

ห้องสมุดงานวิจัย สำนักงานคณะกรรมการวิจัยแห่งชาติ



247906

การวิเคราะห์โครงสร้างคำศัพท์ของภาษาล้านนาโดยใช้เทคนิคการคำนวณเมื่อกราฟ

ณัฐนิชา สุรเดช

จิตวิทยาการศึกษาคณะศึกษาศาสตร์
สาขาวิชาจิตวิทยาการคอมพิวเตอร์

บัณฑิตวิทยาลัย
มหาวิทยาลัยเชียงใหม่
พฤษภาคม 2554



การวิเคราะห์โครงสร้างคำของภาษาล้านนาโดยใช้เทคนิคการทำเหมืองกราฟ



ณัฐนิมา สุรเดช

วิทยานิพนธ์นี้เสนอต่อบัณฑิตวิทยาลัยเพื่อเป็นส่วนหนึ่ง
ของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์

บัณฑิตวิทยาลัย
มหาวิทยาลัยเชียงใหม่
พฤษภาคม ๒๕๕๔

การวิเคราะห์โครงสร้างคำของภาษาล้านนาโดยใช้เทคนิคการทำเหมืองกราฟ

ณัฐมีมา สุรเดช

วิทยานิพนธ์นี้ได้รับการพิจารณาอนุมัติให้นำเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิทยาการคอมพิวเตอร์

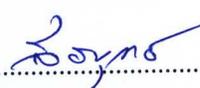
คณะกรรมการสอบวิทยานิพนธ์


.....ประธานกรรมการ
ผู้ช่วยศาสตราจารย์ ดร.รัฐสิทธิ์ สุขะหุด


.....กรรมการ
รองศาสตราจารย์ ดร.อนาโตน โรเจอร์ เป็ลติเยร์


.....กรรมการ
รองศาสตราจารย์ ดร.จिरยุทธ ไชยจรรูวณิช

อาจารย์ที่ปรึกษาวิทยานิพนธ์


.....
รองศาสตราจารย์ ดร.จिरยุทธ ไชยจรรูวณิช

18 พฤษภาคม 2554

© ลิขสิทธิ์ของมหาวิทยาลัยเชียงใหม่

กิตติกรรมประกาศ

วิทยานิพนธ์นี้สำเร็จลงได้ด้วยความกรุณาจาก รองศาสตราจารย์ ดร.จิรยุทธ ไชยจารุณิช อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้ซึ่งกรุณาให้ความรู้ คำแนะนำ คำปรึกษา และตรวจแก้ไขจน วิทยานิพนธ์เสร็จสมบูรณ์ ผู้เขียนขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.รัฐสิทธิ์ สุชะหุด รองศาสตราจารย์ ดร.อนาโตล โรเจอร์ เป็ลติเยร์ และรองศาสตราจารย์ ดร.จิรยุทธ ไชยจารุณิช ที่กรุณาเป็น กรรมการสอบวิทยานิพนธ์ และให้คำแนะนำเป็นอย่างดีตลอดมา

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร.อุดม รุ่งเรืองศรี และอาจารย์เกริก อัครชินเรศ ที่ กรุณาให้ความรู้และคำแนะนำต่างๆเกี่ยวกับภาษาล้านนาเป็นอย่างดีตลอดมา

ขอกราบขอบพระคุณคณาจารย์ประจำภาควิชาวิทยาการคอมพิวเตอร์ทุกท่านที่ได้ให้ความรู้ ทางวิชาการ และทักษะทางคอมพิวเตอร์อันเป็นพื้นฐานที่ดี ส่งผลให้วิทยานิพนธ์นี้สำเร็จได้ด้วยดี ขอขอบพระคุณเจ้าหน้าที่ประจำภาควิชาวิทยาการคอมพิวเตอร์ทุกท่านที่ให้ความสะดวกและความ ช่วยเหลือในเรื่องต่างๆ เป็นอย่างดี

ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ ที่คอยอบรมสั่งสอน ดูแลเอาใจใส่ เป็นกำลังใจและให้ การสนับสนุนในทุกๆด้าน

ท้ายที่สุดนี้ หากมีสิ่งขาดตกบกพร่องหรือผิดพลาดประการใด ผู้เขียนขออภัยเป็นอย่างสูง ในข้อบกพร่องและความผิดพลาดนั้น และผู้เขียนหวังว่าวิทยานิพนธ์นี้คงมีประโยชน์แก่ผู้อ่านบ้าง ไม่มากก็น้อย

ชื่อเรื่องวิทยานิพนธ์	การวิเคราะห์โครงสร้างคำของภาษาล้านนาโดยใช้เทคนิคการทำเหมืองกราฟ
ผู้เขียน	นางสาวณัฐธิมา สุระเดช
ปริญญา	วิทยาศาสตรมหาบัณฑิต (วิทยาการคอมพิวเตอร์)
อาจารย์ที่ปรึกษาวิทยานิพนธ์	รองศาสตราจารย์ ดร. จิรยุทธ ไชยจารุณนิช

บทคัดย่อ

247906

งานวิจัยนี้ นำเสนอการวิเคราะห์โครงสร้างคำภาษาล้านนาด้วยเทคนิคการทำเหมืองกราฟ โดยการนำเสนอคำภาษาล้านนาให้อยู่ในรูปแบบกราฟที่เป็นตัวแทนของคำภาษาล้านนาทั้งหมด 77 รูปแบบ ซึ่งแบ่งเป็น 5 ประเภทคือ กราฟของโครงสร้างคำผสมสระ โครงสร้างคำที่มีตัวสะกดในตำแหน่งด้านล่างพยัญชนะต้น โครงสร้างคำที่มีตัวสะกดในตำแหน่งด้านหลังพยัญชนะต้น โครงสร้างคำที่มีพยัญชนะซ้อน และโครงสร้างคำที่มีตัวไหล ในการวิเคราะห์ได้ทำการแปลงคำภาษาล้านนาให้อยู่ในรูปแบบของกราฟตามที่ได้ออกแบบไว้ จากนั้นนำกราฟที่เป็นตัวแทนของคำภาษาล้านนาดังกล่าวไปทำการวิเคราะห์ด้วยเทคนิคการทำเหมืองกราฟ พบว่าผลที่ได้จากการวิเคราะห์คือลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำทั้ง 5 ประเภท ซึ่งผลที่ได้จากการวิเคราะห์ด้วยเทคนิคการทำเหมืองกราฟแสดงให้เห็นโครงสร้างและความสัมพันธ์ของตัวอักษรภายในคำได้ชัดเจน

Thesis Title	Analysis of Lanna Morphology Using Graph Mining Technique
Author	Miss Natsima Suradet
Degree	Master of Science (Computer Science)
Thesis Advisor	Associate Professor Dr. Jeerayut Chaijaruwanich

ABSTRACT

247906

In this thesis, we propose graphs representation for Lanna words, the novel graph patterns approach for mining the frequent patterns of Lanna word. To analyze the Lanna morphology through the structure and relationship between letters in Lanna words, we designed 77 graph structures which are divided into five groups. In the experiments, the five groups of graph are constructed from Lanna word. Then, the frequent subgraphs are discovered by using graph mining technique. Experimental results demonstrated that the Lanna words can be represented through their related graph patterns.

สารบัญ

	หน้า
กิตติกรรมประกาศ	ก
บทคัดย่อภาษาไทย	ง
บทคัดย่อภาษาอังกฤษ	จ
สารบัญ	ฉ
สารบัญตาราง	ช
สารบัญภาพ	ฉ
บทที่ 1 บทนำ	1
1.1 หลักการและเหตุผล	1
1.2 วัตถุประสงค์ของการวิจัย	2
1.3 ประโยชน์ที่คาดว่าจะได้รับ	2
1.4 แผนดำเนินการ ขอบเขต และวิธีการวิจัย	2
บทที่ 2 ทฤษฎี งานวิจัย และเอกสารที่เกี่ยวข้อง	5
2.1 การทำเหมืองกราฟ	5
2.2 รายละเอียดเกี่ยวกับ SUBDUE โดยสังเขป	8
2.3 การเข้ารหัส Minimum Description Length ของกราฟ	12
2.4 โครงสร้างคำภาษาล้านนา	14
บทที่ 3 วิธีการดำเนินการวิจัย	19
3.1 การออกแบบกราฟที่เป็นตัวแทนของโครงสร้างคำภาษาล้านนา	20
3.2 การแปลงคำภาษาล้านนาให้อยู่ในรูปแบบกราฟ	23
บทที่ 4 ผลการวิจัย	38
4.1 การเตรียมชุดข้อมูล	38
4.2 การวิเคราะห์โครงสร้างคำภาษาล้านนา	40
4.3 ผลการวิจัย	40
4.4 สรุปผลการวิจัย	41

บทที่ 5 บทสรุป	50
5.1 บทสรุป	50
5.2 ปัญหาและอุปสรรค	51
5.3 ข้อเสนอแนะ	52
เอกสารอ้างอิง	53
ภาคผนวก	54
ภาคผนวก ก การออกแบบกราฟที่เป็นตัวแทนของโครงสร้างคำภาษาล้านนา	55
ภาคผนวก ข ภาพเอกสารภาษาล้านนาเรื่องคร่าวขอประวัติครูบาภิไชยขาวปี	82
ภาคผนวก ค โครงสร้างคำที่ปรากฏบ่อยในเอกสารที่ตรวจสอบโดยผู้เชี่ยวชาญ	100
ประวัติผู้เขียน	119

สารบัญตาราง

ตาราง	หน้า
2.1 การเทียบพยัญชนะภาษาล้านนากับพยัญชนะภาษาไทย	15
2.2 การเทียบสระภาษาล้านนากับสระภาษาไทย	15
4.1 ตัวอักษร 10 อันดับแรกที่มีความถี่สูงในคำผสมสระ	46
4.2 ตัวอักษร 10 อันดับแรกที่มีความถี่สูงในคำที่มีตัวสะกดวางไว้ด้านล่างพยัญชนะต้น	47
4.3 ตัวอักษร 10 อันดับแรกที่มีความถี่สูงในคำที่มีตัวสะกดวางไว้ด้านหลังพยัญชนะต้น	47
4.4 ตัวอักษร 10 อันดับแรกที่มีความถี่สูงในคำที่มีพยัญชนะซ้อน	47
4.5 ตัวอักษร 10 อันดับแรกที่มีความถี่สูงในคำที่มีตัวไหล	48
5.1 ตัวอย่างคำภาษาล้านนาที่มีรูปแบบการเขียนแตกต่างกัน	51
5.2 ตัวอย่างคำยากที่เกิดจากการนำคำสองคำมาเขียนรวมกัน	52
5.3 ตัวอย่างคำยากที่เกิดจากการเขียนคำหลายๆคำรวมกัน	52

สารบัญภาพ

รูป	หน้า
2.1 ตัวอย่างชุดข้อมูลกราฟ (graph dataset)	5
2.2 กราฟที่ปรากฏบ่อย	5
2.3 ขั้นตอนวิธี Apriori	6
2.4 ขั้นตอนวิธี PatternGrowth	7
2.5 ขั้นตอนวิธีการค้นหาของ SUBDUE	9
2.6 ตัวอย่างกราฟ	12
2.7 โครงสร้างคำผสมสระ	16
3.1 ขั้นตอนการดำเนินการจับ	19
3.2 ตัวอย่างโครงสร้างคำและรูปแบบกราฟของคำผสมสระ	20
3.3 ตัวอย่างโครงสร้างคำและรูปแบบกราฟของคำที่มีตัวสะกดวางไว้ด้านล่างพยัญชนะต้น	20
3.4 ตัวอย่างโครงสร้างคำและรูปแบบกราฟของคำที่มีตัวสะกดวางไว้ด้านหลังพยัญชนะต้น	21
3.5 ตัวอย่างโครงสร้างคำและรูปแบบกราฟของคำที่มีพยัญชนะซ้อน	21
3.6 ตัวอย่างโครงสร้างคำและรูปแบบกราฟของคำที่มีตัวไหล	21
3.7 ตัวอย่างการทำงานของขั้นตอนวิธีการแปลงคำผสมสระ	24
3.8 ตัวอย่างการทำงานของขั้นตอนวิธีการแปลงคำที่มีตัวสะกดวางไว้ด้านล่างพยัญชนะต้น	26
3.9 ตัวอย่างการทำงานของขั้นตอนวิธีการแปลงคำที่มีตัวสะกดวางไว้ด้านหลังพยัญชนะต้น	28
3.10 ตัวอย่างการทำงานของขั้นตอนวิธีการแปลงคำที่มีพยัญชนะซ้อน	30
3.11 ตัวอย่างการทำงานของขั้นตอนวิธีการแปลงคำที่มีตัวไหล	33
4.1 ลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำผสมสระ	41
4.2 ลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำที่มีตัวสะกดวางไว้ด้านล่างพยัญชนะต้น	42
4.3 ลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำที่มีตัวสะกดวางไว้ด้านหลังพยัญชนะต้น	44
4.4 ลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำที่มีพยัญชนะซ้อน	45

4.5 ลักษณะโครงสร้างคำที่ปรากฏบ่อยของคำที่มีตัวไหล	46
ก.1 รูปแบบกราฟของคำที่มีสระวางไว้ด้านหน้าพยัญชนะต้น	56
ก.2 รูปแบบกราฟของคำที่มีสระวางไว้ด้านหลังพยัญชนะต้น	56
ก.3 รูปแบบกราฟของคำที่มีสระวางไว้ด้านบนพยัญชนะต้น	57
ก.4 รูปแบบกราฟของคำที่มีสระวางไว้ด้านล่างพยัญชนะต้น	57
ก.5 รูปแบบกราฟของคำที่ผสมกับไม้เเกะ ไม้เเกะ ไม้โกะ และไม้โก้ (บาลี)	57
ก.6 รูปแบบกราฟของคำที่ผสมกับไม้เเกะ	58
ก.7 รูปแบบกราฟของคำที่ผสมกับไม้กั๊วะ	58
ก.8 รูปแบบกราฟของคำที่ผสมกับไม้เเกียะ	58
ก.9 รูปแบบกราฟของคำที่ผสมกับไม้เเกีย	58
ก.10 รูปแบบกราฟของคำที่ผสมกับไม้เเกื้อะ	59
ก.11 รูปแบบกราฟของคำที่ผสมกับไม้เเกื้อ	59
ก.12 รูปแบบกราฟของคำที่ผสมกับไม้เเกื้อะ	59
ก.13 รูปแบบกราฟของคำที่ผสมกับไม้เเกื้อ	59
ก.14 รูปแบบกราฟของคำที่ผสมกับไม้เเก้า	60
ก.15 รูปแบบกราฟของคำที่ผสมกับไม้เเก้าย	60
ก.16 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้เเกะ ไม้กั๊ ไม้กั๊ ไม้กั๊ และไม้กั๊	60
ก.17 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้ก้า	61
ก.18 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้เเกะ และไม้เเก้	61
ก.19 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้เเกะ	61
ก.20 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้โก้	61
ก.21 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้โกะ	62
ก.22 รูปแบบกราฟของคำที่มีตัวสะกดกลางเมื่อผสมกับไม้เเกื้อ	62
ก.23 รูปแบบกราฟของคำที่มีตัวสะกดหลังเมื่อผสมกับไม้กั๊ ไม้กั๊ ไม้กั๊ และไม้กั๊	62
ก.24 รูปแบบกราฟของคำที่มีตัวสะกดหลังเมื่อผสมกับไม้เเกีย	63
ก.25 รูปแบบกราฟของคำที่มีตัวสะกดหลังเมื่อผสมกับไม้เเกื้อ	63
ก.26 รูปแบบกราฟของคำที่มีพยัญชนะซ้อนชั้นเดียวที่ไม่มีสระ	63
ก.27 รูปแบบกราฟของคำที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้ก้า	64
ก.28 รูปแบบกราฟของคำที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้กั๊	64

ก.47 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้เก้และมีไม้เก้อยู่หน้าพยัญชนะต้น	70
ก.48 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้กู่และมีไม้เก้อยู่หน้าพยัญชนะต้น	71
ก.49 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้เก้และมีไม้เก้	71
ก.50 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้โก้และมีไม้เก้	71
ก.51 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้โก้(บาลี) และมีไม้เก้	72
ก.52 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้กั้งและมีไม้เก้	72
ก.53 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่มีตัวการ์นต์และมีพยัญออก	73
ก.54 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่มีพยัญออกอยู่บนพยัญชนะต้น	73
ก.55 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้กำและมีพยัญออก	73
ก.56 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้เก้และมีพยัญออก	74
ก.57 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนชั้นเดียวที่ผสมกับไม้กู่และมีไม้เก้	74
ก.58 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนสองชั้น	74
ก.59 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนสองชั้นที่ผสมกับไม้เก้	75
ก.60 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนสองชั้นที่ผสมกับไม้กู่	75
ก.61 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนสองชั้นที่ผสมกับไม้กู่และไม้เก้	76
ก.62 รูปแบบกราฟของค่าที่มีพยัญชนะซ้อนสามชั้น	76
ก.63 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระหน้า	77
ก.64 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระหลัง	77
ก.65 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระบน	77
ก.66 รูปแบบกราฟของค่าที่มีคำตัวไหลที่ผสมกับสระล่าง	78
ก.67 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระหน้าและมีตัวสะกด	78
ก.68 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระหลังและมีตัวสะกด	78
ก.69 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระบนและมีตัวสะกด	79
ก.70 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระล่างและมีตัวสะกด	79
ก.71 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับสระอำ	79
ก.72 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับพยัญออก	80
ก.73 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับไม้หันอากาศ	80
ก.74 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับพยัญออกและไม้กั่ว	80

ฐ

- ก.75 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับไม้เกือ 81
- ก.76 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับไม้เกีย 81
- ก.77 รูปแบบกราฟของค่าที่มีตัวไหลที่ผสมกับพยัญชนะซ้อน 81