SPEX Environment for Real-time Cross-Media Recommendation

Roeland Ordelman
Netherlands Institute for Sound and Vision
Hilversum
rordelman@beeldengeluid.nl

Michiel Hildebrand
Spinque
Utrecht
michiel@spinque.com

ABSTRACT

This demo paper describes a video search environment for developing cross-media video recommendation applications. We demonstrate two of such applications, one specifically geared towards media producers and one for the general public.

1 INTRODUCTION

Content producers are seeking convergence, bringing together the three C's: Content, Computing and Communication. Potentially, they have access to a huge amount of high-quality content, filling up in a diversity of repositories. Content that has substantial value for these *next level information services* includes: materials to interact more or maybe better with audiences, materials to target new demographics, and materials which could be used for completely new services and experience formats.

A key limitation of existing tools for digital content production, management and reuse, is their limited support for exploration of audiovisual content available in large audiovisual archives. They lack the functionality necessary to enable producers to increase engagement and diversity of their media productions, or strategies that stimulate innovation capacity by enabling "*innoversity*": diversity as a source of content innovation to develop new formats.

In addition to the well-known value of individual assets that can be uncovered and selected using traditional 'ad-hoc' search technology, a large potential remains in the "latent information" in large repositories (or a federation of repositories). The SPEX (Smart Production using Enhanced video eXploration) environment aims to stimulate the uncovering of this potential by providing content-based, cross-media recommendation or "video hyperlinking": based on textual "triggers" such as news feeds from news agencies, related videos fragments from a large audiovisual archive are provided. In case a news event occurs, content producers have immediate access to a diverse set of related material that they can quickly explore for reuse in new media productions. This is demonstrated in the "Feit&Fictie" application, a first working prototype that uses the SPEX environment for recommendation.

Evidently, also for other user groups such an application is interesting. For example, a widget version of the application can be used as an extension in a web browser to suggest relevant videos given a web text such as a news item or a blog. We will also demonstrate this 'public' version. Note however that in practice there are viewing restrictions due to IPR on the content.

2 SYSTEM OVERVIEW

We use two datasets from the Netherlands Institute for Sound and Vision. The IMMIX dataset contains metadata about radio and TV programs. In July 2017 this dataset contains 1.8M RTV programs, with metadata such as titles, descriptions, genre, presenters and broadcast information. There are also 4.2M segments defined within
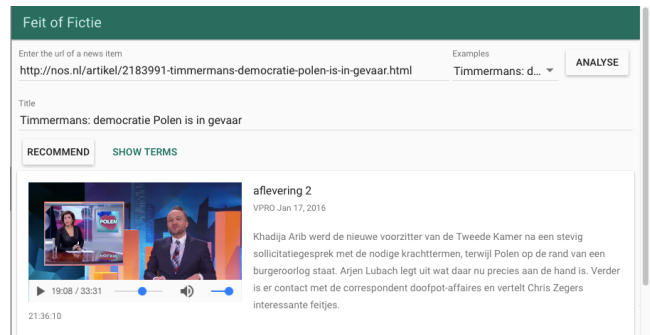
these programs. The segments are manually defined by archivists. They contain a title, start and end time and in most cases a description. The segments are contained within in 600K programs. The second dataset contains the subtitles for the hearing impaired ("teletekst 888"). These speech transcripts are available for TV programs broadcast after 2011 (77K programs), and contains about 24M transcript lines. The two datasets are mapped into a graph model within the Spinque Desk environment [1], integrating the program metadata with the transcripts.

We consider the process of cross-media suggestions (from text to video segments) as a two-step procedure. First we extract query terms from the input text (query formulation). Second we retrieve the segments relevant to the extracted query terms. Within Spinque Desk we are experimenting with various approaches to these two steps. For the query formulation we use a basic approach. We tokenize the input text and compute a weight for each token based on their term frequency (TF) in the input text and the IDF of the term in the program metadata. For the retrieval of the segments we start with the manually defined segments available in the data. For this we use the extracted and weighted terms from the input text as a query for a BM25 retrieval model on the titles and descriptions of the segments. This approach is limited to the manually defined segments which are contained in one third of the RTV programs. Using the transcripts we extend the coverage of the TV programs, and suggest relevant time intervals within a program. For this approach we combine the retrieval of the full TV program based on the transcript, with the retrieval of individual lines in the transcripts (aggregated to 5 minute windows).

---

[1] www.spinque.com