, d.sotiriou@atc.gr, Abdul.Sadka@brunel.ac.uk, ebroul.izquierdo@elec.qmul.ac.uk*

## Abstract

*Multimedia analysis and reuse of raw un-edited audio visual content known as rushes is gaining acceptance by a large number of research labs and companies. A set of research projects are considering multimedia indexing, annotation, search and retrieval in the context of European funded research, but only the FP6 project RUSHES is focusing on automatic semantic annotation, indexing and retrieval of raw and un-edited audio-visual content. Even professional content creators and providers as well as home-users are dealing with this type of content and therefore novel technologies for semantic search and retrieval are required. As a first result of this project, the user requirements and possible user-scenarios are presented in this paper. These results lay down the foundation for the research and development of a multimedia search engine particularly dedicated to the specific needs of the users and the content.*

## 1. Introduction

Even in the area of professional content creation and provision as well as for home-users a tremendous amount of audio-visual data has been and will be produced. This type of content known as rushes has some features which differ significantly from general multimedia content. To handle this amount of un-edited data novel approaches are required to annotate, index as well as search and retrieve. A block diagram of a complete rushes search engine might look as depicted in Fig.1.

The first task for successful research and development is to analyze the requirements by the users. A fully user driven development will increase the acceptance and potential application of such a
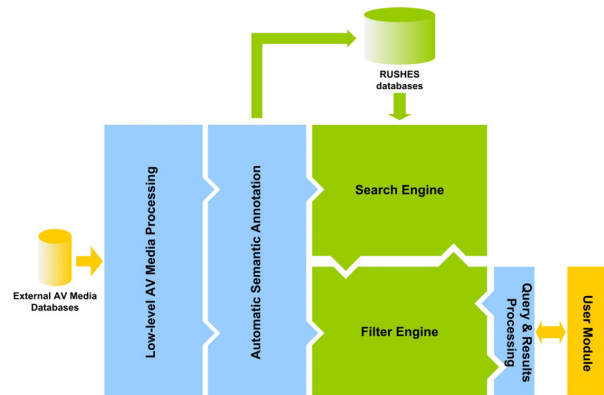


Fig. 1 Block-diagram of a RUSHES search engine

search and retrieval engine. The European FP6 funded research project RUSHES aims to design, implement, and validate through trial a system for indexing, accessing and delivering raw, unedited audio-visual footage, known as rushes, and promote the reuse of such content in the production of new multimedia assets by offering semantic media search capabilities. In this paper we are presenting first results regarding the user requirements for home-users and professional content creators and providers. In the next section we briefly derive the main characteristics of rushes content compared to general purpose multimedia data. In section 3 and 4 the results of the user evaluation are presented. The paper ends with a concluding summary.

## 2. Specific features of rushes

Raw un-edited audio-visual content known as rushes shares many features with general-purpose multimedia data, but it also shows special characteristics:

- Camera setting (camera distance, focus, angle, motion, etc), scene composition and production features (e.g. light) may be a more important retrieval concept than in general media (especially for professionals).

- Shot-boundary detection is not enough as a segment decomposition tool (since most rushes are single-shot sequences anyway), and sub-shot analysis is needed.

- The soundtrack is frequently irrelevant (or inexistent): the soundtracks of stock shots are usually not used (or used just as background sound). So semantic extraction must often relay purely on visual features. Nevertheless in some scenarios, audio/speech may provide helpful information for multi-modal annotation.

- Material shot at one location is usually repetitive; special care must be taken when analysing and summarising to detect redundancies.

- Annotations for rushes, when present, are sparse and correspond only to general categories: some keywords (usually production metadata), geographical location, and shooting date. Moreover, they are usually applied to undifferentiated clusters of raw material.

- No story analysis or subject discovery, or domain-based analysis tool can be appropriately used, since there is usually no storyboard, and fragments are isolated.

These features create special needs when structuring raw material with the purpose of making it available for reuse. Therefore state-of-the-art technology on semantic-based analysis can be used, but it must be extended and adapted as necessary to be able to process raw data.

# 3. Analysis of user requirements

To gain acceptance of potential end-users, their requirements must be understood in order to develop a user driven multimedia indexing, annotation, search and retrieval engine. Hence, the first task is to identify the needs and requirements of professional content creators and providers as well as home-users. The approach to receive latest information from user perspective was two-fold.

For the home-users a detailed questionnaire has been developed within the project. This questionnaire contains a comprehensive set of questions in order to catch in an objective way their requirements. The questionnaire includes a combination of questions regarding the general behaviour of the users with respect to search engines in general and also some specific questions regarding searches for audiovisual material, i.e. how these queries could be made and how users prefer to view results and so on. The type of questions and the resulting feedback is presented in section 3.1.

For the second group of users, the professional content creators and providers, we were able to deploy the direct access to people working at broadcasters. One RUSHES partner, Euskal Telebista Televisión Vasca S.A. (ETB), held meetings with different professionals belonging to this communication enterprise. This group of several professionals has been composed by users with different profiles such as people working in the archive, journalist department, and production. The outcome of the interviews is summarized in section 3.2.

## 3.1 User requirements for home-users

In order to capture the user requirements for RUSHES and to improve the understanding of the user behaviour and expectations with regards to future multimedia search engines for un-edited raw material, we have prepared and distributed a questionnaire. Once the questionnaires were filled, we analysed the results and came up with a number of basic requirements from the user side. The full questionnaire can be found in [1]. All partners have been requested to distribute the questionnaire within their institutions via email. The sample size of our questionnaire was 72 persons. The results are as follows:

The first group of questions was related to the analysis of our sample from a demographic point of view. The first three questions are related to the age, education and work experience of the participants and allow us to see if our sample is representative for the target audience. It has been recognized, that the requested persons are setting a representative sample for this survey. The distribution among age, education level and work experience was equally distributed.

The next set of questions tried to capture the users' behavior with respect to the usage of search engines, not explicitly tied to the search for audiovisual content but general queries for information. The main objective was to stress the general trend in search engines more than to investigate the requirements from audiovisual searches. The results are as follows:

- 93% of the users apply search engines several times a day

- The frequency of usage of search engines for photos or videos is much more different ranging form several times a day (26%) to 3-5 times a week (22%) or every few weeks (19%), the remaining percentages lie in between.

- Two third of the users prefer a single search engine

- Google is the most prominent application (90%) [2]

- Almost all users think that similarity search is very or probably helpful

- Most users considered text-based semantic search to be either necessary or useful.

- Regarding images, they are considered mostly as useful rather than necessary, which can be explained due to current user habits of receiving only text results and as thus considering image as beneficial but not mandatory.

- Video is even considered more useful than text and images, but not mandatory, again probably because only a few sites offer video results (i.e. YouTube, MySpace etc.) [3][4][5]

- Almost 90% consider the results received from search queries as accurate most of the time

- Regarding trustworthiness of the users to the results and information they receive. 78% consider them trusted services but have their concerns with regards to their fairness

In the next set of questions we investigated the willingness of users to receive information regarding audiovisual searches in different forms.

- Almost 80% of the users are searching for public photos or videos sometime or hardly ever.

- The frequency of search for public photos or videos from internet news sites is in the same order of magnitude

- the users are searching regularly for images, occasionally for videos and rarely for sounds

- concerning keyword search the users perceives as important filenames, description words, location/place and year as either important or a must

- 2/3 of the users prefer to have accurate results (quality) over fast responses

- 80% see the possibility of receiving results in thematic groups as very helpful or probably helpful

- 90% see a presentation of results in interactive maps as very/probably helpful

- 2/3 of the users prefer a browsing through the results in category order, the remaining third prefers a browsing by popularity order

By summarizing our findings, a novel multimedia indexing and retrieval system for rushes content must offer the following functionalities:

- Users are able to perform similarity searches in content-based query

- Users receive information in image and video format

- The application allows searching especially for image and video content

- The application allows searching through semantic keywords, especially location/place, people/ activities, time and descriptive words

- The application should aim for the accuracy of the results as opposed to the performance

- Results to be presented as thematic groups (blogs, rumours, official announcements) and if possible in interactive maps

- Users are able to browse results in category order and then in popularity

- The application displays all results with list and preview method

- The application displays individual results in a mix of preview, key frames and details

## 3.2 User requirements for professional content creators and providers

In this context, the professional user has been defined as the user in broadcaster environment or as an external producer. Therefore, a professional user is said to be the user who creates or provides the content that is usually broadcasted. Usually, people have specific motivations associated with their work environment and their activities. Their behaviour is typically goal-driven and users have a clear understanding of what kind of application they are willing to use to achieve their scopes. Hence, it is important to understand their motivation and the context within which they operate in order to design suitable applications for the final users.

User requirements have been gathered through interviews with different relevant user groups within

ETB. These interviews started with an exploration of users' current habits followed by a discussion session to detect needs and the possible potential of the incoming technology. These interviews were performed in different departments: documentalist, journalist and external producers. Five documentalists have been interviewed, whereas 11 journalists joined the survey. In the external producer group, 6 colleagues from two different producers participated.

The analysis of the interviews shows that users wish to have an application which annotates clips automatically. Currently the users apply a manual form of annotation, and this requires most of their time and effort. In the current situation it is the most relevant drawback and the raw material can not be used until someone performs the required annotation. This fact leads to a delay in the use of this raw material. This requirement is not of relevance in terms of real-time processing, but it is necessary to avoid that the material is waiting for a free time slot of a documentalist.

The specific requirements resulting from the interviews for a future annotation tool are as follows:

- The users wish to apply the same annotation to a number of different clips.

- The same word to annotate some content should belong to the same family.

- The user should have the choice of accepting or rejecting the automatic annotation that has been generated.

- Users would like to search using 'person', 'date' and/or 'time'.

- The user will search for further clips based on a set of pre-selected images, by selecting one image and requesting the application to search for clips that contain images similar to the selected one (query by example).

- The users wish to search for specific visual content, as for example: landscape, cityscape, outdoors, indoors, sunset, etc.

- A search by keywords is requested.

- Search capability for clips containing people or not.

- The search result should be presented by thumbnails, which are preferably key frames resulting from summarization.

The initial user requirements gathering activity revealed some very specific user requirements that should be used to focus the technology development of future rushes search engines. There will be a notable change through the incorporation of a novel multimedia indexing, annotation and retrieval engine for a broadcaster environment. The increase of the performance of the workflow will be mainly achieved by significantly improved automatic semantic annotation. By using a novel multimedia indexing, annotation and retrieval engine, the capability of automatic annotation facilitates the work of the documentation department, as they could be in charge of reviewing the automatic annotation, accepting or rejecting it, and making the suitable changes. The annotation can be performed more efficiently by saving time and costs, as the first annotation can be made automatically and the documentalist can use his work time to improve other content annotation.

## 4. Conclusion

The experience as application developers and application users has been a tremendous gain to the community in the sense that it enables researchers and developers to take into account first and foremost the user's needs. The analysis of a questionnaire resulted in a lot of requirements defined by home users and based on their current habits. The professionals analysed their daily workflow and derived requirements for a novel annotation tool. Since both sets of requirements are carefully considered during development of a novel rushes indexing and search engine, the user acceptance and usability of future search engines will be increased and the actual workflow of professionals will be improved, facilitated and accelerated remarkably.

## 5. Acknowledgement

## 6. References

[1] RUSHES project deliverable D5 "Requirement Analysis and Use-Cases Definition for Professional Content Creators or Providers and Home-Users", http://www.rushes-project.eu/upload/Deliverables/D5_WP1_ETB_v04.pdf, August 2007.

[2] http://video.google.com/

[3] http://www.google.com/

[4] http://www.youtube.com/

[5] http://www.myspace.com/