Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia - Social and Behavioral Sciences 64 (2012) 186 – 191

Procedia
Social and Behavioral Sciences

International Educational Technology Conference

Ivia: Interactive Video Intelligent Agent Framework for Instructional Video Information Retrieval

Dr. Emdad Khan *, Dr. Adel AlSalem

Imam University, Riyadh, Saudi Arabia

Abstract

Current use of e-learning management systems(ELMS) in educational institutions is on the rise. These systems are rapidly increasing in regards to volume. Instructional video is one type of content that is inherently large in volume and sequential in nature in accessing the videos which makes it difficult to manage and retrieve information. There has been extensive research on video information retrieval in the past decade. Existing systems need pre-processing by human intervention, are cost prohibitive, or do not exhibit the natural interaction. In this paper, we propose a framework for information retrieval of instructional video content in an ELMS that utilizes Natural Language Understanding / PROCESSING and an Intelligent Agent in a seamless integrated environment to address the key issues of the existing solutions.

Keywords; Human Computer Interaction; E-Learning; Information Retrieval; Multimedia; Natural Language Understanding (NLU).

1. Introduction

With the increase of deployment of e-learning management systems(ELMS) in the educational community, it is currently common for any organization involved in education or training to deploy such

* Corresponding author. Tel.: 96612586711, 96612581888; fax: 9661258889.
E-mail address: emdad@ccis.imamu.edu.sa
alsalem@ccis.imamu.edu.sa

1877-0428 © 2012 Published by Elsevier Ltd. Selection and/or peer-review under responsibility of The Association Science Education and Technology
Open access under CC BY-NC-ND license. doi:10.1016/j.sbspro.2012.11.022

a system as a support to the learning process. The content within the ELMS is initially unfilled. As more users utilize the system, the content becomes large until it has been packed with an abundant amount of content. ELMS content items have many forms. In addition to other items, the ELMS may contain an abundant amount of training videos captured from real classrooms, talks, and seminars. These training videos are then archived in the ELMS for later viewings as instructional resources. Since instructional videos are considered more favourable to the learner than other forms of instructional material, they will continue to grow in number and size. Moreover, Lectures can convey core course material in a more easily digestible form than textbooks. Students often favour the retention of live lectures in contrast to extensive self reading of textbooks[5].

Unfortunately, for the experienced person searching for the needed information in one or more long videos is a challenging task. This is due to the sequential nature in accessing the videos. The search for information involves the playback of the video, and possibly a number of videos, from the beginning to the end in the worst case, if it is found at all. This sequential search of information in videos, in which many of them are long, requires time and effort.

The training video contains, in general, a number of items. First, the voice of the presenter. The portrait of the presenter can also be apparent in the video. Second, the presentation slides which contain text(and possibly images). The training videos serve two purposes; either a learning tool for the novice or a reference to an experienced person. For the novice person, the video will be usually played from the beginning to the end and possibly more than once. For an experienced user to sequence through long hours of video to find a particular clip is a boring and tedious task. What is needed is a way to find the relevant video clip with the least effort on the use's part.

In conventional video retrieval systems a question analysis is performed and then, the system searches the video metadata library for the best match. This places a constraint on the content to hold enough metadata for the query to be satisfied. Moreover, current video indexing and retrieval techniques are based on visual and audio features which are not suitable for lecture videos that have frequent scene changes [6].
We have invented a technique to retrieve information for instructional video content that enables a person browse and retrieve instructional video content in a natural and seamless manner. The proposed approach involves a collection of tools and techniques for instructional video information retrieval in a natural and seamless manner by the integration and extension of several existing techniques into a coherent framework referred as IVIA: Interactive Video Intelligent Agent. IVIA accepts commands in natural

language by means of user's voice or typed commands using the conventional keyboard. The input is analyzed and parsed based on semantics of the input words and sentences. IVIA then creates a sequence of commands (or instructions) to search video content using the indexed search information on the video. Our approach is to facilitate the user to obtain the needed information in a speedy and natural fashion.
This is also made possible without the need to perform pre-processing such as re-indexing or cataloguing. This is especially true if the video content is not indexed with enough information to describe its nature. IVIA achieves this by using a "rendering" function of the indexed content to make the search process more effective. Moreover, the video content may have an incorrect index. IVIA can also be utilized during the processes of restructuring instructional resources within the ELMS.

2. Related Work

Some traditional systems yield as output the query a set of ordered results. Then the user is left with the burden of reordering the results depending on relevance. This puts an immense burden on the user which is an entirely manual process. The new ordering is fedback into the system and a new result is obtained. The user performs a new reordering and so on. This loops many times until the appropriate result is obtained. This puts a constraint on the user in an unnatural way. Moreover, this loop may continue for a many iterations that may result in terminating the search by the user[8].
Other efforts are not proven to retrieve information due to loss of context. This may happen if the same text is available in two different contexts[9][10][11][12].

3. Our Approach

In order to satisfy the user's query, we must first acquire as much text as possible from the instructional video. There are many methods described in the literature to locate and extract text in video recordings. In this process, there exist a number of challenges. This is especially true if the text is embedded in images with complex background. The problem exists because the background may involve counters that vary sharply. This background may have comparable intensities to the text and therefore considered as part of the text.

Since the main source of text in instructional videos are presentation slides (e.g. Microsoft Power Point Presentation®), it makes it much easier to extract such text. In the presentation slides, the text is normally published by the author taking into account a design that is readable and clear. [This is repeated below – so, please take out one ...]

The text extraction is a multi-stage process which involves segregation of frames that contain text from others. A single presentation slide may present in many frames, and thus duplicate frames are discarded. Since the main source of text in instructional videos are presentation slides(e.g. Microsoft Power Point Presentation®), it makes it much easier to extract such text. In the presentation slides, the text is normally published by the author taking into account a design that is readable and clear. Text from Closed caption can also be found in the instructional videos. In the literature, there are a number of established methods of extracting text from videos by utilizing OCR software of a selected frame for some time [1] [2] [3][4]. The objective is to have an Intelligent Agent that attempts to obtain the most desired and accurate results. The form of the query is natural language, which is spoken or typed, and obtains the specific answer from such an Intelligent System. Thus, our approach is the use of Natural Language Understanding (NLU) for Intelligent Information Retrieval. The Semantic Engine of NLU helps to derive the meaning of the input words (see below for more details) and sentence which, in turn, get rendered to equivalent content search commands. To reach to a level of certainty, IVIA resolves ambiguities by resorting to the Question and Answer Module.

While traditional approaches to Natural Language Understanding (NLU) have been applied over the past 50 years, results show insignificant advancement, and NLU, in general, remains a complex open problem. NLU complexity is mainly related to semantics: abstraction, representation, real meaning, and computational complexity. We argue that while existing approaches are great in solving some specific problems, they do not seem to address key Natural Language problems in a practical and natural way. [7] proposed a Semantic Engine using Brain-Like approach (SEBLA) that uses Brain-Like algorithms to solve the key NLU problem (i.e. the semantic problem) as well as its sub-problems. In this paper, we plan to utilize SEBLA to apply NLU for Interactive Video Intelligent Agent (IVIA) in the following ways:

- In the query sentence/string to understand the meaning of each words and sentence.
- Generate all related sets of query strings using semantic meaning of each words (thus generating lot more search results related to the input words).
- Extract the most appropriate and related results from the extended search results.

The main theme of our approach in SEBLA is to use each word as object with all important features, most importantly the semantics. In our human natural language based communication, we understand the meaning of every word even when it is standalone without any context. Sometimes a word may

have multiple meanings which get resolved with the context in a sentence. The next main theme is to use the semantics of each word to develop the meaning of a sentence as we do in our natural language understanding as human. Since learning is an interactive and iterative process, the proposed system involves the learner in a natural dialogue to obtain the required results.

References

[1] A. Amir, G. Ashour, and S. Srinivasan, "Toward automatic real time preparation of online video proceedings for conference talks and presentations," in *34th Hawaii Int. Conf. System Sciences*, Maui, HI, 2001, pp. 1662–1669.

[2] M.A. Smith and T. Kanade, "Video Skimming and Characterization through Language and Image understanding Techniques," technical report, Carnegie Mellon Univ. 1995

[3] A. K. Jain and Y. Zhong, "Page Segmentation in Images and Video Frames," Pattern Recognition, Vol. 31, No. 12, pp. 2055-2-76, 1998

[4] D. Chen, K. Shearer, H. Bourlard, Text enhancement with asymmetric 3lter for video OCR, in: Proceedings of the 11th International Conference on Image Analysis and Processing, 2001, pp. 192–198.

[5] Barker, P.G. and Benest, I.D. The on-line lecture concept: a comparison of two approaches in media technologies. *IEEE Colloquium Digest*, *96* (148), 1996, 9/1 - 9/7.

[6] Zhang, Dongsong. " Virtual mentor and the LAB System- Toward Building An Interactive, Personalized, and Intelligent E-Learning Environment". The Journal of Computer Information Systems. Information Association for Computer Information Systems. 2004.

[7] Khan, E., "Natural Language Understanding Using Brain-Like Approach: Word Objects and Word Semantics Based Approaches help Sentence Level Understanding (to be submitted soon)", a Patent Being Filed in US.

[8] Chekuri Choudary, Tiecheng Liu, and Chin-Tser Huang "Semantic Retrieval of Instructional Videos" ISMW '07 Proceedings of the Ninth IEEE International Symposium on Multimedia Workshops, 2007,

[9] W. Li, S. Gauch, J. Gauch, and K. M. Pua, "Vision: A digital video library," in *1st ACM Int. Conf. Digital Libraries* , Bethesda, MD, 1996, pp. 19–27.

[10] H. D.Wactlar, A.G. Hauptmann, M. G. Christel, R. A. Houghton, and A. M. Olligschlaeger, "Complementary video and audio analysis for broadcast news archives," *Commun. ACM*, vol. 43, pp. 42–47, 2000.

[11] R. Lienhart, "Automatic text recognition for video indexing," in *Fourth ACM Int. Conf. Multimedia* , Boston, MA, 1996, pp. 11–20.

[12] Y. Kuwano, H. A. Taniguchi, S. Mori, and H. K. Kurakake, "Telop-ondemand: Video structuring and retrieval based on text recognition," in *2000 IEEE Int. Conf. Multimedia and Expo (ICME 2000) United States*, New York, 2000, pp. 759–762

[13] Khan E, Alesia E, "e-Services using any Phone & User's Voice: Bridging Digital Divide & help Global Development", **IEEE International Conference on Information Technology and e-Services**, March 24-26, 2012, Tunisia.