**แบบเสนอร่างโครงการวิจัย (Draft proposal)**

**1.ชื่อโครงการ**

ภาษาไทย  ระบบจัดการโครงงานคอมพิวเตอร์สมนาและระเบียบงานวิจัย

**ภาษาอังกฤษ Computer Project,Seminar And Research  Methology Management System**

**2.รายละเอียดของโครงการ**

2.1 หลักการและเหตุผล

เนื่องจากภาควิชาวิทยาการคอมพิวเตอร์ในแต่ละปี นักศึกษาชั้นปีที่ 4 ทุกคน ประกอบไปด้วย 3 สาขาวิชา ได้แก่ สาขาวิทยาการคอมพิวเตอร์ สาขาเทคโนโลยีสารสนเทศและการสื่อสาร สาขาภูมิสารสนเทศศาสตร์ ต้องทำโครงงานคอมพิวเตอร์เพื่อวัดประสิทธิภาพก่อนการจบการศึกษาส่งผลให้

มีโครงงานคอมพิวเตอร์จำนวนมากถูกพัฒนาขึ้นมา ภาควิชาจึงจัดทำเว็บไซต์รวบรวมโครงงาคอมพิวเตอร์

ให้นักศึกษาสามารถค้นหาโครงงานเพื่อเป็นแนวทางในการพัฒนาโครงงานในปีถัดไป และเนื่องจากจะมีการจัดตั้งภาควิชาการคอมพิวเตอร์ไปเป็นคณะใหม่ควรต้องมีระบบอำนวยความสะดวกในการเซ็นส่ง/ตรวจงานโครงงานคอมพิวเตอร์

            ซึ่งระบบค้นหาโครงงานของเว็บไซต์ยังไม่มีความยืดหยุ่นและครอบคลุมมากพอ เนื่องจากสามารถค้นหาได้ค่อนข้างกว้าง ยังไม่มีการแยกหมวดหมู่ โดยผู้ใช้สามารถค้นหาได้จาก สาขาวิชา ปีการศึกษา ภาคการเรียน รหัสโครงงาน ชื่อโครงงานและอาจารย์ที่ปรึกษา ทฤษฎีที่เกี่ยวข้อง เครื่องมือที่ใช้ นิยาม เท่านั้นจึงส่งผลให้ยากต่อการค้นหาและไม่อำนวยความสะดวกมากพอและในการเซ็นส่งงานหรือตรวจงานนักศึกษาต้องนัดและเข้าพบอาจารย์ที่ปรึกษาบ้างครั้งอาจารย์ก็ไม่สะดวกพบจึงต้องรอบางครั้งเป็นเวลานานกว่าจะได้เข้าพบ

จากปัญหาข้างต้นที่กล่าวมาผู้จัดทำจึงเห็นสมควรว่าจะพัฒนาต่อยอดเว็บไซต์รวบรวมโครงงานคอมพิวเตอร์ให้สามารถค้นหาข้อมูลได้จากTagในการแยกหมวดหมู่ เช่น Tag และจากการที่เราอยากเพิ่มระบบในการอำนวยความสะดวกในเซ็นการส่ง/ตรวจงานโครงงานคอมพิวเตอร์ วิธีการในการลงนามเอกสารแบบ Digital Signatureเป็น Subset ของ Electronic Signature เนื่องจากเป็นลายเซ็นที่อยู่ในรูปแบบของอิเล็กทรอนิกส์เหมือนกัน แต่ได้เพิ่มเติมคุณสมบัติด้านความปลอดภัยเข้าไป เพื่อให้มีความน่าเชื่อถือมากยิ่งขึ้น คุณสมบัติดังกล่าวประกอบด้วย 1.Signer Authentication เป็นความสามารถในการพิสูจน์ว่าใครเป็นคนเซ็นเอกสาร ตัวลายเซ็นจะสามารถใช้ในการเชื่อมโยงไปยังบุคคลที่เซ็นเอกสารได้ 2.Data Integrity เป็นความสามารถในการตรวจสอบ หรือพิสูจน์ได้ว่ามีการแก้ไขเอกสารหลังจากที่ได้มีการเซ็นไปแล้วหรือไม่ 3.Non-repudiation การไม่สามารถปฏิเสธความรับผิดชอบได้ เนื่องจากลายเซ็นที่สร้างขึ้นมีเอกลักษณ์ สามารถพิสูจน์ในชั้นศาลได้ว่าใครเป็นผู้เซ็นเอกสารโดยที่เว็บไซต์ยังคงสามารถใช้งานในส่วนของระบบเดิมได้อย่างครบถ้วนดังเดิม

**2.2 วัตถุประสงค์**

1.  เพื่อพัฒนาเว็บไซต์รวบรวมโครงงานคอมพิวเตอร์ของนักศึกษาปี4 ภาควิชาวิทยาการคอมพิวเตอร์

2.  เพื่อพัฒนาเว็บไซต์ที่สามารถค้นหาตัวอย่างโครงงานคอมพิวเตอร์ของนักศึกษาปี 4 จากคำสืบค้นด้วยTag

**2.3 วัตถุประสงค์ของโครงงาน**

1  เพื่อพัฒนาเว็บไซต์รวบรวมโครงงานคอมพิวเตอร์ของนักศึกษาปี4 ภาควิชาวิทยาการคอมพิวเตอร์

2  เพื่อพัฒนาเว็บไซต์ที่สามารถค้นหาตัวอย่างโครงงานคอมพิวเตอร์ของนักศึกษาปี4 จากคำสืบค้นด้วยTag

3 **เ**พื่อพัฒนาเว็บไซต์ที่สามารถลงนามเอกสารแบบ Digital Signatureได้

**2.4.ความต้องการของระบบ**

ความต้องการของระบบ e-project

1 การจัดการข้อมูลผู้มีสิทธิ์ใช้งานระบบ

2 นักศึกษาสามารถเพิ่มหัวข้อโครงงาน upload เอกสารในแต่ละขั้นตอนการทำงานได้

3 สามารถสืบค้นข้อมูลได้หลายรูปแบบเช่น ค้นโดยชื่อเรื่อง ผู้ทำ ที่ปรึกษา เนื้อหาที่เกี่ยวข้อง

สืบค้นข้อมูลตาม tagที่สนใจ (มีการจำกัดเนื้อหาในการดาวโหลด)

4 มีเว็บบอร์ด ประกาศ/ประชาสัมพันธ์

5 มีระบบการตรวจสอบรายงาน เอกสารที่ต้องส่ง โดย อ.ที่ปรึกษา/กรรมการ สามารถตรวจสอบแบบออนไลน์ได้(Digital Signature)

6 มีระบบบันทึกการให้คำปรึกษาออนไลน์(แทนแบบฟอร์มเข้าพบที่ปรึกษา)

บบการจัดกลุ่มสอบ

7 สามารถ Running หมายเลขโครงงานตามเงื่อนไขที่ระบุได้

8 มีระบบจัดตารางสอบ

9 มีระบบการให้คะแนนและประมวลผลคะแนน ตัดเกรด รายงานสรุปเกรด

10 มีระบบพิมพ์ใบสำคัญรับเงินโครงการพิเศษ เพื่อเบิกค่าตอบแทนที่ปรึกษา/กรรมการสอบ

11 มีระบบแจ้งหัวหน้าโครงงานให้ อ.ที่ปรึกษา

           12 รายงานต่างๆ

            -หัวข้อโครงงานแต่ละสาขา/ที่ปรึกษา/ปีการศึกษา

            -แบบฟอร์มให้คะแนนสอบ/การสอบ/ปีการศึกษา

            -กลุ่มสอบ

            -ตารางสอบ

            -จำนวนโครงงานต่อที่ปรึกษา

            -นักศึกษาที่ยังไม่ส่งหัวข้อ

            -นักศึกษาที่ยังไม่ส่งรายงาน(แยกแต่ละรายงงาน)

**2.5**.**ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง**

**ลายเซ็นดิจิตอล (Digital Signature)**

เป็นสิ่งที่ใช้ยืนยันตัวบุคคล (Identification) ในโลกดิจิตอล เปรียบเสมือนกับการลงนามเอกสารด้วยปากกา   
ทำให้สามารถมั่นใจได้ว่าข้อมูลเป็นของใคร ซึ่งข้อมูลนั้นก็จะถูกรับรองว่าเป็นของคน ๆ นั้นจริง ๆ โดยวิธีการในการลงนามแบบดิจิตอลที่ศึกษามีอยู่ 2 วิธี คือ Electronic Signature และ Digital Signature [1] ซึ่งมีความหมายและข้อแตกต่างกันดังนี้

2.1.1 Electronic Signature

เป็นการลงนามเอกสารโดยใช้สัญลักษณ์ หรือลายเซ็นที่อยู่ในรูปแบบอิเล็กทรอนิกส์ โดยเป็นการลงนามโดยบุคคลที่เป็นเจ้าของเอกสารหรือต้องการรับรองเอกสารนั้น สัญลักษณ์ที่นิยมใช้ได้แก่ รูปภาพลายเซ็นที่เซ็นด้วยหมึกปากกาลงในกระดาษแล้วสแกนเข้าสู่ระบบคอมพิวเตอร์ การใช้เมาส์ ลายนิ้วมือ stylus วาดรูปลายเซ็น  
บนหน้าจอคอมพิวเตอร์ ลายเซ็นที่แนบท้ายอีเมลล์ การพิมพ์ชื่อด้วยคีย์บอร์ด รูปภาพลายนิ้วมือ การคลิก   
“I Agree” ใน ข้อตกลง ต่างๆ เป็นต้น [2]

2.1.2 Digital Signature

Digital Signature นั้นมีจุดประสงค์เดียวกันกับ Electronic Signature เนื่องจากเป็นลายเซ็นที่อยู่ในรูปแบบของอิเล็กทรอนิกส์เหมือนกัน การลงนามก็จะใช้วิธีแบบเดียวกัน แต่มีการเพิ่มเติมคุณสมบัติทางด้านความปลอดภัยเข้าไป เพื่อให้มีความน่าเชื่อถือมากยิ่งขึ้น โดยคุณสมบัติดังกล่าวประกอบด้วย

1) Signer Authentication

เป็นคุณสมบัติที่ใช้ในการพิสูจน์ว่าใครเป็นคนลงนามเอกสารนั้น ตัวลายเซ็นจะมีเอกลักษณ์เฉพาะตัวที่สามารถใช้ในการเชื่อมโยงไปยังบุคคลที่ลงนามเอกสารได้

2) Data Integrity

เป็นคุณสมบัติที่ใช้ในการตรวจสอบ หรือพิสูจน์ว่ามีการแก้ไขเปลี่ยนแปลงเนื้อหาของเอกสารหลังจากที่ได้มีการลงนามไปแล้วหรือไม่ หากมีการแก้ไขก็จะทำให้เอกสารนั้นตกเป็นโมฆะ โดยจะมีการแจ้งเตือนไปยังผู้อ่าน หรือทำให้เอกสารนั้นไม่สามารถที่จะเปิดอ่านได้อีกต่อไป

3) Non-repudiation

การไม่สามารถปฏิเสธความรับผิดชอบได้ เนื่องจากลายเซ็นที่สร้างขึ้นมีเอกลักษณ์ สามารถพิสูจน์  
ในชั้นศาลได้ว่าใครเป็นผู้เซ็นเอกสาร และในปัจจุบันลายเซ็นดิจิตอลได้รับการยอมรับตามกฎหมายแล้ว จึงสามารถ  
ที่จะนำมาใช้เป็นหลักฐานในทางกฎหมายได้

**ตารางที่ 1** ข้อดีข้อเสียของ Electronic Signatures และ Digital Signature [1]

|  |  |  |
| --- | --- | --- |
| **วิธีในการลงนามแบบดิจิตอล** | **ข้อดี** | **ข้อเสีย** |
| Electronic Signature | - เป็นการใช้ สัญลักษณ์ หรือลายเซ็นที่อยู่ในรูปแบบอิเล็กทรอนิกส์ ที่มีเอก- ลักษณ์เฉพาะตัวที่ผู้ใช้สามารถเห็นได้ง่าย ทำให้ทราบได้ว่าใครเป็นเจ้าของลายเซ็น | – ถูกคัดลอกและปลอมแปลงไปใส่ในเอกสารอื่นได้ง่าย  – ไม่สามารถตรวจสอบ หรือพิสูจน์ได้ว่ามีการแก้ไขเปลี่ยนแปลงเนื้อหาของเอกสารหลังจากที่มีการลงนามไปแล้วหรือไม่  –ไม่สามารถพิสูจน์ได้ว่าผู้ที่ลงนามในเอกสารนั้นเป็นตัวจริงหรือไม่ จึงโดนปฏิเสธความรับผิดชอบได้ และไม่ได้รับการรับรองตามกฏหมาย |
| Digital Signature | – เอกสารที่ผ่านการลงนามแล้วจะไม่สามารถแก้ไขได้ หากมีการแก้ไขก็จะสามารถตรวจสอบได้  – ผู้เซ็นเอกสารจะไม่สามารถปฏิเสธความรับผิดชอบได้  – สามารถใช้เป็นหลักฐานได้ตามกฏหมายได้ เทียบเท่ากับการเซ็นเอกสารในกระดาษด้วยหมึกปากกา | – ความปลอดภัยขึ้นอยู่กับการเข้า รหัสลับ  - มีความยุ่งยากในการเชื่อมโยงการลงนามแบบดิจิตอลกับลายเซ็นที่แนบไปในเอกสาร |

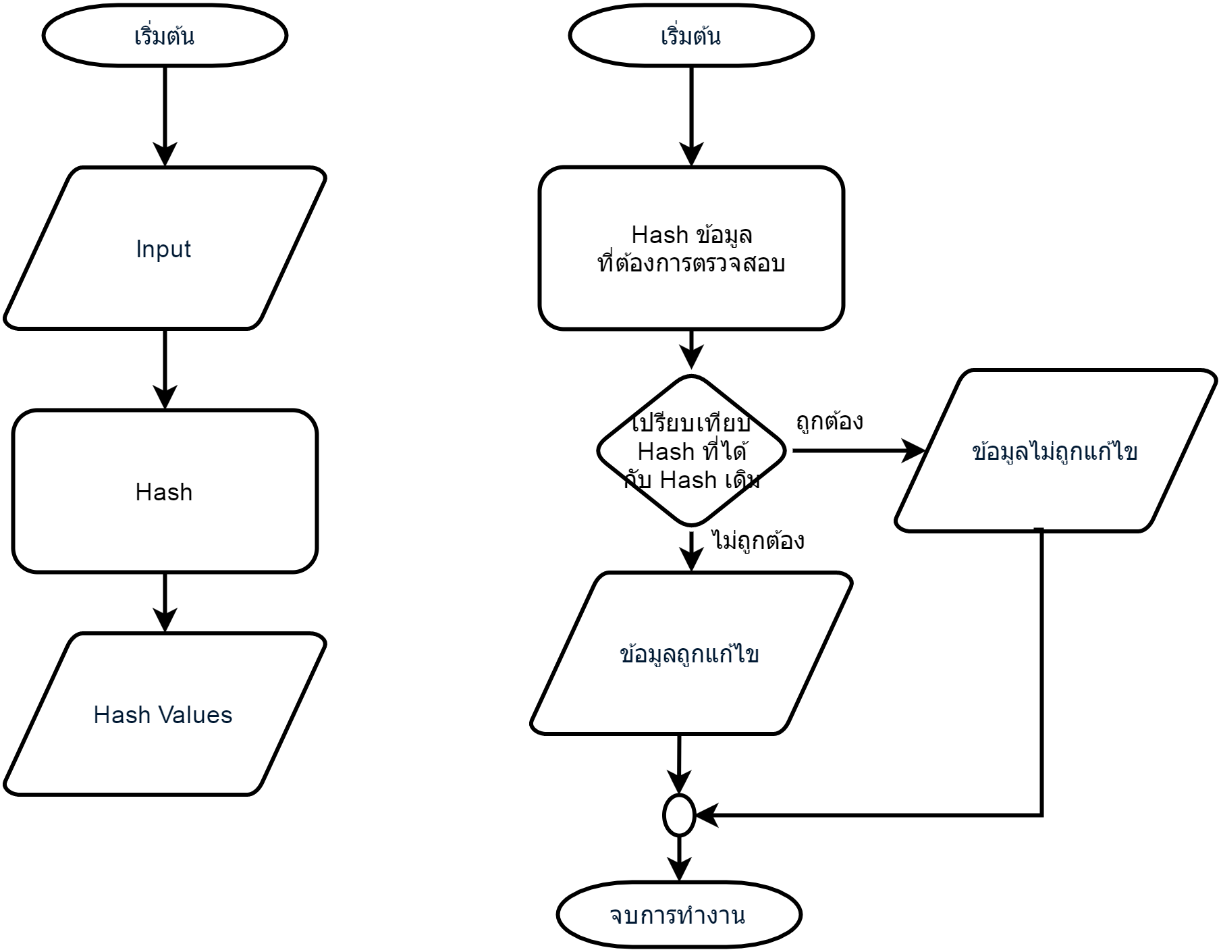
จากตารางที่ 1 จะเห็นว่าการลงนามเอกสารแบบอิเล็กทรอนิกส์ทั้ง 2 รูปแบบนั้น มีข้อดี และข้อเสียที่แตกต่างกันไป การเลือกรูปแบบที่จะมาใช้ในการพัฒนานั้นขึ้นอยู่กับว่า ความต้องการของระบบนั้นมีอะไรบ้าง เมื่อพิจารณาความต้องการของระบบจัดการโครงงานและงานวิจัยคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น พบว่าต้องการความปลอดภัยและน่าเชื่อถือ และจะปฏิเสธความรับผิดชอบไม่ได้   
จึงควรเลือกใช้รูปแบบ Digital Signature

**2.2 การเข้ารหัสข้อมูล (Cryptography)**

การเข้ารหัสข้อมูลโดยพื้นฐานแล้วจะเกี่ยวข้องกับวิธีการทางคณิตศาสตร์เพื่อใช้ในการปกป้องข้อมูล  
หรือข้อความตั้งต้นที่ต้องการส่งไปถึงผู้รับ ข้อมูลตั้งต้นจะถูกแปรเปลี่ยนไปสู่ข้อมูลหรือข้อความอีกรูปแบบหนึ่ง  
ที่ไม่สามารถอ่านเข้าใจได้โดยใครก็ตามที่ไม่มีกุญแจสำหรับเปิดดูข้อมูลนั้น เราเรียกกระบวนการในการแปรรูป  
ของข้อมูลตั้งต้นว่า การเข้ารหัสข้อมูล (Encryption) และกระบวนการในการแปลงข้อความที่ไม่สามารถอ่าน   
และทำความเข้าใจได้ให้กลับไปสู่ข้อความดั้งเดิม ว่าการถอดรหัสข้อมูล (Decryption) [3]

**2.3 การตรวจสอบความถูกต้องของข้อมูล**

เป็นการใช้ขั้นตอนวิธี (Algorithm) บางอย่างในการสร้างค่าสำหรับตรวจสอบความถูกต้อง (Integrity)   
ของข้อมูล ซึ่งถ้าหากข้อมูลถูกแก้ไขก็จะสามารถตรวจสอบได้ โดยการทำขั้นตอนวิธีเดิมซ้ำอีกครั้งจากนั้นนำค่า  
ที่ได้มาเปรียบเทียบกัน หากไม่มีการแก้ไขข้อมูลค่าที่ได้ก็จะเป็นค่าเดียวกัน ดังภาพที่ 1 โดยขั้นตอนวิธีในการสร้าง  
ค่า Hash นั้นมีอยู่หลายวิธี ซึ่งแต่ระวิธีก็จะมีข้อดีและข้อเสียแตกต่างกันไป ยกตัวอย่างเช่น Hash, Message Digest (MD), SHA เป็นต้น [4]



**ภาพที่ 1** แสดงขั้นตอนวิธีการตรวจสอบความถูกต้องของข้อมูล

**2.4 ชุดรหัสผ่านแบบใช้งานครั้งเดียว (One Time Password)**

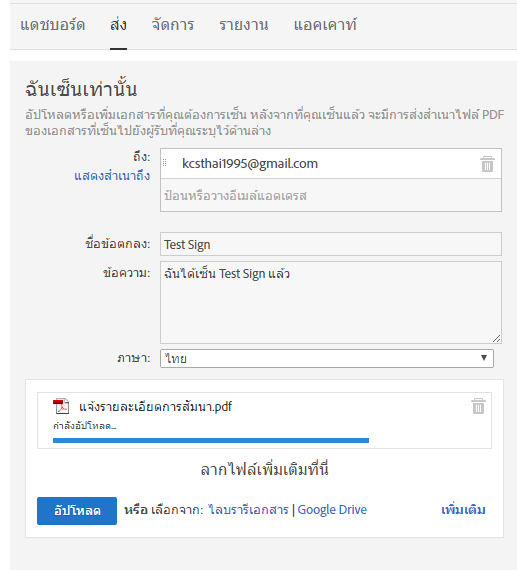
One Time Password (OTP) คือ รหัสผ่าน (Password) ในการเข้าสู่ระบบต่างๆ ที่ต้องการความปลอดภัยมากกว่าปกติ เช่นระบบชำระเงินออนไลน์ ระบบที่เกี่ยวข้องกับธนาคาร เป็นต้น โดยรหัสผ่านจะสามารถใช้ได้  
ครั้งเดียว และเป็นรหัสผ่านที่ถูกสร้างขึ้นให้ไม่ซ้ำกันโดยใช้ขั้นตอนวิธี หรือการสุ่ม (Random) ขึ้นมา ในปัจจุบันนิยมใช้ควบคู่กับการเข้าสู่ระบบด้วยรหัสผ่านแบบเดิมที่เป็นรหัสผ่านแบบตายตัว (Fixed Password) ซึ่งจะทำให้ระบบดังกล่าวมีระดับความปลอดภัยที่สูงขึ้น

ระบบ One-Time Password (OTP) มีหลายรูปแบบ เช่น อุปกรณ์แบบพกพา, Email และ SMS แต่ที่ได้รับความนิยมจะเป็นการใช้อุปกรณ์พกพาหรือ Token ในการสร้างรหัสผ่าน และการส่งรหัสผ่านไปยังโทรศัพท์มือถือที่ได้ลงทะเบียนไว้ [5]

**2.5 ระบบ Adobe Sign**

เป็นระบบการลงนามเอกสารออนไลน์ผ่านเว็บไซต์ที่ถูกพัมนาโดย Adobe Systems Incorporated เป็นการลงนามที่สามารถนำไปใช้อ้างอิงได้ตามกฎหมาย โดยการใช้งานนั้น ผู้ใช้งานจะต้องทำการสมัครสมาชิกกับ  
Adobe ก่อนโดยข้อมูลที่ใช้สมัครนั้นควรจะเป็นข้อมูลจริง เพราะข้อมูลในส่วนนี้จะถูกนำไปเป็นส่วนประกอบของลายเซ็น หลังจากที่สมัครสมาชิกเรียบร้อยแล้ว จะมีให้ทดลองใช้ได้ 30 วัน หากครบกำหนดการทดลองใช้แล้ว จะต้องเสียค่าใช้จ่ายในการใช้งานเป็นรายเดือน โดยจะมีขั้นตอนในการลงนามเอกสารดังต่อไปนี้

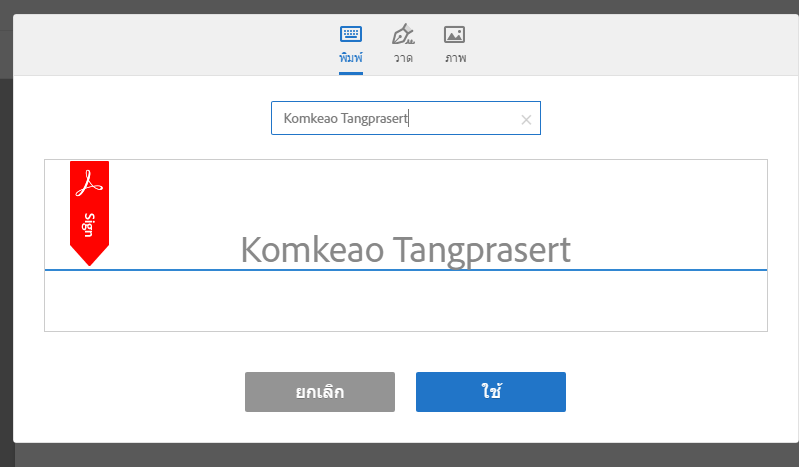
2.5.1 ผู้ใช้งานต้องทำการอัพโหลดเอกสารเข้าไปในระบบ ดังภาพที่ 2 โดยระบบจะให้ระบุ Email ของ  
ผู้ที่ต้องการให้ลงนามในเอกสารนั้น โดยจะสามารถระบุได้มากกว่า 1 คน ผู้ที่จะลงนามได้นั้น จะเป็นต้องสมาชิกกับทาง Adobe ด้วยเช่นกัน



**ภาพที่ 2** ขั้นตอนการอัพโหลดเอกสารเข้าสู่ระบบ

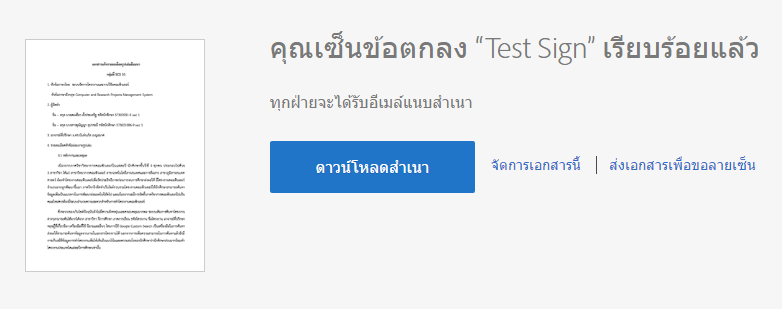
2.5.2 ระบบจะแสดงเนื้อหาของเอกสารที่อัพโหลดเสร็จแล้ว เพื่อให้ผู้ใช้งานเลือกตำแหน่งที่จะทำการเซ็นบนเอกสาร

2.5.3 เมื่อเลือกลงนามเอกสาร จะสามารถเลือกวิธิในการลงนามได้ 3 วิธี คือ การพิมพ์ข้อความเป็นตัวอักษร การใช้ภาพลายเซ็น และการวาดลายเซ็นเอง ดังภาพที่ 3



**ภาพที่ 3** ขั้นตอนการสร้างรายเซ็นดิจิตอล

2.5.4 หากทำการลงนามเอกสารและตรวจสอบข้อมูลแล้ว ก็สามารถเลือกดาวโหลดเอกสารได้ทันที   
หรือในกรณีที่มีผู้ลงนามเอกสารหลายคน ก็จะต้องรอให้ทุกคนทำการลงนามก่อน ดังภาพที่ 4



**ภาพที่ 4** ขั้นตอนการดาวโหลดเอกสารที่ลงนามแล้ว

จะเห็นได้ว่าการลงนามเอกสารแบบดิจิตอลในระบบ Adobe Sign ของ Adobe Systems Incorporated นั้นมีความสะดวกสบายใช้งานง่าย เป็นมิตรต่อผู้ใช้ และยังคงคุณสมบัติทั้งสามข้อคือ Signer Authentication,   
Data Integrity และ Non-repudiation ไว้ได้อย่างครบถ้วน เนื่องจากสามารถเชื่อมโยงไปยังผุ้ลงนาม เมื่อทำการ  
ลงนามไปแล้วก็จะไม่สามารถแก้ไขข้อความในเอกสารได้ รวมทั้งได้รับการยอมรับตามกฎหมายอีกด้วย

**2.6 การค้นคืนสารสนเทศ (Information Retrieval)**

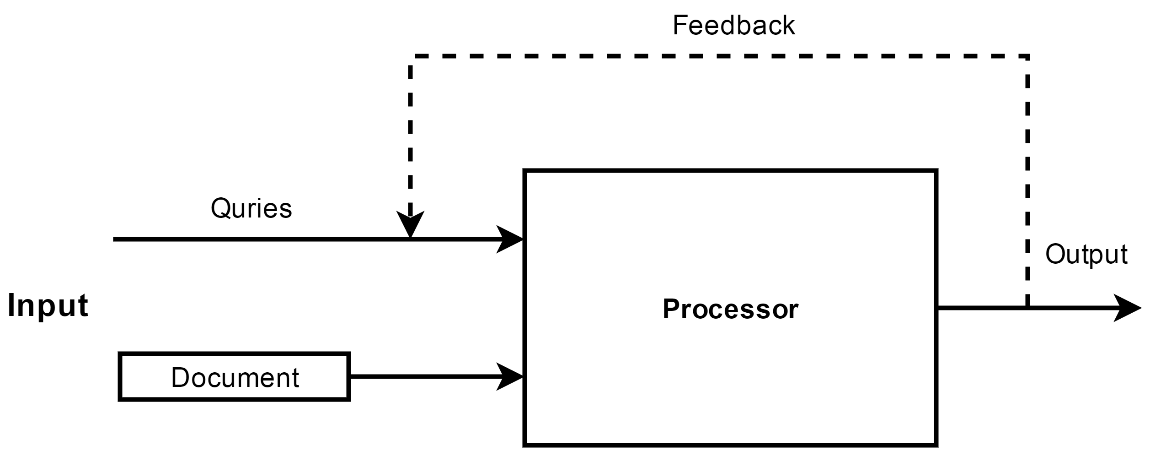
ระบบค้นคืนสารสนเทศ (Information Retrieval System) หรือ IR เป็นระบบที่จัดการและประมวลผลสารสนเทศประเภทเอกสาร (Document) ในรูปแบบต่างๆ เช่น หนังสือ วารสาร บทความ ซึ่งเกี่ยวข้องกับการเก็บบันทึก การดึงเอกสาร และการแสดงรูปแบบ โดยใช้ลักษณะเด่นของเนื้อหาของเอกสารเป็นหลักในการค้นหาและดึงเอาเอกสารที่ต้องการออกมา โดยแยกแยะความเกี่ยวข้อง (Relevant) และความไม่เกี่ยวข้อง (Non-relevant) [6]

2.6.1 ระบบค้นคืนสารสนเทศสามารถแบ่งได้ 3 ประเภทคือ

1) ระบบค้นคืนที่ให้คำถาม คำตอบ เป็นการบริการค้นคำตอบสำหรับคำถามที่ต้องการคำตอบที่  
เป็นข้อเท็จจริง

2) ระบบค้นคืนที่ให้ข้อมูลเป็นตัวเลขหรือสัญลักษณ์ เป็นระบบจัดเก็บข้อมูลทางฟิสิกส์ เคมี   
สำมะโนประชากร เป็นต้น

3) ระบบค้นค้นข้อความจากวารสาร เป็นระบบที่จัดเก็บตัวเนื้อหา เอกสารและสามารถเรียกข้อความส่วนใดส่วนหนึ่งของเอกสารได ้ เช่น ฐานข้อมูลทางกฎหมายเป็นต้น [9]



**ภาพที่ 5** แสดงส่วนประกอบของระบบค้นคืนสารสนเทศ [6]

2.6.2 ส่วนประกอบของระบบค้นคืนสารสนเทศดังภาพที่ 5 แบ่งได้เป็น 3 ส่วนได้แก่

1) ส่วนนําเข้าข้อมูล (Input) เป็นส่วนของการป้อนข้อคําถาม (query) จากผู้ใช้ซึ่งเป็นภาษาธรรมชาติหรืออาจเป็นการนําเข้า Metadata ซึ่งเป็นสารสนเทศเกี่ยวกับเอกสารหรืออาจไม่เป็นส่วนหนึ่งขอเอกสารก็ได้แต่เป็นข้อมูลเกี่ยวกับข้อมูล (data about data) อาทิเช่น

ก) Descriptive Metadata เปนการนําเขาสารสนเทศที่เปนความหมายของเอกสารที่อยูภายนอก เช่น ผูแตง(Author), ชื่อเรื่อง(Title), แหลงที่มา (Source), วันที่ (Date), ISBN, สํานักพิมพ (Publisher), ความยาว (Length)

ข) Semantic Metadata Concerns The Content เปนการนําเขาเนื้อความที่มีความหมายเชน บทคัดย่อ (Abstract), คําสําคัญ (Keywords), รหัสของหัวเรื่อง (SubjectCodes) ซึ่งอาจเปน Library of Congress หรือ Dewey Decimal หรือ UMLS (Unified Medical Language System) ก็ได

ค) เทอมของหัวเรื่อง (Subject terms) ซึ่งอาจมาจาก ontologies พิเศษเปนลําดับขั้นของเทอมมาตรฐาน (hierarchical taxonomies of standardized semantic terms)

ง) อาจเปนสารสนเทศของเว็บ (Web Metadata) ก็ได เชน META tag in HTML

ระบบค้นคืนสารสนเทศจะนําสารสนเทศเหลานี้ผานการประมวลผลแบบเชื่อมโดยตรงกับระบบคอมพิวเตอร ซึ่งผูใชจะมีการโตตอบหรือปฏิสัมพันธกับระบบโดยตรง

2) โปรเซสเซอร (Processor) เปนส่วนของการประมวลผล ไดแก การจัดโครงสรางของสารสนเทศในรูปแบบที่เหมาะสม ประกอบดวย การสรางตัวแทนเอกสาร, การแบงแยกกลุมของเอกสาร, การจัดเก็บสารสนเทศ, การดึงขอมูลตามที่ผูใชตองการ การทํางานนั้นจะนําขอคําถามไปเปรียบเทียบกับตัวแทนเอกสารที่มีอยู เพื่อดึงเอกสารที่ใกล้เคียงนำออกมาให แกผูสอบถาม

3) สวนของผลลัพท์ (Output) ผลลัพธที่ไดจากระบบเป็นข้อความสั้นๆ เชน ชื่อหนังสือ, หมายเลขเอกสาร, ชื่อผูแต่ง, สํานักพิมพ เปนตน ผูใชสามารถพิจารณาจากขอมูลตางๆ ที่ไดจากระบบถาเอกสารที่ไดมีจํานวนมากเกินไปหรือไมใกลเคียงกับสิ่งที่ตองการ ผูใชสามารถปรับปรุงขอคําถามใหมเพื่อใหขอคําถามนั้นสืบคนสารสนเทศไดตรงกับความตองการมากที่สุด เป็นระบบตอบกลับ (feedback) ดังนั้นผลลัพธ์ที่ไดจึงขึ้นอยูกับ ขอคําถาม (Query) ของผูใช้ [6]

2.6.3 หลักการทำงานของการค้นคืนสารสนเทศ

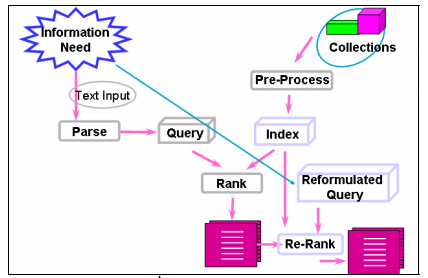
1) การคัดเลือก เป็นการการรวบรวมเอกสารตามเกณฑ์ และนโยบายที่กำหนด ไว้ซึ่งสอดคล้องกับความต้องการของผู้ใช้งาน

2) การวิเคราะห์เอกสาร ได้แก่ การจัดหมวดหมู่ การจัดทำ รายการการทำดรรชนี และการทำสาระสังเขป

3) การจัดระเบียบแฟ้มข้อมูล

4) การค้นคืน วิเคราะห์แนวคิดและกำหนดศัพท์ หลังจากนั้นก็นำคำศัพท์ไปดำเนินการค้น ถ้าคำศัพท์ตรงกับดรรชนี คำค้นของเอกสารนั้นจะได้รับ เอกสารจำนวนหนึ่ง หรือผู้ค้นจะทำการปรับปรุงเอกสารให้เป็นที่พอใจ  
ของผู้ใช้บริการ

5) การแจกจ่าย เป็นการนำส่งผลการค้นคืนให้แกผู้ใช้ ท ี่มีความต้องการเอกสาร เอกสารในเรื่องนั้นๆ [9]



**ภาพที่ 6** แสดงหลักการทำงานของการค้นคืนสารสนเทศ [9]

2.6.4 การสรางระบบคนคืนสารสนเทศ แบงออกเปน 4 ขั้นตอนคือ

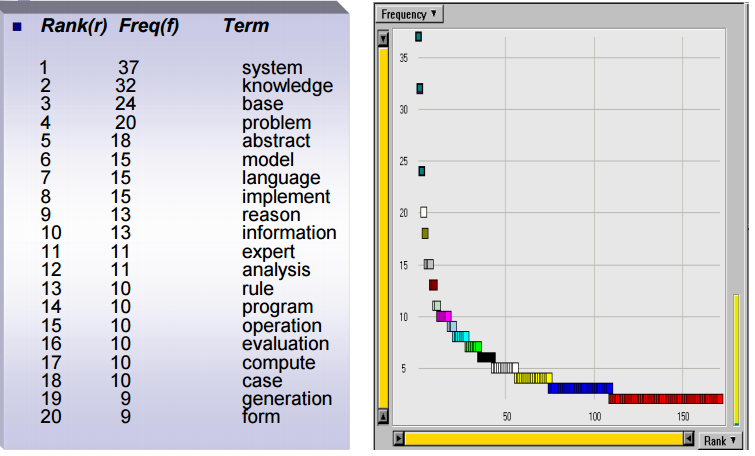
1) การวิเคราะหขอความ (Text Analysis) เป็นวิธี การวิเคราะห์ข้อความให้ได้มาซึ่งตัวแทนเอกสาร

แบ่งออกเป็น 2 วิธีคือ

ก) วิเคราะห์ทางด้านภาษาศาสตร์ วิธีการนี้ มีความยุ่งยากและเสียค่าใช้จ่ายสูงมาก

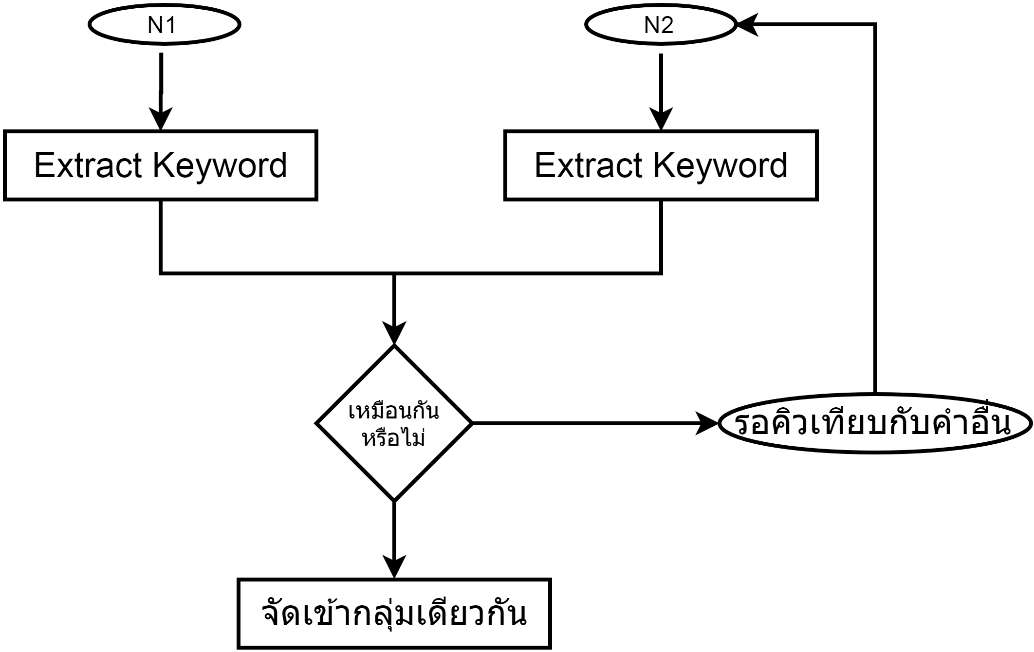
ข) วิเคราะห์ทางด้าน สถิติศาสตร์เป็นวิธีการที่นิยม

ซึ่งวิธีการนี้ได้มีการ ตรวจสอบและทดลองเพื่อให้สามารถนา มาประยุกต์ใช้กับ ระบบค้นคืนสารสนเทศ (IR) ได้ดีซึ้งได้ศึกษาหลัก การพื้นฐานโดยใช้ทฤษฎี ของลุนซ์ (Luhn) เข้ามาช่วยในการในการสร้างตัวแทนเอกสารโดยสมมุติให้ f เป็นความถี่ (Frequency) ของการเกิดขึ้นของคำใหม่ ในตำแหน่ง ต่างๆของข้อความ และให้ r เป็น Rank Order หรือลำดับของระดับ ของการเกิดขึ้นของคำๆ นั้นในเอกสาร และความสัมพันธ์ของ f และค่าของ f สูงจะส่งผลให้ r มีค่าต่ำ ดังภาพที่ 7 ดังนั้นความถี่ของข้อมูล (Frequency Data) สามารถถูกใช้เพื่อนำมาใช้วัด ความสำคัญของคำและประโยคที่ใช้แทนเอกสารหนึ่งได้ [10]



**ภาพที่ 7** แสดงการวิเคราะหขอความโดยใช้ทฤษฎีของลุนซ์ (Luhn) [10]

2) การจัดแบงกลุมขอมูล (Classification)

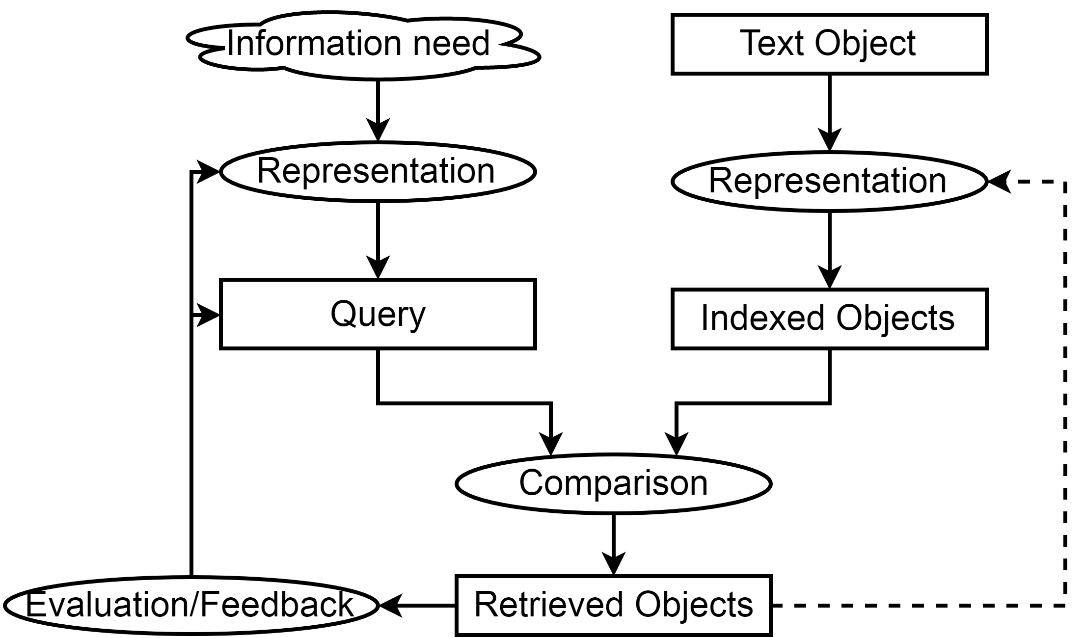


**ภาพที่ 8** แสดงการจัดแบงกลุมขอมูล (Classification) [11]

เมื่อได้ตัวแทนเอกสารแล้ว จากนั้นเราจะนำตัวแทนเอกสารที่ได้ มาเปรียบเทียบกัน ว่าเหมือนกันหรือไม่ ถ้าเหมือนก็จะนำมาจัดเข้ากลุ่มเดียวกัน แต่ถ้าไม่เหมือนกันจะไปรอคิวเพื่อเทียบกับคำอื่นต่อไปดังภาพที่ 8

3) การเก็บบันทกข้อมูลลงในแฟมขอมูล เราก็จะนำตัวแทนเอกสารที่ได้มาจัดลงแฟ้มแทนข้อความ  
ของเอกสารทั้งหมดให้สมบูรณ์ให้อยู่ในรุปแบบดรรชนี (Index)

4) การคนคืนสารสนเทศ



**ภาพที่ 9** แสดงการคนคืนสารสนเทศ

การค้นคืน เราจะนำข้อคำถาม (Query) ไปดำเนินการค้นในแฟ้ม (Indexed Objects) ว่ามีอะไรที่ตรงหรือใกล้เคียงกับตัวแทนเอกสารหรือดรรชนี (Indexed Objects) ไหม ถ้าข้อคำถาม (Query) กับดรรชนี (Indexed Objects) ตรงกัน คำระบบจะได้รับ เอกสารจำนวนหนึ่งนำม

าจัดลำดับและส่งไปแสดงให้ผู้ใช้ หรือผู้ใช้สามารถจะทำการปรับปรุง (Feedback) โดยสงความต้องการมาใหม่ ระบบจะทำการจัดลำดับใหม่ เพื่อให้เอกสารให้เป็นที่พอใจของ ผู้ใช้บริการตามภาพที่ 9

2.6.5 การประเมินผลระบบคนคืนสารสนเทศ (Evaluation of IR System)

ค่าความครบถ้วน (Recall) คือเป็นการวัดความสามารถของระบบในการดึงเอกสารที่เกี่ยวข้องออกมา โดยให้ ค่าของจํานวนเอกสารที่ถูกเรียกคืนออกมาได้และตรงความต้องการ หารด้วยจํานวนเอกสารที่ตรง  
ความต้องการทั้งที่ถูกเรียกคืนและไม่ถูกเรียกคืน ซึ่งสามารถเขียนเป็นสูตรคณิตศาสตร์ [8] ได้ดังนี้

Recall =

ส่วนค่าความแม่นยำ (Precision) คือเป็นการวัดความสามารถของระบบในการขจัดเอกสารที่ไม่เกี่ยวข้องออกไป โดยใช้ค่าของจํานวนเอกสารที่ถูกเรียกคืนและตรงความต้องการ หารด้วยจํานวนเอกสารที่ถูกเรียกคืนออกมาได้ทั้งหมดไม่ว่าจะตรงหรือไม่ตรงความต้องการ ซึ่งสามารถเขียนเป็นสูตรคณิตศาสตร์ได้ดังนี้

Precision =

ระบบค้นคืนสาระสนเทศที่ดีนั้นจะมี Recall สมดุลกับ Precision

**C**

**B**

**A**

**พื้นที่ค้นหา**

**ภาพที่ 10** ภาพอธิบายวิธีการคํานวณค่าความครบถ้วน และค่าความแม่นยำ [8]

จากภาพที่ 10 จะอธิบายวิธีการคํานวณค่าความครบถ้วน และค่าความแม่นยำดังนี้ วงกลม A+B คือ เอกสารที่ตรงกับความต้องการ วงกลม C+B คือ เอกสารที่ระบบเรียกคืนออกมาได้ ส่วน B คือ เอกสารที่ระบบ  
เรียกคืนออกมาได้และตรงกับความต้องการ ค่าความครบถ้วน มีวิธีคิดโดย นําค่าจำนวนเอกสารที่ระบบเรียกคืนออกมาได้และ ตรงกับความต้องการ (B) หารด้วยผลรวมของจํานวนเอกสารที่ตรงกบความต้องการและจํานวนเอกสารที่ระบบเรียกคืนออกมาได้และตรงกับความต้องการ (A+B) ส่วนคาความแม่นยำ มีวิธีคิด โดยนําค่าจำนวนเอกสารที่ระบบเรียกคืนออกมาได้และตรงกบความต้องการ (B) หารด้วยผลรวมของเอกสารที่ระบบเรียกคืนออกมาได้และตรงกับความต้องการและเอกสารที่ระบบเรียกคืนออกมา (B+C) ซึ่งสามารถเขียนเป็นสูตรคณิตศาสตร์ได้ดังนี้

ค่าความครบถ้วน =

ค่าความแม่นยำ =

ค่าความครบถ้วน และค่าความแม่นยำจะเป็นค่าที่มีจํานวนเต็มเท่ากับหนึ่ง แต่เพื่อให้ทําความเข้าใจได้ง่ายขึ้น ผู้วิจัยจึงคูณเลขหนึ่งร้อยเพิ่มเข้าไปในการคํานวณค่าดังกล่าวเพื่อเลื่อนจุดทศนิยมขึ้น 2 ตําแหน่ง ทําให้มีจํานวนเต็มเท่ากับหนึ่งร้อย และเนื่องจากค่าความครบถ้วนและค่าความแม่นยำเป็นค่าที่มองจากมิติที่ต่างกัน หากต้องการผลที่สะท้อนค่าทั้งสองนี้จําเป็นจะต้องนําค่าทั้งสองไปคํานวณหาคาเฉลี่ยที่ เรียกว่า ค่าเอฟ-เมเชอร์ (F-measure) ซึ่งจะให้น้ำหนักของค่าความครบถ้วนและค่าความแม่นย่ำเท่าๆกัน ค่าเอฟ-เมเชอร์ จึงเป็นค่าที่นิยมใช้บอกประสิทธิภาพของระบบ ค่าเอฟ-เมเชอร์สามารถคำนวณได้ โดยนําค่าความครบถ้วนคูณด้วยค่าความแมนยำคูณด้วยสอง แล้วหารด้วยผลรวมของคาความครบถ้วนและค่าความแม่นยำ ซึ่งเขียนเป็นสูตรคณิตศาสตร์ได้ [8] ดังนี้

F =

**ตารางที่ 2** ความแตกต่างระหว่าง Database และการค้นคืนสารสนเทศ [7]

|  |  |  |
| --- | --- | --- |
|  | **Database** | **IR** |
| การค้นข้อมูล | ข้อมูลมีโครงสร้าง  มีความหมายที่ชัดเจน | จะไม่ใช้โครงสร้างข้อมูล  ส่วนใหญ่จะเป็น Metadata |
| การดึงข้อมูลข้อมูล | ตรงไปตรงมา | คลุมเครือ ไม่แน่ชัด  (ตามธรรมชาติของภาษา) |
| ผลลัพท์ที่ได้ | แน่นอนและถูกต้อง | เกี่ยวข้องเป็นบางครั้ง |
| การติดต่อกับระบบ | เป็นการ Query แบบ  ครั้งเดียวจบ | มีการโต้ตอบและปรับปรุงผลลัพท์ให้มีความถูกต้องมากขึ้น |

จากตารางที่ 2 จะเห็นว่าการค้นคืนสารสนเทศ (IR) มีความยืดหยุ่นในการค้นหาข้อมูลมากกว่า Database เพราะเป็นการค้นหาข้อมูลจากลักษณะเด่นภายในเอกสาร และ Metadata นอกจากนี้ยังสามารถที่จะปรับปรุงการค้นหา ให้มีความถูกต้องมากยิ่งขึ้น

**2.6.  ขอบเขตและข้อจำกัดของการวิจัย**

1  ระบบถูกใช้งานผ่านเว็บบราวเซอร์บนคอมพิวเตอร์และโทรศัพท์มือถือ

2  โครงงานที่ทำการรวบรวมเป็นโครงงานของนักศึกษาชั้นปีที่ 4 ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่นเท่านั้น

**2.7.  สถานที่ทำวิจัย**

       ภาควิชาวิทยาการคอมพิวเตอร์  คณะวิทยาศาสตร์  มหาวิทยาลัยขอนแก่น

**2.8 ประโยชน์ที่คาดว่าจะได้รับ**

1 สามารถเป็นตัวอย่าง แนวทางในการทำโครงงานและต่อยอดโครงงานที่มีอยู่แล้วให้ดียิ่งกว่าเดิม

2 สามารถค้นหาโครงงานจากคำคีย์เวิร์ด ตามTag และการระบุเงื่อนไขได้ถูกต้อง ละเอียดมากยิ่งขึ้น

3 สามารถเซ็นลงนามเอกสารผ่านระบบได้

**2.9  แผนและระยะเวลาดำเนินการ**

        ตารางการดำเนินงานและระยะเวลา