

การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 2

The 2nd STOU Graduate Research Conference

ระบบเทียบโอนรายวิชาโดยใช้แบบจำลองเวกเตอร์สเปซ

Course Credit Transfer System Using Vector Space Model

บารมี โอศิริกุล (Baramee Osateerakul)* ธวัชชัย งามสันติวงศ์ (Thavtchai Ngamsantivong) **

บทคัดย่อ

บทความนี้เป็นงานวิจัยที่นำเสนอระบบเทียบโอนรายวิชาและการวัดความคล้ายคลึงของ มหาวิทยาลัยเทคโนโลยีราชมงคลสุวรรณภูมิ โดยทดลองกับสาขาวิชาระบบสารสนเทศและคอมพิวเตอร์ธุรกิจ มีวัตถุประสงค์ (1) เพื่อพัฒนาระบบเทียบโอนรายวิชาจากการวัดความคล้ายคลึงของรายวิชาโดยนำหลักการวัดความคล้ายคลึงเชิงมุมโคไซน์ และการให้ค่าน้ำหนักของคำมาใช้กับหลักการของแบบจำลองเวกเตอร์สเปซ (2) เพื่อวิเคราะห์วิธีการวัดความคล้ายคลึงของภาษาไทย

ในการดำเนินงานประยุกต์ใช้การตัดคำ และนำคำที่ได้นั้นไปเปรียบเทียบกับ Dictionary หลังจากนั้นนำคำที่มีความหมายมาเก็บไว้ เพื่อเป็น Keywords ซึ่ง Model ที่นำมาใช้ในการเปรียบเทียบ คือ Vector Space Model (VSM) เป็นการเปรียบเทียบคำโดยที่ไม่ต้องมีคำที่ตรงกันทั้งหมดขึ้นอยู่กับความคล้ายคลึง และการวัดความคล้ายคลึงของเอกสารภาษาไทย แบ่งออกเป็น 2 ส่วน คือ ส่วนการเตรียมคำอธิบายรายวิชาก่อนการประมวลผลและส่วนการประมวลผล ในการวัดความคล้ายคลึง (similarity) ระหว่างคำอธิบายรายวิชาระดับ(ป.ตรี) และคำอธิบายรายวิชาระดับประกาศนียบัตรวิชาชีพชั้นสูง ถ้ามีค่าความคล้ายคลึงสูงย่อมแสดงว่ารายวิชาระดับประกาศนียบัตรวิชาชีพชั้นสูง ดังกล่าวนั้นมีโอกาสเทียบโอนหน่วยกิตกับระดับปริญญาตรีสูง โดยผู้จัดทำได้ใช้โปรแกรม Visual Studio 2008 เป็นเครื่องมือที่ใช้ในการพัฒนาระบบและใช้ โปรแกรม Microsoft Access 2007 ในการจัดการเกี่ยวกับฐานข้อมูล

หลังจากพัฒนาระบบเสร็จแล้วได้ทำการทดสอบระบบและประเมินประสิทธิภาพของระบบด้วยวิธีการเปรียบเทียบคำร้อยละในการเทียบโอนรายวิชา และวัดความคล้ายคลึงของรายวิชา การทดสอบประสิทธิภาพพบว่าระบบสามารถทำงานได้ถูกต้องคิดเป็นร้อยละ 86.66 เป็นไปตามวัตถุประสงค์ของการค้นคว้า

Abstract

This article present a research on measurement system, compared to similar, and automatic credits transfer of computer courses of Rajamangala University of Technology Suvarnabhumi. It's try on Information System and Business Computer majors. The purposes of study is develop a system of credits transfer by measuring the similarity of the course using the angular cosine similarity measure and the weight of the principles of the vector space model. Secondly, analyze methodology of measurement thai words similarity.

In the application of thai words segmentation, after the words segmentation was taken. It's compared to the dictionary and store the meaningful words as the keywords by using Vector Space Model (VSM) to measure the similarities. The process of comparison composes had two parts: the preparation of course description and comparison process. The system measures the similarity between the course of bachelor degree and the courses of diploma degree. The result of comparison of similarity high. The system is developed by using Visual Studio 2008 and using Microsoft Access 2007.

The testing of system efficiency is taken by using percentage of correction courses transfer and measument of similarity. The result show that 86.66 correct and it confirm the research hypotheses.

คำสำคัญ: เวกเตอร์สเปซ, ความคล้ายคลึง

Keywords: Vector space, Similarity

* หลักสูตรวิทยาศาสตรมหาบัณฑิต มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ YewSystem@gmail.com

บทนำ

เนื่องจากปัจจุบันตามมหาวิทยาลัยต่าง ๆ นักศึกษามีเข้าเรียนเป็นจำนวนมากและการเรียนการสอนปริญญาตรีหลักสูตรเทียบโอน ทำให้เกิดปัญหาในการเทียบโอนผลการเรียนระดับปริญญาตรี ถึงแม้ว่าจะมีหลักการในการเทียบโอนรายวิชาที่ทบวงมหาวิทยาลัยได้ออกข้อแนะนำเกี่ยวกับแนวปฏิบัติที่ดีในการเทียบโอนผลการเรียนระดับปริญญาตรีมาแล้วก็ตาม ตัวอย่างเช่น การเทียบโอนรายวิชาต้องเป็นรายวิชาหรือกลุ่มรายวิชาที่มีเนื้อหาสาระครอบคลุมไม่น้อยกว่าสามในสี่ของรายวิชาหรือกลุ่มรายวิชาที่ขอเทียบ เป็นรายวิชาที่สอบไล่ได้ไม่ต่ำกว่าระดับคะแนน C หรือ 2.00 หรือเทียบเท่าการเทียบโอนผลการเรียนระดับปริญญาตรี นั้น โดยปกติจะจัดทำโดยเจ้าหน้าที่หรืออาจารย์ที่เกี่ยวข้องกับการเทียบโอนรายวิชา หลังจากนั้นนำผลมาพิมพ์ลงเครื่องคอมพิวเตอร์เท่านั้น ซึ่งในการเปรียบเทียบรายวิชาว่าสามารถเทียบกันได้หรือไม่นั้นดำเนินการโดยให้อาจารย์ผู้สอนในรายวิชาดังกล่าวหรือผู้ที่มีความเชี่ยวชาญในรายวิชาดังกล่าวเทียบคำอธิบายรายวิชาให้เพื่อสรุปผลว่าสามารถเทียบกันได้หรือไม่ จากวิธีดังกล่าว ทำให้เกิดปัญหาในการปฏิบัติงาน คือ การที่จะเทียบโอนรายวิชากันได้นั้น ยังไม่มีตัวชี้วัดที่แน่นอนว่าสัดส่วนของคำอธิบายรายวิชาที่นำมาเทียบกันนั้นมีความเหมือนหรือคล้ายคลึงกันเป็นสามในสี่จริง ๆ ของคำอธิบายรายวิชาทั้งหมดทำให้เกิดเป็นข้อเสียอย่างหนึ่งที่ทำให้นักศึกษาที่จะเทียบโอนรายวิชา ต้องสงสัยในการเทียบโอนว่าทำไมรายวิชาบางรายวิชาถึงเทียบกันไม่ได้ ทั้ง ๆ ที่ตัวนักศึกษาเองก็ตรวจสอบดูแล้วว่าเหมือนกันสามในสี่ก็ตาม อีกทั้งยังต้องอาศัยบุคลากรที่มีความรู้และประสบการณ์ในการเทียบโอน จึงจะสำเร็จและไม่เกิดปัญหา หรือเกิดปัญหาน้อยที่สุด

วิธีการแก้ปัญหาในงานวิจัย ได้มีแนวความคิดที่จะใช้วิธีการวัดความคล้ายคลึงของรายวิชา จากคำอธิบายรายวิชาภาษาไทยโดยนำหลักการวัดค่าความคล้ายคลึงเชิงมุมโคไซน์ คือ การวัดค่าความคล้ายคลึงกันระหว่างเวกเตอร์ d และ d' และการให้ค่าน้ำหนักของคำหรือ Keywords มาใช้กับหลักการของแบบจำลองเวกเตอร์สเปซมาวัดความคล้ายคลึงทั้งนี้ค่าความคล้ายคลึงสูงสุดที่วัดด้วยวิธีนี้ มีค่าเท่ากับ 1 โดยมีความหมายคือ เวกเตอร์ทั้งสองทำมุมระหว่างกัน 0 องศา นั่นคือเวกเตอร์ทั้งสองมีทิศทางเดียวกัน เวกเตอร์ในที่นี้หมายถึงคำอธิบายรายวิชาภาษาไทย ซึ่งวิธีนี้จะทำให้ตอบข้อสงสัยของนักศึกษาได้ว่าคำอธิบายรายวิชาที่นำมาเทียบกันนั้น มีความคล้ายคลึงกันตามข้อกำหนดของทบวงมหาวิทยาลัย คือ มีเนื้อหาสาระครอบคลุมไม่น้อยกว่าสามในสี่ของรายวิชาที่ขอเทียบ

วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาระบบเทียบโอนรายวิชาจากการวัดความคล้ายคลึงของรายวิชาโดยนำหลักการวัดความคล้ายคลึงเชิงมุมโคไซน์ และการให้ค่าน้ำหนักของคำมาใช้กับหลักการของแบบจำลองเวกเตอร์สเปซ
2. เพื่อศึกษาวิธีการวัดความคล้ายคลึงของภาษาไทย
3. เพื่อศึกษาวิธีการใช้แบบจำลองเวกเตอร์สเปซ

วิธีดำเนินการวิจัย

วิธีการศึกษาระบบเทียบโอนรายวิชาจากการวัดความคล้ายคลึงของคำอธิบายรายวิชาภาษาไทยโดยใช้แบบจำลองเวกเตอร์สเปซ ประกอบด้วย มีดังนี้

การวัดความคล้ายคลึง (Similarity Measure) การวัดความคล้ายคลึงของเอกสารนั้นสามารถจำแนกได้เป็น 2 ประเภท คือ

(1) การวัดความคล้ายคลึงจากการมีหรือไม่มีคำสำคัญในเอกสาร

(2) การวัดความคล้ายคลึงจากความถี่ของคำสำคัญที่พบในเอกสาร

การวัดความคล้ายคลึงของเอกสารจะอาศัยค่าน้ำหนักคำหลักที่มีอยู่ในแต่ละเอกสารและค่าน้ำหนักคำของเอกสาร โดยค่าน้ำหนักคำหลักในเอกสารคำนวณได้จากขั้นตอนการเตรียมข้อมูล ส่วนค่าน้ำหนักคำของข้อสอบถามจะคำนวณได้จากขั้นตอนการเปรียบเทียบเอกสารในแต่ละครั้งเมื่อคำนวณได้ค่าน้ำหนักคำหลักในเอกสารและค่าน้ำหนักคำของข้อสอบถามได้แล้วจะนำไปแทนที่ในสมการ (1) เพื่อคำนวณหาความคล้ายคลึงของเอกสาร โดยใช้วิธีแบบเวกเตอร์ (Vector Based Similarity Measurement)

$$sim(Q, D_j) = \frac{\sum_i (W_{Q,j} * W_{i,j})}{\sqrt{\sum_j W_{Q,j}^2} \sqrt{\sum_i W_{i,j}^2}} \quad (1)$$

เมื่อ
 $sim(Q, D_j)$ คือ ค่าความคล้ายคลึงระหว่างข้อสอบถามกับเอกสาร j
 $\sum_i (W_{Q,j} * W_{i,j})$ คือ ค่าผลรวมของ Cross Product ระหว่างค่าน้ำหนักคำของข้อสอบถาม Q กับค่าน้ำหนักคำหลัก i ในเอกสาร j
 $\sqrt{\sum_j W_{Q,j}^2}$ คือ ผลรวมของค่าน้ำหนักทั้งหมดของข้อสอบถาม Q
 $\sqrt{\sum_i W_{i,j}^2}$ คือ ผลรวมของค่าน้ำหนักทั้งหมดของคำหลัก i ในเอกสาร j

รูปที่ 1 การวัดความคล้ายคลึง

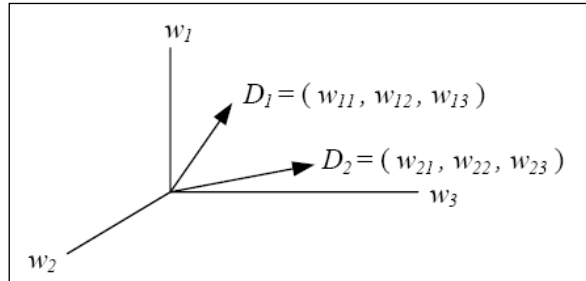
การวัดความคล้ายคลึงเชิงมุม (Cosine Similarity Measurement) การวัดความคล้ายคลึงของเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุมนั้นเป็นวิธีการเปรียบเทียบความคล้ายคลึงของเอกสารสองเอกสาร โดยแต่ละเอกสารจะถูกแทนด้วยเวกเตอร์ขนาดเอ็น (N-Dimensional Vector) ซึ่งเก็บค่าน้ำหนักคำแต่ละคำในเอกสารนั้น (N-Dimensional Vector in Term Space) การเปรียบเทียบความคล้ายคลึงของเอกสารจะเปรียบเทียบโดยดูจากมุมโคไซน์ของมุมระหว่าง 2 เวกเตอร์ของเอกสาร หากเอกสารทั้งสองเอกสารคล้ายคลึงกันมาก เวกเตอร์ของเอกสารทั้ง 2 จะทับกันเกือบสนิท มุมจึงมีค่าน้อย ค่าโคไซน์ที่ได้จะมีค่ามาก ตัวอย่างในการวัดค่าความคล้ายคลึงระหว่างเวกเตอร์ d และเวกเตอร์ d' จะมีค่าดังสมการ (2)

$$Similarity = \frac{d \cdot d'}{|d| |d'|} \quad (2)$$

โดย $d \cdot d'$ คือ เวกเตอร์ของ โปรดัค (Product)
 ตัวอย่างเช่น
 $d = (2, 1, 1, 1, 0)$
 $d' = (0, 0, 0, 1, 0)$
 $d \cdot d' = (2 \cdot 0) + (1 \cdot 0) + (1 \cdot 0) + (1 \cdot 1) + (0 \cdot 0)$
 $= 1$
 $|d| = \sqrt{2^2 + 1^2 + 1^2 + 1^2 + 0^2}$
 $= \sqrt{7}$
 $= 2.646$
 $|d'| = \sqrt{0^2 + 0^2 + 0^2 + 1^2 + 0^2}$
 $= \sqrt{1}$
 $= 1$
 ค่าความคล้ายคลึง
 $= \frac{1}{1 \cdot 2.646}$
 $= 0.378$

รูปที่ 2 ค่าความคล้ายคลึงเชิงมุม

ทั้งนี้ค่าความคล้ายคลึงสูงสุดที่วัดด้วยวิธีนี้จะเท่ากับ 1 โดยมีความหมายคือ เวกเตอร์ทั้งสองทำมุมระหว่างกัน 0 องศา นั่นคือเวกเตอร์ทั้งสองมีทิศทางเดียวกัน



รูปที่ 3 เวกเตอร์เอกสารใน 3-มิติ

ขั้นตอนการดำเนินงาน

1. ศึกษาทฤษฎีและหลักการต่าง ๆ ที่เกี่ยวข้อง
 - 1.1 หลักการตัดคำภาษาไทย
 - 1.2 หลักการแบบจำลองเวกเตอร์สเปซ
 - 1.3 หลักการเปรียบเทียบความหมายของคำกับพจนานุกรมข้อมูลและพจนานุกรมศัพท์คอมพิวเตอร์ของราชบัณฑิตยสถาน
 - 1.4 การให้ค่าน้ำหนักคำ
 - 1.5 การนับค่าความถี่ของคำ
 - 1.6 หลักการวัดความคล้ายคลึงของเอกสาร
2. จัดเตรียมข้อมูลรายวิชาทางด้านคอมพิวเตอร์ที่จะทำการทดสอบ โดยข้อมูลจะแบ่งออกเป็นสองส่วนด้วยกันคือ ส่วนรายวิชาระดับปริญญาตรี และส่วนรายวิชาระดับประกาศนียบัตรชั้นสูง โดยกำหนดให้รายวิชาระดับปริญญาตรีเป็นต้นแบบในการเทียบโอน
3. วิเคราะห์และออกแบบระบบโดยออกแบบฐานข้อมูลเพื่อใช้ในการเก็บข้อมูล
4. กำหนดขอบเขตและขั้นตอนการทำงานของโปรแกรม
 - 4.1 แนววิธีในการตัดคำ เพื่อให้ได้คำที่มีความหมาย
 - 4.2 แนววิธีคำนวณความคล้ายคลึงของเอกสาร
 - 4.3 รูปแบบการรายงานหรือแสดงผลที่ได้จากการคำนวณความคล้ายคลึง
5. พัฒนาระบบ ทดสอบผล และทำการปรับปรุงระบบ
6. สรุปผลการเทียบโอนออกมาในลักษณะทางสถิติ เช่น ร้อยละ

ผลการวิจัย

1. จัดเตรียมข้อมูลรายวิชา

1.1 ในการจัดเตรียมข้อมูลรายวิชานั้น ต้องจัดเตรียมทั้งในระดับปริญญาตรี(ป.ตรี) และในระดับประกาศนียบัตรชั้นสูง(ปวส.) โดยทำการป้อนข้อมูลเข้าไปในระบบที่ได้จัดทำขึ้น ป้อนทั้งรหัสวิชา, ชื่อวิชา, คำอธิบายรายวิชา

ตัวอย่าง ศึกษาขั้นตอนการพัฒนาโปรแกรมคอมพิวเตอร์
แนวคิดทั่วไปเกี่ยวกับการวิเคราะห์

1.2 นำชื่อวิชาและคำอธิบายรายวิชามาทำการตัดคำ โดยอาจจะ มี 1 พยางค์, 2 พยางค์ หรือ 3 พยางค์ ก็ได้ แล้วเก็บคำไว้

ตัวอย่าง ศึกษา,ขั้นตอน,การ,พัฒนา,โปรแกรม,คอมพิวเตอร์,
แนวคิด,ทั่วไป,เกี่ยวกับ,การ,วิเคราะห์

1.3 หลังจากที่ได้ตัดคำเรียบร้อยแล้ว และเก็บคำเรียบร้อยแล้ว ก็จะนำคำดังกล่าวมาทำการเทียบกับ พจนานุกรม (Dictionary) เพื่อเลือกเฉพาะคำที่มีความหมายนำมาใช้ในการวัดความคล้ายคลึง แต่นอกจากนี้แล้วยังมีการนำคำที่ตัดเรียบร้อยแล้วที่คิดกัน 2-3 ตำแหน่งมาทำการต่อกันแล้วเทียบกับพจนานุกรม(Dictionary) อีก เพื่อให้ได้คำใหม่เพิ่มมากขึ้น และทำการเก็บคำทั้งหมดที่มีความหมายไว้

ตัวอย่าง ศึกษา,ขั้นตอน,การ,พัฒนา,โปรแกรม,คอมพิวเตอร์,
แนวคิด,ทั่วไป,เกี่ยวกับ,การ,วิเคราะห์,ขั้นตอน,การ,พัฒนา,
โปรแกรมคอมพิวเตอร์,เกี่ยวกับ

1.4 หลังจากที่ได้คำที่มีความหมายทั้งหมดแล้ว ขั้นตอนต่อไปจะทำการตัดคำหยุด (Stopword) ซึ่งเป็นคำที่ไม่จำเป็น คำเชื่อมประโยค หรือคำที่ใช้บ่อย ๆ เช่น “การ” “ความ” “ควร” “ที่” “จะ” “ไป” “ใช้” “ได้” เป็นต้น เมื่อตัดคำหยุดเสร็จแล้ว ก็จะได้ คำที่เป็น Key words หรือคำที่เป็นใจความสำคัญของแต่ละรายวิชา ก็จะทำการบันทึกข้อมูลลงในฐานข้อมูล เพื่อใช้อ้างอิงในการสร้างเวกเตอร์ของคำที่ตรงกันทั้งฐานข้อมูล และความยาวของเวกเตอร์ คือ ค่าความถี่ของคำที่ถูกเลือกซึ่งปรากฏอยู่ในรายวิชานั้นประกอบร่วมกัน

ตัวอย่าง ขั้นตอน,พัฒนา,โปรแกรม,คอมพิวเตอร์,แนวคิด,
วิเคราะห์,ขั้นตอน,การ,พัฒนา,โปรแกรมคอมพิวเตอร์

2. ขั้นตอนการเปรียบเทียบค่าและวัดความคล้ายคลึง

2.1 การเปรียบเทียบคำก็คือ เลือกรายวิชาในระดับปริญญาตรี(ป.ตรี) และในระดับประกาศนียบัตรชั้นสูง(ปวส.) อย่างละ 1 รายวิชา นำคำที่เป็น Keywords หรือคำที่เป็นใจความสำคัญของแต่ละรายวิชามาเทียบกัน เพื่อสร้างเป็นเวกเตอร์ของคำ รวมถึงนับความถี่ของคำที่ปรากฏอยู่ในรายวิชานั้น ๆ ด้วย

2.2 หลังจากได้เวกเตอร์และความถี่แล้ว ทำการวัดความคล้ายคลึงของรายวิชาด้วยวิธีการวัดความคล้ายคลึงเชิงมุม เป็นวิธีการเปรียบเทียบความคล้ายคลึงของรายวิชาทั้ง 2 รายวิชา การเปรียบเทียบความคล้ายคลึงของรายวิชาจะเปรียบเทียบโดยดูจากมุมโคไซน์ของมุมระหว่าง 2 เวกเตอร์ของรายวิชา หากรายวิชาทั้งสองรายวิชาคล้ายคลึงกันมาก เวกเตอร์ของรายวิชาทั้ง 2 จะทับกันเกือบสนิท มุมจึงมีค่าน้อย ค่าโคไซน์ที่ได้จะมีค่ามาก

3. ผลการทดลอง

3.1 ผลการทดลองที่เทียบกันได้

ลำดับ	(d) Key ป.ตรี	จำนวน ...	(d') Key ปวส.	จำนวน ...	(d * d')	d	d'
1	ชั้น	1	ชั้น	1	1	1	1
2	ตอน	1	ตอน	1	1	1	1
3	พัฒนา	1	พัฒนา	2	2	1	4
4	โปรแกรม	2	โปรแกรม	3	6	4	9
5	คอมพิวเตอร์	2	คอมพิวเตอร์	1	2	4	1
6	แนวคิด	1	แนวคิด	0	0	1	0
7	วิเคราะห์	1	วิเคราะห์	1	1	1	1
8	ปัญหา	1	ปัญหา	2	2	1	4
9	ผัง	1	ผัง	2	2	1	4
10	งาน	1	งาน	2	2	1	4
11	รหัส	1	รหัส	1	1	1	1
12	เทียม	1	เทียม	1	1	1	1
13	คำนวณ	1	คำนวณ	0	0	1	0
14	เงื่อนไข	1	เงื่อนไข	0	0	1	0
15	ขั้นตอน	1	ขั้นตอน	1	1	1	1
16	การพัฒนา	1	การพัฒนา	1	1	1	1
17	โปรแกรมคอมพิวเตอร์...	2	โปรแกรมคอมพิวเตอร์...	0	0	4	0
18	ผังงาน	1	ผังงาน	2	2	1	4
19	การคำนวณ	1	การคำนวณ	0	0	1	0
20	การวน	1	การวน	0	0	1	0

รูปที่ 4 คำสำคัญและตารางความถี่ที่เทียบกันได้

$$\begin{aligned}
 d &= (1,1,1,2,2,1,1,1,1,1,1,1,1,1,2,1,1,1) \\
 d' &= (1,1,2,3,1,0,1,2,2,2,1,1,0,0,1,1,0,2,0,0) \\
 d \cdot d' &= (1 \times 1) + (1 \times 1) + (1 \times 2) + (2 \times 3) + (2 \times 1) + (1 \times 0) + (1 \times 1) + (1 \times 2) + \\
 &\quad (1 \times 2) + (1 \times 2) + (1 \times 1) + (1 \times 1) + (1 \times 0) + (1 \times 0) + (1 \times 1) + (1 \times 1) + \\
 &\quad (2 \times 0) + (1 \times 2) + (1 \times 0) + (1 \times 0) \\
 &= 25 \\
 |d| &= \sqrt{29} \\
 &= 5.385 \\
 |d'| &= \sqrt{37} \\
 &= 6.082 \\
 \text{ค่าความคล้ายคลึง} &= \frac{25}{5.385 \times 6.082} \\
 &= \frac{25}{32.7566} \\
 &= 0.763
 \end{aligned}$$

รูปที่ 5 ค่าความคล้ายคลึง

The 2nd STOU Graduate Research Conference

ผลเทียบโอน		รายละเอียดเทียบโอน					
ลำดับ	(d) Key ป.ตรี	จำนวน ...	(d') Key ป.วส.	จำนวน ...	(d * d')	d	d'
1	ฝึก	1	ฝึก	0	0	1	0
2	ปฏิบัติ	1	ปฏิบัติ	1	1	1	1
3	ภาษา	2	ภาษา	0	0	4	0
4	คอมพิวเตอร์	2	คอมพิวเตอร์	1	2	4	1
5	โปรแกรม	3	โปรแกรม	0	0	9	0
6	ระดับ	1	ระดับ	0	0	1	0
7	สูง	1	สูง	0	0	1	0
8	ข้อมูล	3	ข้อมูล	0	0	9	0
9	ชนิด	1	ชนิด	0	0	1	0
10	นิพจน์	1	นิพจน์	0	0	1	0
11	รับ	1	รับ	0	0	1	0
12	แสดง	1	แสดง	0	0	1	0
13	ผล	1	ผล	0	0	1	0
14	โครงสร้าง	1	โครงสร้าง	0	0	1	0
15	วน	1	วน	0	0	1	0
16	เงื่อนไข	1	เงื่อนไข	0	0	1	0
17	แถว	1	แถว	1	1	1	1
18	ลำดับ	1	ลำดับ	0	0	1	0

รูปที่ 6 คำสำคัญและตารางความถี่ที่เทียบกันไม่ได้

$$\begin{aligned}
 \text{ค่าความคล้ายคลึง} &= \frac{4}{7.615 * 1.732} \\
 &= \frac{4}{13.19} \\
 &= 0.303
 \end{aligned}$$

รูปที่ 7 ค่าความคล้ายคลึง

อภิปรายผลการวิจัย

จากผลการวิจัยพบว่าค่าความคล้ายคลึงของรายวิชาในระดับปริญญาตรี กับในระดับประกาศนียบัตรวิชาชีพชั้นสูงนั้น ถ้ามีค่าความคล้ายคลึงเกิน 75% ถือว่าอยู่ในเกณฑ์ที่สามารถเทียบโอนรายวิชากันได้ ซึ่งก็ขึ้นอยู่กับข้อกำหนดของทบวงมหาวิทยาลัย คือ มีเนื้อหาสาระครอบคลุมไม่น้อยกว่าสามในสี่ของรายวิชาที่ขอเทียบ นั่นเอง จากตัวอย่างของผลการวิจัย มีทั้งที่มีค่าสูงคือ 0.763 คิดเป็น 76.3% และค่าต่ำ คือ 0.303 คิดเป็น 30.3% ดังนั้นจึงทำให้ทราบว่าค่าสูงนั้นสามารถเทียบโอนกันได้ ส่วนค่าต่ำนั้นไม่สามารถเทียบโอนกันได้ ซึ่งค่าความคล้ายคลึงดังกล่าวเป็นไปตามหลักทฤษฎีการวัดความคล้ายคลึงของเอกสารด้วยวิธีการวัดความคล้ายคลึงเชิงมุม ค่าความคล้ายคลึงสูงสุดที่วัดได้ มีค่าเท่ากับ 1 โดยมีความหมายคือ เวกเตอร์ทั้งสองทำมุมระหว่างกัน 0 องศา นั่นคือเวกเตอร์ทั้งสองมีทิศทางเดียวกันหรือเหมือนกันนั่นเอง

การประชุมเสนอผลงานวิจัยระดับบัณฑิตศึกษา มหาวิทยาลัยสุโขทัยธรรมาธิราช ครั้งที่ 2
The 2nd STOU Graduate Research Conference

ข้อเสนอแนะ

การให้น้ำหนักคำที่เท่ากันทุก ๆ คำ ยังไม่ใช่วิธีการที่ดีที่สุด จึงควรเพิ่มเรื่องของน้ำหนักคำของแต่ละคำให้มีความแตกต่างกันหรือมีระดับของความสำเร็จในแต่ละระดับที่ไม่เท่ากัน และควรนำมาตรฐานการศึกษา TQF เข้ามาช่วยในการกำหนดคำสำคัญ

เอกสารอ้างอิง

- สุริรา พลันสังเกตุ และคณะ(2553) “การวิเคราะห์ความคล้ายคลึงของประโยคภาษาอังกฤษจากหน้าเว็บด้วยลูซีน (English Sentences Similarity Analysis on Web Pages using Lucene)” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศและการสื่อสารคณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์
- นางสาวรวีร์ เกษร (2553) “การวัดความคล้ายคลึงของเอกสารภาษาไทยโดยใช้การประมวลผลภาษาธรรมชาติ” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่
- ทินกร คุณาสัทธี และคณะ (2552) “การจัดกลุ่มเอกสารคาอริบายเว็บภาษาไทยสำหรับผลการค้นคืนด้วยอนเนกกาทีฟเมทริกซ์แฟกทอไรเซชัน” วิทยานิพนธ์ปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
- สิทธิโชค ปัญญาฤกษ์ชัย และศศิพาณี นุชิตประสิทธิ์ชัย (2552) “ระบบการค้นคืนสารสนเทศโดยใช้เทคนิค N-Gram” ภาควิชาเทคโนโลยีสารสนเทศ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
- <http://phpir.com/simple-search-the-vector-space-model> March , 22 2012
- <http://www.miislita.com/term-vector/term-vector-3.html> April ,21 2012
- http://en.wikipedia.org/wiki/Vector_space_model April ,21 2012
- <http://www.mua.go.th/users/hecommission/doc/law/ministry%20law/29%20transfer%20credit%202545.pdf> February, 22 2012