



شبكة المعلومات الجامعية

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ





شبكة المعلومات الجامعية



شبكة المعلومات الجامعية

التوثيق الالكتروني والميكروفيلم



شبكة المعلومات الجامعية

جامعة عين شمس

التوثيق الالكتروني والميكروفيلم

قسم

نقسم بالله العظيم أن المادة التي تم توثيقها وتسجيلها
علي هذه الأفلام قد اعدت دون أية تغيرات



يجب أن

تحفظ هذه الأفلام بعيداً عن الغبار

في درجة حرارة من 15 – 20 مئوية ورطوبة نسبية من 20-40 %

To be kept away from dust in dry cool place of
15 – 25c and relative humidity 20-40 %



شبكة المعلومات الجامعية



بعض الوثائق الأصلية تالفة



شبكة المعلومات الجامعية



بالرسالة صفحات

لم ترد بالأصل

**A LARGE-SCALE COMPUTATIONAL PROCESSOR OF THE
ARABIC MORPHOLOGY, AND APPLICATIONS**

by

Mohamed Attia Mohamed Elaraby Ahmed

**A Thesis Submitted to the
Faculty of Engineering, Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
COMPUTER ENGINEERING**

Under the Supervision of

Aly Hassan Fahmy
Professor
Faculty of Engineering
Cairo University

Mohsen Abdul Raziq Rashwan
Associate Professor
Faculty of Engineering
Cairo University

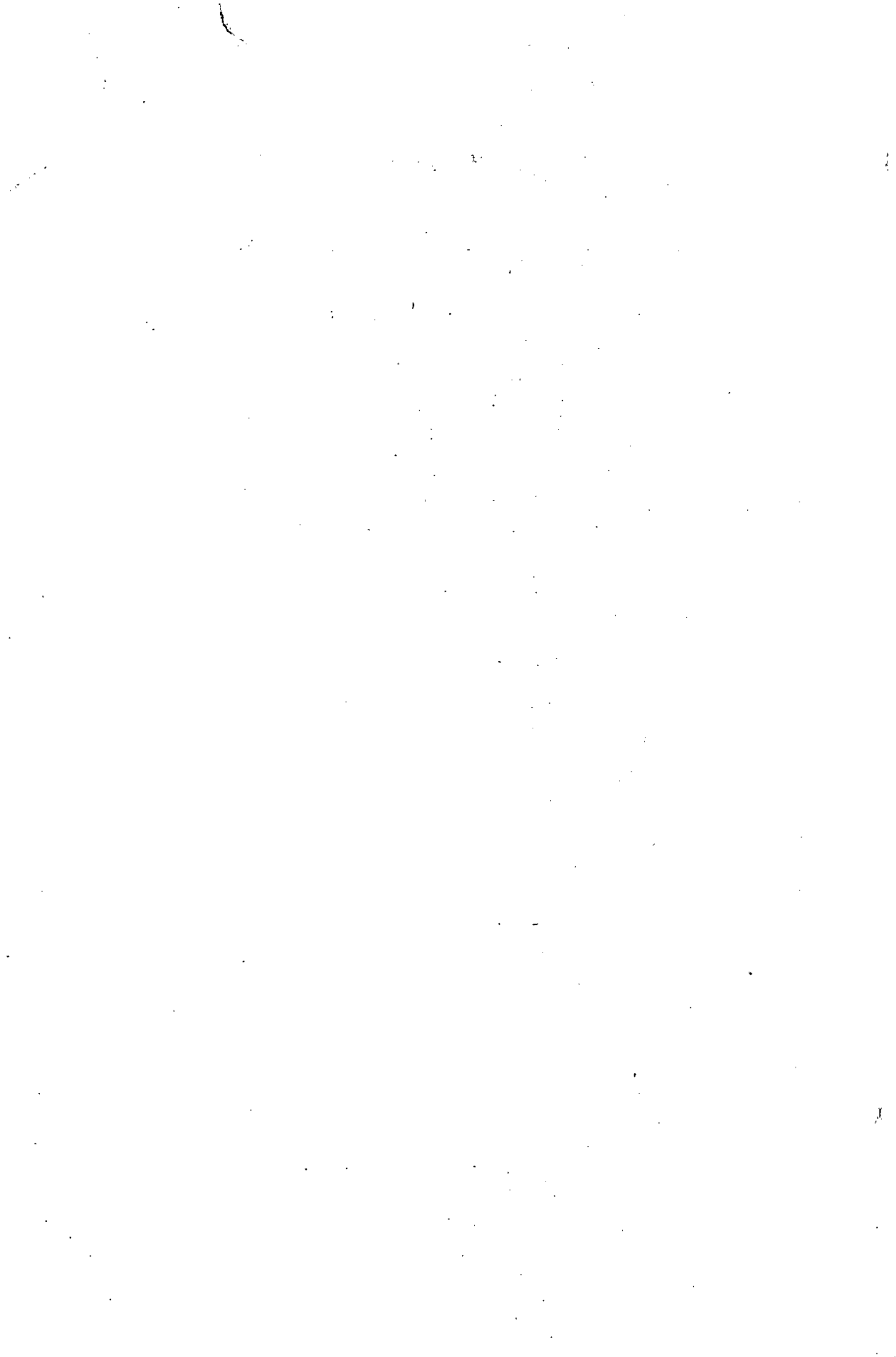
M. Rashwan

Yaseen Al-Shareef
Associate Professor
Faculty of Engineering
Cairo University

B
1-CAN

Faculty of Engineering, Cairo University
Giza, Egypt
January 2000

Aly Hassan Fahmy



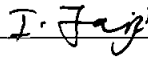
A LARGE-SCALE COMPUTATIONAL PROCESSOR OF THE ARABIC MORPHOLOGY, AND APPLICATIONS

by

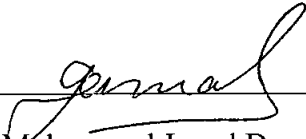
Mohamed Attia Mohamed Elaraby Ahmed

A Thesis Submitted to the
Faculty of Engineering, Cairo University
in Partial Fulfillment of the
Requirements for the Degree of
MASTER OF SCIENCE
in
COMPUTER ENGINEERING

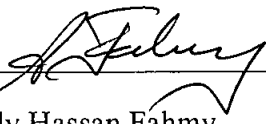
Approved by the
Examining Committee



Prof. Dr. Ibrahim Faraj



Prof. Dr. Mohammad Jamal Darwish



Prof. Dr. Aly Hassan Fahmy

Thesis Supervisor



Dr. Mohsen Abdul Raziq Rashwan

Thesis Supervisor

Faculty of Engineering, Cairo University
Giza, Egypt
January 2000

Acknowledgements

I'm greatly indebted to both of the two honorable gentlemen Prof. Dr. Aly H. Fahmy and Dr. Mohsen Rashwan for not only their generous efforts at supervising this thesis, but also for the many good things that I learnt from them inside and outside the college.

Many other gentlemen deserve more than mere thanks for their appreciable roles that follow:

I regard Eng. Nagy Fatehy Mohammad as the grandfather of this thesis for two things. First, his master thesis was the starting point of mine. Second, my direct long discussions with him highly enriched my understanding of the subject.

I proudly consider myself as a student of Dr. Mohammad Afifi, the brilliant researcher at speech recognition, for he taught me a lot of mathematics and sparked me with a lot of ideas that I used here and in other works.

Two experts at the classical Arabic linguistics Yasser Sarhan and Ahmad Atef shared remarkably at this thesis by guiding me through the jungles of literature on classical Arabic linguistics as well as systematically providing to me their precious notes and corrections.

My close friend Galal Khallaf was a real hero that struggled a lot for achieving a respectable edition of this thesis.

Three bright software engineers Mohammad Abdullah, Ibrahim Sobh and Tamer Atef made my morphological processing engine a useful thing by making it the core of a large and integrated system for Arabic text-retrieval and multimedia-book-authority.

I sincerely feel that only Allah can reward all the good guys mentioned above.

Finally, I hope this thesis be a useful addition to the still-poor research activity at the practical Arabic computational linguistics as well as a starting point of more advanced achievements in this field.

List of Contents

Acknowledgements.....	IV
List of Contents.....	V
List of Figures.....	VII
Abstract.....	VIII
Chapter 1:	
INTRODUCTION.....	1
1.1: Morphological processing.....	2
1.2: Previous work on the Arabic morphological processing.....	2
1.3: This system “Morpho3”.....	6
1.4: Thesis overview.....	7
Chapter 2:	
THE MODEL OF THE ARABIC WORD.....	8
2.1: The definition of a <i>Word</i>	9
2.2: The definition of an <i>Arabic word</i>	9
2.3: Arabic is a diacritized language.....	11
2.4: The Prefix-Body-Suffix structure of the Arabic word.....	14
2.5: The structure of the Arabic word's body.....	16
2.6: A unified formal description of the Arabic word.....	19
2.7: The function of an Arabic morphological analyzer.....	21
Chapter 3:	
APPLICATIONS OF THE COMPUTATIONAL PROCESSING OF THE ARABIC MORPHOLOGY.....	22
3.1: Text as an active medium in multimedia applications.....	23
3.2: Arabic text enhancement.....	24
3.3: On-line Arabic dictionaries.....	24
3.4: Smart text search.....	25
3.5: Support for Arabic Text-to-Speech systems.....	27
3.6: Support for Arabic OCR systems.....	28
3.7: Extracting the morphological print of an Arabic writer.....	28
3.8: Supporting the higher layers of Arabic linguistic analysis.....	29
Chapter 4:	
THE MAIN CHALLENGES BEFORE THE COMPUTATIONAL PROCESSING OF THE ARABIC MORPHOLOGY.....	30
4.1: Coverage.....	31
4.2: Tolerability.....	32
4.3: Disambiguation.....	33
Chapter 5:	
BUILDING THE MORPHOLOGICAL KNOWLEDGE BASE.....	36
5.1: The 9 types of morphological entities.....	37

5.2: The two parts of entity description.....	38
5.3: The description of the P and S entities.....	38
5.4: The description of the F_{rd} entities.....	45
5.5: The description of the R_{rd} entities.....	51
5.6: The description of the F_{id} , F_f and F_a entities.....	52
5.7: The description of the R_f and R_a entities.....	54
5.8: Notes on the authority of the morphological entities.....	55
Chapter 6:	
THE MORPHOLOGICAL ANALYSIS ALGORITHM.....	56
6.1: Extract all the possible prefixes.....	57
6.2: Extract all the possible suffixes.....	57
6.3: Match the possible prefixes with the possible suffixes.....	58
6.4: Extract all the possible body strings.....	60
6.5: Unify the similar body strings.....	61
6.6: Dissolve the bodies into roots and forms.....	61
6.7: Match the forms with the corresponding prefix-suffix pairs.....	63
Chapter 7:	
MORPHOLOGICAL SYNTHESIS.....	65
7.1: The purpose of morphological synthesis.....	66
7.2: The morphological synthesis algorithm.....	67
Chapter 8:	
STATISTICAL DISAMBIGUATION.....	69
8.1: The disambiguation scheme.....	70
8.2: The correlation neighborhood.....	71
8.3: Two hard difficulties.....	72
8.4: The Bayes Turing-Discount Back-off technique.....	74
8.5: Entity $4m$ -grams instead of word m -grams.....	76
Chapter 9:	
CONCLUSIONS AND SUGGESTIONS FOR FUTURE WORK.....	77
9.1: Conclusions.....	78
9.2: Suggestions for future work.....	79
REFERENCES.....	81

List of Figures

2.1: The classification of the Arabic words.....	10
3.1: Dividing the processing of a natural language into layers.....	29
5.1: Classifying the 9 types of morphological entities.....	37
5.2: The tree of prefix <i>IDP</i> 's.....	40
5.3: The tree of suffix <i>IDP</i> 's.....	41
5.3: The tree of suffix <i>IDP</i> 's (cont.'d).....	42
6.1: A part of the binary table that defines the C function.....	58

