

# **SEARCHING SPEECH KEYWORDS FROM VIDEO**

**MOUSTAFA A. M. DALOULL**

**UNIVERSITI UTARA MALAYSIA 2010**

6  
2010  
11455  
2010

# **SEARCHING SPEECH KEYWORDS FROM VIDEO**

A project submitted to Dean of Postgraduate Studies and Research  
in partial fulfillment of the requirements for the degree  
Master of Science of Information Technology  
Universiti Utara Malaysia

**By**

**Moustafa A. M. Daloull**



**KOLEJ SASTERA DAN SAINS  
(College of Arts and Sciences)  
Universiti Utara Malaysia**

**PERAKUAN KERJA KERTAS PROJEK  
(Certificate of Project Paper)**

Saya, yang bertandatangan, memperakukan bahawa  
(I, the undersigned, certifies that)

**MOUSTAFA A. M. DALOULL**  
**(802364)**

calon untuk Ijazah  
(candidate for the degree of) **MSc. (Information Technology)**

telah mengemukakan kertas projek yang bertajuk  
(has presented his/her project of the following title)

**SEARCHING SPEECH KEYWORDS FROM VIDEO**

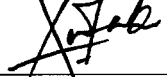
seperti yang tercatat di muka surat tajuk dan kulit kertas projek  
(as it appears on the title page and front cover of project)

bahawa kertas projek tersebut boleh diterima dari segi bentuk serta kandungan  
dan meliputi bidang ilmu dengan memuaskan.  
(that this project is in acceptable form and content, and that a satisfactory  
knowledge of the field is covered by the project).

Nama Penyelia  
(Name of Supervisor) : **DR. SHAHRUL AZMI MOHD YUSOF**

Tandatangan  
(Signature) :  Tarikh (Date) : 17/10/10

Nama Penilai  
(Name of Evaluator) : **MISS ANIZA MOHAMED DIN**

Tandatangan  
(Signature) :  Tarikh (Date) : 17/10/10

## PERMISSION OF USE

---

In presenting this project in partial fulfillment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for copying of this project in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor or, in their absence by the Dean of Postgraduate Studies and Research. It is understood that any copying or publication or use of this project or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my project.

Requests for permission to copy or to make other use of materials in this project, in whole or in part, should be addressed to

**Dean of Postgraduate Studies and Research  
College of Arts and Sciences  
Universiti Utara Malaysia  
06010 UUM Sintok  
Kedah Darul Aman  
Malaysia**

## ABSTRACT

---

*Video contains various types of data that can be extracted using various techniques and tools. The extracted data can be used in developing video retrieving and indexing systems. Video's speech is rich with information and can be used by students as a supporting tool in learning process. Video is divided into many parts depending on their content. Searching for the part of interest may require manual searching through the entire video which may be time consuming. Therefore, this study focuses on investigating the existing approaches of searching and retrieving videos and to develop a method to make video content more easily searchable. A web-application prototype was developed using Java and JSP for searching the video speech using keywords. Finally, the users' satisfaction of the developed prototype was measured using IBM CSUQ questionnaire.*

## DEDICATION

---

*I humbly thank Allah Almighty, the Merciful and the Beneficent, who gave me health, thoughts and co-operative people to enable me achieve this goal.*

*I wish to dedicate this work to Holy Prophet Muhammad (Peace be upon him) and his companions who laid the foundations of Modern civilization and paved the way for social, moral, political, economical, cultural and physical revolution.*

*I also thank my parents (Dr. Adnan, Um Moustafa), my brother (Pharm. Mohammad) and my sisters for their never ending moral support and prayers which always acted as a catalyst in my academic life.*

*To my best friends who support me and pushed me to the top.*

## ACKNOWLEDGEMENT

---

*At the beginning of my speaking, I thank Allah for helping me in my study and guiding me to continue what I have started in my educational life. I thank Allah in every day for giving me the ability and motivation to continue this work.*

*After thanking Allah, I would like to convey my regards to my supervisor Dr. Shahrul Azmi bin Mohd. Yusof for the benefit and precious information that he gave me as one of his students. I thank and honor his for helping me to complete my study in a good way.*

*Finally, I would like to say thankfulness word for the lecturers in the Faculty of Information Technology at Universiti Utara Malaysia.*

*Thank you UUM...*

*Moustafa A. M. Daloull*

*2010*

## LIST OF ACRONYMS

CC	Closed Caption
VCBR	Video Content-Based Retrieval
PDA	Personal Digital Assistant
MIT	Massachusetts Institute of Technology
WMV	Windows Media Video File
MOV	QuickTime Movie
AVI	Video Interleave File
FLV	Flash Video File
GNU	General Public License
UML	Unified Modeling Language
OCR	Optical Character Recognition
RTSP	Real-Time Transport Streaming Protocol
HTTP	Hypertext Transfer Protocol
ZCR	Zero Crossing Rate
GIFT	GNU Image-Finding Tool
IR	Information Retrieval
XML	Extensible Markup Language
RR	Round-Robin
RSV	Ranking Status Value
CSUQ	Computer System Usability Questionnaire
JSP	Java Server Pages
HTML	Hypertext Markup Language
J2SDK	Java 2 Software Development Kit
URL	Uniform Resource Locator
UUM	Univesiti Utara Malaysia



## LIST OF TABLES

Table 2.1: Comparison of Video Data with Other Types of Data .....	8
Table 3.1: Prototype's Requirements.....	30
Table 3.2: Objectives and Followed Methods to Verify Its Outcomes .....	34
Table 5.1: CSUQ Factors and Its Items .....	50
Table 5.2: Descriptive Statistics (Usefulness) .....	51
Table 5.3: Descriptive Statistics (Information Quality).....	53
Table 5.4: Descriptive Statistics (Interface Quality) .....	54
Table 5.5: Descriptive Statistics (Overall Satisfaction Factor).....	55
Table 5.6: Prototype's Requirements Testing Result .....	58

## LIST OF FIGURES

Figure 2.1: The Content of CC File .....	16
Figure 2.2: Architecture of Typical Video Retrieval System .....	20
Figure 2.3: JACOB Architecture .....	22
Figure 2.4: Architecture of The Automatic Video Search System .....	23
Figure 2.5: Block Diagram of The Video Browsing and Retrieval System .....	26
Figure 3.1: Three-Cycle View of Design Science Research. ....	29
Figure 3.2: <i>Google2SRT</i> Main User Interface .....	33
Figure 4.1: Use Case Diagram .....	36
Figure 4.2: View Recently Added Sequence Diagram .....	37
Figure 4.3: Play Video Sequence Diagram.....	38
Figure 4.4: Play Clip Sequence Diagram.....	40
Figure 4.5: Add Video Sequence Diagram.....	41
Figure 4.6: Class Diagram .....	42
Figure 4.7: Main Page.....	43
Figure 4.8: Search Result Page.....	44
Figure 4.9: Play Video Page .....	45
Figure 4.10: Search Video Content Form .....	46
Figure 4.11: Search Video Content Result Page.....	47
Figure 4.12: Play Clip Page .....	48
Figure 5.1: Descriptive Statistics Mean for Usefulness.....	52
Figure 5.2: Descriptive Statistics Mean for Information Quality .....	53
Figure 5.3: Descriptive Statistics Mean for Interface Quality .....	54
Figure 5.4: Descriptive Statistics Mean for Overall Satisfaction Factor .....	56
Figure 5.5: Questionnaire's Factors Percentages .....	56

## TABLE OF CONTENTS

PERMISSION OF USE .....	i
ABSTRACT .....	ii
DEDICATION .....	iii
ACKNOWLEDGEMENT .....	iv
LIST OF ACRONYMS .....	v
LIST OF TABLES .....	vi
LIST OF FIGURES .....	vii
<b>CHAPTER ONE : INTRODUCTION</b>	
1.1 Background .....	1
1.2 Problem Statement .....	2
1.3 Research Questions .....	3
1.4 Research Objectives .....	4
1.5 Research Scope .....	4
1.6 Research Significant.....	5
1.7 Report Organization.....	6
1.8 Summary .....	6
<b>CHAPTER TWO : LITERATURE REVIEW</b>	
2.1 Introduction.....	7
2.2 Video Data .....	7
2.2.1 Characteristics of Video Data .....	7
2.2.2 Content of Video Data .....	9
2.3 Video in Education.....	9
2.4 Video Content Analysis .....	11

2.4.1 Audio Feature Analysis.....	11
2.4.1.1 Frame-Level Features .....	12
2.4.1.2 Clip-Level Features.....	13
2.5 Video Lecture.....	14
2.6 Closed Captions .....	16
2.7 Video Indexing.....	17
2.8 Video Retrieval Approaches .....	18
2.8.1 Visual Feature-Based Approach .....	19
2.8.2 Auditory Feature-Based Approach .....	19
2.8.3 Text-Based Approach.....	19
2.8.4 Task Based Approach .....	20
2.9 Architecture of Typical Video Retrieval System .....	20
2.10 Examples of Video Retrieval System .....	21
2.10.1 JACOB System .....	21
2.10.2 Automatic Video Search System .....	22
2.10.3 Video Browsing and Retrieval System .....	25
2.11 Summary .....	27

**CHAPTER THREE : RESEARCH METHODOLOGY**

3.1 Introduction .....	28
3.2 Research Methodology Design .....	28
3.2.1 The Relevance Cycle.....	30
3.2.2 The Rigour Cycle .....	30
3.2.3 The Design Cycle.....	31
3.2.3.1 Development Technologies.....	32
3.2.3.2 Development Tools .....	32

3.2.3.3 Other Tools .....	32
3.3 Summary .....	33

**CHAPTER FOUR : SYSTEM ANALYSIS AND DESIGN**

4.1 INTRODUCTION .....	35
4.2 System Design.....	35
4.3 User Functionalities .....	36
4.3.1 View Recently Added Sequence Diagram.....	37
4.3.2 Play Video Sequence Diagram .....	38
4.3.3 Play Clip Sequence Diagram .....	39
4.3.4 Add Video Sequence Diagram.....	40
4.4 Class Diagram .....	42
4.5 System Users Interfaces .....	43
4.5.1 The Main Page .....	43
4.5.2 The Search for Video Page .....	44
4.5.3 Play Video Page .....	45
4.5.4 Search Video Content .....	46
4.6 Summary .....	48

**CHAPTER FIVE : EVALUATION**

5.1 Introduction .....	49
5.2 Evaluation Techniques.....	49
5.3 About the Questionnaires.....	50
5.4 Data Analysis .....	51
5.4.1 Descriptive Statistics of Usefulness.....	51
5.4.2 Descriptive Statistics of Information Quality .....	52
5.4.3 Descriptive Statistics of Interface Quality .....	54

5.4.4 Descriptive Statistics of Overall Satisfaction Factor .....	55
5.4.5 Descriptive Statistics of Questionnaire's Factors .....	56
5.4.5.1 System Usefulness .....	57
5.4.5.2 Information Quality.....	57
5.4.5.3 Interfaces Quality .....	57
5.4.5.4 Overall Satisfaction.....	57
5.5 Functional Testing.....	57
5.6 Summary .....	58
<b>CHAPTER SIX : CONCLUSION</b>	
6.1 Introduction .....	59
6.2 Contribution of the Study.....	59
6.3 Results Discussion .....	60
6.4 Limitation and Future Work.....	60
<b>REFERENCES.....</b>	<b>62</b>
<b>APPENDIX A: QUESTIONNAIRE.....</b>	<b>66</b>

## CHAPTER ONE

### INTRODUCTION

#### 1.1 Background

The rapid growth in technology especially in multimedia technologies has led to huge archive of multimedia data in a lot of different areas and applications. Generally, there are four different types of multimedia information, which are video, images, audio and text (Masihi & Charkari, 2005). Furthermore, the multimedia information is much more attractive to browse than plain-text especially videos (Liu & Kender, 2004). So, some of the internet users like to watch videos to get new knowledge. Thus, they are focusing more on their content rather than what the video looks like (Dongsong & Nunamaker, 2004; Jiang & Elmagarmid, 1998).

Online video libraries have become a huge resource for education. This is due to the popularity of video editing software which makes video editing easier. In addition, the advancement of video distributing makes online video widespread because there is also low-cost or free online storage space for hosting video such as *SchoolTube* ([www.schooltube.com](http://www.schooltube.com)), *TeaHerTube* ([www.teachertube.com](http://www.teachertube.com)), and *YouTube* ([www.youtube.com](http://www.youtube.com)). Furthermore, the current streaming media technology allows watching the video while it is downloading (Zhenxiang & Qingsen, 2010).

The contents of  
the thesis is for  
internal user  
only



## REFERENCES

- Abowd, G., Brotherton, J., & Bhalodia, J. (1998). *Classroom 2000: A system for capturing and accessing multimedia classroom experiences. Proceedings of Conference on Human Factors in Computing Systems held on 18 - 23 April 1998 at Los Angeles, California, United States* (pp. 20-21). New York: ACM.
- Ahanger, G., & Little, T. D. C. (2001). Data semantics for improving retrieval performance of digital news video systems. *IEEE Transactions on Knowledge and Data Engineering*, 13(3), 352-360.
- Brezeale, D., & Cook, D. J. (2009). Learning Video Preferences Using Visual Features and Closed Captions. *IEEE Multimedia*, 16(3), 39-47.
- Brown, M. G., Foote, J. T., Jones, G. J. F., Jones, K. S., & Young, S. J. (1996). *Open-vocabulary speech indexing for voice and video mail retrieval. Proceedings of the Fourth ACM International Conference on Multimedia held on 18-22 November 1996 at Boston, Massachusetts, United States* (pp. 307 - 316). New York: ACM.
- Bukvova, H. (2009). Research as a Process: A Comparison between Different Research Approaches. *Sprouts: Working Papers on Information Systems*, 9(29), 9-29.
- Dimitrova, N. (2004). Context and memory in multimedia content analysis. *IEEE Multimedia*, 11(3), 7-11.
- Dongsong, Z., & Nunamaker, J. F. (2004). A natural language approach to content-based video indexing and retrieval for interactive e-learning. *IEEE Transactions on Multimedia*, 6(3), 450-458.
- Dorai, C., Kermani, P., & Stewart, A. (2001). *Elm-n: e-learning media navigator. Proceedings of the Ninth ACM International Conference on Multimedia held on 30 September - 5 October 2001 at Ottawa, Canada* (pp. 634 - 635). New York: ACM.
- Dugonik, B., Brezocnik, Z., Debevc, M., & Maribor, F. (2002). *Video production for distance education. Proceedings of the 24th International Conference on Information Technology Interfaces held on 24-27 June 2002 at Cavtat, Croatia* (pp. 141 - 145). USA: IEEE Press.

- Elmagarmid, A., Jiang, H., Helal, A., Joshi, A., & Ahmed, M. (1997). *Video database systems: issues, products, and applications*. Boston: Kluwer Academic Publishers.
- Fang, H., Wen, Z., Liu, J., & Jiao, B. (2009). *Research on the retrieval process of teaching video resources based on content*. *Proceedings of the 2nd IEEE International Conference on Broadband Network & Multimedia Technology held on 18-20 October 2009 at Beijing, China* (pp. 472 - 476). USA: IEEE Press.
- Flachsbart, J., Franklin, D., & Hammond, K. (2000). *Improving human computer interaction in a classroom environment using computer vision*. *Proceedings of the Fifth International Conference on Intelligent User Interfaces held on 9-12 January 2000 at New Orleans, USA* (pp. 86 - 93). New York: ACM.
- Geetha, P., & Narayanan, V. (2008). A Survey of Content-Based Video Retrieval. *Journal of Computer Science*, 4(6), 474-486.
- Hachman, M. (2010, April 03). YouTube Makes Captioning Available to All. *PC Magazine*. Retrieved 14-August-2010, from <http://www.pcmag.com/article2/0,2817,2360955,00.asp>
- Hartsell, T., & Yuen, S. (2006). Video streaming in online learning. *Association for the Advancement of Computing In Education Journal*, 14(1), 31-43.
- Hevner, A. (2007). The Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19(2), 87-92.
- Jiang, H., & Elmagarmid, A. (1998). WVTDq-A Semantic Content-Based Video. *IEEE Transactions On Knowledge And Data Engineering*, 10(6), 947-966.
- KOM. (2008). *Google2SRT* 0.4. Retrieved 19-June-2010, from <http://google2srt.sourceforge.net>
- La Cascia, M., & Ardizzone, E. (1996). *JACOB: just a content-based query system for video databases*. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing held on 7-10 May 1996 at Georgia, USA* (pp. 1216 - 1219). USA: IEEE Press.

- Lee, C., & Tan, T. (2003). *Humanizing e-learning. Proceedings of the International Conference on Cyberworlds held on 3-5 December 2003 at Singapore* (pp. 418 - 422). USA: IEEE Computer Society.
- Lewis, J. (1995). IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. *International Journal of Human-Computer Interaction*, 7(1), 57-78.
- Liu, T., & Kender, J. (2004). *Lecture videos for e-learning: Current research and challenges. Proceedings of the IEEE International Workshop on Multimedia Content-based Analysis and Retrieval held on 13-15 December 2004 at Miami, Florida, USA* (pp. 574 - 578). USA: IEEE Computer Society.
- Masihi, Z. G., & Charkari, N. M. (2005). *Content based Video Retrieval based on Approximate String Matching. Proceedings of the International Conference on Computer as a Tool held on 21-24 November 2005 at Belgrad, Serbian* (pp. 1300 - 1303). USA: IEEE Press.
- Mozo, A. J., Obien, M. E., Rigor, C. J., Rayel, D. F., Chua, K., & Tangonan, G. (2009). *Video steganography using Flash Video (FLV). Proceedings of the IEEE Conference on Instrumentation and Measurement Technology held on 5-7 May 2009 at Singapore* (PP. 822 - 827). USA: IEEE Press.
- Perea-Ortega, J. M., Martin-Valdivia, M. T., Montejo-Raez, A., & Urea-Lopez, L. A. (2009). *A content-based information retrieval system for video searching. Proceedings of the Eighth International Symposium on Natural Language Processing held on 20-22 October 2009 at Bangkok, Thailand* (pp. 21 - 25). USA: IEEE Press.
- Petkovic, M. (2000). *Content-based video retrieval*. Paper presented at the 7th International Conference on Extending DataBase Technology held on 27-31 March 2000 at Konstanz, Germany.
- Repp, S., Gross, A., & Meinel, C. (2008). Browsing within Lecture Videos Based on the Chain Index of Speech Transcription. *IEEE Transactions on Learning Technologies*, 1(3), 145-156.
- Snoek, C., & Worring, M. (2005). Multimodal video indexing: A review of the state-of-the-art. *Multimedia Tools and Applications*, 25(1), 5-35.

- Sood, A., Sarthi, D., & Mittal, A. (2006). *A Multi-Language System for Knowledge Extraction in E-Learning Videos. Proceedings of the IEEE Annual India Conference held on 15-17 September 2006 at New Delhi, India* (pp. 1 - 6). USA: IEEE Press.
- Tokito, H., Sasabe, M., & Nakano, H. (2007). *Improvement of the capacity of wireless networks by optimizing communication distance. Proceedings of the 9th International Conference on Mobile Wireless Communications Networks held on 19-21 September 2007 at Cork, Ireland* (pp. 11 - 15). USA: IEEE Press.
- Yao, W., Zhu, L., & Jin-Cheng, H. (2000). Multimedia content analysis-using both audio and visual clues. *IEEE Signal Processing Magazine*, 17(6), 12-36.
- Yingying, Z., & Dongru, Z. (2003). *Video browsing and retrieval based on multimodal integration. Proceedings of the International Conference on Web Intelligence held on 13-17 October 2003 at Beijing, China* (pp. 650 - 653). USA: IEEE Press.
- Zhenxiang, S., & Qingsen, C. (2010). *Online educational video: Why it could be a huge resource library of educational television in the web 2.0 era. Proceedings of the 2nd International Conference on Networking and Digital Society held on 30-31 May 2010 at Wenzhou, China* (pp. 98 - 100). USA: IEEE Press.