

**Design of Task Queue and Scheduling Models for
Web Based Machine Translation System**

by

Fong Pui Chi

Master of Science Degree in Software Engineering



2013



**Faculty of Science and Technology
University of Macau**

Design of Task Queue and Scheduling Models for Web Based Machine Translation System

by

Fong Pui Chi

A thesis submitted in partial fulfillment of the
requirements for the degree of

Master of Science Degree in
Software Engineering

Faculty of Science and Technology
University of Macau

2013

Approved by _____
Supervisor

Date _____

In presenting this thesis in partial fulfillment of the requirements for a Master's degree at the University of Macau, I agree that the Library and the Faculty of Science and Technology shall make its copies freely available for inspection. However, reproduction of this thesis for any purposes or by any means shall not be allowed without my written permission. Authorization is sought by contacting the author at

Address:
AV. CORONEL MESQUITA, 41
1 ANDAR B
ED. YEE FU
MACAU

Telephone: 28560309
Fax:
E-mail: fongpuichi@yahoo.com.hk



Signature _____

Date _____

University of Macau

Abstract

Design of Task Queue and Scheduling Models for Web Based Machine Translation System

by Fong Pui Chi

Thesis Supervisor: Dr. Wong Fai
Master of Science Degree in
Software Engineering

Different web-based Machine Translation systems have their own design in handling the translation task. In particular, we made a careful study in how to better handle many users' translation requests at the same time, and the problems of client request's starvation and server's bottleneck. We proposed a Client-Server architecture based on Moses Statistical Machine Translation (SMT) toolkit for multi-language translation environment with a better way to handle users' requests. We first created a bilingual corpus by catching bilingual context from the Internet and we made use of this information to be aligned with each other for training of SMT model. In our design, internal queues are used to store users' requests, and a multi-threading approach is applied in controlling user's translation request. Different experiments are performed and showed that our proposed design can minimize client request's starvation and server's bottleneck, and maximize system's efficiency.

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES.....	ix
LIST OF ABBREVIATIONS.....	xi
CHAPTER 1: INTRODUCTION	14
1.1 RULE-BASED APPROACH	14
1.2 EXAMPLE-BASED MACHINE TRANSLATION.....	16
1.3 STATISTICAL MACHINE TRANSLATION.....	18
1.4 PRINCIPLE OF STATISTICAL MACHINE TRANSLATION	18
1.4.1 LANGUAGE MODEL	19
GOOD-TURING SMOOTHING	20
BACK OFF	21
INTERPOLATION	21
1.4.2 RECENT CHANGES IN SMT.....	22
1.4.3 TRANSLATION MODEL	24
WORD ALIGNMENT	25
IBM-MODEL 1	25
IBM-MODEL 2	26
IBM-MODEL 3	27
IBM-MODEL 4	28
IBM-MODEL 5	29
HIDDEN MARKOV MODEL	29
PHRASE ALIGNMENT	30
REORDERING	33
1.4.4 DECODING.....	35
TRANSLATION PROCESS	35
TRANSLATION OPTIONS	36
HYPOTHESIS EXPANSION	37
PRUNING	38

1.5 Realted Toolkits	39
GIZA++	39
MOSES	39
SRILM, IRSTLM AND KENLM	39
1.6 PROBLEM STATEMENTS.....	41
1.6.1 LIMITED RESOURCE IN PORTUGUESE-CHINESE MACHINE TRANSLATION.....	41
1.6.2 OS ENVIRONMENT FOR MULTI-LANGAUGE MACHINE TRANSLATION.....	41
1.7 MOTIVATION	41
1.8 WEB MT SYSTEMS.....	42
1.8.1 SYSTEM MODULES	42
1.9 BRIEF OVERVIEW OF CHAPTERS	43
CHAPTER 2: MOSES	44
2.1 TRAINING	45
2.1.1 PREPARE DATA.....	45
2.1.2 RUNNING GIZA++	50
2.1.3 ALIGN WORDS.....	53
2.1.4 GET LEXICAL TRANSLATION TABLE.....	54
2.1.5 EXTRACT PHRASES.....	54
2.1.6 SCORE PHRASES	55
2.1.7 BUILD REORDERING MODEL	56
2.1.8 FACTORED MODEL	60
2.1.9 GENERATION MODEL.....	61
2.1.10 CREATE CONFIGURATION FILE.....	61
2.2 DECODING.....	62
2.3 MULTI-THREADED MOSES.....	63
2.3.1 TECHNIQUE OF MULTI-THREADED PROGRAMMING.....	64
2.3.2 INTRODUCTION OF MULTI-THREADED MOSES	64
2.4 CHAPTER CONCLUSION.....	68
CHAPTER 3: MOSES SERVER.....	69

3.1 WORKING STRUCTURE.....	69
3.2 MULTI-THREADED MOSES SERVER	70
3.3 USING MULTIPLE TRANSLATION SYSTEMS IN THE SAME SERVER	71
3.4 COMPARISON OF TWO MULTI-THREADED TOOLS: MULTI-THREADED TRADITIONAL MOSES AND MULTI-THREADED MOSES SERVER	73
3.4.1 ADVANTAGES AND DISADVANTAGES.....	73
3.5 CHAPTER CONCLUSION.....	75
CHAPTER 4: WEB ENVIRONMENT DESIGN	76
4.1 WEB-BASED DESIGN.....	76
4.2 WEB DESIGN OF SINGLE MOSES SERVER.....	78
4.3 WEB DESIGN OF MULTIPLE MOSES SERVERS	78
4.4 APPLICATION	79
4.4.1 COMMUNICATION BETWEEN DIFFERENT TRANSLATION PAIRS	79
4.4.2 SYSTEM INTERFACE.....	81
4.5 CHAPTER CONCLUSION.....	82
CHAPTER 5: THREAD SCHEDULING	83
5.1 SIDE EFFECTS OF USING THREAD SCHEDULING.....	83
5.2 IMPROVEMENT OF THREAD SCHEDULING IN THE WEB DESIGN.....	83
5.3 THREAD SCHEDULING TECHNIQUES.....	83
5.3.1 FIFO	84
5.3.2 ROUND-ROBIN SCHEDULING	84
5.3.3 SHORTEST JOB FIRST (SJF).....	85
NON-PREEMPTIVE SJF	85
PREEMPTIVE SJF	86
5.3.4 PREEMPTIVE SHORTEST REMAINING JOB FIRST (SRJF)	87
5.3.5 MULTI-LEVEL QUEUE	87
5.3.6 MULTI-LEVEL FEEDBACK QUEUE	88

5.3.7 PREEMPTIVE SHORTEST REMAINING JOB FIRST (SRJF)	
+E_TIME	89
5.3.8 HYBRID MULTI-LEVEL QUEUE AND SHORTEST	
REMAINING JOB FIRST (MLQ + SRJF)	90
5.3.9 WORST CASE ESTIMATED TIME (WCET).....	90
WCET-ALL	91
WCET-EACH	91
WCET = $ax-b$	92
WCET: N Phrases	93
5.3.10 HYBRID SCHEDULING	94
5.3.11 WEIGHTED FAIR QUEUEING (WFQ).....	94
5.3.12 CLASS-BASED WEIGHTED FAIR QUEUEING (CI-WFQ)	95
5.3.13 HYBRID WEIGHTED FAIR QUEUEING AND SHORTEST	
REMAINING JOB FIRST (SRJF)	97
5.3.14 WEIGHTED FAIR MULTI-THREADING (WFM).....	97
5.4 COMPARISON OF DIFFERENT THREAD SCHEDULING	
TECHNIQUES CONSIDERED IN WEB TRANSLATION	99
5.4.1 ADVANTAGES AND DISADVANTAGES	99
5.4.2 SELECTED THREAD SCHEDULING TECHNIQUE	100
5.5 APPLICATION OF SELECTED SCHEDULING TECHNIQUES IN	
WEB-BASED MOSES SYSTEM	101
5.5.1 APPLICATION OF WEIGHTED FAIR MULTI-THREADING IN	
WEB-BASED MOSES SYSTEM	101
5.6 CHAPTER CONCLUSION.....	105
CHAPTER 6: EXPERIMENTS.....	106
6.1 Evaluation in web application of single moses server	107
6.2 Evaluation in web application of multiPLE moses server	117
6.3 CHAPTER CONCLUSION.....	122
CHAPTER 7: CONCLUSION	124
7.1 RESEARCH CONTRIBUTIONS	124
7.2 FUTURE DIRECTIONS	124

BIBLIOGRAPHY	126
--------------------	-----



LIST OF FIGURES

<i>Number</i>	<i>Page</i>
Figure 1: Transfer-Based Machine Translation	15
Figure 2: Interlingual Machine Translation	15
Figure 3: Example-Based Machine Translation	17
Figure 4: Statistical Machine Translation	18
Figure 5: The procedures of creating a SMT system	19
Figure 6: Example of IBM-Model 3	27
Figure 7: Example 1 of Phrase Alignments	31
Figure 8: Chinese-to-English	31
Figure 9: English-to-Chinese	32
Figure 10: Intersection of these two word alignments	32
Figure 11: Union of these two word alignments	32
Figure 12: non-consistent of phrase extraction	33
Figure 13: distance-based reordering	35
Figure 14: Step 1 in example of translation process	36
Figure 15: Step 2 in example of translation process	36
Figure 16: Step 3 in example of translation process	36
Figure 17: Translation Options	37
Figure 18: The first step of Hypothesis Expansion	37
Figure 19: The whole steps in Hypothesis Expansion	38
Figure 20: System Modules	42
Figure 21: Architecture of Moses	44
Figure 22: Search Page of Catching Context of Pages	46
Figure 23: Three different orientations of phrases in the lexicalized reordering model: (m) monotone; (s) swap; (d) discontinuous	57
Figure 24: Single-Threaded Moses	64
Figure 25: UML Sequence diagram for multi-threaded Moses mainline	65
Figure 26: Three-Threaded Moses	66

Figure 27 : Example of multi-threaded Moses Step 1	66
Figure 28: Example of multi-threaded Moses Step 2	67
Figure 29: Example of multi-threaded Moses Step 3	67
Figure 30: Example of multi-threaded Moses Step 4	67
Figure 31: Architecture of Moses Server	69
Figure 32: Moses Server without assigning parameter –threads	70
Figure 33: Moses Server assigning parameter 3-threads	71
Figure 34: Four Translation Systems in four individual processes of Moses Server	72
Figure 35: Four Translation Systems in the same Moses Server	73
Figure 36: Our Web Server Design with FIFO	77
Figure 37: Our Web Server Design with other scheduling	78
Figure 38: Web Server Design of Multiple Moses Servers	79
Figure 39: Communication between translation pairs in Single Moses Server	80
Figure 40: Communication between translation pairs in Multiple Moses Servers	80
Figure 41: Interface of the translation system	81
Figure 42: FCFS example	84
Figure 43: Round-Robin example	85
Figure 44: Non-Preemptive SJF example	85
Figure 45: Preemptive SJF example	86
Figure 46: SRJF example	87
Figure 47: Multi-Level Queue Example	88
Figure 48: Multi-Level Feedback Queue Example	89
Figure 49: One example of WCET-ALL	91
Figure 50: One example of WCET-EACH	92
Figure 51: Weighted Fair Queuing	95
Figure 52: Weight Fair Multi-threading	98
Figure 53: WFM used in our web design	101
Figure 54: Example of WFM Step 1	102
Figure 55: Example of WFM Step 2	102
Figure 56: Example of WFM Step 3	103
Figure 57: Example of WFM Step 4	103

Figure 58: Example of WFM Step 5.....	103
Figure 59: Example of WFM Step 6.....	104
Figure 60: Example of WFM Step 7.....	104
Figure 61: Example of WFM Step 8.....	104



LIST OF TABLES

<i>Number</i>	<i>Page</i>
Table 1: Addresses of pages used to prepare corpus	47
Table 2: Example of vocabulary file (.vcb)	51
Table 3: Example of numbered format file (.snt)	51
Table 4: Example of .vcb.classes file.....	52
Table 5: Example of .vcb.classes.cats file	52
Table 6: Example of final generated word alignments	53
Table 7: Example of lexical translation table file	54
Table 8: Example of alignment file named extract	55
Table 9: Example of phrase table file	56
Table 10: Example of reordering file named <i>extract.o</i>	58
Table 11: Example of reordering table file	59
Table 12: Example of generation model file.....	61
Table 13: Comparison between Moses Server and Traditional Moses.....	74
Table 14: Calculation of thread number used in two users.....	98
Table 15: Comparison between different types of thread scheduling (Part 1)	99
Table 16: Comparison between different types of thread scheduling (Part 2)	100
Table 17: Size of LM in four translation-systems	106
Table 18: Information of training corpora (Part 1)	106
Table 19: Information of training corpora (Part 2)	107
Table 20: Information of test sets in Portuguese-to-Chinese	108
Table 21: Information of test sets in Chinese-to-Portuguese	109
Table 22: Information of test sets in English-to-Chinese	110
Table 23: Information of test sets in Chinese-to-English	111
Table 24: Evaluations in 10-Thread Single Moses: Portuguese-to-Chinese	113
Table 25: Evaluations in 10-Thread Single Moses: Chinese-to-Portuguese	114
Table 26: Evaluations in 10-Thread Single Moses: Chinese-to-English.....	115
Table 27: Evaluations in 10-Thread Single Moses: English-to-Chinese	116

Table 28: Evaluations in 10-Thread Multiple Moses: Portuguese-to-Chinese.....	118
Table 29: Evaluations in 10-Thread Multiple Moses: Chinese-to-Portuguese.....	119
Table 30: Evaluations in 10-Thread Multiple Moses: Chinese-to-English	120
Table 31: Evaluations in 10-Thread Multiple Moses: English-to-Chinese	121



LIST OF ABBREVIATIONS

MT	Machine Translation
RBMT	Rule-Based Machine Translation
EBMT	Example-Based Machine Translation
SMT	Statistical Machine Translation
FIFO	First In First Out
FCFS	First Come First Served
RR	Round Robin
SJF	Shortest Job First
SRJF	Shortest Remaining Job First
LRSF	Least Remaining Sentences First



ACKNOWLEDGMENTS

I wish to give my greatest gratitude to my thesis supervisor, Dr. Wong Fai, who gave me valuable advices and suggestions. He also gave me a lot of helps and support during my project.

I also would like to give my greatest appreciation to Mr. Francisco de Oliveira who gave me helpful comments.

In addition, I would like to thank Mr. Tian Liang who gave me much technical support.

I also would like to acknowledge my boss and my colleagues and they gave me a lot of helpful suggestion.

DEDICATION

I wish to dedicate this thesis to my thesis supervisor, my boss, my classmates and my colleagues, who gave me a lot of encouragement during my thesis research.

