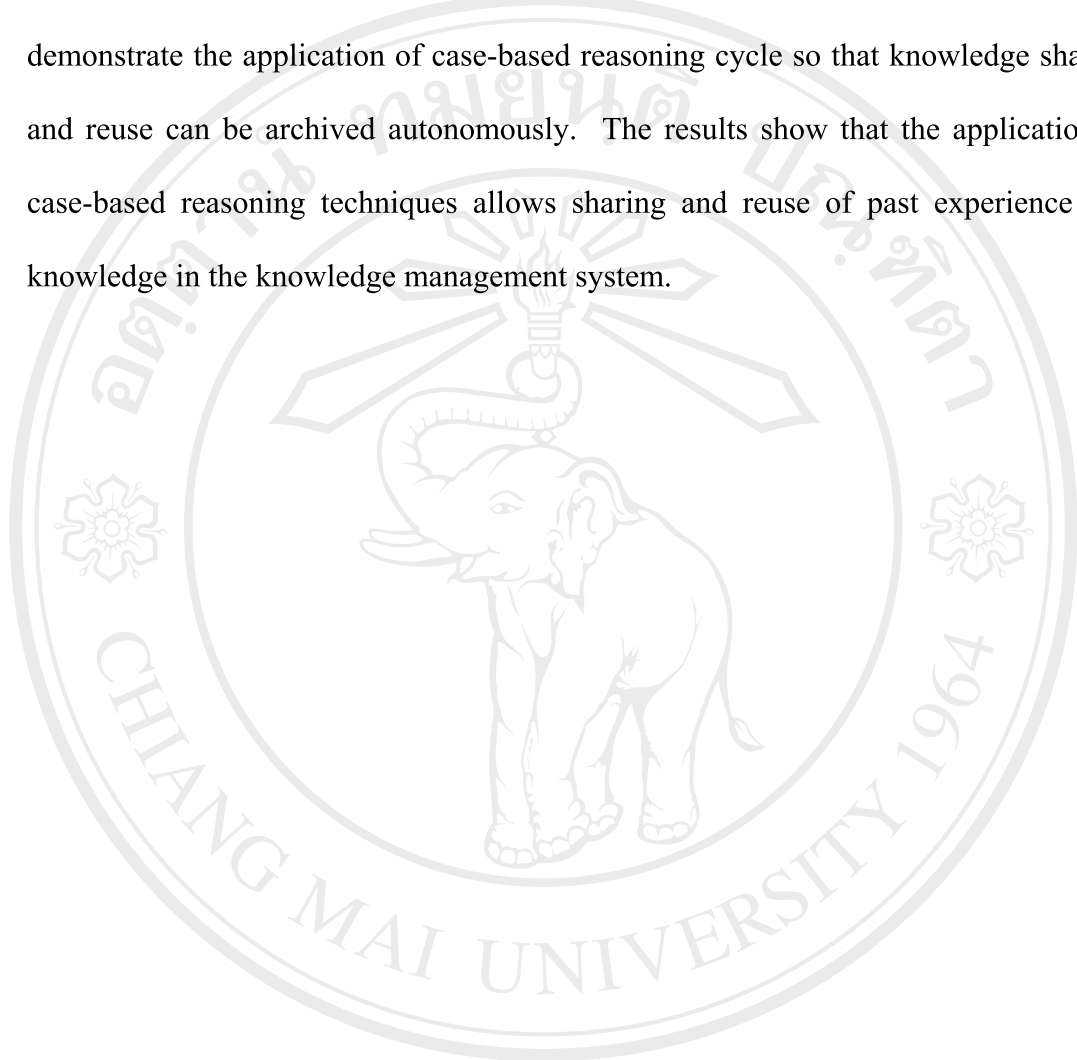


Thesis Title	Metadata Extraction Using Case-based Reasoning for Thai Knowledge Repository	
Author	Mr. Krisda Khankasikam	
Degree	Doctor of Philosophy (Knowledge Management)	
Thesis Advisory Committee	Lect. Dr. Nopasit Chakpitak	Chairperson
	Asst. Prof. Dr. Napat Harnpornchai	Member
	Lect. Dr. Tirapot Chandarasupsang	Member

ABSTRACT

This research aims to investigate the feasibility of techniques in case-based reasoning for Thai metadata extraction in the development of knowledge management system to allow knowledge sharing and reuse between Thai domain expert and librarian. Case-based reasoning is one area of artificial intelligence that allows us to deal with a situation that is similar to the current ones and used them to solve new problem. The application of case-based reasoning techniques allows the knowledge management system to acquire new knowledge by adopting the knowledge gained in new cases and reusing the old one in the case repository. This research also proposes a metadata driven mechanism that provides standardized vocabularies and conceptualization of knowledge domain.

A prototype for Thai metadata extraction was developed in form of web based to demonstrate the application of case-based reasoning cycle so that knowledge sharing and reuse can be archived autonomously. The results show that the application of case-based reasoning techniques allows sharing and reuse of past experience and knowledge in the knowledge management system.



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ชื่อเรื่องวิทยานิพนธ์ การสกัดเมทาเดทา โดยใช้การให้เหตุผลแบบฐานกรณี

สำหรับคลังความรู้ภาษาไทย

ผู้เขียน

นายกฤษดา ชันกสิกรรม

ปริญญา

วิทยาศาสตรดุษฎีบัณฑิต (การจัดการความรู้)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์

อ. ดร. ณพิศมัญ จักรพิทักษ์ ประธานกรรมการ

ผศ. ดร. นภัษฐ์ หาญพรชัย กรรมการ

อ. ดร.ธีรพจน์ จันทรสุมแสง กรรมการ

บทคัดย่อ

วิทยานิพนธ์ฉบับนี้ จัดทำขึ้นเพื่อพิสูจน์ความเป็นไปได้ ของการให้เหตุผลแบบฐานกรณี

สำหรับการสกัดเมทาเดทาภาษาไทย โดยการพัฒนากระบวนการจัดการความรู้ซึ่งก่อให้เกิดการ
แบ่งปันความรู้ และการนำความรู้กลับมาใช้ใหม่ระหว่างผู้เชี่ยวชาญเฉพาะด้านและบรรณารักษ์
การให้เหตุผลแบบฐานกรณีเป็นสาขาหนึ่งของศาสตร์ทางด้านปัญญาประดิษฐ์ ซึ่งสามารถนำ

ขั้นตอนและวิธีการแก้ ปัญหาของปัญหาที่เคยเผชิญมาแล้วและมีความคล้ายคลึงกับปัญหาที่กำลัง

เผชิญอยู่ มาประยุกต์ใช้กับการแก้ปัญหาที่กำลังเผชิญอยู่ได้ โดยโปรแกรมอรรถประโยชน์ที่มี

ความสามารถของการให้เหตุผลแบบฐานกรณี สามารถส่งผลให้ระบบการจัดการความรู้ได้รับ
ความรู้ใหม่ โดยใช้วิธีการเพิ่มความรู้จากฐานกรณีใหม่ และการนำความรู้จากคลังความรู้กลับมาใช้
ใหม่ นอกจากนี้ วิทยานิพนธ์ฉบับนี้ยังนำเสนอการใช้งานเมทาเดตาเป็นกลไกในการสร้างคำศัพท์
มาตรฐาน เพื่อกำหนดกรอบแนวคิดของขอบเขตความรู้อีกด้วย
โปรแกรมต้นแบบสำหรับการสกัดเมทาเดตาภาษาไทยที่ทำงานบนระบบเครือข่าย ถูกพัฒนา
เพื่อพิสูจน์ว่าโปรแกรมอรรถประโยชน์ที่มีความสามารถของวัฏจักรการให้เหตุผลแบบฐานกรณี
ก่อให้เกิดการแบ่งปันความรู้ และการนำความรู้กลับมาใช้ใหม่ได้อย่างอัตโนมัติ ผลลัพธ์จากการ
ทดลองแสดงให้เห็นว่า โปรแกรมอรรถประโยชน์ที่มีความสามารถในการให้เหตุผลแบบฐานกรณี
ส่งผลให้เกิดการแบ่งปันและการนำกลับมาใช้ใหม่ของประสบการณ์และความรู้ ในระบบการ
จัดการความรู้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENT	iii
ENGLISH ABSTRACT	iv
THAI ABSTRACT	vi
TABLE OF CONTENTS	viii
LIST OF TABLES	xii
LIST OF FIGURES	xiii
CHAPTER 1 INTRODUCTION	1
1.1 Introduction	1
1.2 Research Problems	5
1.3 Overview of Research	7
1.4 Research Objectives	10
1.5 A Brief Overview of Research Approach	10
1.6 Organization of Thesis	11
CHAPTER 2 LITERATURE REVIEW	12
2.1 Document Technology	13
2.1.1 Character Set Standard	13
2.1.2 Markup and Stylesheets	23
2.1.3 Web-based Document Technology	27

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่

Copyright © by Chiang Mai University

All rights reserved

TABLE OF CONTENTS (CONTINUED)

	Page
2.2 Metadata	31
2.2.1 The Nature of Metadata	33
2.2.2 Categorization of Metadata	36
2.2.3 Metadata Standardization	39
2.2.4 Summary	49
2.3 Approaches for Metadata Generation	50
2.3.1 Harvesting	51
2.3.2 Extraction Based On Visual Characteristics	52
2.3.3 Extraction Based On Natural Language	53
2.3.4 Extraction Based On Document Code	55
2.3.5 Summary	57
2.4 Information Extraction	59
2.4.1 Information Extraction System Architecture	60
2.4.2 Measuring Information Extraction System Performance	62
2.4.3 Cross Validation	65
2.4.4 Summary	66

TABLE OF CONTENTS (CONTINUED)

	Page
2.5 Case-based Reasoning	67
2.5.1 Case-based Reasoning Architecture	68
2.5.2 Examples of Case-based Reasoning in Real World	74
2.5.3 Examples of Case-based Reasoning in KMS	76
2.5.4 Summary	77
2.6 Knowledge Management Systems	78
2.7 Conclusion	84
 CHAPTER 3 METHODOLOGY	 86
3.1 Finding Requirements of Problem Solving	87
3.2 Context Analysis of Raw Data	87
3.3 Knowledge Capture from Expert Librarian	88
3.4 KMS Design	88
3.5 Algorithm Design	95
3.5.1 Case Retrieval Module	96
3.5.2 Metadata Generating Module	97
3.5.3 Metadata Verification Module	99
3.6 Software Implementation	99

TABLE OF CONTENTS (CONTINUED)

	Page
3.7 Domain Expert Involvement	104
3.8 Result Analysis	105
3.8.1 Evaluation of System Using Qualitative Approach	105
3.8.2 Evaluation of System Using Quantitative Approach	105
CHAPTER 4 EXPERIMENTAL RESULTS	109
4.1 Results Evaluation by Using Qualitative Approach	109
4.2 Results Evaluation by Using Quantitative Approach	111
4.3 Results Discussion	117
4.4 Application	120
4.5 Conclusion	122
CHAPTER 5 CONCLUSION	123
5.1 Research Conclusion	123
5.2 Generalization of this Thesis	124
5.3 Future Works	125
BIBLIOGRAPHY	127
CURRICULUM VITAE	143

LIST OF TABLES

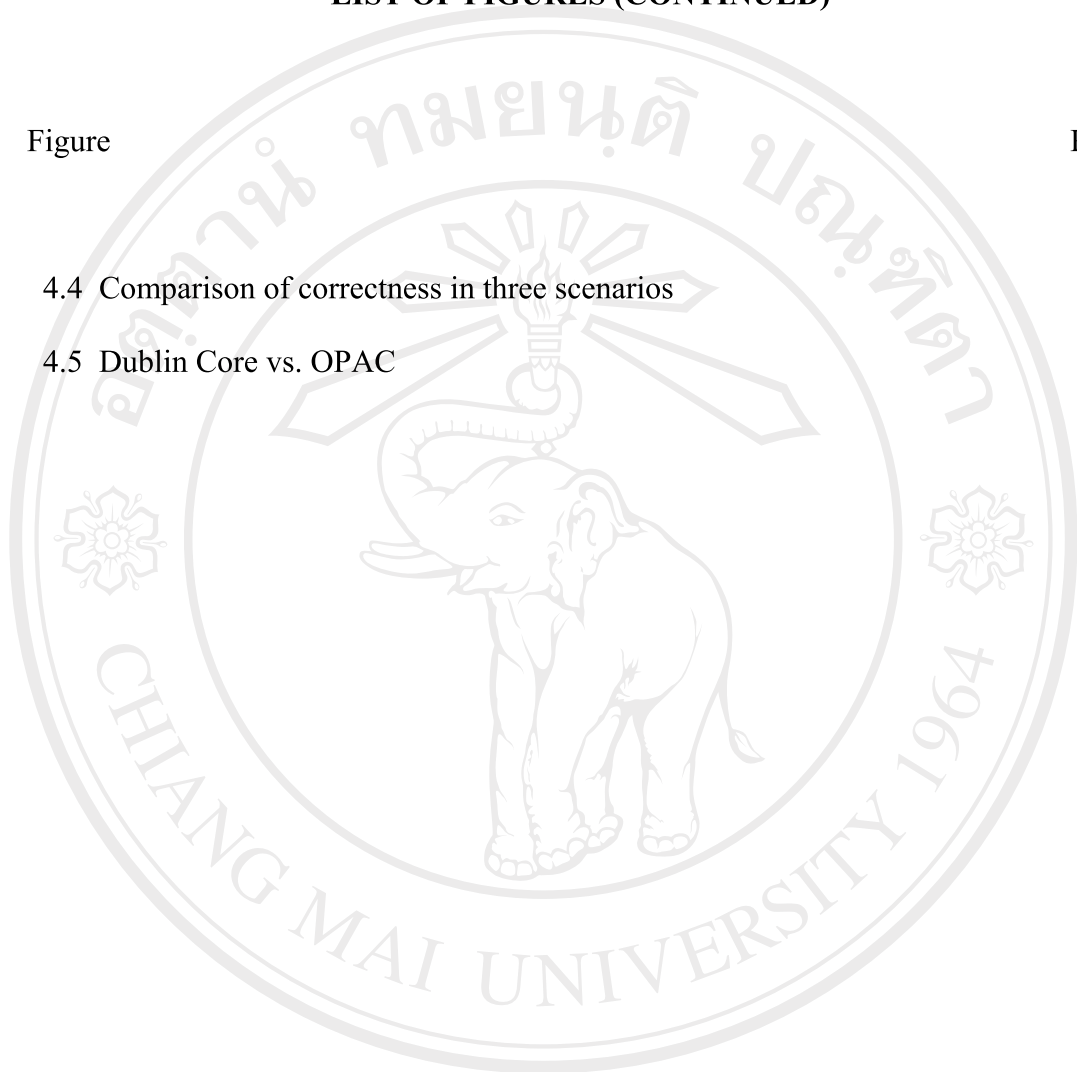
Table	Page
2.1 Detail of element set	34
2.2 Some metadata standardization	41
3.1 CBR cycle VS. KMS cycle	89
4.1 Results of satisfaction	110
4.2 Results from expert librarian	112
4.3 Results from prototype system without guiding from expert	113
4.4 Results from prototype system with 0 retrieved case	113
4.5 Results from prototype system with 63 retrieved cases	114
4.6 Comparison of Prototype and OPAC	120

LIST OF FIGURES

Figure	Page
2.1 Information Extraction system as a black box	59
2.2 Basic modules in Information Extraction system	60
2.3 Example of precision versus recall plot	64
2.4 Example of precision versus recall plot approximated to curve	65
2.5 CBR cycle	69
2.6 KMS development cycle	81
3.1 Prototype processes which agents are used	91
3.2 Flow chart for case retrieval	92
3.3 Architecture of the prototype system	95
3.4 Example of Thai theses abstract and its structure	97
3.5 Example of the thesis abstract header analyzed structure	98
3.6 A reasoning screen	100
3.7 An attribute of metadata screen	101
3.8 A user input screen	102
3.9 A user metadata retrieval screen	103
3.10 A proposed result screen	104
4.1 Comparison of satisfaction in four scenarios	110
4.2 Comparison of precision in four scenarios	115
4.3 Comparison of recall in four scenarios	116

LIST OF FIGURES (CONTINUED)

Figure	Page
4.4 Comparison of correctness in three scenarios	117
4.5 Dublin Core vs. OPAC	121



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved