SINAI at VideoCLEF 2008

José M. Perea-Ortega, Arturo Montejo-Ráez, M. Teresa Martín-Valdivia,
Manuel C. Díaz-Galiano, L. Alfonso Ureña-López
SINAI Research Group. Computer Science Department. University of Jaén
Campus Las Lagunillas, Ed. A3, E-23071, Jaén, Spain
{jmperea,amontejo,maite,mcdiaz,laurena}@ujaen.es

Abstract

This paper describes the first participation of the SINAI research group in the VideoCLEF 2008 track. We have only submitted runs for the classification task on Dutch and English languages. Our approach has consisted in the use of a particular Information Retrieval system as classification architecture, using the speech transcriptions as textual queries and generating textual corpus for each topic class. In order to generate this textual corpus we have used the Google[1] search engine. The experiments show that an IR system can perform well as classifier of multilingual videos, using their speech transcriptions and obtaining good results.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.4 Systems and Software

General Terms

Measurement, Performance, Experimentation

Keywords

Content-based retrieval, Multilingual video classification, Information Retrieval, VideoCLEF

1 Introduction

This paper describes the first participation of the SINAI[2] research group from University of Jaén in the VideoCLEF 2008 campaign. VideoCLEF is a new track for Cross Language Evaluation Forum (CLEF) 2008 and it aims to develop and evaluate tasks in processing video content in a multilingual environment. The overall objective is to combine and extend these subtasks and perform completely automatic generation of RSS feeds specific to a particular information need and personalized to a particular language preference. The main task for VideoCLEF 2008 involves assigning topic class labels to videos of television episodes[3].

The aim of our first participation in VideoCLEF has been the study of the problem of this task, and the development of a basic architecture which approaches it. We have some experience in the field of multimedia video retrieval [1] and in image retrieval [2, 3, 4].

[1] http://www.google.com
[2] http://sinai.ujaen.es
[3] http://ilps.science.uva.nl/Vid2RSS/Vid2RSS08/Vid2RSS08.html

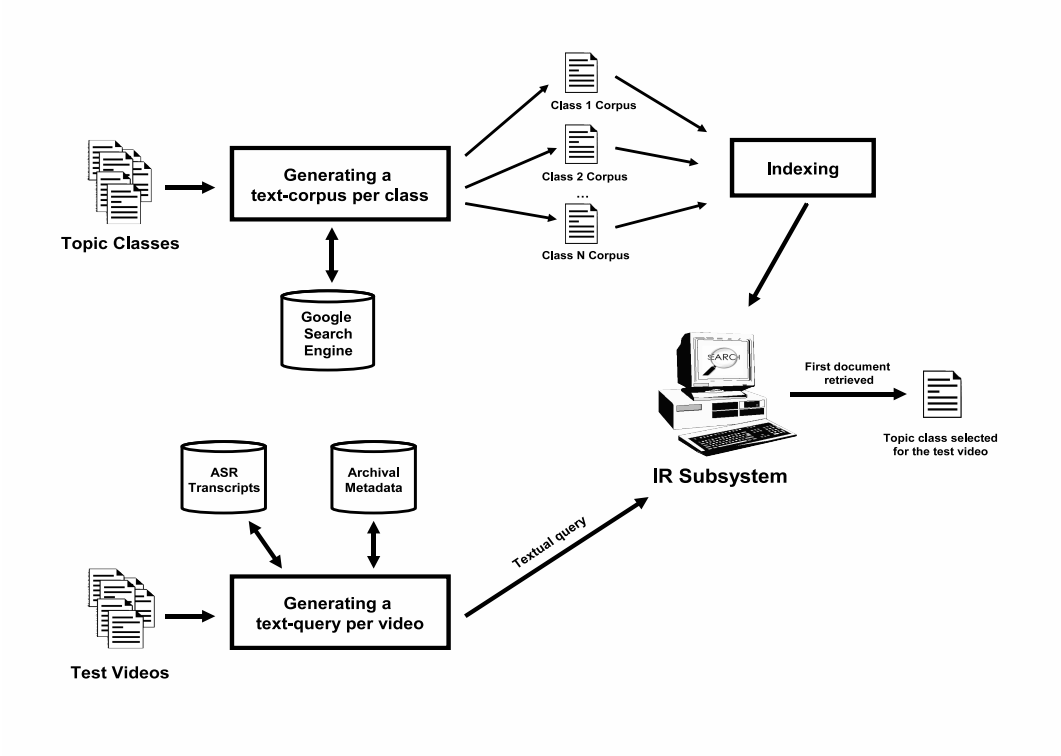

Figure 1: Basic architecture of the SINAI video classifier

This paper is organized as follows: section 2 describes the whole system. Then, in the section 3 experiments, results and resources employed are described. Finally, in the section 4, the conclusions are presented.

2 System description

The overall architecture of our automatic video classifier is based on the use of a particular IR system as classification system. In our experiments we have used Lemur[4] as IR system.

In our approach we have two main processes:

- **Generating a text-corpus per class**. We have generated a textual corpus per topic class. This corpus corresponds to the ten top retrieved results by Google, using the topic word (e.g. *Architecture*) as search query. The use of Google has been found useful in other tasks like Robust Retrieval[5]. These results have been combined into one single document. Therefore, a document per class is obtained after this process. Each document per class is indexed by means of IR system. This index will be used for retrieving the speech transcriptions preprocessed of each test video (textual queries).

- **Generating a text-query per video**. We have used the textual information available from Automatic Speech Recognition (ASR) output in order to generate the textual queries for each test video. This data has been preprocessed using the *Dutch stemmer* from Snowball[5] for Dutch language and *Porter stemmer*[6] for English. We have also discarded the *stop-words* for both languages.

[4]http://www.lemurproject.org
[5]http://snowball.tartarus.org/algorithms/dutch/stemmer.html

Experiment	Language	Using metadata	P	R	F1
SINAI-JEAN-Class-I	Dutch	no	0.65	0.42	0.51
SINAI-JEAN-Class-I-Trans	English	no	0.13	0.09	0.10
SINAI-JEAN-Class-II	Dutch	yes	0.68	0.44	0.53

Table 1: Micro-averaging results

Experiment	Language	Using metadata	P	R	F1
SINAI-JEAN-Class-I	Dutch	no	0.91	0.34	0.49
SINAI-JEAN-Class-I-Trans	English	no	0.68	0.28	0.40
SINAI-JEAN-Class-II	Dutch	yes	0.89	0.36	0.51

Table 2: Macro-averaging results

For each textual query generated the IR subsystem retrieves the document class more relevant, using the standard *TF·IDF* [7] weighting scheme. The Figure 1 shows this basic architecture of our approach.

3 Experiments and Results

In our first participation in VideoCLEF, we have only submitted three runs for the classification subtasks. There is one mandatory classification task (classification task I) which uses speech transcriptions only. We have submitted two runs for this task, one for each language (Dutch and English). The name of these experiments are SINAI-JEAN-Class-I and SINAI-JEAN-Class-I-Trans respectively. For the third run we have used the archival metadata supplied (classification task II), only for Dutch language. This run is identified as SINAI-JEAN-Class-II. The results using micro-averaged and macro-averaged measures are shown in Tables 1 and 2.

Analyzing the results obtained, we can observe that English language leads to worse results than Dutch. This could be an expected behavior, due to the lower relevance of this language in the corpus compared to Dutch. In micro-averaged values, the use of metadata brings slightly better precision and recall measurements, although in macro-averaged results this improvement is only present in recall.

3.1 Resources

The ASR transcriptions of video files have been supplied by the organization of VideoCLEF[6]. This video data are Dutch television documentaries and contain Dutch as dominant language, but also contain a high proportion of spoken English, such as interviewed guests. On the other hand, the archival metadata resource contains program titles and short descriptions of the content of video files.

4 Conclusions

We have described here the approach followed by the SINAI research group in its participation in the VideoCLEF 2008 campaign. Our results show that, despite the simplicity of our system, transcriptions are a good source of information for video classification. Anyhow, some enhancements on the system can be performed, by selecting additional sources of learning data: we are working on a system that uses Wikipedia articles too.

[6]The data for VideoCLEF 2008 has been provided by The Netherlands Institute of Sound and Vision (Beeld & Geluid). Thank you also to the University of Twente for providing the speech recognition transcripts.

References

[1] Díaz-Galiano, M.C., Perea-Ortega, J.M., Martín-Valdivia, M.T., Montejo-Ráez, A., Ureña-López, L.A. SINAI at TRECVID 2007. In *Proceedings of the TRECVID 2007 Workshop (TRECVID 2007)*, 2007.

[2] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Ráez, A., Ureña-López, L.A. SINAI at ImageCLEF 2007. In *Proceedings of the Cross Language Evaluation Forum (CLEF 2007)*, 2007.

[3] Díaz-Galiano, M.C., García-Cumbreras, M.A., Martín-Valdivia, M.T., Montejo-Ráez, A., Ureña-López, L.A. SINAI at ImageCLEF 2006. In *Proceedings of the Cross Language Evaluation Forum (CLEF 2006)*, 2006.

[4] Martín-Valdivia, M.T., García-Cumbreras, M.A., Díaz-Galiano, M.C., Ureña-López, L.A., A. Montejo-Ráez. SINAI at ImageCLEF 2005. In *Proceedings of the Cross Language Evaluation Forum (CLEF 2005)*, 2005.

[5] F. Martínez-Santiago, M.A. García-Cumbreras, and Arturo Montejo-Ráez. SINAI at CLEF 2007 Ad Hoc Robust track 2007: Applying google search engine for robust cross-lingual retrieval. In *Proceedings of the Cross Language Evaluation Forum (CLEF 2007)*, 2007.

[6] M.F. Porter. An algorithm for suffix stripping. In *Program 14*, pages 130–137, 1980.

[7] G. Salton and M. J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, London, U.K., 1983.