



UNIVERSITI PUTRA MALAYSIA

**HEARTBEAT DISEASE DIAGNOSIS USING TEXT-BASED
APPROACHES**

EHSAN SAFAR KHORASANI

FSKTM 2011 14

**HEARTBEAT DISEASE DIAGNOSIS USING TEXT-BASED
APPROACHES**

By

EHSAN SAFAR KHORASANI

**Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia,
in Fulfillment of the Requirement for the Degree of Master of Science**

December 2011

Dedicated to my dear parents



© COPYRIGHT UPM

Abstract of thesis to be presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirements for the degree of Master of Science

HEARTBEAT DISEASE DIAGNOSIS USING TEXT-BASED APPROACHES

By

EHSAN SAFAR KHORASANI

December 2011

Chairman: Associate Prof. Shyamala Doraisamy, PhD

Faculty: Computer Science and Information Technology

Heart sound signals are the important asset for heart examination in primary healthcare centers to aid significantly in the diagnosis of heart diseases. Interpretation of heart sounds is a problematic and difficult skill that requires cardiology specialists. The diagnosis of heart disease from heart sound can differ between cardiologists and would require more detailed and expensive tests. However, heart disease diagnosis by heartbeat is preferable and still widely used as the first step to diagnosis. Computer aided auscultation has emerged as a cost-effective technique to analyze and interpret the heart sounds. Digital heart sound recordings with background noise, similarity among heart diseases, recording environment conditions, auscultation body points makes detection of heart diseases complicated.

There are several methods for automated detection and classification of heart diseases and heart sound analysis that have been proposed. Some of them used Artificial Neural Network method for detection and classification of heart sounds. Another technique that it used for diagnosis the heart problem is Hidden Markov Model (HMM) that they suggest HMM for segmentation of heart sound recorded for clinical and classification purpose. However, to the best knowledge of the researcher, no prior study has encoded heart sound to text string.

In this study, we propose a feasible technique for developing a heartbeat sound retrieval system using text encoding techniques which is useful towards automated heart disease detection. The audio format of heart sound recordings are preprocessed and transcribed into the MIDI format. The MIDI files are then encoded to text strings using the pitch and duration information based on n -gram, these text strings then form musical words. These text strings are then indexed and tested for retrieval using both database and Information Retrieval (IR) systems. The Longest Common Subsequence (LCS) matching algorithm was used for identifying similarities from the database. With IR, full text indexing of the recordings was used and retrieved using known item searches from a search engine.

The feasibility of these text encoding techniques were shown from retrieval experiments with around 100 digital heart sound recordings. Overall, experimental results performed clearly showed the feasibility of using proposed text encoding techniques for diagnosing heart problems. Thus, it can be said that the results

presented for heart sound retrieval system are very promising for queries in Normal and Abnormal heart sound categories.



© COPYRIGHT UPM

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

**DIAGNOSIS PENYAKIT DEGUPAN JANTUNG
MENGUNAKAN PENDEKATAN TEXT-BASED**

Oleh

EHSAN SAFAR KHORASANI

Disember 2011

Pengerusi: Profesor Madya. Shyamala Doraisamy, PhD

Fakulti: Sains Komputer dan Teknologi Maklumat

Pentafsiran degupan jantung merupakan kemahiran yang sukar dan bermasalah serta memerlukan pakar kardiologi. Diagnosis penyakit jantung daripada degupan jantung boleh berbeza di kalangan pakar kardiologi dan memerlukan ujian yang lebih terperinci dan mahal. Walau bagaimanapun, diagnosis penyakit jantung melalui degupan jantung adalah lebih baik dan digunakan secara meluas sebagai langkah pertama untuk diagnosis. *Auscultation* berbentuk komputer telah muncul sebagai teknik kos yang berkesan untuk menganalisis dan mentafsir degupan jantung. Rakaman digital degupan jantung dengan latar belakang bunyi bising, kesamaan antara penyakit jantung, keadaan rakaman persekitaran, dan *auscultation* tubuh badan telah merumitkan lagi pengesanan penyakit jantung.

Beberapa kaedah untuk pengesanan automatic, klasifikasi penyakit jantung, dan analisis degupan jantung telah dicadangkan. Sebahagian daripada mereka

menggunakan kaedah *Artificial Neural Network* untuk mengesan dan klasifikasi degupan jantung. Satu lagi teknik yang digunakan untuk diagnosis masalah jantung ialah Hidden Markov Model (HMM) yang mencadangkan HMM untuk segmentasi degupan jantung dirakam untuk tujuan klinikal dan klasifikasi. Kajian ini merupakan pengetahuan penyelidik yang terbaik, kerana tiada penyelidikan yang merekodkan degupan jantung kepada rentetan teks (text string).

Dalam kajian ini, kami mencadangkan satu teknik yang membangunkan sistem dapatan semula degupan jantung dengan menggunakan kaedah rentetan teks untuk pengesanan penyakit jantung secara automatik. Format audio rakaman degupan jantung diproses dan disimpan dalam format MIDI. Setelah itu fail MIDI tersebut dienkod kepada rentetan teks berdasarkan N-gram dengan maklumat yang didapatkan daripada nada dan jangka masa. Selepas ini, rentetan teks yang dienkod akan membentuk bunyi muzik. Rentetan teks ini kemudiannya diindeks dan diuji untuk dapatan semula dengan menggunakan kedua-dua cara, iaitu pangkalan data dan sistem dapatan semula maklumat (IR). *Longest Common Subsequence* (LCS) dipadankan dengan algoritma yang hampir sama untuk mengenal pasti persamaannya daripada pangkalan data. Dengan IR, indeks teks lengkap daripada rakaman dan dapatan semula dengan carian item yang diketahui melalui enjin carian telah digunakan.

Kemungkinan pendekatan dapatan semula berdasarkan teks ini boleh dilihat daripada eksperimen dapatan semula melalui lebih kurang 100 rakaman digital

degupan jantung. Keseluruhannya, keputusan uji kaji yang dilakukan menunjukkan keboleh laksanaan yang jelas dengan menggunakan pendekatan pengekodan teks yang dicadangkan untuk mendiagnosis masalah jantung. Oleh itu, ia boleh dikatakan bahawa keputusan yang ditunjukkan untuk sistem dapatan semula degupan jantung sangat memberi harapan untuk pertanyaan dalam kategori bunyi jantung yang normal dan tidak normal.



ACKNOWLEDGEMENTS

I would like to thank my supervisor, Associate Professor Dr. Shyamala Doraisamy for her valuable comments and advice through the course of this research. Her encouragement and professional review helped this thesis and other technical papers to be further improved.

My further gratitude goes to Dr. Azreen Azman for his great help and technical advices.

Also, my eternal gratitude is owed to my family who have been supportive in everything I have done. In particular, I would like to thank my mother, Shokooh for her never-ending love and support. I am highly indebted to my sister, Neda for her understanding, encouragement and support during my study.

I also want to thank of all my second family members in Malaysia, including all my friends for providing me with great friendship and experience in my academic and social life. Specially, I owe gratitude to my friend, Ghazal for her impressive help in my thesis and his expressions when things seemed not to be in track.

Finally, thanks God for giving me another opportunity to know myself by living in Malaysia.

I certify that an Examination Committee met on 6 September 2011 to conduct the final examination of Ehsan Safar Khorasani on his Master degree thesis entitled “HEARTBEAT DISEASE DIAGNOSIS USING TEXT-BASED APPROACHES” in accordance with the Universities and University Collage Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The committee recommends that the student be awarded the Master of Science.

Chairman, PhD

Faculty of Computer Science and Information Technology
University Putra Malaysia
(Chairman)

Examiner 1, PhD

Faculty of Computer Science and Information Technology
University Putra Malaysia
(Internal Examiner)

Examiner 2, PhD

Faculty of Computer Science and Information Technology
University Putra Malaysia
(Internal Examiner)

External Examiner, PhD

Faculty of Computer Science and Information Technology
University Putra Malaysia
(External Examiner)

NORITAH OMAR, PhD

Associate Professor and Deputy
Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the senate of Universiti Putra Malaysia and has been accepted as fulfillment of the requirement for the degree of Master of Science. The members of the Supervisory Committee were as follows:

Shyamala Doraisamy, PhD

Associate Professor
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Chairman)

Azreen Azman, PhD

Senior Lecturer
Faculty of Computer Science and Information Technology
University Putra Malaysia
(Member)

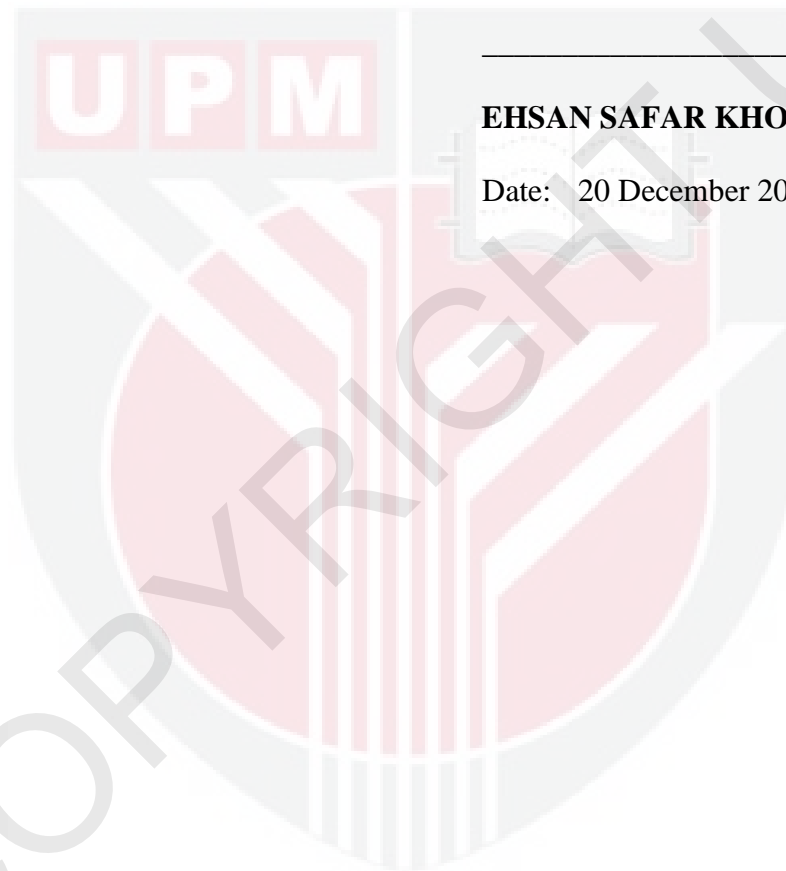
Bujang Bin Kim Huat, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

I declare that the thesis is my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously and is not concurrently, submitted for any other degree at Universiti Putra Malaysia or other institutions.



EHSAN SAFAR KHORASANI

Date: 20 December 2011

TABLE OF CONTENTS

	Page
DEDICATION	ii
ABSTRACT	iii
ABSTRAC	vi
ACKNOWLEDGEMENTS	xii
DECLARATION	xii
LIST OF FIGURES	xvii
LIST OF ABBREVIATIONS	xix
CHAPTER	
1 INTRODUCTION	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Research Objectives	6
1.4 Research Scopes and Limitations	6
1.5 Research Contribution	8
1.6 Organization of Thesis	9
2 LITERATURE REVIEW	11
2.1 Introduction	11
2.2 Heart sound Classification	11
2.2.1 Preprocessing and heart sound signal analysis	12
2.2.2 Heart Sound Analysis	16
2.3 Heart sound retrieval systems	23
2.3.1 Heart Sound Transcription	23
2.3.2 Information Retrieval	25
2.3.3 Music String-matching Using LCS	29
2.4 Summary	31

3	RESEARCH METHODOLOGY	32
3.1	Introduction	32
3.2	Research Design	32
3.3	Data Collection	34
3.4	Performance Evaluation Metrics	38
3.5	Experiment Setup	40
3.6	Summary	41
4	HEART SOUND TEXT ENCODING AND RETRIEVAL	43
4.1	Introduction	42
4.2	Physical Examination	43
4.2.1	Cardiac Auscultation	44
4.3	Preprocessing	45
4.4	Data Encoding	50
4.4.1	Data Transcription (MIDI Transcription)	51
4.4.2	Text Encoding Techniques	54
4.5	Indexing and Database Construction	60
4.5.1	Storing text-strings in Database	61
4.5.2	Indexing Musical words	65
4.6	Retrieval Using database and information retrieval systems	68
4.6.1	String-Matching approach	68
4.6.2	IR Approach	70
5	RESULTS AND DISCUSSION	72
5.1	Introduction	72
5.2	Document Length	72
5.3	Query Documents	73
5.4	Interval Distribution	74
5.5	IR Experiments	75
5.6	Comparison of Retrieval Techniques	88
5.6.1	Result Discussion	89

5.7	Summary	91
6	CONCLUSION AND FUTURE WORK	92
6.1	Conclusion	92
6.2	Future Work	93
	REFERENCES	94
	BIODATA OF STUDENT	109
	LIST OF PUBLICATIONS	110

