

NEW TOOL FOR APPROACHING E-LEARNING: VIDEORDER™

Videorder™ Voice-based Speech Recognition and Language Processing Search Technology with Finder™ Engine

Ferenc Kiss, Lia Bassa

*Department of Information and Knowledge Management
Budapest University of Technology and Economics
Foundation for Informations Society*

Viktor Justin

Your Service Media Communication Agency

Keywords: Speech recognition, voice based search technology, language processing, teaching application for E-learning.

Abstract: Videorder™ Voice-based Speech Recognition and Language Processing search technology with Finder™ engine from @Your Service Media Communication Agency offers a market ready solution today, for speech recognized searching at any kind of film material on the basis of audible information inside, like words, sentences. All this independent of the language involved! The main strength of the system is Language Processing and Speech Recognition in one application package. Videorder™ allows you to search any video or audio clips that are relevant to your query. The method is a good example how the basic aim of ICEIS can be implemented: bringing together the achievements of a research team with trainers of information and knowledge management as well as with a Foundation for Information Society becoming the practitioner of the program.

1 FILM AS SOURCE OF INFORMATION

The film is the most concise form for communicating information. In this case it also means various sorts of audiovisual recordings like lectures, presentations, illustrations to be communicated, real life or fictional recordings. The film aims at creating a new 'frame of reference' in order to enhance the full participation and collaboration between media researchers, education authorities, media professional bodies, social and political institutions and – of course – the users. In order to encounter this tendency, citizens – especially young people – should start exploring the real possibilities of enabling themselves and the future generations to give a critical and creative answer to value-loaded image building. One of the ways this can be achieved is the new medium: **the**

Internet. We strongly believe that future citizens, individually or collectively, will use this new communication medium to distribute their own material and offer alternatives to corporate 'mainstream' images.

As a Hungarian writer said in the early 1920ies about movie:

“Eye is like a faithful herald: as soon as it gets to know something, it informs his master about it in a short and succinct sentence. If the amount of the new knowledge is so much that he cannot make a full stop in his breathtaking report, the whole information becomes a single stream as it happens to a chain rushing in front of our eyes, we see it as a single line. The chain of the subsequent sentences turn into one straight line: movement. Eye handles impressions as Melanesian indigenous do figures: what is more than five, is all the same, it is too

much. All the movements happening within one second in reality are the consequences of infinite number of functions. For our eyes, this infinite number means the same as six subsequent pictures coming from Edison's projector: six equals infinity."

The information provision by the means of recorded audio and video messages has had a serious disadvantage until now: its real time duration, its linearity meaning that it cannot be broken into searchable basic units independently of the sounding or viewed effects. Therefore

Research and development of external tools to make film information searchable is being carried on in our days with great efforts.

The main idea is thus to develop new and innovative analytical tools that will allow film, television or any broadcasting consumers, teachers engaged in media literacy programs and industry professionals to better understand and fully appreciate the potential in the media output. Our aim is to present a tool and guideline for innovative possibilities - e-participation channels - permitted by the use of the new Information Communication Technologies (ICTs).

If the above problem is solved, it also provides a device for E-Learning projects. As the teaching material can be handled by the learners, a tailor made curriculum can be set up.

2 THE PROLIFERATION OF MULTIMEDIA CONTENT

The volume of multimedia content is expanding at an exponential rate. Terrestrial, Satellite and Cable Television, Radio and streaming media (audio and video) from the World Wide Web means that users can choose from a wide variety of media. But the growing mountain of content raises a critical question: how do users find the information they want, on demand? Until now this question was unanswered and avoided.

3 VOICE-BASED SEARCH TECHNOLOGY – SPEECH RECOGNITION - OVERVIEW

Current speech recognition technology has high word error rates for large vocabulary sizes. There is very little repetition in queries, providing the small

amount of information that could be used to guide the speech recognizer. In speech recognition applications, the recognizer can use context, such as a dialogue history, to set up certain expectations and guide the recognition. Voice search queries lack such context. Voice queries can be very short (on the order of only a few words or single word), so there is very little information in the utterance itself upon which a voice recognition determination can be made. The hunt for the right videoclip on the Web today is almost entirely dependent on metadata, the tags and descriptions that identify the who, what and where. Many video site operators believe that metadata is the best tool for searching and all we may need. Nevertheless metadata needs to be recognized first. Speech recognition is the field of computer science dealing with designing computer systems that can recognize spoken words. Note that voice recognition implies only the computer taking dictation, not that it understands what is being said. Comprehending human languages falls under a different field of computer science called natural language processing.

A number of voice recognition systems are available on the market. The most powerful can recognize thousands of words. However, they generally require an extended training session during which the computer system becomes accustomed to a particular voice and accent. Such systems are said to be speaker dependent. Many systems also require that the speaker speak slowly and distinctly and separate each word with a short pause. These systems are called discrete speech systems. Recently, great strides have been made in continuous speech systems - voice recognition systems that allow you to speak naturally. There are now several continuous-speech systems available for personal computers. Because of their limitations and high cost, voice recognition systems have traditionally been used only in a few special situations.

4 VIDEORDER™

Videorder™ Voice-based Speech Recognition and Language Processing search technology with Finder™ engine from @Your Service Media Communication Agency offers a market ready solution today, for speech recognized searching at any kind of film material on the basis of audible information inside, like words, sentences. All this independent of the language involved!

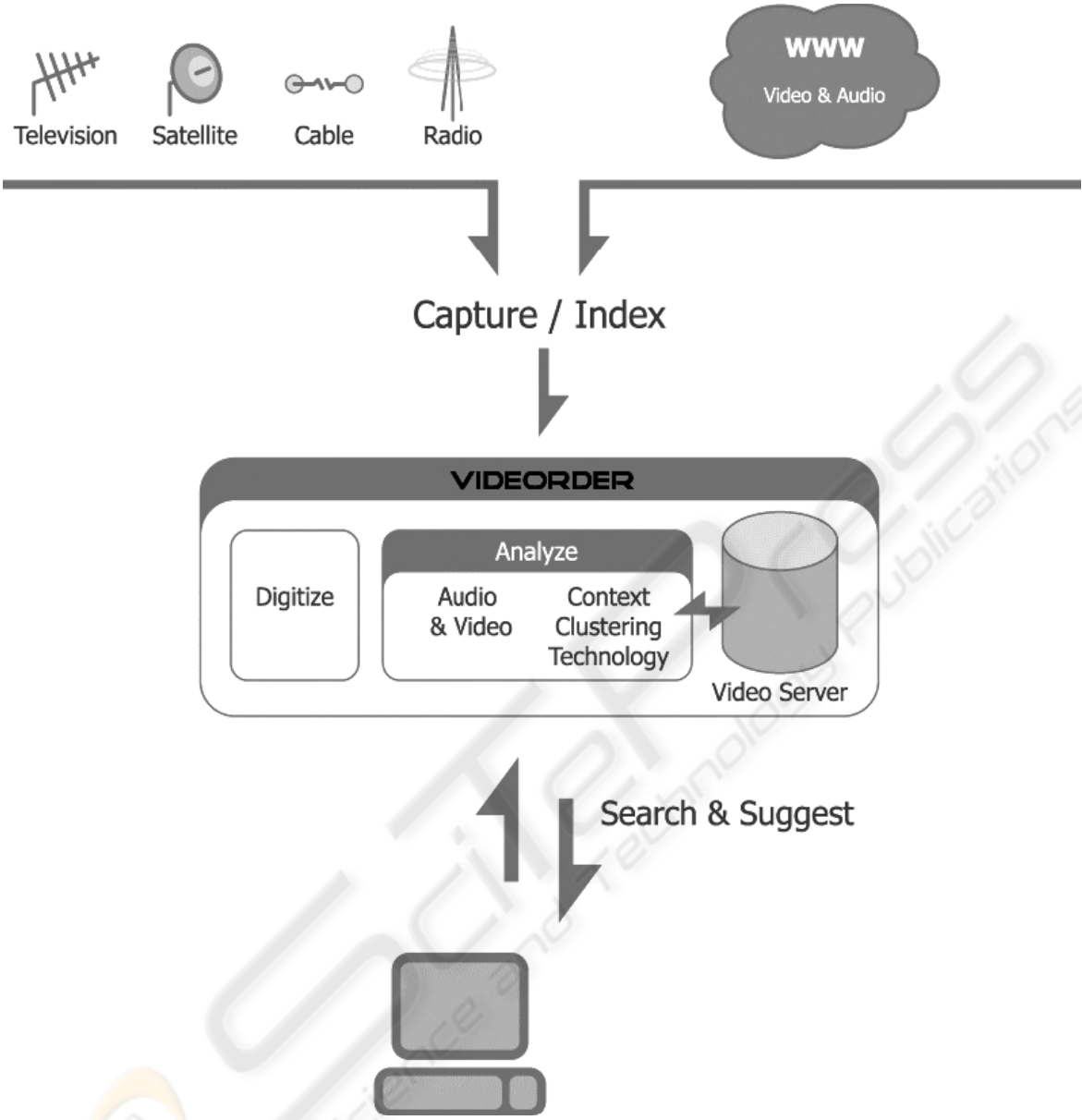


Figure 1: The system of Videorder.

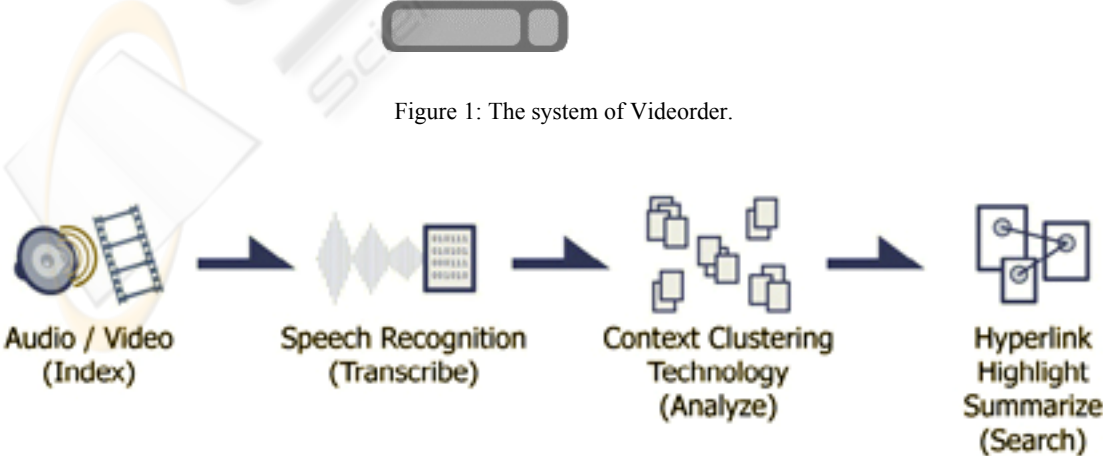


Figure 2: The process of the technology.

The main strength of the system is Language Processing and Speech Recognition in one application package.

The advantages of this system are the following:

- Language independent speech recognition
- 100% error free queries
- Use with mobile services
- Very short search times (especially compared to real-time-viewing)
- Vast searchable databases of any kind of filmed material
- Educational use for corporate trainings or at universities
- Educational use for distance learning systems
- Logical add-on for internet based film-sharing portals
- Wide development capacity

5 TECHNICAL FEATURES

- Language independent speech recognition
- 100% error free queries
- Xml based
- Integrated metadata GUI
(means: real time search parallel with watching video stream, playing presentations, or any kind of additional media file)
- Result exported to any kind of database.

Videorder™ combines the most advanced speech recognition technology with intelligent text analysis and synchronization technologies to deliver unparalleled automation and retrieval of multimedia content.

Once feeding the input video or audio stream into Videorder™, it is using advanced image and audio analysis techniques to extract information of a video or audio file in real time and to create a rich index about the video or audio content. This complete, precise, time-stamped index provides fine-grained access to the video content that you can use for searching and efficiently locate a specific video segment for playback. For indexing Videorder™ uses advanced indexing technology, so the user can quickly locate specific segments within the video content or audio clip. Videorder™ generates metadata tracks to save information generated by the media analysis process. The information in all metadata tracks are time stamped and synchronized with the associated digital video file. Videorder™

platform controls both the indexing and encoding processes to ensure synchronization between the metadata captured from the video asset and the associated digital file.

6 ADVANCED VIDEO/ AUDIO ANALYSIS

Videorder™ utilizes advanced speech recognition techniques and synchronization technologies to analyze and understand the actual content (spoken words) of an audio/video file, delivering supreme accuracy, and access to multimedia content in any form. The output of these analytical sub-processes are stored as further metadata tracks, alongside with the digitally encoded content itself; not only does Videorder™ know what was said, Videorder™ knows exactly when it was said.

7 CORE AUDIO TECHNOLOGY

Videorder™ uses audio analysis technology that applies advanced methods to deal with all aspects of processing digital audio signals of an audio or video stream. In order to analyse the spoken words of an audio or video stream, the audio analysis techniques used by Videorder™ are based on neural network technology that is able to provide a fast, accurate and dynamic solution within variable and rapidly changing acoustic environments.

Rather than just relying on the existing metadata for the description of an audio or video clip, Videorder™ actually provides the ability to retrieve a wide range of multimedia content, based on the spoken words that were actually said in the television or radio clip. This opens a whole new realm of possibilities for accessing multimedia online or in a database.

8 NON-DICTATED SPEECH

Some information feeds, like e.g. news broadcasts and radio, are often intrinsically difficult to transcribe due to noisy conditions and less than perfect articulation. Videorder™ having an additional, special feature, the sophisticated “Human Ears” technology enables it to filter out extraneous noise, to compensate low volume levels predict intended dialogue with high probability. Approaches

to searching through information using simple keyword query mechanisms fail because the context provided is not enough. In such cases users are flooded by hundreds or thousands results, all of which match the query and many representing the different contexts within which the query term or terms can appear in. Videorder™ is unique in not second-guessing a user's need but seamlessly provides the automated guidance required to enrich the overall experience. Once the video or audio stream has been indexed, encoded and analyzed the digital files are stored on the Videorder™ server from where you can retrieve and play the digital files through the Videorder™ web browser. The Videorder™ video server provides fast storage and retrieval for all digital files of the indexed video/audio channel.

9 VIDEORDER™ SEARCH

Videorder™ allows you to search any video or audio clips that are relevant to your query. The Videorder™ query refinement feature allows selecting the best results of a query in order to produce even more relevant results. Sorting by relevance or dating Videorder™ enables the user to sort audio and video results by conceptual relevance or date.

10 CONCLUSION

The next step is to prepare a pilot of the training program where supposedly the participants will be actively involved in building and maintaining (moderating) their own transnational communities in order to share knowledge and insights into political, social and cultural collective decision-making for the purpose of the introduction of this new technique in training. In order to establish responsibility for these new media of exchange, participants will analyze existing virtual communities and the discursive behavior observed in these fora. By disseminating results and providing information for people, we would like to promote awareness of the E-Learning program actions and co-operation between involved institutions, players and students, the latter both interested in the further development of the tool or in using it for their studies.

The above described method is a good example how the basic aim of ICEIS can be implemented: bringing together the achievements of the research

team of Videorder™ with trainers of information and knowledge management as well as with the Foundation for Information Society becoming the practitioner of the program by buying, using and disseminating it to its partners.

REFERENCES

- Bindé, Jérôme, *Towards knowledge societies*. Paris, UNESCO, 2005.
- Illeris, Knud, *Adult education and adult learning*, Malabar, FL, Krieger, 2004. 245 p. Bibl.: p.231-242
- James, Kathryn; Nightingale, Christine, *Discovering potential: a practitioner's guide to supporting improved self-esteem and well-being through adult learning*. Leicester, UK, NIACE, 2004. 85 p.,
- Longworth, Norman, *Enthusiasing and empowering learners*. Leicester, UK. pp.149-159. (In: Journal of adult and continuing; vol.10, 2004, no.2).
- Madhukar, Indira, *Lifelong learning in the information society*. New Delhi, Authorspress, 2004.
- Singh, Madhu, *The social recognition of informal learning in different settings and cultural contexts*. Köln, Germany, Böhlau, 2005.
- Thinesse-Demel, Jutta, *Museums, libraries and cultural heritage: democratising culture, creating knowledge and building bridges; report on the workshop held at the CONFINTEA V Midterm Review Conference, Bangkok, 2003*. Hamburg, Germany, UNESCO Institute for Education, 2005
- Welton, Michael, *Designing the just learning society: a critical inquiry*. Leicester, UK, NIACE, 2005. 250 p.
- Youngman, Frank, *Changer les choses: les programmes de développement et la formation des éducateurs d'adultes* Bonn. pp.137-152. (In: Education des adultes et développement; 65,2005). Bibl.: p.152.
- Contribuer à un avenir plus viable: éducation de qualité, compétences nécessaires dans la vie courante et éducation du développement durable* UNESCO Paris, UNESCO, Division de la promotion d'une éducation de qualité, 2005. 6 p.,
- If any, the appendix should appear directly after the references without numbering, and not on a new page.