

การพัฒนาผังความสัมพันธ์ของวิจัยจากผลงานเผยแพร่วิทยานิพนธ์นักศึกษาระดับบัณฑิตศึกษา
ของมหาวิทยาลัยขอนแก่นด้วยเทคโนโลยีเหมืองข้อความ

Implementing KGU Research Map from Graduate Publications by Text Mining

อภิเชษฐ์ บุญจง (Aphichet Boonjaung)* ดร.วนิดา แก่นอากาศ (Dr.Wanida Kanarkard)**

บทคัดย่อ

ในงานวิจัยนี้ได้นำเสนอการพัฒนาผังความสัมพันธ์ของวิจัยจากผลงานเผยแพร่วิทยานิพนธ์นักศึกษาระดับบัณฑิตศึกษาของมหาวิทยาลัยขอนแก่นระหว่างปี 2550-2556 โดยมีจำนวน 29,067 ฉบับ ซึ่งนำเอาองค์ความรู้ในการนำเสนอศักยภาพของผลงานวิทยานิพนธ์ในรูปแบบที่สามารถดูได้ง่ายในรูปแบบของผังความสัมพันธ์ (topic maps) ที่สามารถทำได้ในเวลาอันสั้นด้วยเทคโนโลยีการทำเหมืองข้อมูล (data mining) และสามารถเป็นแนวทางเลือกหนึ่งที่สะท้อนศักยภาพด้านการวิจัยของมหาวิทยาลัยขอนแก่นได้เป็นอย่างดี แนวทางการพัฒนาต่อจากผังความสัมพันธ์สามารถนำเครื่องมือการสืบค้นมาเพิ่มศักยภาพในการค้นหาความสัมพันธ์ที่ต้องการให้เร็วยิ่งขึ้น ถึงแม้จะมีการนำเสนอข้อมูลมหาศาลในรูปแบบกราฟที่ง่ายต่อการทำความเข้าใจ แต่ยังไม่สามารถสืบค้นได้ลงไปยังละเอียดเฉพาะเรื่องได้

ABSTRACT

This research presents the implementation of research map, from the total number of 29,067 Khon Kaen University's graduate publications in 2007 – 2013. The implementation adopts knowledge in presenting publication data in a form of accessible topic maps, which could be completed within a short period of time using data mining technology, and could be an alternative which well reflects research capabilities of Khon Kaen University. Such implementation of research map is able to apply searching tools to increase the speed of searching abilities in data relationships needed. In spite of the accessibility of using graphs to present a large volume of data; however, it is unable to search for detailed specific topics.

คำสำคัญ: เหมืองข้อความ ผังสัมพันธ์หัวข้อ ผังสัมพันธ์งานวิจัย

Key Words: Text mining, Topic map, Research map

* นักศึกษา หลักสูตรวิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น

** รองศาสตราจารย์ ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยขอนแก่น

บทนำ

ในปี 2552 บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น ได้เริ่มกลยุทธ์ที่สำคัญในกำหนดทิศทางการวิจัยของ บัณฑิตศึกษาให้เชื่อมโยงกับกรอบและทิศทางการวิจัย ของมหาวิทยาลัยและศูนย์วิจัยเฉพาะทางของ มหาวิทยาลัยขอนแก่น และเชื่อมโยงผลงานวิจัยของ บัณฑิตศึกษากับฝ่ายวิจัยในการทำผลงานการวิจัยไป ถ่ายทอดองค์ความรู้สู่ชุมชนและการสร้างนวัตกรรม ใหม่ โดยมีโครงการดังนี้ 1) โครงการพัฒนารอบทิศ ทางการวิจัยของกลุ่มสาขาให้สอดคล้องกับมหาวิทยาลัย 2) วิเคราะห์ห้องสมุดความรู้และความเชี่ยวชาญของ มหาวิทยาลัยขอนแก่นจากข้อมูลวิทยานิพนธ์ เพื่อเป็น ข้อมูลสะท้อนศักยภาพด้านงานวิจัยของนักศึกษาและ คณาจารย์ระดับบัณฑิตศึกษา แต่ใช้เวลานานถึง 3 ปีใน การพัฒนาโดยใช้นักสารสนเทศ ทำการสกัดข้อมูลด้วย มือ และป้อนข้อมูลที่ได้จากการสกัดองค์ความรู้นี้ไปยัง ตัวคณาจารย์เพื่อให้มีการรับรองความถูกต้อง แต่ก็ยังไม่ สามารถดำเนินการได้สำเร็จลุล่วงตามที่วาง วัตถุประสงค์ในข้างต้น รวมถึงการนำเสนอในรูปแบบ ตารางที่มีปริมาณมาก การค้นหาเฉพาะเจาะจงเป็นไปได้ ลำบาก ดังนั้นเพื่อให้ได้ข้อมูลที่ถูกต้อง รวดเร็ว

ดังนั้นจึงได้มีการมองหาทางเลือกในการพัฒนา องค์ความรู้ที่ได้จากการสกัดจากข้อมูลวิทยานิพนธ์ที่ จัดเก็บโดยบัณฑิตวิทยาลัย และการเผยแพร่ผลงานของ นักศึกษาในฐานะข้อมูลตีพิมพ์ในแหล่งตีพิมพ์ต่างๆ ทั้ง นอคมหาวิทยาลัยและในวารสารของบัณฑิตวิทยาลัย เอง ดังนั้นด้วยคุณลักษณะของปัญหาที่พบ จึงได้ มองเห็นประโยชน์ของเทคโนโลยีเหมืองข้อความ (Text Mining) ในการสกัดองค์ความรู้ และจัดกลุ่มหัวข้อโดย กำหนดคำสำคัญ ในการสกัดองค์ความรู้และแสดงผล ในรูปแบบผังความสัมพันธ์ (Topic Map)

วัตถุประสงค์การวิจัย

1. พัฒนาระบบเหมืองข้อความในการคัดแยกคำสำคัญ (Keywords) จากงานวิทยานิพนธ์ของนักศึกษา

บัณฑิตศึกษามหาวิทยาลัยขอนแก่น และจากการ เผยแพร่ผลงานวิทยานิพนธ์

2. พัฒนาความสัมพันธ์ของคำสำคัญ เชื่อมโยงไปยัง ข้อมูลอาจารย์ เวลา คณะ สถานที่ การนำไปใช้
3. เพื่อสะท้อนศักยภาพของผลงานวิทยานิพนธ์ใน รูปแบบที่สามารถดูได้ง่าย (Visualization)

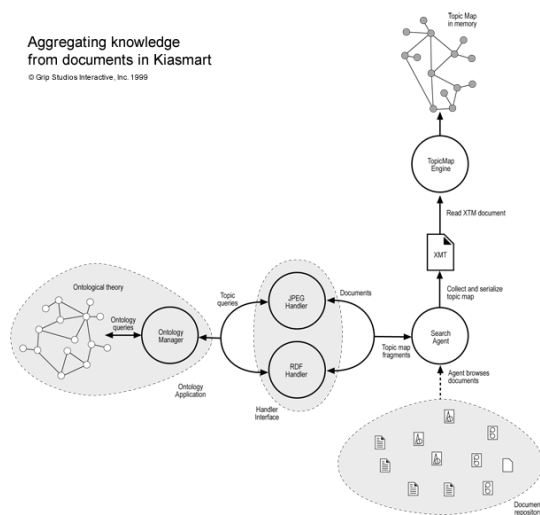
งานวิจัยที่เกี่ยวข้อง

เหมืองข้อความ (Archibugi, Pianta, 1996) เป็นการ ประยุกต์ใช้เทคนิคการทำเหมืองข้อมูล (data mining) กับ แฟ้มข้อมูลแบบตัวอักษรและเป็นแบบไม่มีโครงสร้าง (non- structured) หรือ มีโครงสร้างน้อย (less-structured) เป็นการค้นหาความรู้ใหม่จากข้อมูลประเภทข้อความที่ มีปริมาณมากโดยอัตโนมัติ โดยการสกัดคำ ค้นหา รูปแบบ และความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อความ เอกสาร เพื่อให้เกิดความหมายและสามารถนำมาใช้ ประโยชน์ได้

คุณลักษณะเด่นของเทคโนโลยีนี้คือ (Cañas, Carff, Hill, Carvalho, Arguedas, Eskridge, 2005)

- ค้นหาเนื้อหาที่แฝงเร้น ("hidden" content) ของ เอกสารต่างๆ
- รวมทั้งแสดงความสัมพันธ์กันที่มีประโยชน์ เพิ่มเติมเข้ามาด้วย
- จัดกลุ่มเอกสารให้อยู่ในเรื่องเดียวกัน (common themes)

ซึ่งได้มีงานวิจัยหลายชิ้น (Chen, 1999) (Chen, Kinshuk, Wei, Chen, 2008) (Jia, Wang, Ran, Yang, Liao, Chiu, 2011) (Pepper, 2000) ที่นำเอาพัฒนางานในสาขา ต่างๆ ที่ประสบปัญหาเกี่ยวกับจำนวนเอกสารจำนวนมาก ที่มีลักษณะความต้องการในการหาความสัมพันธ์ของ ข้อความกับงานวิจัยนี้ รูปที่ 1 แสดงตัวอย่างงานวิจัยที่ สกัดองค์ความรู้เพื่อนำไปใช้ในการวางแผนเชิงธุรกิจ จากเอกสารสื่อประชาสัมพันธ์



รูปที่ 1 งานวิจัยที่เกี่ยวข้องที่ใช้เทคนิคสร้างผังความสัมพันธ์ของ Kiasmart (Wandora)

วิธีการวิจัย

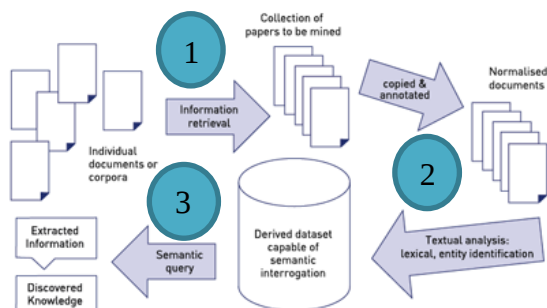
1. ข้อมูลที่ใช้ในการศึกษา

งานวิจัยระดับบัณฑิตศึกษาของมหาวิทยาลัยขอนแก่น ที่นำมาศึกษาในระหว่างปี 2550-2556 ได้มาจากแหล่งข้อมูลดังต่อไปนี้

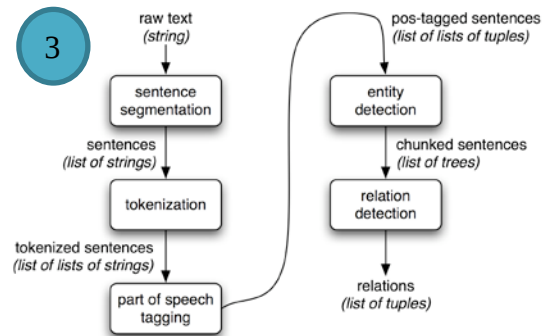
1. ฐานข้อมูลตีพิมพ์เผยแพร่วิทยานิพนธ์
2. ฐานข้อมูลวิทยานิพนธ์
3. ฐานข้อมูลวารสารวิจัย มหาวิทยาลัยขอนแก่น ฉบับบัณฑิตศึกษา

ที่มีการจัดเก็บในรูปแบบไฟล์ pdf, และบริหารจัดการข้อมูลรายการบนเว็บ

2. หลักการทำเหมืองข้อความ (Text Mining)



รูปที่ 2 ขั้นตอนวิธีการเหมืองข้อมูลเอกสาร (text mining) (Fayyad, Piatetsky-Shapiro, Smyth, Uthurasamy, 1996)



รูปที่ 3 ขั้นตอนวิธีการสกัด topic จากข้อความในเอกสาร (Cañas, Carff, Hill, Carvalho, Arguedas, Eskridge, 2005)

3. การสร้างผังสัมพันธ์

งานวิจัยนี้ได้เลือกใช้เครื่องมือ wandora (Wandora) ในการจัดเก็บโครงสร้างความสัมพันธ์ของข้อมูลวิจัยระดับบัณฑิตศึกษา (ontology) ในรูปของ tree และสร้างผังความสัมพันธ์โดยกำหนดความเชื่อมโยง (association) ระหว่างกลุ่ม และนำเสนอข้อมูลผังความสัมพันธ์ด้วยกราฟ (visualization) เพื่อสร้างความเข้าใจได้ง่ายขึ้นในแต่ละความสัมพันธ์

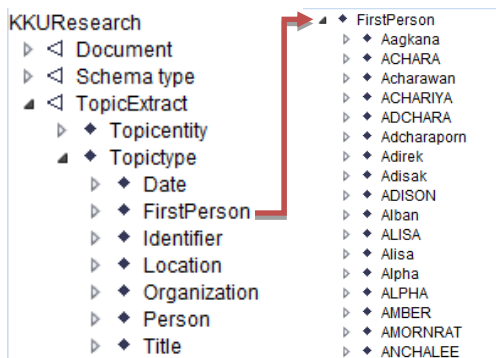
ผลการวิจัย

จากแหล่งข้อมูล 3 แห่งมีเอกสารที่นำมาศึกษาทั้งสิ้น 29,067 ฉบับ ดังแสดงในตารางที่ 1

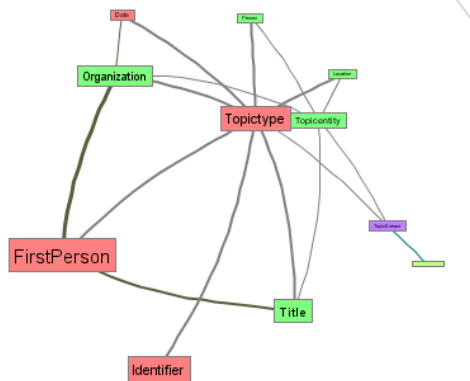
ตารางที่ 1 ข้อมูลผลงานวิทยานิพนธ์และค่าความสัมพันธ์ในงานวิจัย

จำนวนผลงาน (documents)	29,067
จำนวนความสัมพันธ์ระหว่างหัวข้อ (association)	165,126
จำนวนวันที่ (date)	59
จำนวนชื่อผู้แต่ง (first person)	671
จำนวนสถานที่ (location)	576
จำนวนหน่วยงาน (organization)	1,523
จำนวนคำนำหน้าชื่อ (title)	19
จำนวนประเภทข้อมูลงานวิจัย	7

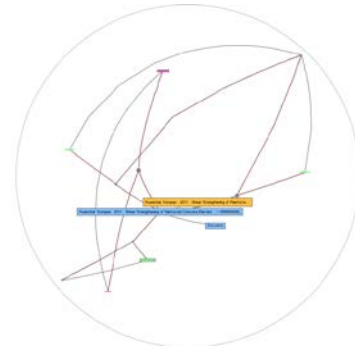
ผลงานวิทยานิพนธ์ที่อยู่ในรูปเอกสาร .doc, .pdf, .txt ที่ประกอบด้วย ชื่อบทความ ชื่อผู้แต่ง ชื่อผู้แต่งร่วม ชื่อสถานที่ ชื่อหน่วยงาน คำนำหน้าชื่อผู้แต่ง คำสำคัญและวันที่ที่เผยแพร่เอกสาร ที่จัดเก็บในระบบสารสนเทศบนเว็บ เมื่อผ่านขั้นตอนการทำเหมืองข้อความจะได้ประเภทข้อมูลวิจัย 7 ประเภท ดังแสดงในรูปที่ 3



รูปที่ 3 โครงสร้างความสัมพันธ์ของงานวิจัย (ontology) ในรูปของ Tree



รูปที่ 4 แผนผังความสัมพันธ์ระหว่างประเภทข้อมูลงานวิจัยทั้ง 7 ประเภท

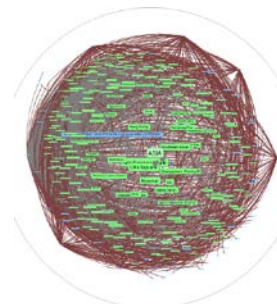


รูปที่ 5 ตัวอย่างเอกสารที่เผยแพร่ในวารสารวิจัยที่ถูกสกัดคำสำคัญและความเชื่อมโยงระหว่างหัวข้อกับคำสำคัญ

ในการสะท้อนศักยภาพงานวิจัยสามารถพิจารณาได้จาก keywords ข้อมูลในตารางที่ 3 นำเสนอจำนวน keywords และเส้นความสัมพันธ์ระหว่างคำสำคัญ และในรูปที่ 6 เป็นการนำเสนอในรูปของกราฟ

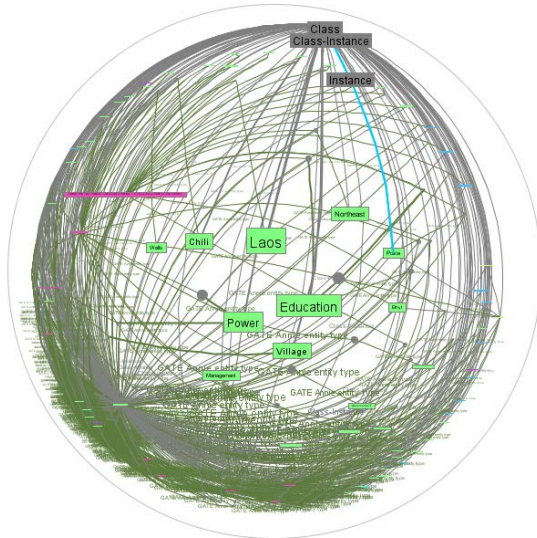
ตารางที่ 3 คำสำคัญของงานวิจัย (keywords)

จำนวนหัวข้อ (topics)	425
ความสัมพันธ์ระหว่างคำ (edge)	1,091



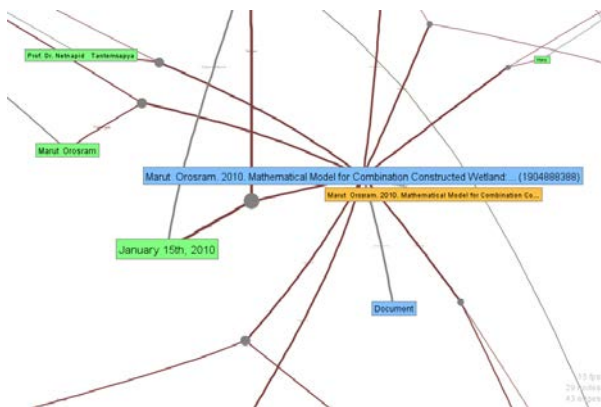
รูปที่ 6 กราฟแสดงความสัมพันธ์ระหว่างหัวข้อ

แสดงการขยายหัวข้อที่มีจำนวนเส้นเชื่อมโยงที่มากที่สุด แสดงให้เห็นถึงความเข้มแข็งในงานวิทยานิพนธ์ที่นักศึกษานิเทศศาสตร์



รูปที่ 7 บางส่วนของคำสำคัญงานวิจัยระดับ
บัณฑิตศึกษามหาวิทยาลัยขอนแก่นที่มีการทำ
วิจัยมากในงานวิทยานิพนธ์

ผลงานที่ได้สุดท้ายจากฝั่งความสัมพันธ์งานวิจัยจากผลงานวิทยานิพนธ์สามารถทำให้เลือกดูหัวข้อวิจัยที่นักศึกษาทำมากที่สุด และยังโยงไปยังผู้วิจัยหลัก หรือวิจัยร่วมคนอื่น ที่ยังมีข้อมูลสถานที่ และหน่วยงานที่ผู้วิจัยสังกัดได้ด้วย เครื่องมือการนำเสนอด้วยกราฟดังแสดงในรูปที่ 8



รูปที่ 8 แผนภูมิความสัมพันธ์ระหว่างหัวข้อที่เชื่อมโยง
ไปยังแต่ละประเภทของข้อมูลวิจัย

อภิปรายและสรุปผลการวิจัย

ผลลัพธ์ที่ได้จากงานวิจัยนี้คือ องค์ความรู้ในการนำเสนอศักยภาพของผลงานวิทยานิพนธ์ในรูปแบบที่สามารถดูได้ง่ายในรูปของผังความสัมพันธ์ (topic maps) ที่สามารถทำได้ในเวลาอันสั้นด้วยเทคโนโลยีการทำเหมืองข้อมูล (data mining) เป็นแนวทางเลือกหนึ่งที่เหมาะสมศักยภาพด้านการวิจัยของมหาวิทยาลัยขอนแก่นได้เป็นอย่างดี แนวทางการพัฒนาต่อยอดจากผังความสัมพันธ์สามารถนำเครื่องการสืบค้นมาเพิ่มศักยภาพในการค้นหาคำความสัมพันธ์ที่ต้องการให้เร็วยิ่งขึ้น ถึงแม้จะมีการนำเสนอข้อมูลมหาศาลในรูปแบบกราฟที่ง่ายต่อการทำความเข้าใจ แต่ยังไม่สามารถสืบค้นได้ลงไปยังละเอียดเฉพาะเรื่องได้

เอกสารอ้างอิง

- Archibugi D., Pianta Measuring M. technological change through patents and innovation survey Technovation, 16 (9) (1996), pp. 451–468
- Cañas AJ., Carff R., Hill G., Carvalho M., Arguedas M., Eskridge, TC. et al. Concept maps: integrating knowledge and information visualization Lecture Notes in Computer Science, 3426/2005 (2005), pp. 181–184
- Chen B., 1999. Introduction to patent map. Lecture notes for the training of patent mapping and patent analysis. Taipei, National Science Council (in Chinese).
- Chen, Kinshuk NS., Wei CW., Chen HJ., Mining e-learning domain concept map from academic articles Computers & Education, 50 (3) 2008, pp. 1009–1021

Fayyad U., Piatetsky-Shapiro G., Smyth P.,

Uthurasamy R. Advances in knowledge
discovery and data mining AAAI Press/The
MIT Press (1996)

Jia H., Wang M., Ran W., Yang S.J.H., Liao J., Chiu
D.K.W. Design of a performance-oriented
workplace e-learning system using ontology
Expert Systems with Applications, 38
(2011), pp. 3372–3382

Pepper S. The TAO of topic maps Proceedings of
XML Europe (2000)

Wandora, <http://www.wandora.org>