

วิธีที่เหมาะสมสำหรับการเทียบโอนรายวิชาโดยการหาค่าความคล้ายคลึงเชิงความหมาย

A Proper Method for a Courses Transfer by Semantic Similarity

วุฒิชัย วิเชียรไชย¹

Vuttichai Vichianchai¹

Received: 30 May 2016 ; Accepted: 5 October 2016

บทคัดย่อ

บทความวิจัยนี้ได้เสนอการเทียบโอนรายวิชาโดยวิธีการหาค่าความคล้ายคลึงเชิงความหมาย เพื่อหาวิธีที่เหมาะสมที่สุดในการเทียบโอนคำอธิบายรายวิชาโดยการประยุกต์ใช้วิธีหาค่าความคล้ายคลึงเชิงความหมาย 5 วิธี ได้แก่ วูและพาล์มเมอร์, เลียค็อคคักและโคโดโรว์, เรสนิค, ลิน และเจียงและคอนราธ โดยแบ่งขั้นตอนการเทียบโอนรายวิชาออกเป็น 3 ขั้นตอนคือ การแบ่งคำแล้วคัดกรองคำ การหาค่าความคล้ายคลึงเชิงความหมายของคำอธิบายรายวิชาของแต่ละวิธีและเรียงลำดับเอกสารที่มีความคล้ายคลึงจากมากไปหาน้อย งานวิจัยนี้ได้ทดสอบประสิทธิภาพ โดยใช้คำอธิบายของรายวิชาของหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง สาขาวิชาเทคโนโลยีสารสนเทศและคำอธิบายรายวิชาของหลักสูตรวิทยาศาสตรบัณฑิต สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม พบว่าวิธีที่เหมาะสมมากที่สุดในการเทียบโอนรายวิชาคือ วิธีการของเจียงและคอนราธ มีค่าความแม่นยำเท่ากับร้อยละ 72 และวิธีที่เหมาะสมน้อยที่สุดในการเทียบโอนรายวิชาคือ วิธีการของเรสนิค มีค่าความแม่นยำเท่ากับร้อยละ 5

คำสำคัญ: การเทียบโอนรายวิชา การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ ความคล้ายคลึงเชิงความหมาย เวิร์ดเน็ต

Abstract

This research project was under taken to find a proper similarity method for a course transfer by algorithms in Word Net, The research included Wu and Palmer, Leacock and Chodorow, Resnik, Lin and Jiang and Conrath. The concept of research, 3 main steps, consisting of word segmentation and stop word, calculation of the semantic similarity of course description and ranking courses by semantic similarity respectively. The evaluation used an English description of the course in the Diploma course, and the description of the course in the Bachelor of Science degree, at Mahasarakham University. It was found that the most proper method is the Jiang and Conrath method, which provides precision equal to 84%, and the least proper method is the Resnik method, which provides precision equal to 10%.

Keywords: Courses transfer, Cosine Similarity Measurement, Semantic similarity, WordNet

บทนำ

มหาวิทยาลัยในประเทศไทยมีการเปิดรับนิสิต นักศึกษาเพื่อเข้าเรียนในหลักสูตรระบบพิเศษเทียบเข้า ซึ่งหลักสูตรดังกล่าวสามารถนำรายวิชาจากสถาบันที่ตนเองจบเข้ามาเทียบโอนกับรายวิชาในหลักสูตรที่สอบเข้าศึกษาได้ โดยมีเงื่อนไขคือ เนื้อหาของคำอธิบายรายวิชาที่นำมาเทียบโอนจะต้องครอบคลุมเนื้อหาของรายวิชาที่ทำการขอเทียบ และทำการนำเสนอขอเทียบโอนให้อาจารย์ที่ปรึกษาและอาจารย์ประจำวิชาหรือ

ผู้เชี่ยวชาญในรายวิชาดังกล่าวเป็นผู้พิจารณาการเทียบโอน แต่เนื่องจากรายวิชาในหลักสูตรมีจำนวนมาก ทำให้เกิดความล่าช้าในการพิจารณา จึงจำเป็นต้องมีวิธีการเทียบโอนที่ช่วยสนับสนุนการพิจารณาบนระบบคอมพิวเตอร์เพื่อให้เกิดความสะดวกและรวดเร็วในการเทียบโอนรายวิชา

จากการศึกษาวิธีการหาค่าความคล้ายคลึงเชิงความหมายวิธีการของวูและพาล์มเมอร์ (Wu and Palmer)¹, เลียค็อคคักและโคโดโรว์ (Leacock and Chodorow)², เรสนิค

¹ อาจารย์, คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ตำบลขามเรียง อำเภอกันทรวิชัย จังหวัดมหาสารคาม 44150

¹ Lecturer, Intellect Laboratory, Faculty of Informatics, Mahasarakham University, Khamriang Sub-District, Kantharawichai District, Mahasarakham Province 44150, Thailand

(Resnik)³, ลิน (Lin)⁴ และเจียงและคอนราท (Jiang and Conrath)⁵ เป็นวิธีที่ได้รับความนิยมอย่างมากในการนำมาใช้ในงานวิจัยหลายงาน ดังนั้นจึงมีแนวคิดในการนำเอาวิธีการดังกล่าวมาใช้ประยุกต์ใช้ในการเทียบโอนรายวิชา เพื่อหาวิธีการที่เหมาะสมที่สุดในการพัฒนาโปรแกรมต้นแบบสำหรับการเทียบโอนรายวิชา

ทฤษฎีที่เกี่ยวข้อง

1) เวิร์ดเน็ตคือ ฐานข้อมูลคำศัพท์ภาษาอังกฤษขนาดใหญ่ของคำนาม คำกริยา คำคุณศัพท์และกริยาวิเศษณ์ โดยมีการจัดกลุ่มเป็นคำ ซึ่งจะเชื่อมโยงความสัมพันธ์และความหมายของคำศัพท์ เป็นเครื่องมือที่มีประโยชน์สำหรับภาษาศาสตร์คอมพิวเตอร์และการประมวลผลภาษาธรรมชาติ⁶ สามารถใช้งานได้ 2 ทางหลักๆ คือ แบบออนไลน์ สามารถใช้งานผ่านเครือข่ายอินเทอร์เน็ตได้ และแบบโปรแกรมสำเร็จรูปสามารถใช้งานผ่านเครื่องที่ติดตั้งโปรแกรมได้เลย ประโยชน์ของเวิร์ดเน็ต สามารถบอกความสัมพันธ์ทางความหมายแบบต่างๆ ใช้สำหรับการประมวลผลภาษาธรรมชาติ แก้ปัญหาความกำกวมในความหมายของคำ (Word Sense Disambiguation) การค้นคืนสารสนเทศ (Information Retrieval) การสรุปย่อเอกสาร(Text Summarization) และการประยุกต์ใช้กับความรู้ทางด้านวิศวกรรมความรู้ (Knowledge Engineering)

2) แบบจำลองเวกเตอร์สเปซ (Vector Space Model) มีหลักการเปลี่ยนเอกสารแต่ละฉบับให้เป็นเวกเตอร์ของคำ โดยที่ขนาดของเวกเตอร์ขึ้นกับจำนวนของคำที่ปรากฏอยู่ในเอกสารฉบับนั้นแบบจำลองเวกเตอร์สเปซ⁷ เพื่อสร้างแบบจำลองทางพีชคณิตในการนำเสนอเอกสารข้อความที่ใช้เวกเตอร์เป็นตัวระบุ ตัวอย่างเช่น การใช้ดัชนีคำศัพท์ โดยส่วนมากจะใช้แบบจำลองเวกเตอร์สเปซในงานการคัดกรองข้อมูล ระบบงานค้นคืนสารสนเทศ การทำเหมืองข้อมูล การทำดัชนี เป็นต้น โดยขั้นตอนการหาความคล้ายคลึงของเอกสารกับคำค้น (Query) ซึ่งคำค้นก็คือได้ว่าเป็นเอกสารเช่นกัน ได้แก่ การนับความถี่ของคำที่ปรากฏในเอกสาร (Term Frequency) ขั้นตอนนี้จะต้องทำการตัดคำ (Word Segmentation) และคัดกรองคำก่อน (Stop word and Stemming) โดยคำที่ผ่านขั้นตอนดังกล่าวจะถูกเรียกว่า ดัชนี (Index) และการหาค่าส่วนกลับความถี่ของเอกสาร (Inverse Document Frequency: IDF) ดังสมการที่ (1)

$$idf_i = \log \left(\frac{N}{n_i} \right) \quad (1)$$

โดย N คือจำนวนเอกสารทั้งหมดที่ถูกเก็บรวบรวมไว้
df คือจำนวนความถี่ของเอกสารที่มีคำแต่ละคำ

ปรากฏอยู่ ถ้าคำนั้นปรากฏในเอกสารทุกเอกสารที่ถูกเก็บรวบรวมไว้ คำนี้จะมีค่าเท่ากับ N

จากนั้นทำการหาค่าน้ำหนักค่าแต่ละในเอกสารและคำค้น ดังสมการที่ (2)

$$W_{ij} \text{ OR } W_{iq} = TF * idf \quad (2)$$

โดย i คือ ลำดับของคำของดัชนี

j คือ ลำดับของเอกสาร

q คือ เอกสารคำค้น

W_{ij} คือ น้ำหนักของคำลำดับที่ i ของดัชนีในเอกสารลำดับที่ j

W_{iq} คือ น้ำหนักของคำลำดับที่ i ของดัชนีในคำค้น
จากนั้นทำการหาค่าความคล้ายคลึง ดังสมการที่ (3)

$$Sim(d_j, q) = \frac{\sum_i (w_{ij} * w_{iq})}{\sqrt{\sum_i (w_{ij})^2} * \sqrt{\sum_i (w_{iq})^2}} \quad (3)$$

โดย n คือจำนวนคำในดัชนีทั้งหมด

3) การวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ (Cosine Similarity Measurement) วิธีการนี้เป็นการเปรียบเทียบเนื้อหาระหว่างเอกสาร โดยการนับความถี่ของคำที่ปรากฏในเอกสาร (Term Frequency)⁸ แต่ละเอกสารที่ได้นำมาเปรียบเทียบกัน จากนั้นนำเอาคำหรือดัชนีทั้งหมดในเอกสารทั้ง 2 มารวมกัน ถ้าเอกสารมีคำเหมือนกันให้นำคำที่เหมือนกันมาสร้างเป็นดัชนีแค่คำเดียวและนำเอาความถี่ของคำในแต่ละเอกสารเข้าเวกเตอร์ทั้ง 2 เวกเตอร์ จากนั้นทำการคำนวณค่าความคล้ายคลึงมุมโคไซน์ได้ ตามสมการที่ (4)

$$Sim(d_1, d_2) = \frac{\sum_i (d_{i1} * d_{i2})}{\sqrt{d_{i1}^2} * \sqrt{d_{i2}^2}} \quad (4)$$

โดย d_1 คือ เวกเตอร์ของรายวิชาที่นำมาเทียบโอน

d_2 คือ เวกเตอร์ของรายวิชาที่ถูกเทียบ

i คือ ลำดับคำในเวกเตอร์

4) การวัดประสิทธิภาพของการค้นคืนข้อมูล (Retrieval Effectiveness) การวัดประสิทธิภาพของระบบค้นคืนข้อมูลที่ได้รับนิยามกันมากคือ การใช้เทคนิคการหาค่าความแม่นยำ (Precision) และการหาค่าความระลึก (Recall) ค่าความแม่นยำ (Precision)⁹ เป็นการประเมินประสิทธิภาพ

ของความแม่นยำของการค้นคืนเอกสาร โดยหาได้จากจำนวนเอกสารที่ค้นคืนได้ถูกต้องหารด้วยเอกสารที่ค้นคืนได้ถูกต้องร่วมกับเอกสารที่ค้นคืนไม่ถูกต้อง โดยสามารถแสดงได้ดังสมการ (5)

$$P = \frac{A}{A+B} \quad (5)$$

โดย P คือ ค่าความแม่นยำ

A คือ จำนวนเอกสารที่ค้นคืนได้ถูกต้อง

B คือ จำนวนเอกสารที่ค้นคืนไม่ถูกต้อง

ค่าความระลึกได้ (Recall) เป็นการประเมินประสิทธิภาพของการค้นคืนเอกสาร โดยหาได้จากเอกสารที่ค้นคืนได้ถูกต้องหารด้วยเอกสารที่ค้นคืนได้ถูกต้องร่วมกับจำนวนเอกสารที่ไม่ถูกค้นคืนแต่ในความเป็นจริงต้องถูกค้นคืนออกมาเป็นผลลัพธ์ โดยสามารถแสดงได้ดังสมการ (6)

$$R = \frac{A}{A+C} \quad (6)$$

โดย R คือ ค่าความระลึก

A คือ จำนวนเอกสารที่ค้นคืนได้ถูกต้อง

C คือ จำนวนเอกสารที่ไม่ถูกค้นคืนแต่ในความเป็นจริงต้องถูกค้นคืนออกมาเป็นผลลัพธ์

งานวิจัยที่เกี่ยวข้อง

1) งานวิจัย¹⁰ ได้เสนอระบบเทียบโอนรายวิชาและการวัดความคล้ายคลึงของคำอธิบายภาษาไทยของรายวิชา โดยทดลองใช้กับสาขาวิชาการระบบสารสนเทศและคอมพิวเตอร์ธุรกิจ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ ซึ่งมีขั้นตอนคือ นำคำอธิบายภาษาไทยของรายวิชาในระดับปริญญาตรีและในระดับประกาศนียบัตรชั้นสูงอย่างละ 1 วิชามาทำการตัดคำด้วยมนุษย์ แล้วนำคำที่เป็นใจความสำคัญของแต่ละรายวิชามาเปรียบเทียบกันเพื่อนับความถี่ของคำที่ปรากฏในรายวิชานั้นและสร้างเวกเตอร์ จากนั้นทำการคำนวณค่าความคล้ายคลึงของรายวิชาทั้ง 2 ด้วยวิธีการวัดความคล้ายคลึงเชิงมุมโคไซน์ ซึ่งจะแทนค่าในสมการที่ ซึ่งผลการทดสอบประสิทธิภาพของงานวิจัยนี้พบว่าสามารถทำงานได้ถูกต้องร้อยละ 86.66

2) งานวิจัย¹¹ ได้เสนอวิธีการด้วยเทคนิคการนำเอาออนโทโลยีของโดเมนห้องสมุดมารวมกัน เพื่อแก้ไขปัญหาความหลากหลายของคุณสมบัติที่เขียนต่างกันแต่อาจจะมีความ

ความหมายคล้ายคลึงกันให้ง่ายต่อการสืบค้นและแลกเปลี่ยนข้อมูลระหว่างห้องสมุดทั้งสอง โดยมีขั้นตอนคือ สร้างสถาปัตยกรรมโดยนำฐานข้อมูล ทั้งสองฐานข้อมูลที่แตกต่างกันมาสร้างเป็นออนโทโลยี โดยจะทำการหาความคล้ายคลึงเชิงความหมายของคุณสมบัติในฐานข้อมูลห้องสมุดด้วยเวิร์ดเน็ต ซึ่งจะภาษา OWL และใช้มาตรฐานดับลินคอร์ เมทาดาตา (Metadata) เป็นเครื่องมือที่ใช้อธิบายและกำหนดรูปแบบโครงสร้างของข้อมูลที่หลากหลาย จากการทดสอบประสิทธิภาพการสืบค้นของงานวิจัย ได้ค่าความแม่นยำ (Precision) เท่ากับ 88% และค่าความระลึกได้ (Recall) เท่ากับ 100% จากการทดลองการใช้คำค้น (Query) 30 ครั้ง

3) งานวิจัย¹² ได้เสนอการค้นคืนข้อมูลบนเว็บโดยวิธีหาค่าความคล้ายคลึงเชิงความหมายของคำ (Semantic Similarity Retrieval Model) จากแนวคิดที่ว่าคำอาจจะเขียนต่างกันแต่มีความหมายคล้ายคลึงกันโดยอาศัยเวิร์ดเน็ต ซึ่งการวัดผลจะทำการสืบค้นรูปภาพและเอกสารบนเว็บ โดยผลการทดลองแสดงให้เห็นว่าประสิทธิภาพของวิธีการค้นคืนเชิงความหมายสูงกว่าวิธีการค้นคืนข้อมูลโดยใช้แบบจำลองเวกเตอร์สเปซ (Vector Space Model:VSM) งานวิจัยนี้ได้ปรับปรุงวิธีการหาค่าความคล้ายคลึงในการค้นคืนข้อมูลโดยใช้แบบจำลองเวกเตอร์สเปซ ให้กลายเป็นแบบจำลองการค้นคืนเชิงความหมาย (Semantic Similarity Retrieval Model:SSRM) โดยการใช้ฐานข้อมูลเวิร์ดเน็ตมาช่วยในการหาความคล้ายคลึงเชิงความหมาย โดยการปรับปรุงขั้นตอนการในการหาเวกเตอร์ของคำค้น จากนั้นจะทำการคำนวณหาความคล้ายคลึงระหว่างคำค้นกับเอกสาร จากสมการที่ (3) ซึ่งผลการประเมินค่าความเหมาะสมจากวิธีการหาค่าความคล้ายคลึงเชิงความหมายหลายๆ วิธีตามประเภทของแต่ละวิธี ผลการทดสอบประสิทธิภาพของงานวิจัยนี้พบว่าวิธี SSRM มีค่าความแม่นยำ (Precision) และค่าความระลึกได้ (Recall) สูงกว่าวิธี VSM

4) งานวิจัย¹³ ได้เสนอการสร้างฐานข้อมูลคำศัพท์เวิร์ดเน็ตภาษาไทยจากเครื่องอ่านพจนานุกรมเพื่อพัฒนาต้นแบบพจนานุกรมภาษาไทยที่สามารถหาความสัมพันธ์เชิงความหมายของคำ ซึ่งสามารถนำไปใช้ในการสืบค้นข้อมูล การจัดกลุ่มเอกสารและการให้ความหมายคำกำกวม สามารถสร้างแบบจำลองเวิร์ดเน็ตภาษาไทยที่มีเนื้อหาครอบคลุมถึงร้อยละ 80 และให้ความถูกต้องร้อยละ 76 มีการเชื่อมโยง 44,844 เส้น ความสัมพันธ์ ในการอธิบายความสัมพันธ์เชิงความหมายจำนวน 13,730 คำ และคำที่มีความหมายเหมือนกันจำนวน 19,582 คำ

ขั้นตอนการประมวลผล

ขั้นตอนการประมวลผลในงานวิจัยนี้ สามารถอธิบายขั้นตอนการประมวลผล (Figure 1)

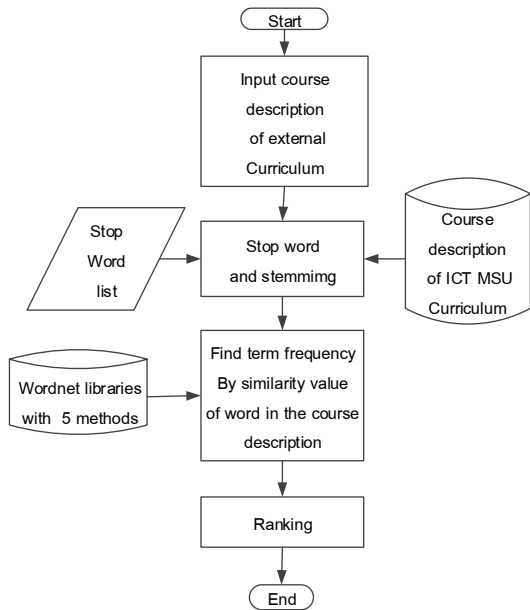


Figure 1 The process of course transfer

จาก (Figure 1) แสดงถึงขั้นตอนการเทียบโอนรายวิชา โดยเริ่มต้นจากการป้อนคำอธิบายภาษาอังกฤษของรายวิชาที่ต้องการจะเทียบโอนเข้าสู่กระบวนการเพื่อทำการหาความคล้ายคลึงเชิงความหมายกับคำอธิบายภาษาอังกฤษของรายวิชาในหลักสูตรวิทยาศาสตรบัณฑิต พุทธศักราช 2555 สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม ซึ่งจะแบ่งขั้นตอนการทำงานออกเป็นสามส่วนคือ การคัดกรองคำหยุด และการหาค่าความคล้ายคลึงเชิงความหมาย ขั้นตอนนี้จะนำคำทั้งหมดที่ได้จากการกรองคำหยุดจากคำอธิบายรายวิชาที่นำมาเทียบไปหาความถี่ของคำ โดยจะไม่นับเฉพาะคำที่เขียนเหมือนกันแต่จะอาศัยวิธีการหาความคล้ายคลึงเชิงความหมายของวิธีการของวูและพาล์มเมอร์, , เลียค็อคกและโคโดโรว์, เรสเนค, ลิน และเจียงและคอนราท์ ซึ่งในการประมวลผลนั้นจะทำการแยกประมวลผลของวิธีการทั้งห้าวิธีอย่างชัดเจน เมื่อทำการเทียบโอนกับคำอธิบายของรายวิชาในหลักสูตรเทคโนโลยีสารสนเทศและการสื่อสารครบทุกรายวิชาแล้วก็จะทำการเรียงลำดับรายวิชาที่มีคำอธิบายรายวิชาคล้ายคลึงจากมากไปหาน้อย ซึ่งรายวิชาที่มีค่าความคล้ายคลึงร้อยละมากกว่าหรือเท่ากับ 70 ขึ้นไปจะถือว่าสามารถเทียบเคียงกันได้

ผลการศึกษาและสรุปผลการทดลอง

จากการทดสอบการเทียบโอนรายวิชา สามารถแสดงการเปรียบเทียบประสิทธิภาพทั้งห้าวิธี โดยเรียงลำดับจากวิธีการที่มีความเหมาะสมมากที่สุดไปหาวิธีการที่มีความเหมาะสมน้อยดัง (ตารางที่ 1)

Table 1 Effective comparison of the five methods for a course transfer

Method	Precision	Recall	F-measure
Jiang and Conrath	72	100	84
Lin	67	100	80
Wu and Palmer	50	100	67
Leacock and Chodorow	17	100	29
Resnik	5	100	10

จากตารางสามารถสรุปได้ว่าวิธีการเทียบโอนรายวิชาของเจียงและคอนราท์ คือวิธีที่เหมาะสมที่สุด โดยได้ค่าประสิทธิภาพพื้นฐานในการเทียบโอนรายวิชาเท่ากับร้อยละ 84 ค่าความแม่นยำเท่ากับร้อยละ 72 และค่าความระลึกเท่ากับร้อยละ 100 ทดสอบ แม้ว่าประสิทธิภาพพื้นฐานในการเทียบโอนรายวิชา มีค่าเป็นที่น่าพอใจ แต่ยังมีข้อผิดพลาด ทั้งนี้เนื่องมาจากคำอธิบายรายวิชาที่นำมาเทียบโอนมีการใช้ทั้งภาษาไทยและภาษาอังกฤษผสมกัน ทำให้บางรายวิชาไม่สามารถนำมาใช้ในการทดสอบได้ เช่น รายวิชาการวิเคราะห์และออกแบบระบบสารสนเทศของหลักสูตรวิทยาศาสตรบัณฑิต พ.ศ. 2555 สาขาวิชาเทคโนโลยีสารสนเทศ คณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคาม มีคำอธิบายของรายวิชาเป็น "The systems and patterns of information systems; the life cycles of software development; searching and analyzing of the users' needs; data flow diagram; data dictionary; description of the processing; designs of file, database, and user interface; software development, software testing, software practical implementation; documentation; software assessment and evaluation; software maintenance; and related case studies" ซึ่งเป็นรายวิชาที่สามารถเทียบโอนกับรายวิชาการวิเคราะห์และออกแบบระบบของหลักสูตรสถาบันเทคนิคหรืออาชีวศึกษาที่มีคำอธิบายของรายวิชาเป็น "ศึกษาและปฏิบัติเกี่ยวกับองค์ประกอบของระบบ วัฏจักรของระบบ ระเบียบวิธีวิเคราะห์ระบบ เครื่องมือที่ใช้ในการวิเคราะห์ระบบ การศึกษาความเหมาะสมของระบบ DFD ERD การออกแบบการแสดงผล การออกแบบฐานข้อมูล การ

เขียนเอกสารและการนำเสนอผลการวิเคราะห์” จากกรณีดังกล่าวจะเห็นว่าไม่มีคำอธิบายที่เป็นภาษาไทยอังกฤษล้วนๆ ดังนั้นจึงขอเสนอให้สถาบันเทคนิคหรืออาชีวศึกษาทำการเขียนคำอธิบายรายวิชาภาษาไทยและภาษาอังกฤษแยกออกจากกัน เพื่อให้ได้คำอธิบายที่เป็นภาษาอังกฤษที่ครบถ้วนและเพิ่มประสิทธิภาพของงานวิจัยนี้มากยิ่งขึ้น

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณคณะวิทยาการสารสนเทศ มหาวิทยาลัยมหาสารคามที่ให้ทุนอุดหนุนงานวิจัยนี้

เอกสารอ้างอิง

- [1] Z. Wu and M. Palmer. “Verb Semantics and Lexical Selection”. In Annual Meeting of the Associations for Computational Linguistics (ACL'94), pages 133–138, Las Cruces, New Mexico, 1994.
- [2] C. Leacock and M. Chodorow. “Combining Local Context and WordNet Similarity for Word Sense Identification in WordNet”. In C. Fellbaum, editor, An Electronic Lexical Database, pages 265–283. MIT Press, 1998.
- [3] O. Resnik. “Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity and Natural Language”. *Journal of Artificial Intelligence Research*, 11:95–130, 1999.
- [4] D. Lin. “Principle-Based Parsing Without Overgeneration”. In Annual Meeting of the Association for Computational Linguistics (ACL'93), pages 112–120, Columbus, Ohio, 1993.
- [5] J. Jiang and D. Conrath. “Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy”. In Intern. Conf. on Research in Computational Linguistics, Taiwan, 1998.
- [6] Java API for WordNet Searching (JAWS), From: <http://lyle.smu.edu/~tspell/jaws/index.html> 2009.
- [7] G. Salton and M. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, 1983.
- [8] R. B. Yates and B. R. Neto. *Modern Information Retrieval*. ADDISON-WESLEY, New York, 1999
- [9] บารมี ไอสรีกุลและรัชชัย งามสันติวงศ์. ระบบเทียบโอนรายวิชาโดยใช้แบบจำลองเวกเตอร์สเปซ. การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษามหาวิทยาลัยสุโขทัย ธรรมาราช ครั้งที่ 2, กันยายน 4-5, 2555.
- [10] อรวรรณ อุไรเรื่องพันธุ์และสมจิตร อาจอินทร์. Semantic Text Summarization Using Ontology. The 5th National Conference on Computing and Information Technology NCCIT 2009.
- [11] G. Varelas, E. Voutsakis, P. Raftopoulou, Euripides G.M. Petrakis and Evangelos E. Milios. Semantic Similarity Methods in WordNet and their Application to Information Retrieval on the Web, WIDM '05 Proceedings of the 7th annual ACM international workshop on Web information and data management, Pages 10-16, 2005.
- [12] S. Patanakul and C. Pluempitiwiriwajewj. “Construction of Thai WordNet lexical database from machine readable dictionaries.” Proc. 10th Machine Translation Summit, Phuket, Thailand, 2005.
- [13] V. Vichianchai, “Documents Comparison Using WordNet”, ICITA 2014. 1-4 July, 2014 The 9th International Conference on Information Technology and Applications. From: <http://www.icita.org/2014/CD/abstracts/th-Vichianchai.htm>.

คำแนะนำสำหรับผู้พิมพ์

วารสารวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยมหาสารคาม กำหนดพิมพ์ปีละ 6 ฉบับ ฉบับที่ 1 (มกราคม-กุมภาพันธ์) ฉบับที่ 2 (มีนาคม-เมษายน) ฉบับที่ 3 (พฤษภาคม-มิถุนายน) ฉบับที่ 4 (กรกฎาคม-สิงหาคม) ฉบับที่ 5 (กันยายน-ตุลาคม) ฉบับที่ 6 (พฤศจิกายน-ธันวาคม) ผู้พิมพ์ทุกท่านสามารถส่งเรื่องมาพิมพ์ได้ โดยไม่ต้องเป็นสมาชิก และไม่จำเป็นต้องสังกัดมหาวิทยาลัยมหาสารคาม ผลงานที่ได้รับการพิจารณาในวารสารจะต้องมีสาระที่น่าสนใจ เป็นงานที่ทบทวนความรู้เดิมหรือองค์ความรู้ใหม่ที่ทันสมัย รวมทั้งข้อคิดเห็นทางวิชาการที่เป็นประโยชน์ต่อผู้อ่าน และจะต้องเป็นงานที่ไม่เคยถูกนำไปตีพิมพ์เผยแพร่ในวารสารอื่นมาก่อนและไม่อยู่ในระหว่างพิจารณาผลงานพิมพ์ในวารสารใด บทความอาจถูกดัดแปลง แก้ไขเนื้อหา รูปแบบ และสำนวน ตามที่กองบรรณาธิการเห็นสมควร ทั้งนี้เพื่อให้วารสารมีคุณภาพในระดับมาตรฐานสากลและนำไปอ้างอิงได้

การเตรียมต้นฉบับ

1. ต้นฉบับพิมพ์เป็นภาษาไทยหรือภาษาอังกฤษก็ได้ แต่ละเรื่องจะต้องมีบทคัดย่อทั้งภาษาไทยและภาษาอังกฤษ การใช้ภาษาไทยให้ยึดหลักการใช้คำศัพท์การเขียนทับศัพท์ภาษาอังกฤษตามหลักของราชบัณฑิตยสถานให้หลีกเลี่ยงการเขียนภาษาอังกฤษปนภาษาไทยในข้อความ ยกเว้นกรณีจำเป็น เช่น ศัพท์ทางวิชาการที่ไม่มีทางแปล หรือคำที่ใช้แล้วทำให้เข้าใจง่ายขึ้น คำศัพท์ภาษาอังกฤษที่เขียนเป็นภาษาไทยให้ใช้ตัวเล็กทั้งหมด ยกเว้นชื่อเฉพาะ สำหรับต้นฉบับภาษาอังกฤษควรได้รับความตรวจสอบที่ถูกต้องด้านการใช้ภาษาจากผู้เชี่ยวชาญด้านภาษาอังกฤษก่อน
2. ขนาดของต้นฉบับ ใช้กระดาษขนาด A4 (8.5x11 นิ้ว) และพิมพ์โดยเว้นระยะห่างจากขอบกระดาษด้านละ 1 นิ้ว จัดเป็น 2 คอลัมน์ ระยะห่างระหว่างบรรทัดในภาษาที่ใช้ double space ภาษาอังกฤษล้วนให้เป็น single space
3. ชนิดของขนาดตัวอักษร ทั้งภาษาไทยและภาษาอังกฤษให้ใช้ตัวอักษร Browallia New ชื่อเรื่องให้ใช้อักษรขนาด 18 pt. ตัวหนา ชื่อผู้พิมพ์ใช้อักษรขนาด 16 pt. ตัวปกติ หัวข้อหลักใช้อักษรขนาด 16 pt. ตัวหนา หัวข้อรองใช้ตัวอักษรขนาด 14 pt. ตัวหนา บทคัดย่อและเนื้อเรื่องใช้ตัวอักษรขนาด 14 pt. ตัวหนา เชิงอรรถหน้าแรกที่เป็นชื่อตำแหน่งทางวิชาการ และที่อยู่ของผู้พิมพ์ ใช้อักษรขนาด 12 pt. ตัวหนา
4. การพิมพ์ต้นฉบับ ผู้เสนองานจะต้องพิมพ์ส่งต้นฉบับในรูปแบบของแฟ้มข้อมูลต่อไปนี้ อย่างไม่อย่างหนึ่ง ได้แก่ ".doc" (MS Word) หรือ ".rtf" (Rich Text)
5. จำนวนหน้า ความยาวของบทความไม่ควรเกิน 15 หน้า รวมตาราง รูป ภาพ และเอกสารอ้างอิง
6. จำนวนเอกสารอ้างอิงไม่เกิน 20 หน้า
7. รูปแบบการเขียนต้นฉบับ แบ่งเป็น 2 ประเภท ได้แก่ ประเภทบทความรายงานผลวิจัยหรือบทความวิจัย (research article) และบทความจากการทบทวนเอกสารวิจัยที่ผู้อื่นทำเอาไว้ หรือบทความทางวิชาการ หรือบทความทั่วไป หรือบทความปริทัศน์ (review article)

บทความรายงานผลวิจัย ให้เรียงลำดับหัวข้อดังนี้

ชื่อเรื่อง (Title) ควรสั้น กระชับ และสื่อเป้าหมายหลังของงานวิจัย ไม่ใช้คำย่อ ความยาวไม่เกิน 100 ตัวอักษร ชื่อเรื่องให้มีทั้งภาษาไทยและภาษาอังกฤษ

ชื่อผู้พิมพ์ [Author(s)] และที่อยู่ ให้มีทั้งภาษาไทยและภาษาอังกฤษ และระบุตำแหน่งทางวิชาการ หน่วยงาน หรือสถาบันที่สังกัด และ E-mail address ของผู้พิมพ์ไว้เป็นเชิงอรรถของหน้าแรก เพื่อกองบรรณาธิการสามารถติดต่อได้

บทคัดย่อ (Abstract) เป็นการย่อเนื้อความงานวิจัยทั้งเรื่องให้สั้น และมีเนื้อหาครบถ้วนตามเรื่องเดิม ความยาวไม่เกิน 250 คำ หรือไม่เกิน 10 บรรทัด และไม่ควรใช้คำย่อ

คำสำคัญ (Keyword) ให้ระบุไว้ท้ายบทคัดย่อของแต่ละภาษาประมาณ 4-5 คำสั้น ๆ

บทนำ (Introduction) เป็นส่วนเริ่มต้นของเนื้อหา ที่บอกความเป็นมา เหตุผล และวัตถุประสงค์ ที่นำไปสู่งานวิจัยนี้ ให้ข้อมูลทางวิชาการที่เกี่ยวข้องจากการตรวจสอบเอกสารประกอบ