- ชื่อของอิลิเมนต์ต้องเป็นไปตามกฎของชื่อใน XML
- XML ให้ความสำคัญกับตัวอักษรที่เป็นตัวเล็ก- ใหญ่
- XML จะไม่ตัดส่วนที่เป็น White Space (ช่องว่าง) ทิ้ง

แอตตริบิวต์ (Attribute)

หมายถึง คู่ของชื่อและค่า ที่สัมพันธ์กันและสัมพันธ์กับอิลิเมนต์ด้วย

ส่วนประกอบใน XML สามารถใช้แอตตริบิวต์ได้ รูปแบบไวยากรณ์เป็นดังนี้:

<element attribute-name = "attribute-value">....</element>

ทุกๆแอตตริบิวต์ต้องมีค่าของแอตตริบิวต์อยู่เสมอแม้ว่าค่าของแอตตริบิวต์จะเป็นเพียงสตริงว่างๆ เช่น "" และ ค่าของแอตตริบิวต์ต้องอยู่ในระหว่างเครื่องหมายคำพูด แม้จะมีเพียงเลขจำนวนก็ตาม

การกำหนดความถูกต้องของ XML ด้วย DTD

DTD มาจากคำว่า Document Type Definition เป็นภาษาที่ทำหน้าที่ในการกำหนดกฎ กติกา ให้กับอีลิเมนต์ (Element), แอตทริบิวต์ (Attribute) และ เอ็นทิตี้(Entity) ที่อยู่ในแหล่งข้อมูล XML

1. วิธีการกำหนดความถูกต้องของข้อมูลสามารถทำได้ 2 ลักษณะคือ

- **แบบฝังส่วนของภาษา DTD ไว้ในXML**

DTD อาจถูกรวมอยู่ในตัวเอกสาร XML เองก็ได้ หลังจากบรรทัด <?xml version="1.0"?> จะต้องพิมพ์ว่า

<!DOCTYPE name of root-element [ตามด้วยนิยามของส่วนประกอบ ส่วนของ DTD จะปิดท้ายด้วย]>

- **แบบอ้างอิงชื่อไฟล์ .dtd**

DTD อาจเป็นเอกสารภายนอกที่ถูกอ้างถึงก็ได้ DTD ดังกล่าวจะเริ่มต้นด้วยข้อความ

<!DOCTYPE name of root-element SYSTEM "address">

แอดเดรสหมายถึง URL ซึ่งชี้ไปยัง DTD นั้น

ในเอกสาร XML ต้องระบุให้ชัดเจนว่าคุณจะใช้ DTD นี้ด้วยบรรทัดข้อความ

<!DOCTYPE name of root-element SYSTEM "address"> ซึ่งควรอยู่ต่อจากบรรทัด

<?xml version="1.0"?>

- เช่น <!DOCTYPE COVER_PAGE SYSTEM "cover.dtd"> เป็นการเรียกใช้ ไฟล์ cover.dtd

โดยทั่วๆไปแล้วนิยมทำการกำหนดความถูกต้องของข้อมูล แบบอ้างอิงชื่อไฟล์ ทั้งนี้เพื่อความสะดวกในการนำไปใช้กับไฟล์อื่นๆ

2. ส่วนประกอบของ DTD

DTD อธิบายส่วนประกอบ โดยใช้ไวยากรณ์ดังต่อไปนี้

ข้อความ <!ELEMENT ตามด้วยชื่อของส่วนประกอบ ตามด้วยคำอธิบายของส่วนประกอบนั้น ตัวอย่างเช่น

<!ELEMENT AUTHOR (#PCDATA)>

คำอธิบาย DTD นี้ นิยามแท็ก <AUTHOR> ของ XML

คำอธิบาย (#PCDATA) ย่อมาจาก parsed character data หมายถึงเนื้อหาของอิลิเมนต์ที่เป็นเพียงข้อมูล

XQuery

XQuery เป็นภาษาที่ใช้สำหรับสืบค้นข้อมูลจากเอกสาร XML โดยW3C ได้เป็นผู้เสนอมาตรฐานภาษาการสืบค้นข้อมูล โดยถูกพัฒนามาจากภาษา SQL ซึ่งทำให้ XQuery สามารถใช้งานได้ง่ายและเป็นที่ยอมรับกันอย่างแพร่หลาย โดย จะมีแนวคิดของ path expression เพื่อทำการระบุไปยังตำแหน่งต่างๆในเอกสาร ซึ่งเป็นวิธีที่สะดวก และมีประสิทธิภาพสำหรับโครงสร้างเอกสารที่มีความซับซ้อน

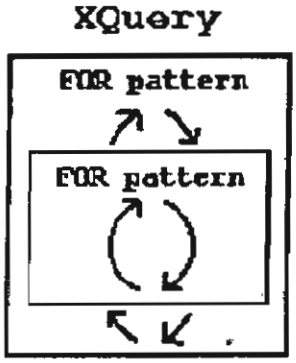
การทำงานของXQuery นั้น เนื่องจาก มีต้นแบบมาจากภาษา SQL จึงทำให้โครงสร้างของประโยคจึงมีความคล้ายคลึงกัน กล่าวคือ จะประกอบไปด้วย ประโยค FOR-LET-WHERE โดยที่ คำสั่ง FOR จะหมายถึง การระบุข้อมูลโดย path expression และนำข้อมูลนั้นมากำหนดให้กับตัวแปร ในแต่ละรอบของการทำงาน ส่วนคำสั่ง LET จะใช้เพื่อกำหนดค่าตัวแปรซึ่งมีความสัมพันธ์กับ FOR ส่วน WHERE จะเป็นส่วนที่ทำการตรวจสอบเงื่อนไขให้กับส่วนประโยค และส่วนสุดท้ายคือ RETURN คือส่วนของการสร้างผลลัพธ์ที่ต้องการออกมา ซึ่งจะเห็นได้ว่า XQuery นั้นมีความกะทัดรัด เข้าใจได้ง่าย

โครงสร้างการทำงานพื้นฐานของ XQuery

การทำงานในการเรียกค้นข้อมูลของ XQuery นั้นจะใช้ประโยค FOR - LET - WHERE - RETURN จากตัวอย่างต่อไปนี้

```
FOR $b IN document("bib.xml")/bib/book
RETURN <book year={ $b/@year }>
{ $b/title }
</book>
```

ซึ่งสามารถสรุปเป็นรูปที่ 3.3 ได้ดังนี้



รูปที่ 3.3 แสดงรูปแบบการทำงานของ XQuery

การเรียกค้นข้อมูลที่เป็นเอกสาร

<ABSTRACT_THAI>ฐานข้อมูลที่อยู่ในรูปของ XML นั้นแม้จะมีความซับซ้อนได้มากกว่าฐานข้อมูลแบบสัมพันธ์ที่มีเพียง 2 มิติ โครงสร้างของข้อมูลจะสามารถซ้อนกันได้หลายชั้น แต่ละชั้นอาจประกอบด้วยโครงสร้างย่อยอีกมากมาย เช่นจากตัวอย่างของข้อมูลบทคัดย่อดังนี้

<PROJECT_CODE>BGJ / 11 / 2543</ PROJECT_CODE >

<TITLE_THAI> การศึกษาบทบาทของ cell wall hydrolases องค์ประกอบของผนังเซลล์และการแสดงออกของยีนที่ควบคุมเอนไซม์ที่เกี่ยวข้องกับการหลุดร่วงของผลกล้วยระหว่างการสุก </ TITLE_THAI >

<AUTHOR>

<TITLE_NAME>ศ.ดร.</TITLE_NAME>

<FIRST_NAME>สายชล</FIRST_NAME>

<LAST_NAME>เกตุษา</LAST_NAME>

</AUTHOR>

<AUTHOR>

<TITLE_NAME>น.ส.</TITLE_NAME>

<FIRST_NAME>วชิรญา</FIRST_NAME>

<LAST_NAME>อิ่มสบาย</LAST_NAME>

</AUTHOR>

<AUTHOR>

<TITLE_NAME>น.ส.</TITLE_NAME>

<FIRST_NAME>อภิวรา</FIRST_NAME>

<LAST_NAME>ประยูรวงศ์</LAST_NAME>

</AUTHOR>

<DEPARTMENT>ภาควิชาพืชสวน</DEPARTMENT>

<FACULTY>คณะเกษตร </FACULTY>

<UNIVERSITY>มหาวิทยาลัยเกษตรศาสตร์</UNIVERSITY>

<Date> วันที่ 30 กันยายน 2543 ถึง วันที่ 29 กันยายน 2544</Date>

<Date>ระยะเวลาโครงการ : วันที่ 30 กันยายน 2543 ถึง วันที่ 29 กันยายน 2544</Date>

<PARAGRAPH>การศึกษาเกี่ยวกับการหลุดร่วงของผลกล้วยระหว่างการสุก ประกอบด้วย 5 การทดลองย่อยคือ 1) การเปลี่ยนแปลงเอนไซม์ cell wall hydrolases และองค์ประกอบของผนังเซลล์ในเปลือกและเนื้อของผลกล้วยไข่ระหว่างอ่อนตัว 2) บทบาทของเอนไซม์ที่ทำให้เกิดการอ่อนตัวของเนื้อเยื่อต่อการหลุดร่วงของผลกล้วยระหว่างการสุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูง 3) การศึกษาเปรียบเทียบการหลุดร่วงของกล้วยหอมและกล้วยน้ำหว้าระหว่างการสุก 4) การศึกษาชั้นบริเวณการหลุดร่วงของผลกล้วยระหว่างการสุก และ 5) การแสดงออกของยีนที่ควบคุมเอนไซม์เกี่ยวข้องกับการหลุดร่วงของผลกล้วยระหว่างการสุก ในการศึกษาครั้งนี้พบว่า การอ่อนตัวของเปลือกและเนื้อของผลกล้วยไข่ระหว่างการสุกเกิดขึ้นในลักษณะที่คล้ายกัน เพคตินที่ละลายน้ำได้มีการเพิ่มขึ้นในเนื้อแต่ไม่ได้เพิ่มในเปลือก กิจกรรม pectin methylesterase ลดลงในเปลือกแต่เพิ่มในเนื้อ ขณะที่กิจกรรม polygalacturonase ลดลงในเปลือกแต่เพิ่มในเนื้อ กิจกรรม ?-galactosidase มีการเพิ่มในเปลือกมากกว่าในเนื้อ ส่วนกิจกรรม cellulase ทั้งในเปลือกและเนื้อไม่มีการเปลี่ยนแปลงระหว่างการสุก

การศึกษาการหลุดร่วงของผลกล้วยไข่และกล้วยหอมทองระหว่างการสุกพบว่า ทั้งผลกล้วยไข่และกล้วยหอมทองที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์สูง (90%) มีการหลุดร่วงของผลมากกว่าและเร็วกว่าผลกล้วยที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำ (60%) กิจกรรม polygalacturonase และ pectin methylesterase ในเปลือกตรงกลางผล และตรงขั้วผลบริเวณที่เกิดการหลุดร่วง ของผลกล้วยไข่ที่บ่มภายใต้ความชื้นสัมพัทธ์สูง มีมากกว่าผลกล้วยที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำ และเนื้อเยื่อบริเวณตรงขั้วผลมีกิจกรรม polygalacturonase และ pectin methylesterase มากกว่าในเนื้อเยื่อเปลือกบริเวณตรงกลางผล ขณะที่การเปลี่ยนแปลงกิจกรรม ?-galactosidase ในเนื้อเยื่อบริเวณกลางผลและขั้วผลของผลกล้วยไข่ที่บ่มภายใต้ความชื้นสัมพัทธ์ต่ำและสูง ไม่สัมพันธ์กับการหลุดร่วงของผล กิจกรรม polygalacturonase ในเปลือกผลตรงกลางและตรงขั้วผลของผลกล้วยหอมทองบ่มให้สุกภายใต้ความชื้นสัมพัทธ์สูงมีมากกว่าในผลกล้วยหอมทองที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำ ขณะที่กิจกรรม pectin methylesterase ในเนื้อเยื่อบริเวณดังกล่าวของผลกล้วยไข่หอมทองที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูงไม่มีความแตกต่างกัน กิจกรรมทั้ง polygalacturonase และ pectin methylesterase ในบริเวณขั้วผลมีมากกว่าในเนื้อเยื่อของเปลือก

บริเวณกลางผลของผลกล้วยที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูง ขณะที่การเปลี่ยนแปลงกิจกรรม ?-galactosidase ในเนื้อเยื่อบริเวณทั้งสองส่วน ไม่สัมพันธ์กับการหลุดร่วงผลกล้วยหอมทองที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูง

การศึกษาเปรียบเทียบการหลุดร่วงผลกล้วยหอมทองและกล้วยน้ำว้าระหว่างการบ่มให้สุกภายใต้ความชื้นสัมพัทธ์ 85% พบว่าผลกล้วยหอมทองมีการหลุดร่วง 100% ขณะที่ผลกล้วยน้ำว้าไม่มีการหลุดร่วงของผล ภายในเวลา 5 วันระหว่างการบ่มให้สุก ผลกล้วยน้ำว้ามีความแน่นเนื้อเปลือกและแรงต้านทานการเกิดหลุดร่วงของผลมากกว่าผลกล้วยหอมทอง กิจกรรม polygalacturonase ในเนื้อเยื่อขั้วผลบริเวณการหลุดร่วงของผลกล้วยหอมทอง มีการเพิ่มขึ้นอย่างรวดเร็วและมีมากกว่าในบริเวณเปลือกตรงกลางผลของผลกล้วยน้ำว้าและหอมทอง และกิจกรรม polygalacturonase ในบริเวณขั้วผลของผลกล้วยหอมทอง มีมากกว่าของผลกล้วยน้ำว้า กิจกรรม pectin methylesterase ในเปลือกผลกล้วยหอมทองเพิ่มขึ้นเล็กน้อยขณะที่ในผลกล้วยน้ำว้ากลับลดลงในระหว่างการสุก ขณะที่กิจกรรม pectin methylesterase ในขั้วผลบริเวณหลุดร่วงของผลกล้วยน้ำว้ามีมากกว่าในผลกล้วยหอมทอง

การศึกษาชั้นของเนื้อเยื่อบริเวณที่เกิดการหลุดร่วงของผลกล้วยหอมทองพบว่า ไม่มีการสร้างชั้นของบริเวณการร่วง (abscission zone) เกิดขึ้นในบริเวณเนื้อเยื่อที่จะเกิดการหลุดร่วงของผล การศึกษาการแสดงออกของยีนที่ควบคุมเอนไซม์ polygalacturonase พบว่า การแสดงออกของยีนในบริเวณขั้วผลที่เกิดรอยหลุดร่วงในผลกล้วยหอมทองมีมากกว่าในผลกล้วยน้ำว้าและการแสดงออกของยีนในบริเวณขั้วผลมีมากกว่าในบริเวณเปลือกผล</PARAGRAPH>

<Keyword>คำสำคัญ : การหลุดร่วงของผล, การสุก, กล้วยไข่, กล้วยหอมทอง, กล้วยน้ำว้า, บริเวณการร่วง, การอ่อนตัว, polygalacturonase, pectin methylesterase, ?-galactosidase

</Keyword>

จะเห็นได้ว่ามีการแบ่งลำดับชั้นของข้อมูลไว้ เป็นส่วนๆ เพื่อให้ง่ายต่อการQueryข้อมูล แต่เนื่องจากเอกสารนั้นในบางครั้งไม่มีรูปแบบที่ตายตัวที่แน่นอนจึงทำให้ในบางครั้งมีข้อจำกัดในการค้นหา

การปรับปรุงโครงสร้างข้อมูลภายหลังจากการทำการวิเคราะห์และประเมินผลการสร้างโครงสร้างในช่วงแรกของงานวิจัยพบว่ายังมีปัญหาในการสืบค้นหาข้อมูลต่าง ๆ อยู่มาก จึงได้ทำการวิเคราะห์และหาโครงสร้างที่เหมาะสมกับงานใหม่ ซึ่งโครงสร้างใหม่ที่จัดทำขึ้นมีความคล้ายคลึงกับโครงสร้างเก่ามีการเปลี่ยนแปลงเพียงบางส่วนเพื่อที่จะให้การทำงานสามารถทำงานต่อไปได้อย่างไม่ติดขัดมากนัก โครงสร้างใหม่ มีดังนี้

โครงสร้างปกหน้ารายงานวิจัย

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
```

```
            <title>ไทย</title>
            <title type="alternative">อังกฤษ </title>
</titleInfo>
<name>
        <namePart>=ชื่อคนที่1 </namePart>
</name>
<name>
        <namePart>ชื่อคนที่2 </namePart>
</name>

<location>
    <url> </url>
</location>
<originInfo>
        <project_code></project_code>
        <publisher>สถาบัน</publisher>
        <dateCreated encoding="w3cdtf">เริ่ม</dateCreated>
        <dateEnd encoding="w3cdtf">สิ้นสุด</dateEnd>
        <dateModified encoding="w3cdtf">แก้ไขครั้งล่าสุด</dateModified>
</originInfo>
<subject>
        <topic> คำหลัก1</topic>
</subject>
<subject>
        <topic>คำหลัก2</topic>
</subject>
</mods>
```

จากการ Tag ข้อมูลใหม่ ทำให้เราได้เห็นถึงสิ่งที่เราสามารถที่จะนำมาใช้ในการสืบค้นข้อมูลได้ง่ายขึ้นและสามารถแบ่งแยกการค้นหาต่าง ๆ ได้ง่ายขึ้นด้วย

นอกจากนี้จากโครงสร้างที่ได้กำหนดไว้ยังได้มีการนำโครงสร้างดังกล่าวไปทำการวิเคราะห์ในส่วนต่าง ๆ โดยใช้พื้นฐานของโปรแกรม Weka ช่วยในการวิเคราะห์

WEKA3 (Data Mining Software in Java)

จากการศึกษา Data mining พบว่า ได้มีงานที่นำเสนอเกี่ยวกับการทำเหมืองข้อมูลที่เกี่ยวข้องกับเอกสารอยู่หลายรูปแบบด้วยกัน ซึ่งอาจจะมีวัตถุประสงค์ที่แตกต่างกันไป โดยมีการประยุกต์เอาหลักการของ Data mining มาสร้างเป็น Application Programในรูปแบบต่างๆ จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่าได้มีการสร้าง Application Program ขึ้นมาเพื่อทำการวิเคราะห์ข้อมูลในรูปแบบต่างๆตามหลักการของ Data mining ชื่อว่า WEKA3 (Data Mining Software in Java)โดย Program นี้ได้พัฒนาขึ้นโดย The university of Waikato ซึ่งได้พัฒนาขึ้นมาเพื่อศึกษาเทคนิคและวิธีการของ Data mining ซึ่งภายหลังเมื่อได้ทำการสร้างโปรแกรมแล้ว ได้มีการนำเอาข้อมูลตัวอย่างหลายๆ ประเภท เข้าไปวิเคราะห์และประมวลผล เพื่อการดูข้อมูลโดยสรุป หรือ เพื่อศึกษาในเรื่องของการก่อให้เกิดความรู้ (Knowledge)

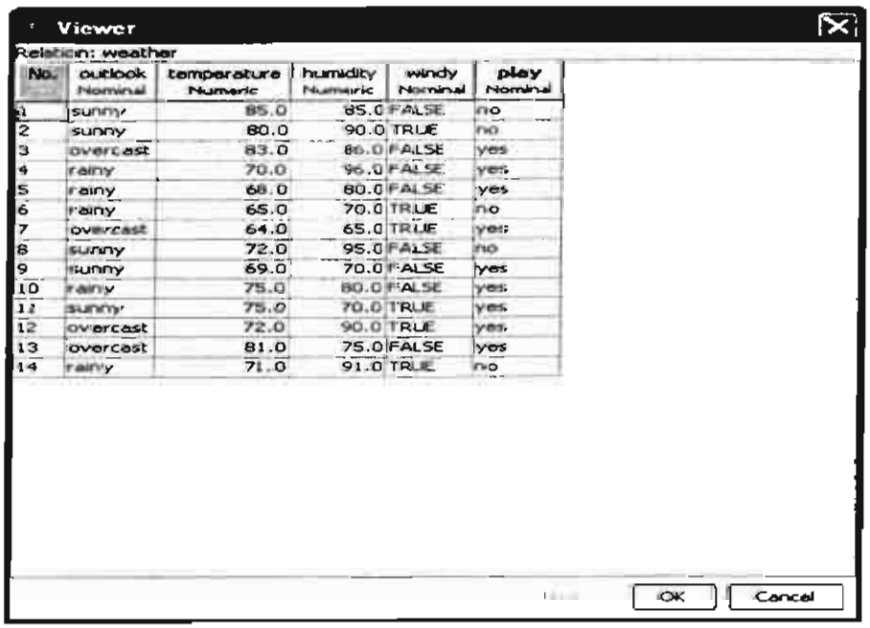
ดังนั้นจึงได้ทำการศึกษาโปรแกรม Weka ถึงการทำงานของโปรแกรม และข้อมูลที่เหมาะสมในการประมวลผลและทำการวิเคราะห์ เพื่อเป็นแนวทางในการประยุกต์ใช้กับข้อมูลงานวิจัยที่มีอยู่

การทำงานของ Weka

จากการศึกษา พบว่า Weka เป็นโปรแกรมที่รวบรวมเอาหลักการทาง Data mining เพื่อใช้ในการวิเคราะห์ข้อมูล โดยโปรแกรม Weka Explorer พัฒนาโดยใช้ภาษา JAVA โดย Weka จะประกอบไปด้วยเครื่องมือสำหรับ การจัดเตรียมข้อมูล(data pre-processing), การจำแนกข้อมูล (classification) ,การจัดกลุ่ม (clustering), การค้นหากฎความสัมพันธ์ (association rules),และนำเสนอข้อมูลที่มีความสัมพันธ์กัน (visualization) การใช้งาน Weka มีขั้นตอนดังรูปที่ 3.4

รูปที่ 3.4 แสดงขั้นตอนการใช้งาน Weka

Weka สามารถที่จะทำการสรุป และ ทำการวิเคราะห์ พร้อมนำเสนอข้อมูลออกมาได้หลายรูปแบบ ไม่ว่าจะเป็นค่าทางสถิติ หรือการแสดงข้อมูลในรูปแบบของกราฟ เริ่มต้นตั้งแต่การเตรียมข้อมูล(data pre-processing) จากรูปเป็นการเตรียมข้อมูล ซึ่งเป็นข้อมูลเกี่ยวกับ สภาพอากาศ ดังรูปที่ 3.5



รูปที่ 3.5 แสดงการเตรียมข้อมูลของ Weka

ซึ่งสามารถนำเข้าได้ในรูปแบบของ ไฟล์ .arff ดังรูป

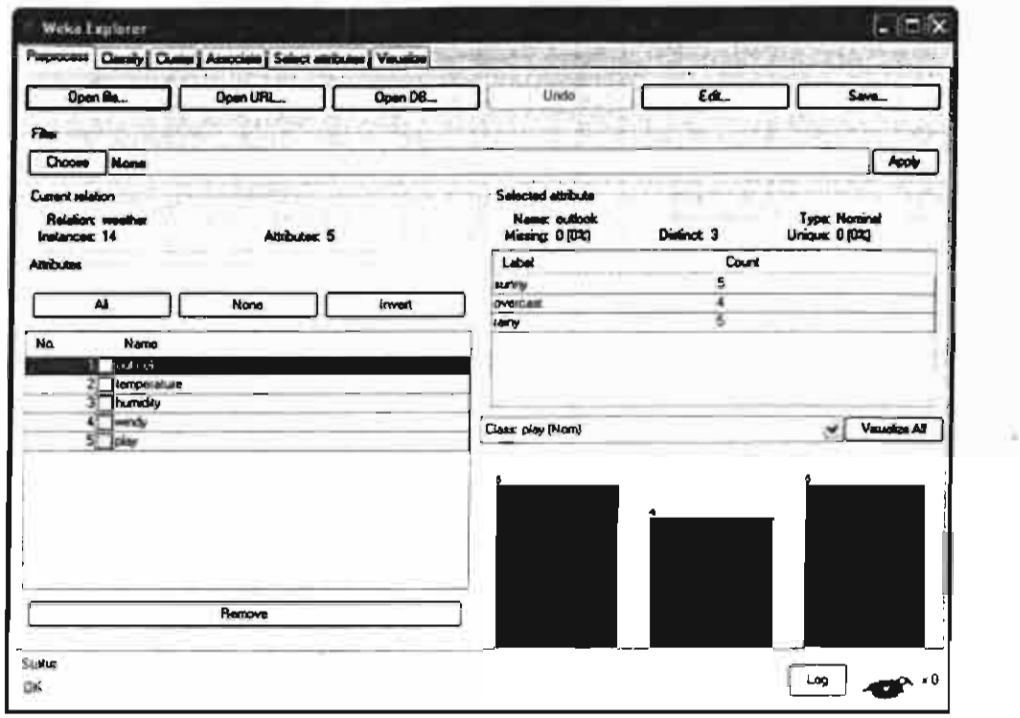
```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

นอกจากนี้ยังสามารถประมวลผลได้ในรูปแบบของ ไฟล์ csv ได้เช่นกัน ดังรูปที่ 3.6

	A	B	C	D	E	F
1	outlook	temperatur	humidity	windy	play	
2	sunny	85	85	FALSE	no	
3	sunny	80	90	TRUE	no	
4	overcast	83	86	FALSE	yes	
5	rainy	70	96	FALSE	yes	
6	rainy	68	80	FALSE	yes	
7	rainy	65	70	TRUE	no	
8	overcast	64	65	TRUE	yes	
9	sunny	72	95	FALSE	no	
10	sunny	69	70	FALSE	yes	
11	rainy	75	80	FALSE	yes	
12	sunny	75	70	TRUE	yes	
13	overcast	72	90	TRUE	yes	
14	overcast	81	75	FALSE	yes	
15	rainy	71	91	TRUE	no	
16						

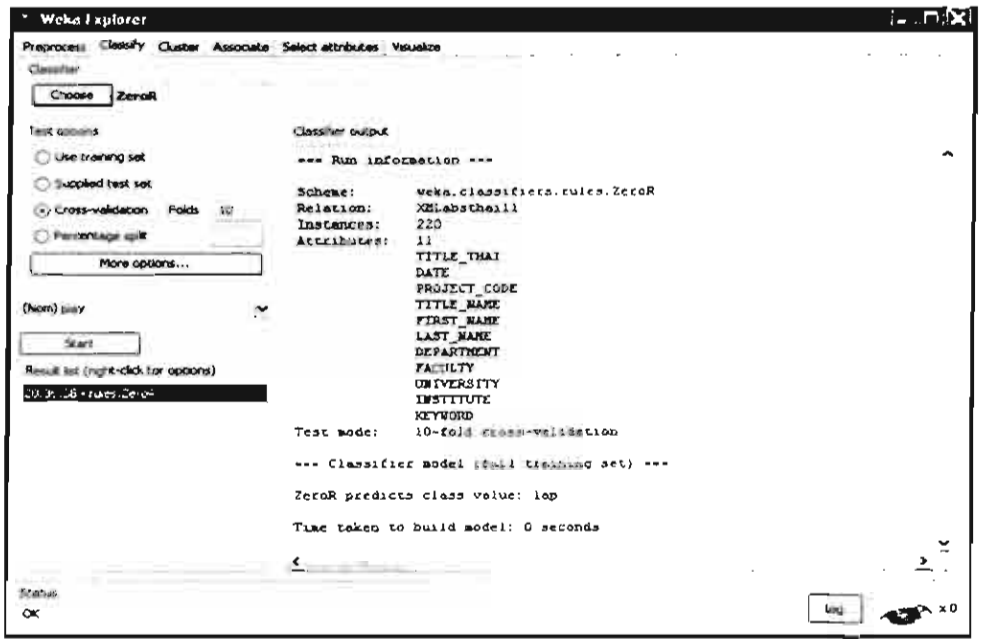
รูปที่ 3.6 แสดงการแปลงไฟล์ข้อมูล csv

เมื่อทำการจัดเตรียมข้อมูลเสร็จแล้วก็สามารถนำไปประมวลผลเพื่อดู ข้อมูลที่สนใจได้ต่อไปดังรูปที่ 3.7



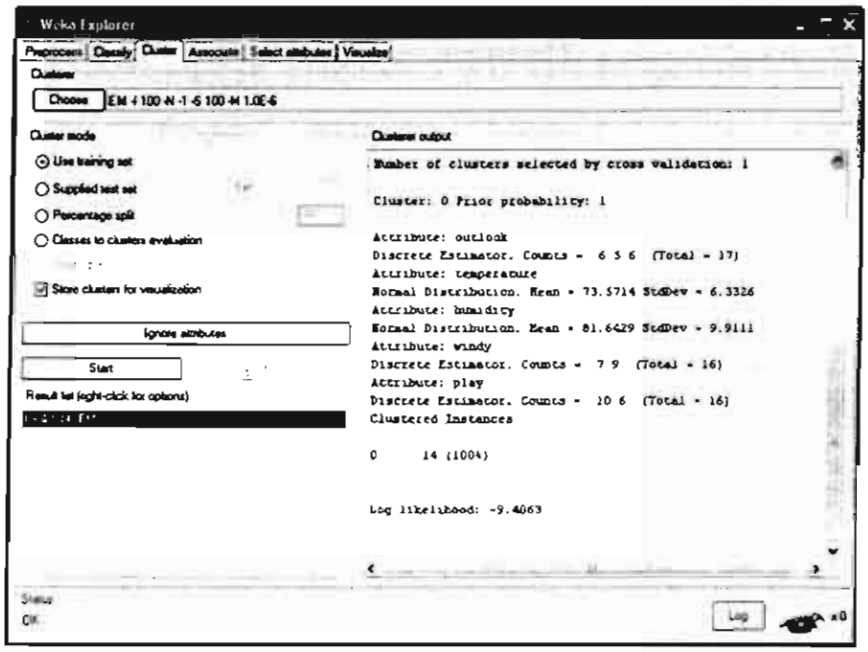
รูปที่ 3.7 แสดงผลการประมวลผลข้อมูลของ weka

การทำงานของโปรแกรม weka ยังสามารถที่จะช่วยในการการจำแนกข้อมูล(classification)ที่มีอยู่โดยมีการคิดคำนวณค่าออกมาในรูปแบบของ ค่าทางสถิติ ดังรูปที่ 3.8



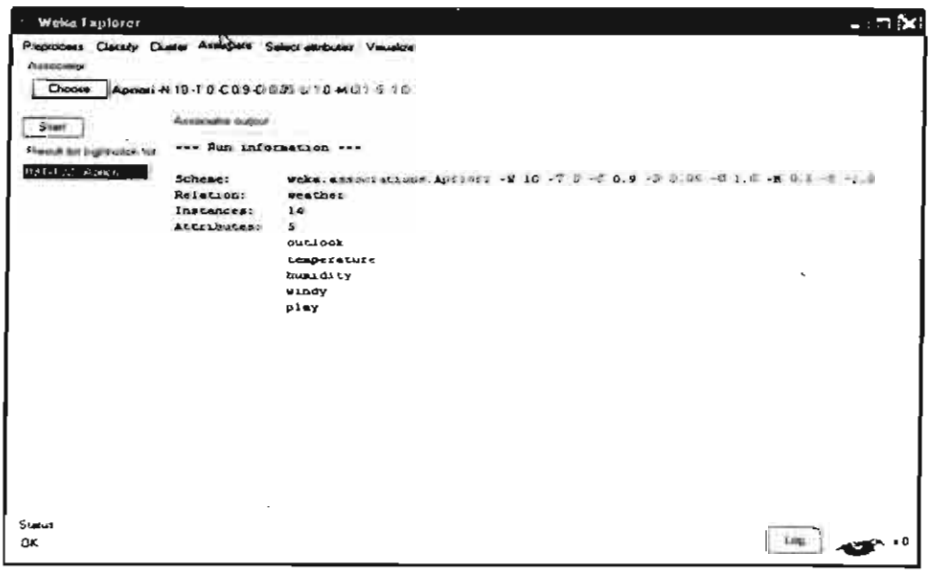
รูปที่ 3.8 แสดงการจำแนกข้อมูลของ weka

และยังสามารถช่วยในการจัดกลุ่มของข้อมูล (clustering) ด้วยวิธีการต่างๆทาง Data mining ดังรูปที่ 3.9



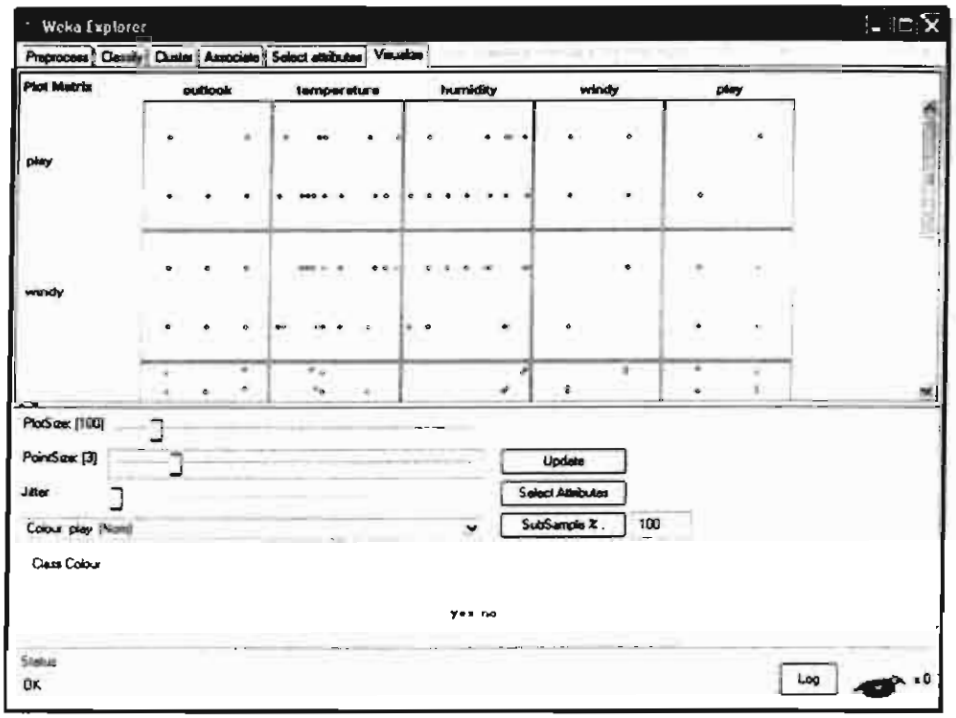
รูปที่ 3.9 แสดงการทำ Data mining ข้อมูลใน weka

ถ้าหากข้อมูลมีความสัมพันธ์หรือเกี่ยวข้องกันในลักษณะของรูปแบบ (pattern) ก็สามารถนำมาสร้างเป็นกฎได้จากข้อมูลที่มีอยู่ ดังรูปที่ 3.10



รูปที่ 3.10 แสดงการสร้างกฎใน weka

นอกจากนี้ในส่วนของการทำงานของ Visualize นั้นสามารถดูข้อมูลที่ลักษณะ Cross กันได้ โดยจะมี การนำเสนอข้อมูลที่มีความสัมพันธ์กันในลักษณะ ของกราฟดังรูปที่ 3.11



รูปที่ 3.11 แสดงการนำเสนอข้อมูลในรูปแบบกราฟ

การประยุกต์นำข้อมูลจากผลงานวิจัยมาวิเคราะห์โดยใช้ Weka

จากการศึกษา weka พบว่า สามารถนำเอาข้อมูลจากงานวิจัยเข้ามาประมวลผลเพื่อดูข้อมูลโดยสรุปได้ โดยจากโครงสร้างของเอกสารข้อมูลงานวิจัยนั้น มีอยู่ หลายโครงสร้างด้วยกัน แต่เนื่องจากว่า ใน weka นั้น จะเน้นการให้ความสำคัญกับข้อมูล หรือว่า attribute ที่สนใจ (หลักการของ Data mining: พิจารณาเฉพาะข้อมูลที่เกี่ยวข้อง) ดังนั้นจึงได้ พิจารณาเฉพาะส่วนของไฟล์ที่เป็น บทคัดย่อไทย (abstract Thai) เท่านั้น เนื่องจากว่า เนื้อหาของข้อมูลนั้น มี attribute ที่สนใจครบถ้วน โดยมีขั้นตอนในการประยุกต์ดังนี้

- ทำการเตรียมข้อมูลโดยเลือกเฉพาะ attribute ที่สนใจ
- ทำการแปลงไฟล์ xml ที่มีอยู่ให้อยู่ในรูปแบบที่ weka ยอมรับ
- ทำการนำข้อมูลเข้าประมวลผล
- พิจารณาผลลัพธ์ที่ได้จากการประมวลผล

แนวคิด และ การเตรียมข้อมูล

ในการเตรียมข้อมูล ได้ทำการศึกษาอาศัยการสืบค้นความรู้จากฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database:KDD) ในส่วนแรก คือ การเตรียมข้อมูลให้เหมาะสม (Pre-Processing)

จากโครงสร้างของเอกสารในส่วนของบทคัดย่อที่มีการนำ XML เข้ามาช่วยในการกำหนด ดังโครงสร้างต่อไปนี้

```
<?xml version="1.0" encoding="UTF-8"?>
<ABSTRACT_THAI>
<PROJECT_CODE> ... </PROJECT_CODE>
<TITLE_THAI>... </TITLE_THAI>
<AUTHOR>
    <NAME> <TITLE_NAME>... </TITLE_NAME>
            <FIRST_NAME>...<FIRST_NAME>
            <LAST_NAME>...</LAST_NAME>
        </NAME>
</AUTHOR>
<DEPARTMENT>... </DEPARTMENT>
<FACULTY>... </FACULTY>
<UNIVERSITY>...</UNIVERSITY>
<ADDRESS>... </ADDRESS>
<E-MAIL>... </E-MAIL>
<DATE>... </DATE>
<PARAGRAPH>...</PARAGRAPH>
<KEYWORD>...</KEYWORD>
</ABSTRACT_THAI>
```

จากโครงสร้างที่มีอยู่ของบทคัดย่อ ภาษาไทย ที่ได้ทำการ tag ด้วย XML แต่ละ Tag มีความหมาย ดังตาราง 3.3

ชื่อ Tag	ความหมาย
<ABSTRACT_THAI> </ABSTRACT_THAI>	หัวข้อระบุว่าเป็นบทคัดย่อภาษาไทย
<PROJECT_CODE> ... </PROJECT_CODE>	ข้อมูลเลขที่สัญญางานวิจัย
<TITLE_THAI>... </TITLE_THAI>	ชื่องานวิจัยที่เป็นภาษาไทย
<AUTHOR> ...</AUTHOR>	ทีมงานวิจัย
<NAME> ... </NAME>	ชื่อผู้ทำงานวิจัย
<TITLE_NAME>... </TITLE_NAME>	คำนำหน้า
<FIRST_NAME>...<FIRST_NAME>	ชื่อ
<LAST_NAME>...</LAST_NAME>	นามสกุล
<DEPARTMENT>... </DEPARTMENT>	ภาควิชา
<FACULTY>... </FACULTY>	คณะ
<UNIVERSITY>...</UNIVERSITY>	มหาวิทยาลัย
<ADDRESS>... </ADDRESS>	ที่อยู่
<E-MAIL>... </E-MAIL>	อีเมลล์
<DATE>... </DATE>	ระยะเวลาดำเนินการ
<PARAGRAPH>...</PARAGRAPH>	เนื้อหาข้อมูลบทคัดย่อ
<KEYWORD>...</KEYWORD>	คำหลัก

ตาราง 3.3 แสดงการสร้าง Tag ของ XML

เมื่อพิจารณาแล้วทำการเลือก ข้อมูลที่สนใจ ได้แก่

<TITLE_THAI>... </TITLE_THAI>

<TITLE_NAME>... </TITLE_NAME>

<FIRST_NAME>...<FIRST_NAME>

<LAST_NAME>...</LAST_NAME>

<DEPARTMENT>... </DEPARTMENT>

<FACULTY>... </FACULTY>

<UNIVERSITY>...</UNIVERSITY>

<INSTITUTE>... </INSTITUTE>

<KEYWORD>...</KEYWORD>

เมื่อทำการเลือกข้อมูลที่สนใจแล้วก็จะเข้าสู่กระบวนการเตรียมข้อมูลในขั้นตอนต่อไปได้แก่ การแปลง ไฟล์ XML ให้อยู่ในรูปแบบประมวลผลสำหรับโปรแกรม Weka

การแปลงข้อมูล XML เพื่อประมวลผลใน Weka

สำหรับไฟล์ที่สามารถนำเข้าสู่โปรแกรม Weka เพื่อทำการวิเคราะห์ มีอยู่หลายรูปแบบด้วยกัน ได้แก่ไฟล์นามสกุล arff , csv ซึ่งจากข้อมูลที่มีอยู่ในรูป XML ต้องทำการแปลงให้อยู่ในรูปแบบที่ Weka กำหนดไว้ ซึ่งในการดำเนินการได้เลือกใช้การแปลงให้อยู่ในรูปทั้ง arff และ csv เพื่อทำการวิเคราะห์ข้อมูลเป็นลำดับต่อไป ยกตัวอย่างรูปแบบของ arff ดังรูปที่ 3.12

@relation XMLabsthai

@attribute TITLE_THAI {การสังเคราะห์อนุพันธ์ของ_4-คลอโร-2,5-ไดฟีนิลออกซาโซลและพอลิเมอร์_,การพัฒนาวิธีการตรวจวัด_แอล-กลูตาเมต_แบบใหม่_โดยใช้หลักการปฏิกิริ
@attribute DATE {2_ปี,30_กันยายน__2545_ถึง_31_สิงหาคม_2547,1_กรกฎาคม_2546_ถึง_30_มิถุนายน_2547,31_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547,1_กรกฎาคม_
@attribute PROJECT_CODE {BGJ_/_35_/_2543,BGJ_4580016,BGJ4680021,MRG4580010,MRG4580011,MRG4580015,MRG4580042,MRG468010
@attribute TITLE_NAME {นางสาว,รศ.ดร.,ผศ.ดร.,ศ.ดร.,นาย,ดร.,รศ.,Professor,นาง,Dr.,ดร.คุณหญิง,พญ.,ผศ.,ศ.นพ.,ท่านผู้หญิง,Prof.Dr.,ศ.}
@attribute FIRST_NAME {ณัฏฐพร,สุภวรรณ,วนิดา,สุเทพ,วิทยา,สมพล_,อรรถชัย,สุดสาย,ปรียา,นิยม,พรสวรรค์_,สุภากรณ์,[illegible],CarolA_,[illegible],Theo
@attribute LAST_NAME {จินาะ,ตันตยานนท์,คำพา,ไวยครุฑธา,มิ่วฒิสม,ประคองพันธ์_,หอมหวน,ตรีวานิช,วิบูลย์ศรษฐ์,เทพศิริ,[illegible],Rut
@attribute DEPARTMENT {ภาควิชาจุลชีววิทยา,ภาควิชาเภสัชกรรม,ภาควิชาชีวเคมี,[illegible],ภาควิชาเคมี,ภาควิชาวิศวกรรมเครื่องกล,Mechanical_Engineerin
@attribute FACULTY {คณะวิทยาศาสตร์,คณะเภสัชศาสตร์,คณะแพทยศาสตร์,[illegible]
@attribute UNIVERSITY {จุฬาลงกรณ์มหาวิทยาลัย,มหาวิทยาลัยมหิดล,มหาวิทยาลัยเกษตรศาสตร์,มหาวิทยาลัยขอนแก่น,มหาวิทยาลัยเชียงใหม่,Vanderbilt_University,Utrecht_
@attribute INSTITUTE {สถาบันธรรมรัฐเพื่อการพัฒนาสังคมและสิ่งแวดล้อม_,สถาบันเทคโนโลยีแห่งเอเชีย_,กรมชลประทาน_}
@attribute KEYWORD {ซินทิลเลเตอร์,4-คลอโร-2,5-ไดฟีนิลออกซาโซล,พอลิเมอร์,ฟลูออเรสเซนต์,สัมประสิทธิ์ในการเปล่งแสงฟลูออเรสเซนต์,กลูตาเมต,การวิเคราะห์ด้วยระบบโฟลอินเจคชั่

@data

การสังเคราะห์อนุพันธ์ของ_4-คลอโร-25-ไดฟีนิลออกซาโซลและพอลิเมอร์_,2_ปี,BGJ_/_35_/_2543,?,?,?,?,?,?,?,?,?
?,?,?,นางสาว,ณัฏฐพร,จินาะ,?,?,?,?,?
?,?,?,รศ.ดร.,สุภวรรณ,ตันตยานนท์,?,?,?,?,?
?,?,?,?,?,?,?,?,จุฬาลงกรณ์มหาวิทยาลัย,?,?
?,?,?,?,?,?,?,?,?,?,?,ซินทิลเลเตอร์
?,?,?,?,?,?,?,?,?,?,?,4-คลอโร-2
?,?,?,?,?,?,?,?,?,?,?,5-ไดฟีนิลออกซาโซล
?,?,?,?,?,?,?,?,?,?,?,พอลิเมอร์
?,?,?,?,?,?,?,?,?,?,?,ฟลูออเรสเซนต์
?,?,?,?,?,?,?,?,?,?,?,สัมประสิทธิ์ในการเปล่งแสงฟลูออเรสเซนต์
การพัฒนาวิธีการตรวจวัด_แอล-กลูตาเมต_แบบใหม่_โดยใช้หลักการปฏิกิริยาหมุนเวียนของเอนไซม์_ดี-ฟีนิลกลัยซีน_อะมิโนทรานสเฟอเรส_และ_แอล-กลูตาเมต_ดีไฮโดรจีเนส_ด้วยระบบโฟลอิ
?,?,?,นางสาว,วนิดา,คำพา,?,?,?,?,?
?,?,?,ผศ.ดร.,สุเทพ,ไวยครุฑธา,?,?,?,?,?
?,?,?,ศ.ดร.,วิทยา,มิ่วฒิสม,?,?,?,?,?
?,?,?,?,?,?,ภาควิชาจุลชีววิทยา,คณะวิทยาศาสตร์,มหาวิทยาลัยมหิดล,?,?
?,?,?,?,?,?,?,?,?,?,?,กลูตาเมต
?,?,?,?,?,?,?,?,?,?,?,การวิเคราะห์ด้วยระบบโฟลอินเจคชั่น
?,?,?,?,?,?,?,?,?,?,?,ดี-ฟีนิลกลัยซีน
?,?,?,?,?,?,?,?,?,?,?,อะมิโนทรานสเฟอเรส
?,?,?,?,?,?,?,?,?,?,?,ปฏิกิริยาหมุนเวียนของสับสเตรท
คุณสมบัติการเป็นแอนติเจนของโปรตีนเมมเบรนผิวนอกของเชื้อ_พาสเจอเรลล่า_มัลโตซิดา_โดยใช้เปลือกหุ้มไวรัสนิวคาสเซิลเป็นระบบนำส่งวัคซีน_,1_กรกฎาคม_2546_ถึง_30_มิถุนายน_25
?,?,?,ศ.ดร.,สมพล_,ประคองพันธ์_,?,?,?,?,?
?,?,?,นาย,อรรถชัย,หอมหวน,?,?,?,?,?
?,?,?,?,?,?,ภาควิชาเภสัชกรรม,คณะเภสัชศาสตร์,มหาวิทยาลัยมหิดล,?,?
?,?,?,?,?,?,?,?,?,?,?,อิสคอม
?,?,?,?,?,?,?,?,?,?,?,ไวโรโซม
?,?,?,?,?,?,?,?,?,?,?,โปรตีนผิวนอก
?,?,?,?,?,?,?,?,?,?,?,พาสเจอเรลล่า
?,?,?,?,?,?,?,?,?,?,?,มัลโตซิดา
?,?,?,?,?,?,?,?,?,?,?,วัคซีนหน่วยย่อย
การขาดเจ็บเนื่องจากการสปอร์ออกซิเดทีฟ__ของ__Salmonella_sp._ซึ่งดื้อ________________________ต่อสารต้านจุลชีพหลายชนิดที่แยกได้จากเนื้อไก่และผลิตภัณฑ์เนื้อไก่_,31_
?,?,?,ดร.,สุดสาย,ตรีวานิช,?,?,?,?,?
?,?,?,รศ.ดร.,ปรียา,วิบูลย์ศรษฐ์,?,?,?,?,?
?,?,?,?,?,?,?,?,มหาวิทยาลัยเกษตรศาสตร์,?,?
?,?,?,?,?,?,?,?,?,?,?,ซัลโมเนลลา
?,?,?,?,?,?,?,?,?,?,?,การดื้อต่อสารต้านจุลชีพ
?,?,?,?,?,?,?,?,?,?,?,เซลล์บาดเจ็บ
?,?,?,?,?,?,?,?,?,?,?,ภาวะเปอร์ออกซิเดทีฟ
การค้นหากลไกการดื้อยาเคมีบำบัดของเซลล์เพาะเลี้ยงมะเร็งท่อน้ำดีในหลอดทดลอง_,1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547,MRG4580010,?,?,?,?,?,?,?,?,?
?,?,?,ผศ.ดร.,นิยม,เทพศิริ,?,?,?,?,?
?,?,?,?,?,?,ภาควิชาชีวเคมี,คณะแพทยศาสตร์,มหาวิทยาลัยขอนแก่น,?,?
?,?,?,?,?,?,?,?,?,?,?,เซลล์เพาะเลี้ยงมะเร็งท่อน้ำดี
?,?,?,?,?,?,?,?,?,?,?,ยาเคมีบำบัด
?,?,?,?,?,?,?,?,?,?,?,ความไวต่อยา
?,?,?,?,?,?,?,?,?,?,?,การดื้อยา
?,?,?,?,?,?,?,?,?,?,?,การแสดงออกของยีน

รูปที่ 3.12 แสดงการแปลงข้อมูลจาก XML เป็น arff

การแปลงไฟล์ในรูปแบบของ csv

การแปลงไฟล์ XML ให้อยู่ในรูปแบบของ CSV นี้สามารถทำได้โดยการเขียนโปรแกรมเพื่อดึงเอาข้อมูลที่สนใจมาอยู่ในรูปแบบของตาราง ดังรูปที่ 3.13

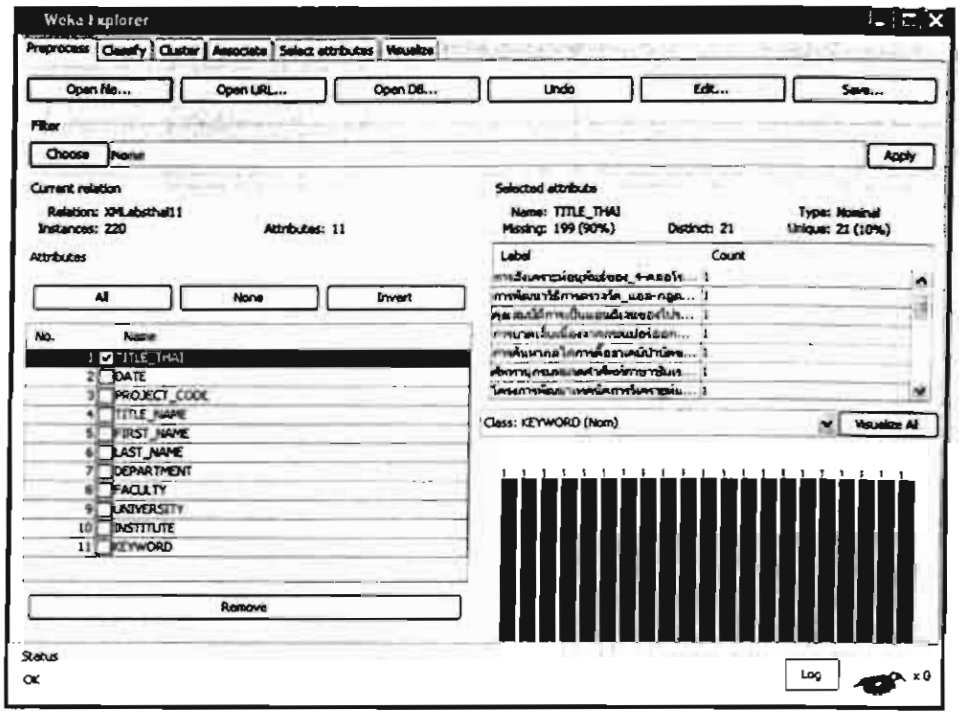
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TITLE_THAI	DATE	PROJECT_CODE	TITLE_	FIRST_	LAST_	DEPARTMENT	FACULTY	UNIVERSITY	INSTITUTE	KEYWORD			
2														
3	การสังเคราะห์	2 0	BGJ / 35 / 2543	?	?	?	?	?	?	?	?			
4	?	?	?	นางสาว	ณัฐพร	[illegible]	?	?	?	?	?			
5	?	?	?	รศ.ดร.	ศุภวรรณ	[illegible]	?	?	?	?	?			
6	?	?	?	?	?	?	?	?	จุฬาลงกรณ์มหา	?	?			
7	?	?	?	?	?	?	?	?	?	?	[illegible]			
8	?	?	?	?	?	?	?	?	?	?	4-คลอโร-2			
9	?	?	?	?	?	?	?	?	?	?	[illegible]			
10	?	?	?	?	?	?	?	?	?	?	[illegible]			
11	?	?	?	?	?	?	?	?	?	?	[illegible]			
12	?	?	?	?	?	?	?	?	?	?	[illegible]			
13														
14	การพัฒนาวิธีก	30 กันยา	BGJ 4580016	?	?	?	?	?	?	?	?			
15	?	?	?	นางสาว	[illegible]	[illegible]	?	?	?	?	?			
16	?	?	?	ผศ.ดร.	[illegible]	[illegible]	?	?	?	?	?			
17	?	?	?	ศ.ดร.	[illegible]	[illegible]	?	?	?	?	?			
18	?	?	?	?	?	?	[illegible]	[illegible]	[illegible]	?	?			
19	?	?	?	?	?	?	?	?	?	?	[illegible]			
20	?	?	?	?	?	?	?	?	?	?	[illegible]			
21	?	?	?	?	?	?	?	?	?	?	[illegible]			
22	?	?	?	?	?	?	?	?	?	?	[illegible]			
23	?	?	?	?	?	?	?	?	?	?	[illegible]			
24														
25	คุณสมบัติการ	1 กรกฎา	BGJ4680021	?	?	?	?	?	?	?	?			
26	?	?	?	ศ.ดร.	สมพล	ประคอง	?	?	?	?	?			
27	?	?	?	นาย	[illegible]	[illegible]	?	?	?	?	?			
28	?	?	?	?	?	?	[illegible]	[illegible]	[illegible]	?	?			
29	?	?	?	?	?	?	?	?	?	?	[illegible]			
30	?	?	?	?	?	?	?	?	?	?	[illegible]			
31	?	?	?	?	?	?	?	?	?	?	[illegible]			
32	?	?	?	?	?	?	?	?	?	?	[illegible]			
33	?	?	?	?	?	?	?	?	?	?	[illegible]			
34	?	?	?	?	?	?	?	?	?	?	[illegible]			

รูปที่ 3.13 แสดงการแปลงไฟล์ XML ให้อยู่ในรูปแบบของ CSV

ผลลัพธ์ที่ได้จาก Weka

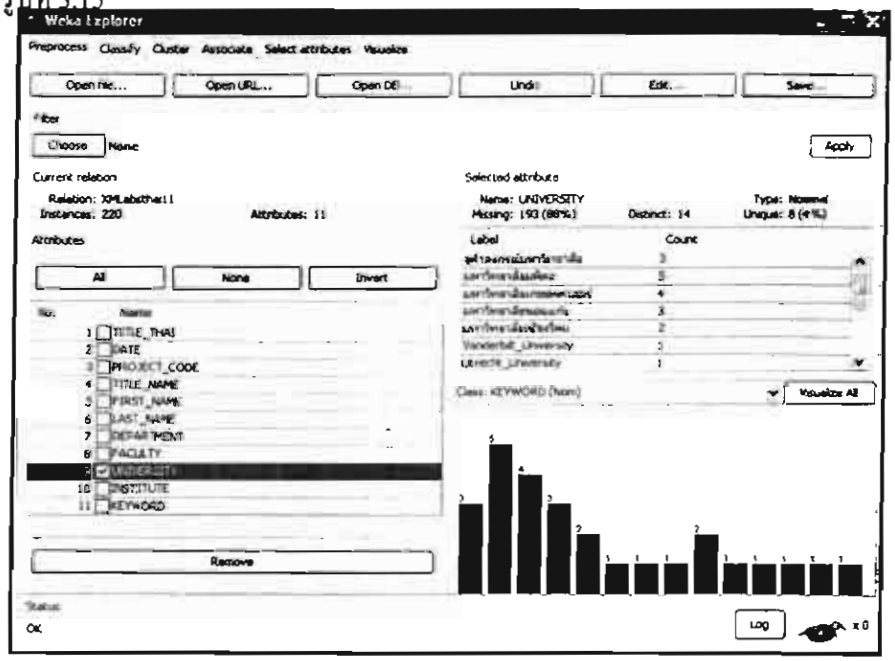
เนื่องจาก ข้อมูลที่มีอยู่ยังมีข้อจำกัดอยู่มากในเรื่องของความครบถ้วนสมบูรณ์ จึงทำให้ไม่สามารถที่จะนำไปวิเคราะห์ทางด้าน Data mining ได้ในระดับสูง ดังนั้นจึงพิจารณาเฉพาะในสองส่วนคือ Preprocess และ การดูข้อมูลแบบ Visualize

เมื่อทำการเลือก preprocess โดยนำข้อมูลเข้าไปทำการประมวลผล สามารถที่จะเลือกดู attribute ที่สนใจ โดยโปรแกรมจะทำการนับ และประมวลผลดูข้อมูลได้ในรูปแบบของกราฟ ดังรูปที่ 3.14



รูปที่ 3.14 แสดงผลการประมวลผลในรูปแบบของกราฟ

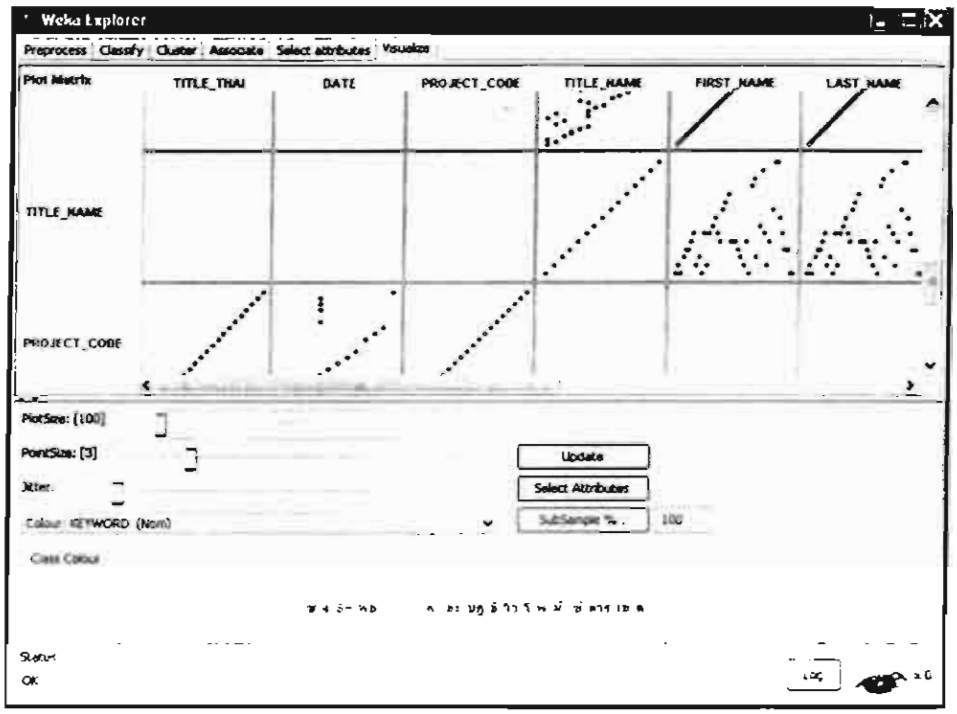
จากรูปเป็นการเลือกดูข้อมูลในส่วนของหัวข้องานวิจัยทั้งหมด พร้อมทั้งแสดงให้อยู่ในรูปของกราฟแท่ง ดังรูปที่ 3.15



รูปที่ 3.15 แสดงการประมวลผลในรูปแบบกราฟแท่ง

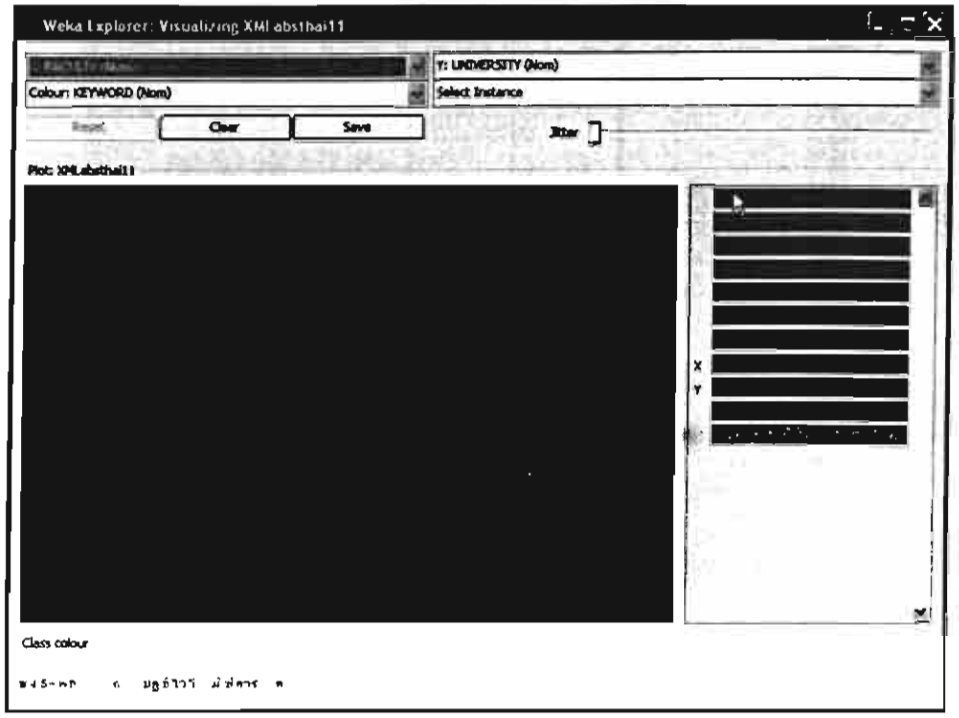
จากรูปเป็นการเลือกดูข้อมูลในส่วนของมหาวิทยาลัยที่มีผลงานวิจัยทั้งหมด พร้อมทั้งแสดงให้อยู่ในรูปของกราฟแท่ง

ส่วนในลักษณะของ Visualize นั้น สามารถเลือกดู attribute ที่มีความสัมพันธ์กันได้ โดยสามารถเลือกพิจารณาได้จากแกน x และ แกน y ซึ่งผลลัพธ์จะออกมาในรูปแบบของกลุ่มข้อมูล ที่มีความสัมพันธ์กัน ดังรูปที่ 3.16



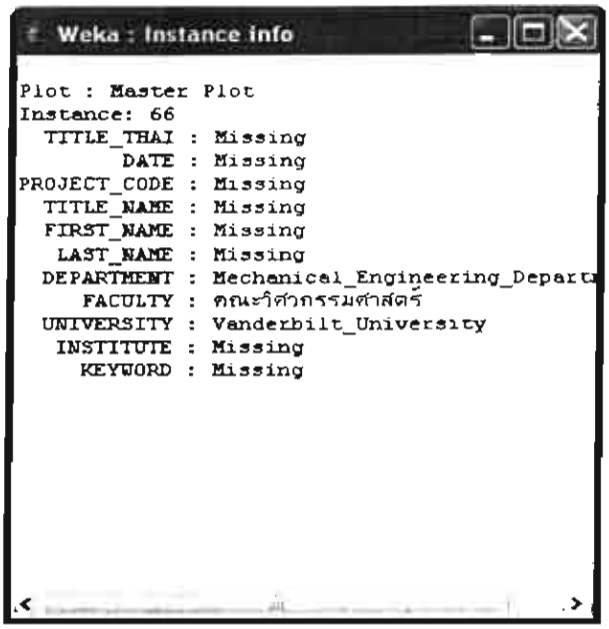
รูปที่ 3.16 แสดงผลลัพธ์ออกมาในรูปแบบของกลุ่มข้อมูล

ซึ่งสามารถเลือกดูข้อมูล ที่มีความสัมพันธ์กันในแบบ Visualize จะแสดงผลดังรูปที่ 3.17



รูปที่ 3.17 แสดงความสัมพันธ์ข้อมูลในแบบ Visualize

โดยสามารถเลือกดูรายละเอียดของข้อมูลจากกราฟได้ดังรูปที่ 3.18



รูปที่ 3.18 แสดงรายละเอียดของข้อมูลที่ได้จากกราฟ

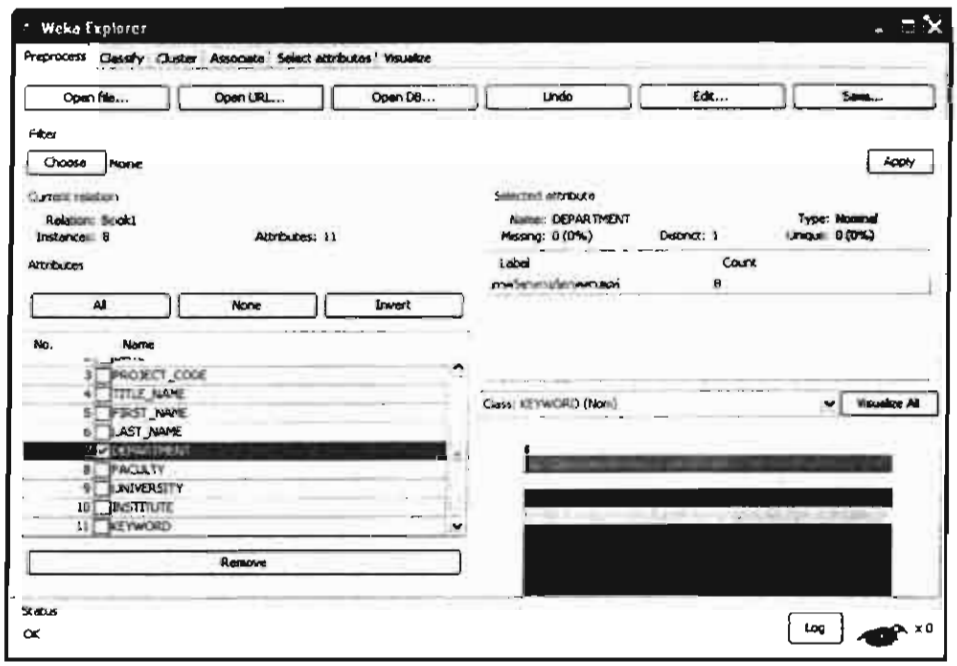
จากผลลัพธ์ที่ได้จากการนำข้อมูลไปประมวลผลด้วย Weka นั้น สามารถสรุปได้คือ ข้อมูลหรือ attribute สามารถที่จะถูกนับ และ แสดงให้อยู่ในรูปแบบกราฟได้ ทำให้สามารถทราบถึงข้อมูลในเชิงปริมาณได้ว่า มีแนวโน้มเป็นอย่างไร ในแง่ของการทำวิจัย แต่เนื่องจาก ปัญหาทางด้านข้อมูลดิบที่มีทำให้ข้อมูลไม่ถูกต้องเท่าที่ควร สาเหตุ เพราะในไฟล์ XML บทคัดย่อไทยนั้น สำหรับข้อมูลเดิมในเอกสารไม่มีส่วนประกอบของข้อมูลหรือว่า attribute ที่ครบถ้วน ทำให้การประมวลผลไม่ถูกต้อง หรือ แม้แต่การไม่ตรงกันของเนื้อหาข้อมูลใน tag ก็ก่อให้เกิดปัญหาได้เช่นเดียวกัน ยกตัวอย่างเช่น ในบางเอกสาร ใช้คำว่า นส. , นางสาว , รองศาสตราจารย์, รศ. เป็นต้น ทำให้ข้อมูลไม่ตรงกัน ซึ่งทำให้ขั้นตอนในการเตรียมข้อมูลนั้นค่อนใช้เวลาในการ Cleaning ข้อมูลที่มีอยู่ และด้วยสาเหตุนี้เองจึงทำให้เมื่อทำการแปลงข้อมูล จาก XML มาอยู่ในรูปแบบของ CSV นั้นยุ่งยากมากขึ้นอีกด้วย จากโครงสร้าง ในไฟล์ CSV จะเห็นว่า ยอมให้มีการ Missing Data (?)ได้ แต่การ Missing Data นี้ก็จะส่งผลในการแสดงความสัมพันธ์ใน Visualize ทำให้ข้อมูลไม่มีความสัมพันธ์กัน ซึ่งถ้าไม่อยากให้มีการ Missing Data เกิดขึ้น ก็จำเป็นจะต้องใส่ค่าข้อมูลที่ซ้ำกันลงไปในแต่ละ เรคคอร์ดของข้อมูล ซึ่ง ในความสัมพันธ์ใน Visualize สามารถที่จะดูความสัมพันธ์ได้ แต่ในส่วนของการนับจำนวนหรือว่าปริมาณใน Preprocess จะเกิดความผิดพลาด ให้ เนื่องจากว่า Weka จะทำการนับข้อมูลที่แยกกันแต่ละ record เป็นคนละอย่างกัน ไม่ถือเป็นตัวเดียวกันถ้าหากทำการใส่ข้อมูลให้ครบแล้วดังไฟล์ arff และ CSVตัวอย่างดังรูปที่ 3.19

```
@relation XMLabsthai

@attribute TITLE_THAI {การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู}
@attribute DATE {'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547'}
@attribute PROJECT_CODE {TRG4580022}
@attribute TITLE_NAME {นางสาว}
@attribute FIRST_NAME {สุดนฮา_}
@attribute LAST_NAME {เจริญวิทย์_}
@attribute DEPARTMENT {ภาควิชากายวิภาคศาสตร์}
@attribute FACULTY {คณะทันตแพทยศาสตร์}
@attribute UNIVERSITY {จุฬาลงกรณ์มหาวิทยาลัย}
@attribute INSTITUTE numeric
@attribute KEYWORD {HNK-1,rat,small,eye,Pax-6,neural,crest,cell}

@data

การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,HNK-1
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,rat
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,small
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,eye
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,Pax-6
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,neural
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,crest
การศึกษาบทบาทของHNK-1ต่อการอพยพของนิวรัลเครสเซลล์ในตัวอ่อนของหนู,'\t1_กรกฎาคม_2545_ถึง_30_มิถุนายน_2547',TRG4580022,
นางสาว,สุดนฮา_เจริญวิทย์_,ภาควิชากายวิภาคศาสตร์,คณะทันตแพทยศาสตร์,จุฬาลงกรณ์มหาวิทยาลัย,?,cell
```

	A	B	C	D	E	F	G	H	I	J	K
1											
2	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				HNK-1
3	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				rat
4	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				small
5	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				eye
6	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				Pax-6
7	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				neural
8	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				crest
9	การศึกษา 1 กรกฎาค		TRG4580022	นางสา	สุดเธา	เจริญใหม่	ภาควิชากายวิภาคศา คณะทันตแพทย จุฬาลงกรณ์มหาวิท ?				cell
10											
11											

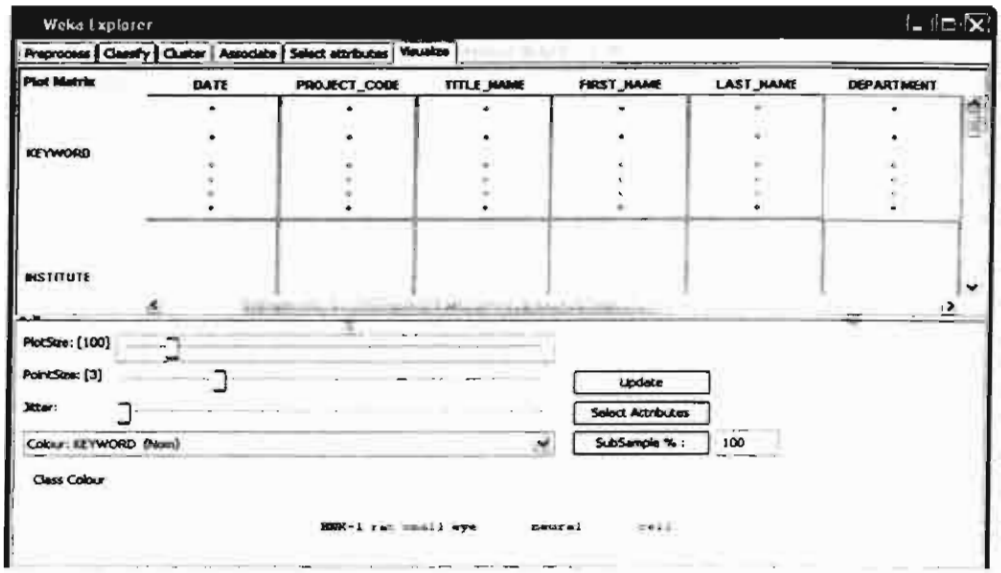
รูปที่ 3.19 แสดงไฟล์ arff และ CSV ของข้อมูล

ผลลัพธ์ที่ได้ให้จำนวนนับที่ได้จาก Weka ไม่ถูกต้องเพราะจากผลลัพธ์ที่ได้จาก**ไฟล์** CSV ควรจะมีเพียง 1 งานวิจัย 1 เลขที่สัญญา 1 ผู้วิจัย จากภาควิชากายวิภาคศาสตร์ คณะ**ทันตแพทย์** จุฬาลงกรณ์มหาวิทยาลัย แต่ผลลัพธ์กลับ นับจำนวนงานวิจัยเป็น 8 เรื่อง 8 เลขที่สัญญา 8 ผู้วิจัย จาก 8 ภาควิชากายวิภาคศาสตร์ 8 คณะทันตแพทย์ และ 8 จุฬาลงกรณ์มหาวิทยาลัย ดังรูปที่ 3.20



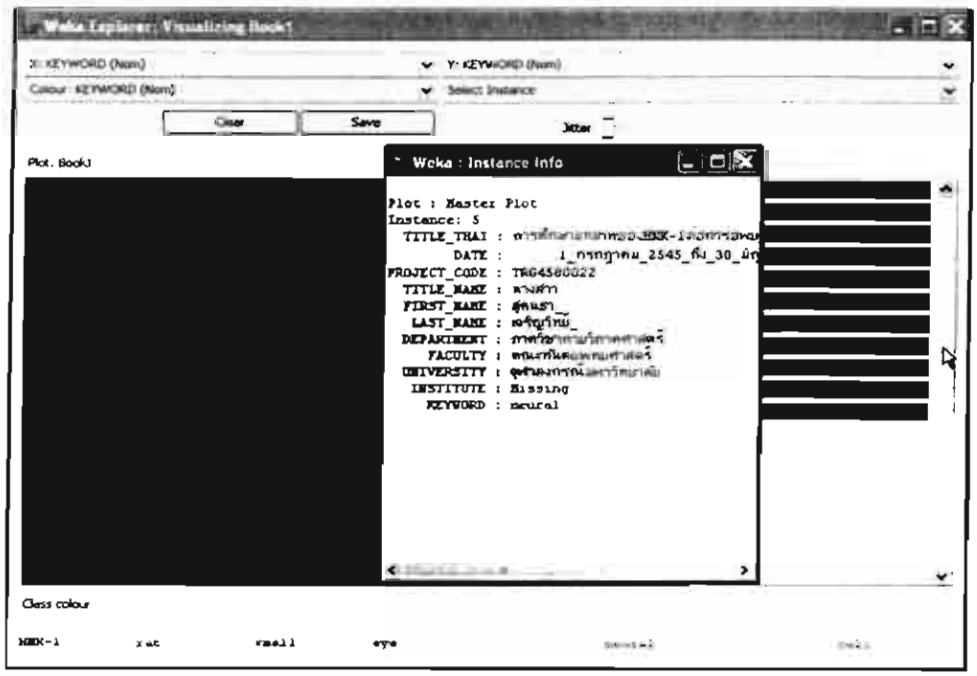
รูปที่ 3.20 แสดงผลลัพธ์ที่ได้ให้จำนวนนับจาก Weka

แต่ส่วนของ Visualize นั้นผลลัพธ์ที่ได้สามารถดูข้อมูลได้ เปอร์เซ็นต์ของ Missing data น้อยลงดังรูปที่ 3.21



รูปที่ 3.21 แสดงผลลัพธ์ที่ได้สามารถดูข้อมูลได้ เปอร์เซ็นต์ของ Missing data

จะเห็นได้ว่าข้อมูลมีความสัมพันธ์กันมากขึ้น Data Missing น้อยลง ดังรูปที่ 3.22



รูปที่ 3.22 แสดงความสัมพันธ์กันของ Data Missing ที่น้อยลง

ผลจากการศึกษา Weka

จากการศึกษาการนำไฟล์จากงานวิจัยไปประมวลผลในโปรแกรม Weka ได้ดังนี้

1. สามารถนำเสนอข้อมูลในเชิงสรุป ได้ และสามารถดูข้อมูลที่มีความสัมพันธ์กันได้
2. การแปลงข้อมูลค่อนข้างจะมีความซับซ้อน จะต้องมีการจัดรูปแบบของไฟล์ข้อมูลที่ต้องการวิเคราะห์ ให้อยู่ในรูปแบบ ของ ไฟล์นามสกุลรูปแบบ arff, csv ก่อนจึงสามารถที่จะประมวลผลได้
3. ถ้าเปรียบเทียบตัวอย่างของข้อมูลอื่นที่นำมาประมวลผลใน weak แล้ว พบว่า ข้อมูลที่ครบถ้วนสมบูรณ์ หรือมี Missing data น้อยๆ สามารถที่จะนำไปประมวลผลได้ผลลัพธ์ที่ถูกต้องมากกว่า สามารถนำไปประมวลผลโดยใช้เทคนิคอื่นๆสำหรับ Data mining ได้

การพัฒนาโปรแกรมสำหรับการดูข้อมูลจากหมืองข้อมูลงานวิจัย

จากการศึกษาโปรแกรม weka ถึงลักษณะการทำงาน จึงได้นำแนวคิดของ การทำ Data Mining มาประยุกต์ โดยอาศัยเทคนิคของการทำ Text Mining มาใช้เพื่อพัฒนาโปรแกรมสำหรับดูข้อมูลสรุปที่สนใจ โดยใช้ เครื่องมือในการพัฒนาได้แก่ Visual Basic โดยจะมีการนำข้อมูลที่อยู่ใน tag XMLจากเอกสารงานวิจัยที่มีอยู่ มาเก็บเป็นความถี่ และ นำไปคำนวณผลทางสถิติเพื่อให้นำข้อมูลไปใช้ในการวิเคราะห์และเป็นแนวทางในการช่วยตัดสินใจต่อไป

ขั้นตอนการออกแบบโปรแกรม

ในการออกแบบโปรแกรมได้ทำการศึกษาถึงแนวคิดในการทำ Text Mining พบว่าเป็นการศึกษา Data Mining มุ่งเน้นเกี่ยวกับข้อมูลที่มีโครงสร้างชัดเจน (Structured Data) แต่อย่างไรก็ตามข้อมูลสารสนเทศที่เก็บอยู่บางส่วนจะอยู่ใน Text Database หรือ Document Database ได้แก่ เอกสารบทความ ข่าว เอกสารวิชาการ เป็นต้น ข้อมูลที่เก็บอยู่ใน Document Database จะเป็นข้อมูลที่มีลักษณะโครงสร้างไม่ชัดเจน (Semi-Structured Data) ลักษณะโครงสร้างข้อมูลแบบ Semi-Structured Data คือโครงสร้างข้อมูลที่ประกอบด้วย Structured data รวมอยู่กับ Unsturctured Data เช่น Title, Author, Publication_date เป็น Structured data ส่วน Abstract และ Body หรือ contents จัดเป็น Unstructured Dataซึ่ง ในวิธีการต่างๆของ Text Mining เช่น Statistical Methodology

Text Mining คือขบวนการทำงานที่เรียกว่า process ที่สกัดข้อมูล (Extract data) จากฐานข้อมูลขนาดใหญ่ (Large Textual Information) เพื่อให้ได้สารสนเทศ (Usefull Textual Information) โดยข้อมูลที่ถูกนำมา Mining เป็นข้อมูลที่มีลักษณะเป็น Text data sets

Text Mining สามารถเรียกสั้นๆว่า TM โดยมี operation ในการทำ Text Mining หลายแบบ เช่น Document Clustering, Document Classification , Summarizing Text เป็นต้น
ด้วยเหตุนี้เองจึงได้นำเอาแนวทางต่างที่คิดว่าเหมาะสมมาประยุกต์ในการทำวิจัยในครั้งนี้ โดย ได้มี

มีการออกแบบการทำงานของโปรแกรมออกเป็น 4 ส่วนด้วยกัน คือ

- ส่วนของการนำเข้าข้อมูล (Open File) เป็นส่วนของการเลือกเอาข้อมูลที่สนใจ (Cleaning data) และ ทำการ แบ่งแยก(Classificaiton)แต่ละ Token ของข้อมูล พร้อมทั้ง มีการสรุปข้อมูลออกมา(Extract)ในรูปแบบของจำนวนนับของความถี่
- ส่วนของการนำเสนอข้อมูลในรูปของกราฟ (View Graph) เป็นส่วนของการสรุปผลข้อมูลแต่ละประเภทที่ได้ทำการ แบ่งแยกเอาไว้แล้วมาทำการนำเสนอในลักษณะของข้อมูลโดยสรุป (Summarizing Text)
- ส่วนของการนำเสนอข้อมูลที่เกี่ยวข้องกัน (Visualization) เป็นส่วนของการนำเสนอถึงมุมมองของข้อมูลที่มีความสัมพันธ์กัน ในรูปแบบของกราฟที่มีแกน x และ แกน Y เพื่อดูลักษณะของข้อมูลที่สอดคล้องกันในแต่ละประเภท
- ส่วนของการดูรายละเอียดของข้อมูลที่สัมพันธ์กัน (Detail) ส่วนนี้เป็นการเลือกดูข้อมูล เฉพาะ ที่สนใจโดยจะแสดงส่วนประกอบของข้อมูลทั้งหมดที่มีอยู่Tag XML

1. ส่วนของการนำเข้าข้อมูล (Open File)

จากข้อมูลที่มีอยู่ในรูปแบบของเอกสารที่มี Tag XML ได้มีการออกแบบให้มีการนำเข้าไฟล์ได้โดยตรง โดยอันดับแรก ทำการอ่านไฟล์ เข้ามาเก็บไว้ จากนั้นทำการแยกข้อมูล ออกจากกันเป็น Token แล้วทำการจัดกลุ่มของTag XML แต่ละชนิด จากนั้นทำการค้นหา Tag XML ที่ต้องการนำมาประมวลผล ดังตาราง 3.4

Tag เปิด	Tag ปิด
<TITLE_THAI>	</TITLE_THAI>
<TITLE_NAME>	</TITLE_NAME>
<FIRST_NAME>	<FIRST_NAME>
<LAST_NAME>	</LAST_NAME>
<DEPARTMENT>	</DEPARTMENT>
<FACULTY>	</FACULTY>
<UNIVERSITY>	</UNIVERSITY>
<INSTITUTE>	</INSTITUTE>
<DATE>	</DATE>
<KEYWORD>	</ KEYWORD>

ตาราง 3.4 แสดงการกำหนด Tag ข้อมูล

โดยจะใช้คำที่แทนความหมายของแต่ละ Tag XML ดังตาราง 3.5

Tag XML	คำที่ใช้แทน
<TITLE_THAI>	ชื่อโครงการ
<TITLE_NAME>	คำนำหน้าชื่อ
<FIRST_NAME>	ชื่อ
<LAST_NAME>	นามสกุล
<DEPARTMENT>	ภาควิชา
<FACULTY>	คณะ
<UNIVERSITY>	มหาวิทยาลัย
<INSTITUTE>	สถาบัน
<DATE>	ระยะเวลาโครงการ
<KEYWORD>	Keyword

ตาราง 3.5 แสดงการใช้คำแทนความหมายของแต่ละ Tag

และจะทำการพิจารณาข้อมูลจากไฟล์ โดยจะดึงเอาข้อมูลที่อยู่ระหว่าง Tag เปิด และ Tag ปิด แล้วมาทำการนับความถี่ของข้อมูลที่อยู่ระหว่าง Tag แต่ละประเภท แล้วทำการแสดงผลของจำนวนข้อมูลดังกล่าว จากเอกสารของการทำเหมืองงานวิจัยที่มีอยู่นั้นมีโครงสร้างที่ไม่เหมือนกันบางเอกสารก็อาจจะมี Tag XML ที่ไม่ตรงกัน ดังตัวอย่างของไฟล์เอกสารต่อไปนี้

ไฟล์ที่ 1

```
<?xml version="1.0" encoding="UTF-8"?>
<ABSTRACT_THAI>
<PROJECT_CODE>BGJ / 11 / 2543 </PROJECT_CODE>
<TITLE_THAI>การศึกษาบทบาทของ cell wall hydrolases องค์ประกอบของผนังเซลล์ และการ
แสดงออกของยีนที่ควบคุมเอนไซม์ที่เกี่ยวข้องกับการหลุดร่วงของผลกล้วยระหว่างการสุก
</TITLE_THAI>
<AUTHOR>
    <NAME> <TITLE_NAME>ศ.ดร. </TITLE_NAME><FIRST_NAME>สายชล
</FIRST_NAME><LAST_NAME>เกตุษา </LAST_NAME></NAME>
    <NAME><TITLE_NAME>นางสาว</TITLE_NAME><FIRST_NAME>วชิรญา
</FIRST_NAME><LAST_NAME>อิ่มสบาย </LAST_NAME> </NAME>
```

<NAME> <TITLE_NAME>นางสาว</TITLE_NAME> <FIRST_NAME>อภิวรา
</FIRST_NAME><LAST_NAME>ประยูรวงศ์ </LAST_NAME> </NAME>
</AUTHOR>
<DEPARTMENT>ภาควิชาพืชสวน </DEPARTMENT><FACULTY> คณะเกษตร
</FACULTY>
<UNIVERSITY>มหาวิทยาลัยเกษตรศาสตร์ </UNIVERSITY>
<ADDRESS>จตุจักร กรุงเทพฯ 10900 </ADDRESS>
<E-MAIL>agrsck@ku.ac.th (ศ.ดร. สายชล เกตุษา) </E-MAIL>
<DATE> ระยะเวลาโครงการ : วันที่ 30 กันยายน 2543 ถึง วันที่ 29 กันยายน 2544 </DATE>
<PARAGRAPH>กล้วยผลมีมากกว่าในเนื้อเยื่อของเปลือกบริเวณกลางผลของผลกล้วยที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูง ขณะที่การเปลี่ยนแปลงกิจกรรม ?-galactosidase ในเนื้อเยื่อบริเวณทั้งสองส่วน ไม่สัมพันธ์กับการหลุดร่วงผลกล้วยหอมทองที่บ่มให้สุกภายใต้ความชื้นสัมพัทธ์ต่ำและสูง

การศึกษาชั้นของเนื้อเยื่อบริเวณที่เกิดการหลุดร่วงของผลกล้วยหอมทองพบว่า ไม่มีการสร้างชั้นของบริเวณการร่วง (abscission zone) เกิดขึ้นในบริเวณเนื้อเยื่อที่จะเกิดการหลุดร่วงของผล การศึกษาการแสดงออกของยีนที่ควบคุมเอนไซม์ polygalacturonase พบว่า การแสดงออกของยีนในบริเวณขั้วผลที่เกิดรอยหลุดร่วงในผลกล้วยหอมทองมีมากกว่าในผลกล้วยน้ำว้าและการแสดงออกของยีนในบริเวณขั้วผลมีมากกว่าในบริเวณเปลือกผล
</PARAGRAPH>
<KEYWORD>คำสำคัญ : การหลุดร่วงของผล, การสุก, กล้วยไข่, กล้วยหอมทอง, กล้วยน้ำว้า, บริเวณการร่วง, การอ่อนตัว, polygalacturonase, pectin methylesterase, ?-galactosidase
</KEYWORD>
</ABSTRACT_THAI>

ไฟล์ที่ 2

<?xml version="1.0" encoding="UTF-8"?>
<ABSTRACT_THAI>
<TITLE_THAI>ชุดโครงการประวัติศาสตร์ท้องถิ่นภาคอีสาน : การขยายตัวของชุมชนลุ่มน้ำชี" วิวัฒนาการและการเปลี่ยนแปลงของพื้นที่ภายในท้องถิ่นริมฝั่งลุ่มน้ำชี" </TITLE_THAI>
<NAME> <TITLE_NAME>นาย </TITLE_NAME><FIRST_NAME>สักรินทร์
</FIRST_NAME><LAST_NAME> แซ่ภู่ </LAST_NAME> </NAME>

<NAME><TITLE_NAME>นาย </TITLE_NAME><FIRST_NAME>อภิชัย
</FIRST_NAME><LAST_NAME>กาบทอง </LAST_NAME></NAME>
<NAME><TITLE_NAME>นาย </TITLE_NAME><FIRST_NAME>มงคล
</FIRST_NAME><LAST_NAME> คาร์น </LAST_NAME></NAME>
<NAME><TITLE_NAME>นาย </TITLE_NAME><FIRST_NAME>สมพงค์
</FIRST_NAME><LAST_NAME>โสทะรักษ์ </LAST_NAME></NAME>
<DEPARTMENT>สาขาสถาปัตยกรรมเมืองและชุมชน </DEPARTMENT>
<FACULTY>คณะสถาปัตยกรรมศาสตร์ ผังเมืองและนฤมิตศิลป์ </FACULTY>
<UNIVERSITY>มหาวิทยาลัยมหาสารคาม </UNIVERSITY>
<DATE>ระยะเวลาโครงการ : พ.ศ. 2546 </DATE>
<PARAGRAPH>

กระบวนการทำความเข้าใจเกี่ยวกับ "พื้นที่ (Space)ริมฝั่งลุ่มน้ำชี"ปรับตัวเชิงพื้นที่ภายในท้องถิ่น ได้แก่ การปรับตัวและเปลี่ยนแปลงเกี่ยวกับกลุ่มคนกับพื้นที่ทางทรัพยากร การปรับตัวและเปลี่ยนแปลงเกี่ยวกับความสัมพันธ์ของกลุ่มคนกับกลุ่มคน การปรับตัวและการเปลี่ยนแปลงระหว่างกลุ่มคนกับความเชื่อและประเพณี และ6) มิติที่ซ้อนทับเชิงพื้นที่กับแนวทางการจัดการพื้นที่ โดยแบ่งออกเป็นสองมิติคือ มิติเชิงซ้อนบนพื้นที่ริมฝั่งลุ่มน้ำชีกับมิติทางการจัดการที่ซ้อนทับมิติทางการจัดการที่ยังคงดำรงอยู่ในพื้นที่ริมฝั่งลุ่มน้ำชี เชื่อมโยงไปสู่คนท้องถิ่น แต่ในท้ายที่สุดสาระที่ได้จะไม่เป็นผลอันใดหากไม่เกิด กระบวนการการ นำข้อมูลที่ได้กลับมาใช้ประโยชน์โดยเฉพาะอย่างยิ่งเพื่อเอื้อ ประโยชน์ให้กับคน ท้องถิ่นอีกครั้ง

</PARAGRAPH>
</ABSTRACT_THAI>

จากโครงสร้างไฟล์เอกสารงานวิจัยทั้งสอง มีโครงสร้างที่แตกต่างกัน ในไฟล์ที่ 1 มี tag keyword แต่ในไฟล์ที่ 2 ไม่มี tag keyword ซึ่งในความเป็นจริงแล้ว ยังมีความแตกต่างทางด้านโครงสร้างอีกมากมายหลายรูปแบบ ทำให้ผลลัพธ์ข้อมูลที่ได้ บางส่วนขาดหายไป ดังนั้นจึงออกแบบให้โปรแกรมจะแสดงเฉพาะข้อมูลใน Tag ที่มีเท่านั้น

2. ส่วนของการนำเสนอข้อมูลในรูปของกราฟ (View Graph)

จากข้อมูลที่ได้ทำการ Extract แล้ว ในส่วนที่ 1 สามารถนำข้อมูลที่ได้มีการนับความถี่แต่และ Tag XML นั้นมาทำการนำเสนอให้อยู่ในรูปของกราฟวงกลมและกราฟแท่ง โดยในการนำเสนอจะทำการการแทนข้อมูลด้วยสีต่างๆ เพื่อบ่งบอกชนิดของข้อมูลที่มีความหลากหลายเพื่อความง่ายต่อการดูข้อมูล พร้อมทั้งทำการแสดงสัดส่วนของ ข้อมูลจากเอกสารงานวิจัยทั้งหมด

3. ส่วนของการนำเสนอข้อมูลที่เกี่ยวข้องกัน (Visualization)

เป็นส่วนที่สามารถทำการเลือกดูข้อมูลที่มีความสัมพันธ์กันได้ โดยสามารถกำหนดข้อมูลที่สนใจ โดยที่สามารถเลือกที่จะกำหนดให้ข้อมูลแสดงบนแกน x หรือ แกน y ได้ ในส่วนนี้ จะเป็นส่วนที่ช่วยให้สามารถดูข้อมูลโดยสรุปได้ โดยที่จะเป็นลักษณะของการมองข้อมูลในลักษณะที่เป็นมิติระหว่างข้อมูลได้ สำหรับส่วนที่1 และ 2 ที่ได้กล่าวไปแล้วนั้นจะเป็นการนำเสนอของข้อมูลเฉพาะในแต่ละ Tag XML ที่สนใจ และนำเสนอในรูปแบบของกราฟ วงกลม และกราฟแท่ง ดังนั้นในส่วนที่ 3 นี้จะช่วยให้เข้าใจได้มากยิ่งขึ้น

4. ส่วนของการดูรายละเอียดของข้อมูลที่สัมพันธ์กัน (Detail)

หากต้องการที่จะดูรายละเอียดของข้อมูลแต่ละจุดบนกราฟ สามารถทำการเลือกเพื่อดูได้โดยจะแสดง ส่วนของข้อมูลทั้งหมด (ทุกๆ Tag XML)

ในการทดสอบโปรแกรมจะทำการนำเข้าข้อมูลเอกสารงานวิจัยที่มีการTag ไว้เรียบร้อย เพื่อทำการทดสอบประสิทธิภาพการทำงานของโปรแกรม ดังจะกล่าวไว้ในบทต่อไป

บทที่ 4

ผลดำเนินการโครงการ

ผลการดำเนินงาน (ช่วงที่ 1)

1. การวิเคราะห์เอกสารข้อความด้วยการประมวลผลภาษาธรรมชาติทำให้เครื่องคอมพิวเตอร์เข้าใจภาษามนุษย์ในระดับคำ และคุณสมบัติของคำโดยอาศัยหน้าที่ของคำ

2. การจัดทำระบบสืบค้นโดยอาศัยอินเด็กซ์ และการหาค่าน้ำหนัก (term weight) เพื่อการจัดลำดับเอกสาร

3. การพัฒนาระบบสืบค้นเอกสารจากคำสำคัญ (keyword) โดยอาศัย Boolean Operation ที่สามารถทำงานผ่านทางเครือข่ายอินเตอร์เน็ต

จากการวิเคราะห์โครงสร้างเอกสารและกระบวนการต่าง ๆ ของเอกสารงานวิจัย สามารถแสดงโครงสร้างและการกำหนดความสำคัญให้กับเอกสารได้โดยใช้ XML เป็นเครื่องมือที่ช่วยในการ Tagging รูปแบบการ Tagging โดยแบ่งตามโครงสร้างที่ได้ทำการกำหนดไว้มี ดังนี้

1. การ Tagging ในส่วนของปกงานวิจัย

```
<?xml version="1.0" encoding=" UTF -8"?>
<!DOCTYPE COVER_PAGE  SYSTEM "cover.dtd">
<COVER_PAGE>
        <TITLE_THAI>   </TITLE_THAI>
        <TITLE_ENGLISH>  </TITLE_ENGLISH>
        <COMPLETE_DATE>  </COMPLETE_DATE>
        <CONTRACT_NUMBER> </CONTRACT_NUMBER>
        <RESEARCHER>
                <AUTHOR_CONSULT>
                        <NAME>
                                <TITLE_NAME>   </TITLE_NAME>
                                <FIRST_NAME>  </FIRST_NAME>
                                <LAST_NAME>   </LAST_NAME>
                        </NAME>
                        <DEPARTMENT>  </DEPATRMENT>
                        <FACULTY>   </FACULTY>
```

```
                    <UNIVERSITY>  </UNIVERSITY>
                    <INSTITUTE> </INSTITUTE>
                </ AUTHOR_CONSULT >
                <AUTHOR>
                    <NAME>
                        <TITLE_NAME>  </TITLE_NAME>
                        <FIRST_NAME>  </FIRST_NAME>
                        <LAST_NAME>  </LAST_NAME>
                    </NAME>
                    <DEPARTMENT>  </DEPATRMENT>
                    <FACULTY>  </FACULTY>
                    <UNIVERSITY>  </UNIVERSITY>
                    <INSTITUTE> </INSTITUTE>
                </ AUTHOR >
        </RESEARCHER>
        <ADDRESS>  </ADDRESS>
</COVER_PAGE>
```

จากโครงสร้างข้างต้น สามารถทำเป็น DTD ได้ดังนี้

```
<?xml version="1.0" encoding=" utf-8"?>
<!ELEMENT COVER_PAGE (TITLE_THAI,TITLE_ENGLISH ,COMPLETE_DATE
,CONTRACT_NUMBER,RESEARCHER,AUTHOR_CONSULT,AUTHOR,NAME, TITLE_NAME ,
FIRST_NAME , LAST_NAME ,DEPARTMENT, FACULTY,UNIVERSITY,ADDRESS)>
<!ELEMENT TITLE_THAI (#PCDATA)>
<!ELEMENT TITLE_ENGLISH (#PCDATA)>
<!ELEMENT COMPLETE_DATE (#PCDATA)>
<!ELEMENT CONTRACT_NUMBER (#PCDATA)>
<!ELEMENT RESEARCHER (#PCDATA)>
<!ELEMENT AUTHOR_CONSULT (#PCDATA)>
<!ELEMENT AUTHOR (#PCDATA)>
<!ELEMENT NAME (#PCDATA)>
<!ELEMENT TITLE_NAME (#PCDATA)>
<!ELEMENT FIRST_NAME (#PCDATA)>
```

<!ELEMENT LAST_NAME (#PCDATA)>

<!ELEMENT DEPARTMENT (#PCDATA)>

<!ELEMENT FACULTY (#PCDATA)>

<!ELEMENT UNIVERSITY (#PCDATA)>

<!ELEMENT ADDRESS(#PCDATA)>

และอธิบายรายละเอียดต่าง ๆ ได้ดังนี้

COVER_PAGE	การกำหนดช่วงเปิดและปิดของปกรายงานวิจัย
TITLE_THAI	ชื่อโครงการวิจัยภาษาไทย
TITLE_ENGLISH	ชื่อโครงการวิจัยภาษาอังกฤษ
COMPLETE_DATE	วันที่รายงานสรุป
CONTRACT_NUMBER	สัญญาเลขที่
RESEARCHER	กลุ่มผู้วิจัยและสถานบัน
AUTHOR_CONSULT	นักวิจัยที่ปรึกษา
AUTHOR	นักวิจัย
NAME	ชื่อนักวิจัย
TITLE_NAME	คำนำหน้านาม
FIRST_NAME	ชื่อ
LAST_NAME	นามสกุล
DEPARTMENT	ภาควิชา
FACULTY	คณะวิชา
UNIVERSITY	มหาวิทยาลัย
INTITUTE	สถาบัน
ADDRESS	ที่อยู่

2. การ Tagging ในส่วนของกิตติกรรมประกาศ

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ACKNOWLEDGEMENT SYSTEM "ack.dtd">
< ACKNOWLEDGEMENT >
        <PARAGRAPH>  </ PARAGRAPH >
</ ACKNOWLEDGEMENT >
```

จากโครงสร้างข้างต้น สามารถทำเป็น DTD ได้ดังนี้

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT ACKNOWLEDGEMENT (PARAGRAPH)>
<!ELEMENT PARAGRAPH (#PCDATA)>
```

และอธิบายรายละเอียดต่าง ๆ ได้ดังนี้

ACKNOWLEDGEMENT	กำหนดช่วงเปิด และ ปิด ของกิตติกรรมประกาศ
PARAGRAPH	เนื้อหาในกิตติกรรมประกาศ

3. การ Tagging ในส่วนของบทคัดย่อภาษาไทย

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE  ABSTRACT_THAI  SYSTEM "abstract_thai.dtd">
< ABSTRACT_THAI >
        <PROJECT_CODE> </ PROJECT_CODE >
        <TITLE_THAI>  </TITLE_THAI>
        <AUTHOR>
                <NAME>
                        <TITLE_NAME>  </TITLE_NAME>
                        <FIRST_NAME>  </FIRST_NAME>
                        <LAST_NAME>  </LAST_NAME>
                </NAME>
                <DEPARTMENT>  </DEPATRMENT>
                <FACULTY>  </FACULTY>
                <UNIVERSITY>  </UNIVERSITY>
                <INSTITUTE> </INSTITUTE>
```

```
        </ AUTHOR >
        <E-MAIL>  </E-MAIL>
        <DATE>  </DATE>
        <PARAGRAPH>  </ PARAGRAPH >
        <KEYWORD>  </ KEYWORD >
  </ ABSTRACT_THAI >
```

จากโครงสร้างข้างต้น สามารถทำเป็น DTD ได้ดังนี้

```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT ABSTRACT_THAI (PROJECT_CODE,TITLE_THAI,AUTHOR, NAME ,
TITLE_NAME , FIRST_NAME , LAST_NAME , DEPARTMENT , FACULTY , UNIVERSITY ,E-
MAIL ,DATE,PARAGRAPH,KEYWORD)>
<!ELEMENT PROJECT_CODE (#PCDATA)>
<!ELEMENT TITLE_THAI(#PCDATA)>
<!ELEMENT AUTHOR(#PCDATA)>
<!ELEMENT NAME (#PCDATA)>
<!ELEMENT TITLE_NAME (#PCDATA)>
<!ELEMENT FIRST_NAME (#PCDATA)>
<!ELEMENT LAST_NAME (#PCDATA)>
<!ELEMENT DEPARTMENT (#PCDATA)>
<!ELEMENT FACULTY (#PCDATA)>
<!ELEMENT UNIVERSITY (#PCDATA)>
<!ELEMENT E-MAIL(#PCDATA)>
<!ELEMENT DATE(#PCDATA)>
<!ELEMENT PARAGRAPH(#PCDATA)>
<!ELEMENT KEYWORD(#PCDATA)>
```

และอธิบายรายละเอียดต่าง ๆ ได้ดังนี้

ABSTRACT_THAI	กำหนดช่วงเปิด และ ปิด ของบทคัดย่อภาษาไทย
PROJECT_CODE	รหัสโครงการ
TITLE_THAI	ชื่อโครงการภาษาไทย
AUTHOR	นักวิจัย
NAME	ชื่อนักวิจัย

TITLE_NAME	คำนำหน้านาม
FIRST_NAME	ชื่อ
LAST_NAME	นามสกุล
DEPARTMENT	ภาควิชา
FACULTY	คณะวิชา
UNIVERSITY	มหาวิทยาลัย
INTITUTE	สถาบัน
E-MAIL	อีเมล์ติดต่อ
DATE	ระยะเวลาที่ทำโครงการ
PARAGRAPH	เนื้อหาบทคัดย่อ
KEYWORD	คำหลักที่ใช้ค้นหา

4. การ Tagging ในส่วนของบทคัดย่อภาษาอังกฤษ

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE ABSTRACT_ENGLISH SYSTEM "abstract_english.dtd">
 < ABSTRACT_ENGLISH >
      <PROJECT_CODE> </ PROJECT_CODE >
      <TITLE_ ENGLISH ></TITLE_ ENGLISH >
      <AUTHOR>
           <NAME>
                <TITLE_NAME>  </TITLE_NAME>
                <FIRST_NAME>  </FIRST_NAME>
                <LAST_NAME>  </LAST_NAME>
           </ NAME >
           <DEPARTMENT> </DEPATRMENT>
           <FACULTY></FACULTY>
           <UNIVERSITY> </UNIVERSITY>
           <INSTITUTE></INSTITUTE>
           <ADDRESS></ADDRESS>
      </ AUTHOR >
      <E-MAIL> </E-MAIL>
      <DATE> </DATE>
```

```
            <PARAGRAPH></ PARAGRAPH >
        <KEYWORD> </ KEYWORD >
</ ABSTRACT_ ENGLISH >
```

จากโครงสร้างข้างต้น สามารถทำเป็น DTD ได้ดังนี้

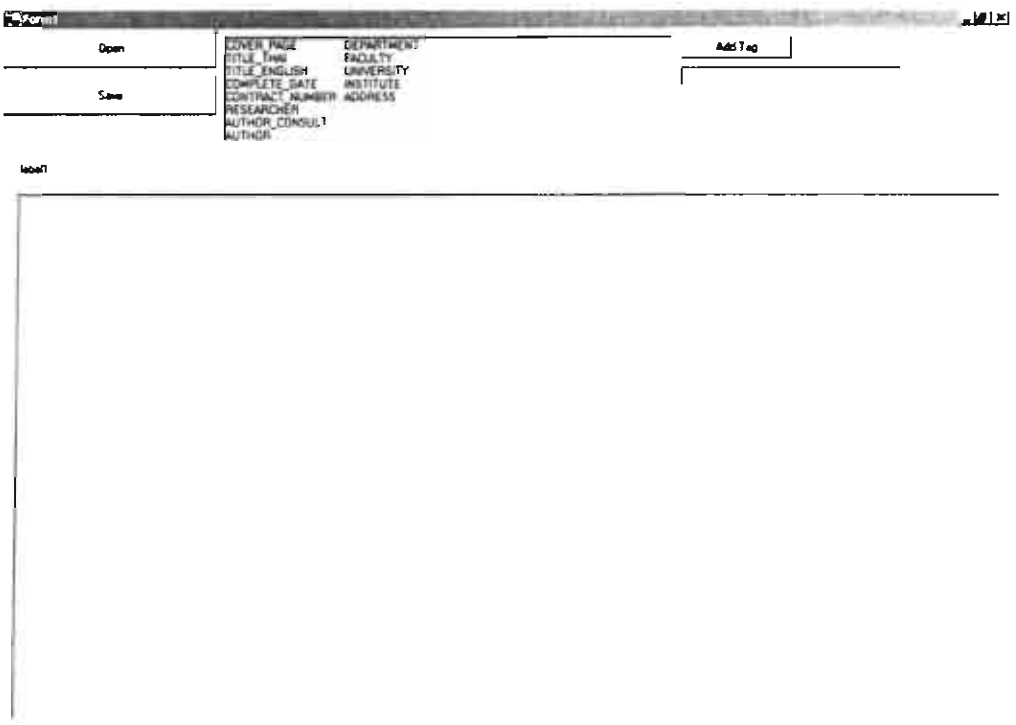
```
<?xml version="1.0" encoding="UTF-8"?>
<!ELEMENT ABSTRACT_ ENGLISH (PROJECT_CODE,TITLE_ ENGLISH, AUTHOR, NAME,
TITLE_NAME , FIRST_NAME , LAST_NAME , DEPARTMENT,FACULTY,UNIVERSITY,
INSTITUTE ,ADDRESS,  E-MAIL , DATE, PARAGRAPH , KEYWORD)>
<!ELEMENT PROJECT_CODE (#PCDATA)>
<!ELEMENT TITLE_ ENGLISH (#PCDATA)>
<!ELEMENT AUTHOR(#PCDATA)>
<!ELEMENT NAME(#PCDATA)>
<!ELEMENT TITLE_NAME (#PCDATA)>
<!ELEMENT FIRST_NAME (#PCDATA)>
<!ELEMENT LAST_NAME (#PCDATA)>
<!ELEMENT DEPARTMENT(#PCDATA)>
<!ELEMENT FACULTY(#PCDATA)>
<!ELEMENT UNIVERSITY(#PCDATA)>
<!ELEMENT INSTITUTE (#PCDATA)>
<!ELEMENT E-MAIL(#PCDATA)>
<!ELEMENT DATE(#PCDATA)>
<!ELEMENT PARAGRAPH(#PCDATA)>
<!ELEMENT KEYWORD(#PCDATA)>
```

และอธิบายรายละเอียดต่าง ๆ ได้ดังนี้

ABSTRACT_ ENGLISH	กำหนดช่วงเปิด และ ปิด ของบทคัดย่อภาษาอังกฤษ
PROJECT_CODE	รหัสโครงการ
TITLE_ ENGLISH	ชื่อโครงการภาษาอังกฤษ
AUTHOR	นักวิจัย
NAME	ชื่อนักวิจัย
TITLE_NAME	คำนำหน้านาม
FIRST_NAME	ชื่อ

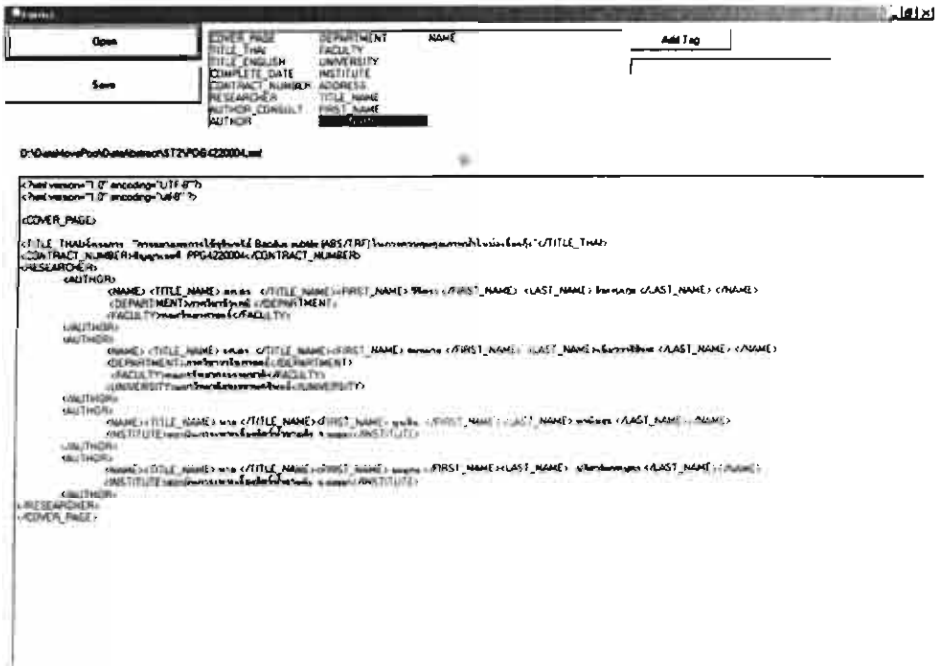
LAST_NAME	นามสกุล
DEPARTMENT	ภาควิชา
FACULTY	คณะวิชา
UNIVERSITY	มหาวิทยาลัย
INTITUTE	สถาบัน
E-MAIL	อีเมล์ติดต่อ
DATE	ระยะเวลาที่ทำโครงการ
PARAGRAPH	เนื้อหาบทคัดย่อ
KEYWORD	คำหลักที่ใช้ค้นหา

จากโครงสร้างทั้ง 4 โครงสร้างการ Tagging ซึ่งเป็นส่วนของเอกสารหลักในการค้นหา ทำให้เราสามารถกำหนดโครงสร้างกลางให้กับเอกสารงานวิจัยทั้งหมดได้ ซึ่งยังมีบางงานที่ไม่มีโครงสร้างที่ครบตามโครงสร้างที่ได้กำหนดไว้ ทำให้เวลาทำงานต้องมีการคัดแยกเอกสารออกเป็นกลุ่มดังที่ได้กล่าวมาแล้ว ลักษณะของหน้าจอโปรแกรมที่ใช้ในการทำการ Tagging ข้อมูลตัวอย่างข้อมูลทั้ง 4 ส่วนที่ได้แสดงตัวอย่างไว้ข้างต้น เป็นดังนี้

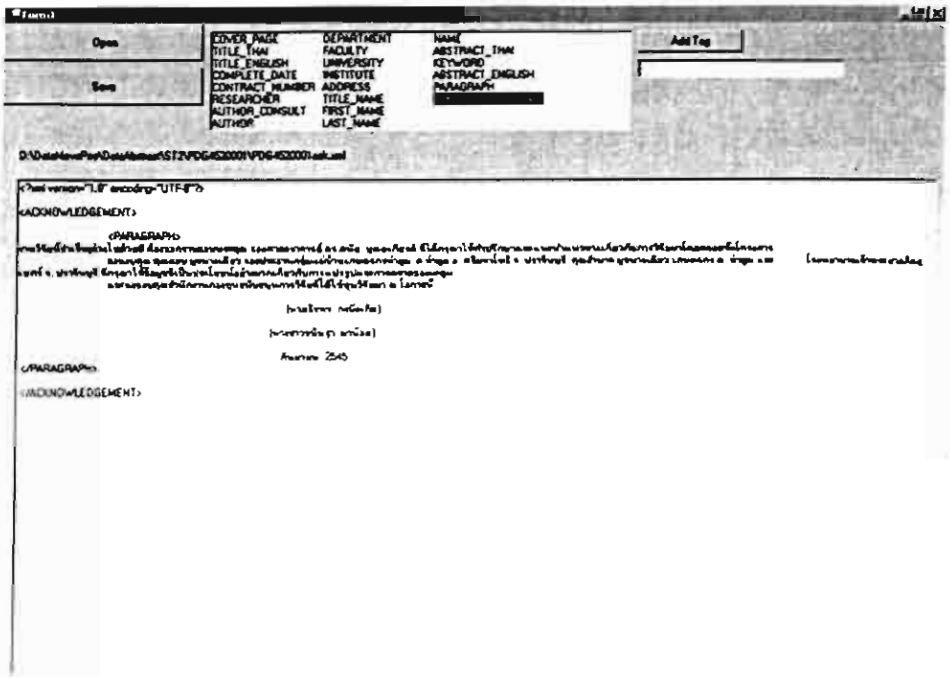


รูปที่ 4.1 แสดงหน้าจอการทำงานในการ Tagging

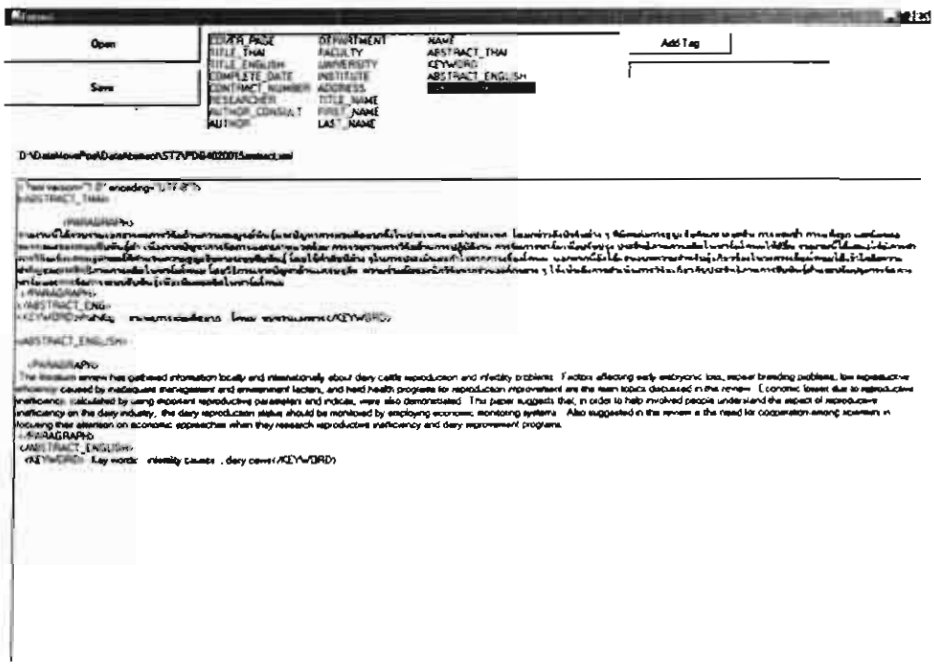
เมื่อเข้าสู่โปรแกรมการทำงานในการ Tagging ข้อมูล ก็จะทำการเลือกปุ่ม Add Tag เพื่อทำการเพิ่ม Tag ที่เราต้องการเข้าสู่โปรแกรมเพื่ออำนวยความสะดวกในการทำงานครั้งต่อไป เมื่อทำการเพิ่ม Add Tag ได้เรียบร้อยก็จะทำการคลิกปุ่ม Open เพื่อทำการเปิดไฟล์ข้อมูลที่ต้องการจะ Tag ข้อมูลแล้วทำการเลือก Tag ที่ต้องการใช้งานตามความโครงสร้างการ Tag ที่ได้ทำการกำหนดไว้แล้ว หลังจากนั้นทำการคลิกปุ่ม Save เพื่อทำการบันทึกข้อมูลเป็นไฟล์ที่มีนามสกุลเป็น xml ได้ดังรูปที่ 4.2



รูปที่ 4.2 แสดงการ Tagging ข้อมูลหน้าปกรายงานวิจัย

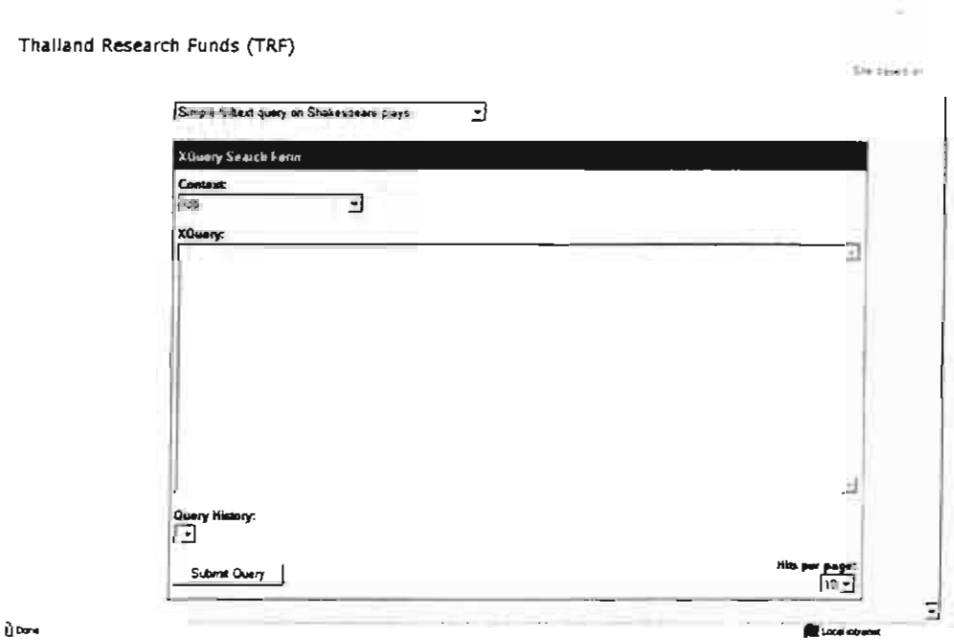


รูปที่ 4.3 แสดงการ Tagging ข้อมูลหน้ากิตติกรรมประกาศ



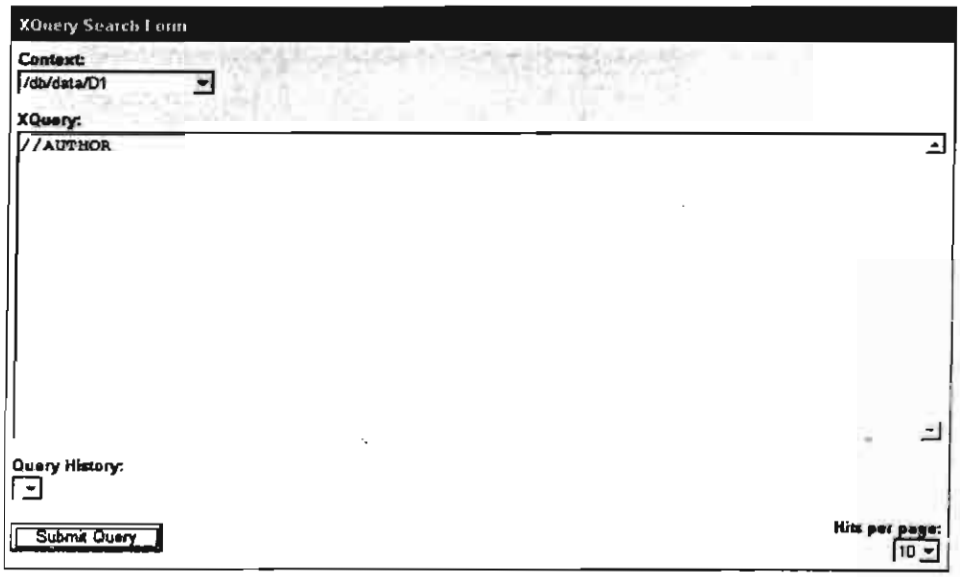
รูปที่ 4.4 แสดงการ Tagging ข้อมูลบทคัดย่อภาษาไทยและอังกฤษ

เมื่อทำการ Tagging ข้อมูลเรียบร้อยแล้วจะทำการเก็บข้อมูลเข้าสู่ฐานข้อมูลเพื่อทำการสร้างดัชนีในการค้นหาข้อมูล ซึ่งการทำงานในช่วงนี้จะอาศัยโปรแกรมอีกโปรแกรมหนึ่งเข้าช่วย โปรแกรมที่ใช้งานเป็นโปรแกรม eXist ซึ่งเป็นโปรแกรมที่ใช้ช่วยในภาษา XML ที่ใช้เก็บข้อมูลและดึงข้อมูลที่ต้องการค้นหาออกมาให้ตามคำที่เราต้องการค้นหาข้อมูลนั้น ๆ ซึ่งขั้นตอนการใช้งาน จะเริ่มต้นด้วยการเปิดโปรแกรม eXist XML DataBase หลังจากนั้นเลือก eXist Database Startup เพื่อทำการเปิดฐานข้อมูลขึ้นใช้งาน หลังจากนั้นก็จะทำการเปิด eXist Client Shell เพื่อทำการโหลดไฟล์ข้อมูลที่ต้องการใช้งานเข้าเก็บไว้ในฐานข้อมูล เมื่อทำการโหลดเรียบร้อยแล้ว ก็จะเข้าสู่หน้าจอการ ค้นหาข้อมูลโดยโปรแกรม eXist ดังรูปที่ 4.5

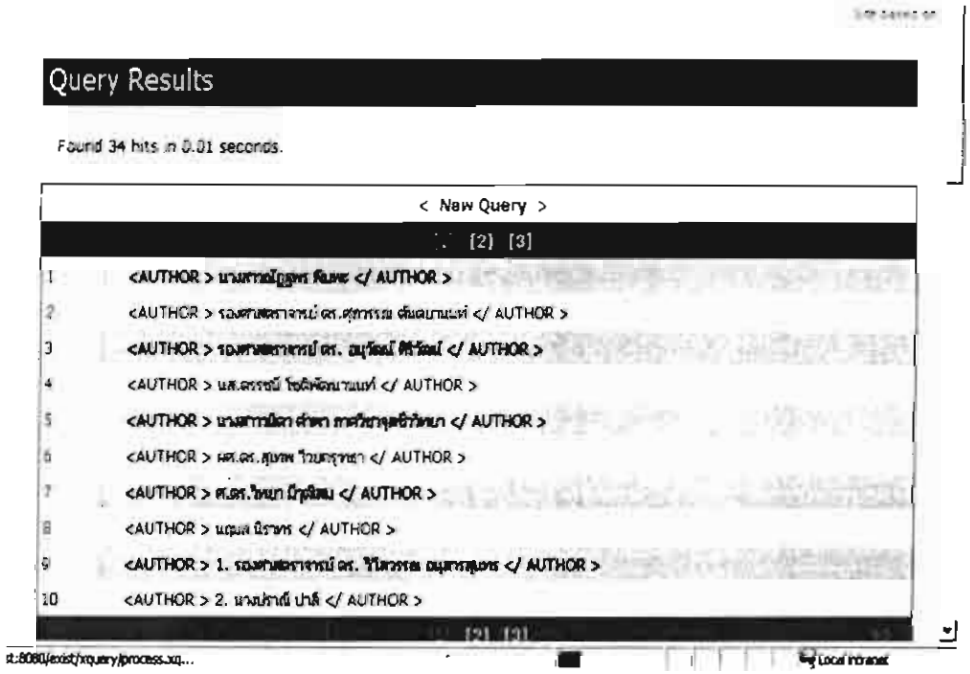


รูปที่ 4.5 แสดงหน้าจอการค้นหาข้อมูลด้วย eXist

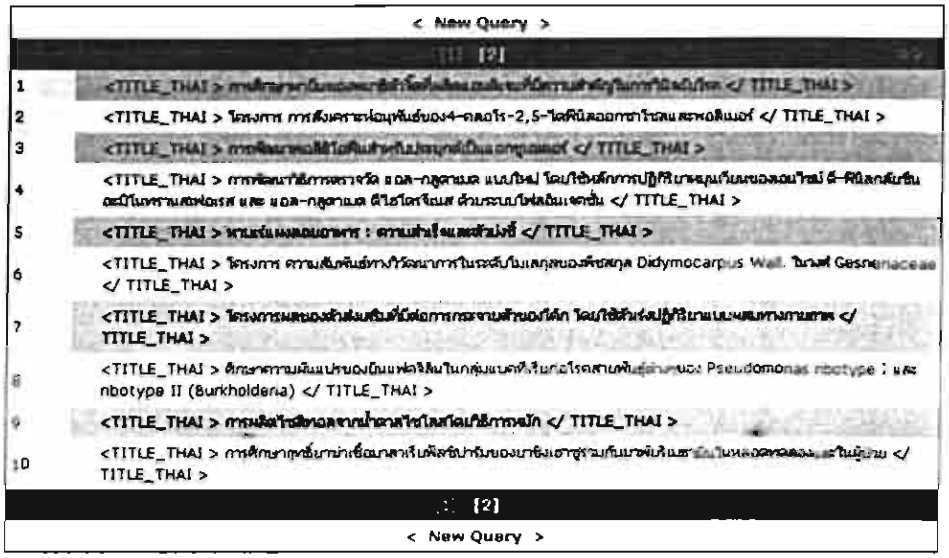
เมื่อเข้าสู่หน้าจอการค้นหาข้อมูลด้วย eXist แล้วเราสามารถกำหนดการค้นหาได้ว่าต้องการค้นหาข้อมูลอะไร ซึ่งเวลาจะค้นหาเราก็เพียงใส่ข้อมูลที่ต้องการค้นหาว่าเป็นอะไรในช่อง XQuery หลังจากนั้นทำการคลิกที่ปุ่ม Submit Query ก็จะปรากฏผลการค้นหาสิ่งที่เราต้องการค้นหาออกมาให้ ได้ดังรูปที่ 4.6แสดงคำสั่งในการค้นหาข้อมูลจากดัชนี รูปที่ 4.7 แสดงผลของการรันคำสั่ง ในการหา AUTHOR รูปที่ 4.8 แสดงผลของการรันคำสั่ง ในการหา TITLE_THAI AUTHOR รูปที่ 4.9 แสดงผลของการรันคำสั่ง ในการหา ACKNOWLEDGEMENT รูปที่ 4.10 แสดงผลของการรันคำสั่ง ในการหา ABSTRACT_ THAI รูปที่ 4.11 แสดงผลของการรันคำสั่ง ในการหา ABSTRACT_ ENG



รูปที่ 4.6 แสดงคำสั่งในการค้นหาข้อมูลจากดัชนีที่ได้สร้างไว้

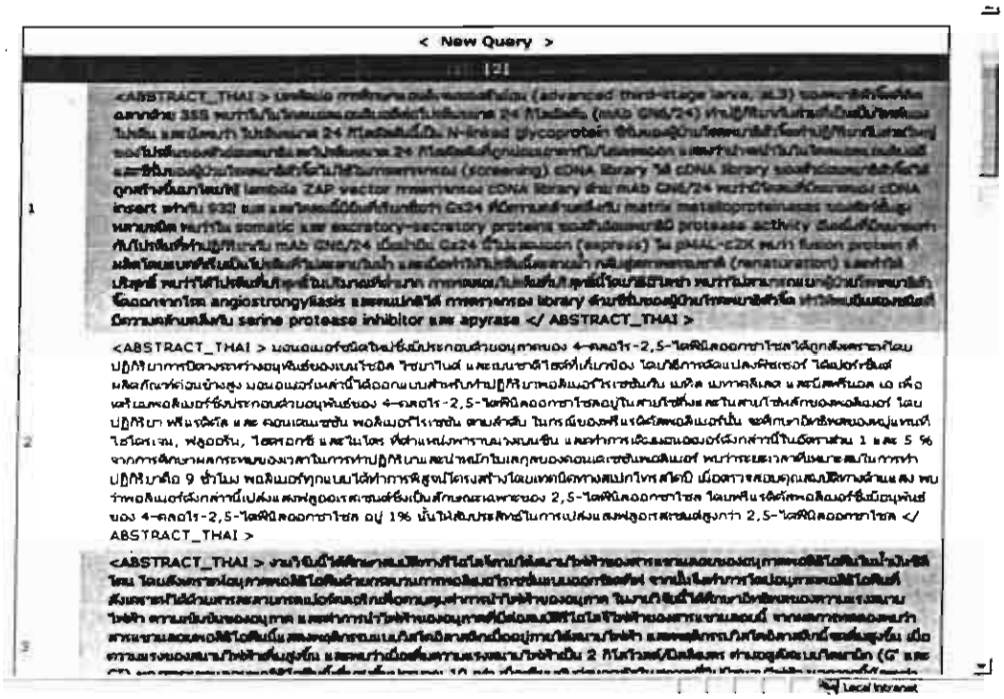


รูปที่ 4.7 แสดงผลของการรันคำสั่ง ในการหา AUTHOR

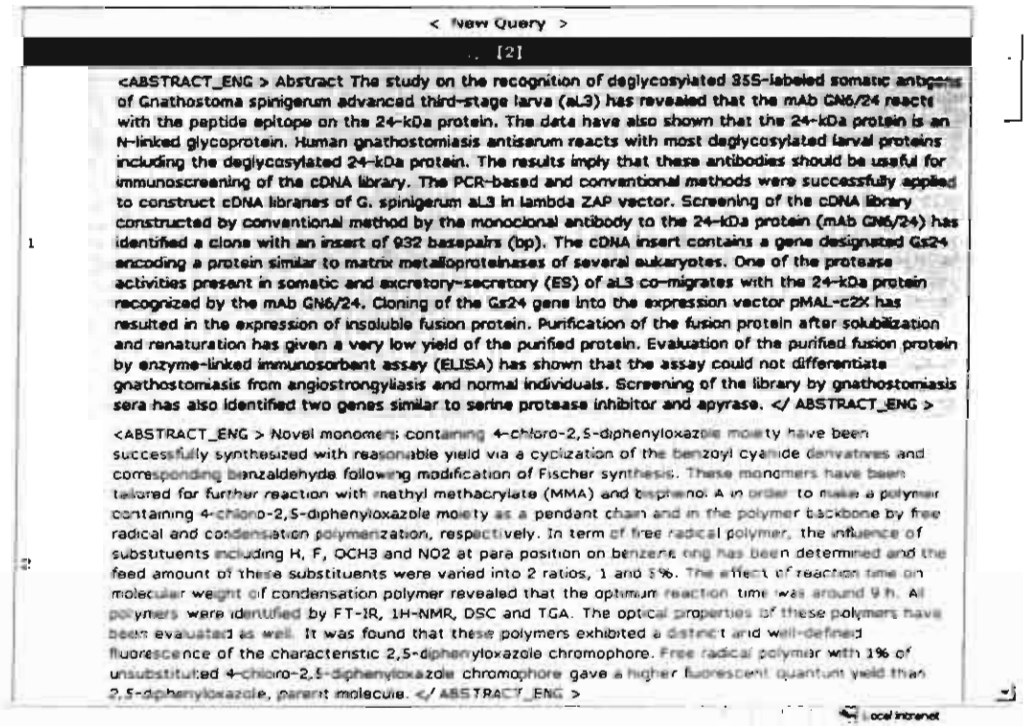


รูปที่ 4.8 แสดงหน้าจอการค้นหาข้อมูลที่เป็น TITLE_THAI

รูปที่ 4.9 แสดงหน้าจอการค้นหาข้อมูลที่เป็น ACKNOWLEDGEMENT



รูปที่ 4.10 แสดงหน้าจอการค้นหาข้อมูลที่เป็น ABSTRACT_THAI



รูปที่ 4.11 แสดงหน้าจอการค้นหาข้อมูลที่เป็น ABSTRACT_ENG

ผลการดำเนินงาน (ช่วงที่ 2)

จากการวิเคราะห์โครงสร้างเอกสารและกระบวนการต่าง ๆ ของเอกสารงานวิจัย สามารถแสดงโครงสร้างและการกำหนดความสำคัญให้กับเอกสารได้โดยใช้ XML เป็นเครื่องมือที่ช่วยในการ Tagging ในช่วงการดำเนินงานช่วงที่ 1 ทำให้มีการพัฒนารูปแบบการ Tagging ข้อมูลขึ้นมาใหม่เพื่อให้อำนวยความสะดวกในการสืบค้นข้อมูลได้ง่ายมากยิ่งขึ้น โดยแบ่งโครงสร้างที่ได้ทำการกำหนดไว้มี ดังนี้

ในส่วนของปกหน้า ดังนี้

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
        <title>  ชื่องานวิจัยภาษาไทย</title>
        <title type="alternative">ชื่องานวิจัยภาษาอังกฤษ  </title>
</titleInfo>
<name>
    <namePart>=ชื่อผู้วิจัยคนที่1 </namePart>
</name>
<name>
    <namePart>ชื่อผู้วิจัยคนที่2 </namePart>
</name>
<location>
   <url> </url>
</location>
<originInfo>
    <project_code>เลขสัญญา</project_code>
    <publisher>      สถาบัน </publisher>
    <dateCreated encoding="w3cdtf"> เริ่ม  </dateCreated>
    <dateEnd encoding="w3cdtf">   สิ้นสุด  </dateEnd>
    <dateModified encoding="w3cdtf">แก้ไขครั้งล่าสุด</dateModified>
</originInfo>
```

ตัวอย่างการ Tag ข้อมูลปกหน้า

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
```

<titleInfo>

<title>

โครงการการศึกษาบทบาทของ cell wall hydrolases องค์ประกอบของผนังเซลล์ และการแสดงออกของยีน ที่เกี่ยวข้องกับเอนไซม์ในการหลุดร่วงของผลกล้วยระหว่างการสุก

</title>

<title type="alternative">

A Study on the Role of Cell Wall Hydrolases, Cell Wall Components and Gene Expression of Enzyme(s) Involved in Finger Drop of Ripening Bananas

</title>

</titleInfo>

<name>

<namePart>

ศ.ดร. สายชล เกตุษา

</namePart>

</name>

<name>

<namePart>

น.ส. วชิรญา อิ่มสบาย

</namePart>

</name>

<name>

<namePart>

น.ส. อภิวรา ประยูรวงศ์

</namePart>

</name>

<originInfo>

<publisher> ภาควิชาพืชสวน คณะเกษตร มหาวิทยาลัยเกษตรศาสตร์ </publisher>

<dateCreated encoding="w3cdtf"></dateCreated>

<dateEnd encoding="w3cdtf"></dateEnd>

<dateModified encoding="w3cdtf"></dateModified>

</originInfo>

</mods>

ในส่วนของบทคัดย่อ แสดงได้ดังนี้

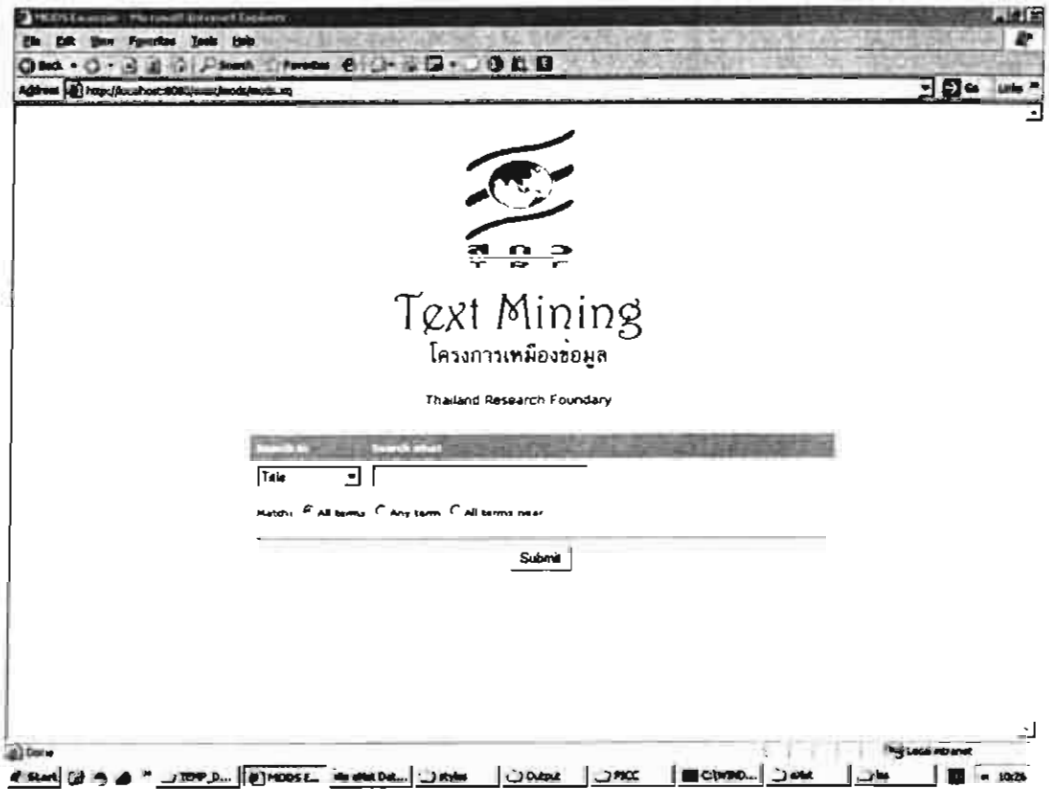
```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
                <title>  ชื่องานวิจัยภาษาไทย</title>
                <title type="alternative">ชื่องานวิจัยภาษาอังกฤษ  </title>
</titleInfo>
<name>
        <namePart>=ชื่อผู้วิจัยคนที่1 </namePart>
</name>
<name>
        <namePart>ชื่อผู้วิจัยคนที่2 </namePart>
</name>
<location>
     <url> </url>
</location>
<originInfo>
        <project_code>เลขสัญญา</project_code>
        <publisher>      สถาบัน </publisher>
        <dateCreated encoding="w3cdtf">  เริ่ม  </dateCreated>
        <dateEnd encoding="w3cdtf">    สิ้นสุด   </dateEnd>
        <dateModified encoding="w3cdtf">แก้ไขครั้งล่าสุด</dateModified>
</originInfo>
<subject>
        <topic> คำหลัก1</topic>
</subject>
<subject>
        <topic>คำหลัก2</topic>
</subject>
</mods>
```

ตัวอย่างการ Tag ข้อมูลในบทคัดย่อ

```
<?xml version="1.0" encoding="UTF-8"?>
<mods xmlns="http://www.loc.gov/mods/v3">
<titleInfo>
</titleInfo>
<name>
        <namePart>ศ.ดร. สายชล เกตุษา  </namePart>
</name>
<name>
        <namePart>น.ส. วชิรญา อิ่มสบาย</namePart>
</name>
<name>
        <namePart>น.ส. อภิวรา ประยูรวงศ์</namePart>
</name>
<originInfo>
        <project_code>BGJ / 11 / 2543 </project_code>
        <dateCreated encoding="w3cdtf"></dateCreated>
        <dateEnd encoding="w3cdtf"></dateEnd>
        <dateModified encoding="w3cdtf"></dateModified>
</originInfo>
<subject>
        <topic>คำสำคัญ</topic>
</subject>
<subject>
        <topic>การหลุดร่วงของผล</topic>
</subject>
<subject>
        <topic>การสุก</topic>
</subject>
<subject>
        <topic>กล้วยไข่</topic>
</subject>
```

<subject>
<topic>กล้วยหอมทอง</topic>
</subject>
<subject>
<topic>กล้วยน้ำว้า</topic>
</subject>
<subject>
<topic>บริเวณการร่วง</topic>
</subject>
<subject>
<topic>การอ่อนตัว</topic>
</subject>
<subject>
<topic>polygalacturonase</topic>
</subject>
<subject>
<topic>pectin</topic>
</subject>
<subject>
<topic>methylesterase</topic>
</subject>
<subject>
<topic>?-galactosidase</topic>
</subject>
</mods>

การแสดงหน้าจอการค้นหาข้อมูลต่าง ๆ ได้ดังรูปที่ 4.12



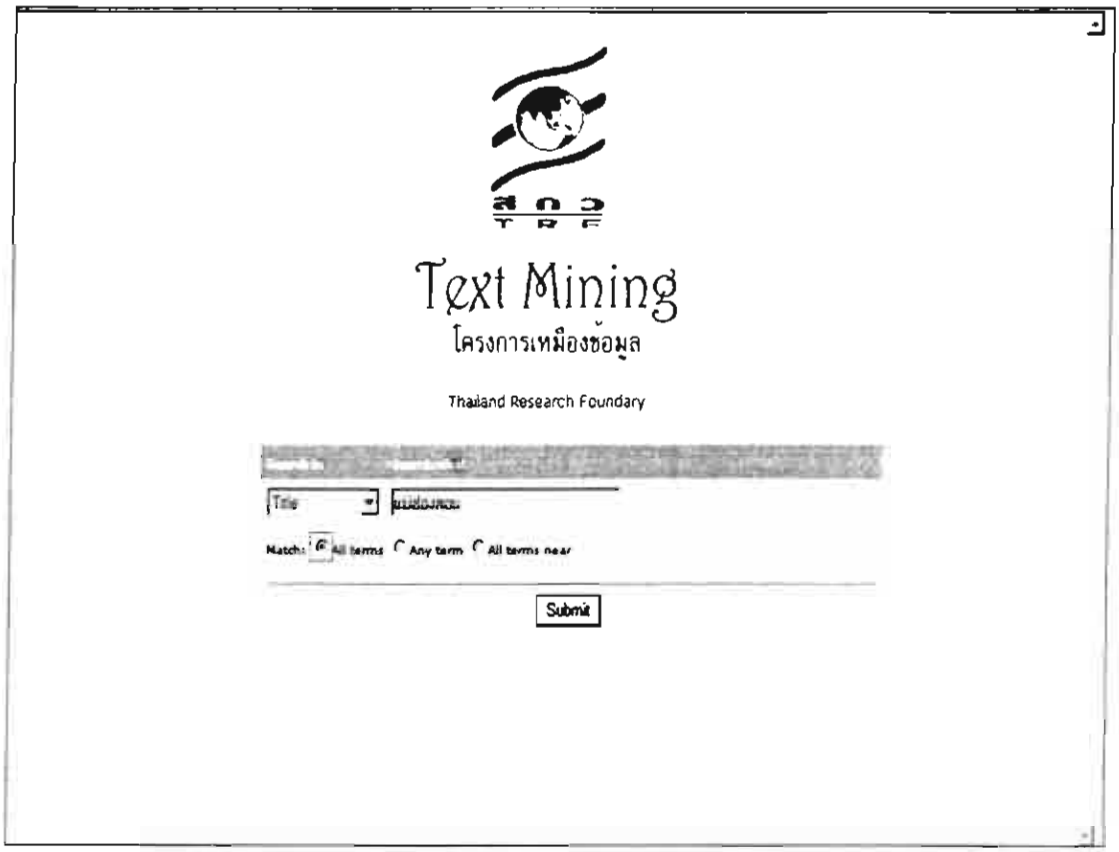
รูปที่ 4.12 แสดงหน้าจอการค้นหาข้อมูลต่าง ๆ

ในหน้าจอนี้จะเป็นหน้าจอแรกในการใช้งานโปรแกรมที่ได้จัดทำขึ้น จะแสดงครั้งแรกเมื่อเปิดใช้โปรแกรม ซึ่งจะมีการถามโดยให้ผู้ใช้สามารถเลือกว่าจะเลือกดูข้อมูลจากส่วนใด เช่น จาก Title, author , keyword ซึ่งเมื่อเลือกอย่างใดอย่างหนึ่งแล้วก็จะทำการใส่ข้อมูลลงในช่องต่อมาว่าจะใส่ข้อความอะไร เพื่อให้สอดคล้องกับสิ่งที่ต้องการค้นหา เช่นตัวอย่างเป็นการค้นหาข้อมูลจากคำว่า Title ในช่อง Search in และคำสำคัญคือ แม่ฮ่องสอน ในช่อง Search what และในบรรทัดต่อมาจะเป็นการเลือกว่าจะให้แสดงข้อมูลที่พบในลักษณะไหนบ้าง

All terms เป็นการแสดงรายการทั้งหมดที่ค้นพบ

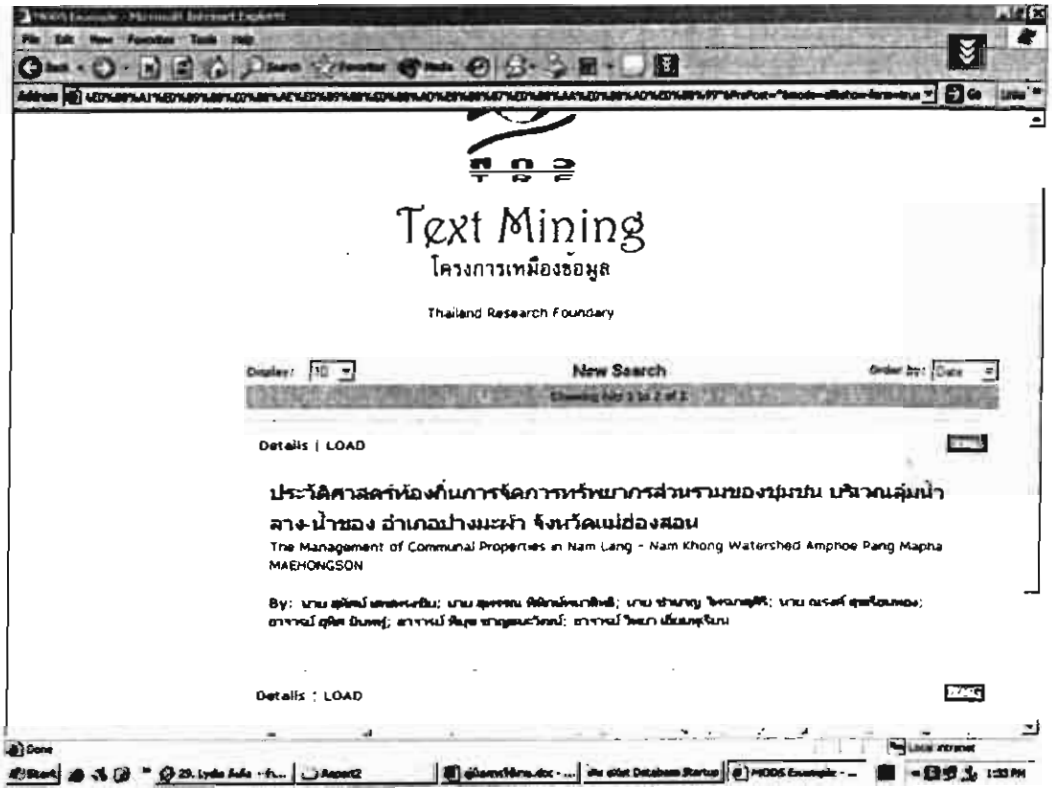
Any term เป็นการแสดงเฉพาะที่ค้นหา

All Terms near เป็นการแสดงรายการทั้งหมดและที่ใกล้เคียงกัน ซึ่งเราสามารถเติมคำที่ต้องการค้นหาได้ในช่องที่ 2 ทางขวามือ เช่นเราต้องการค้นหาข้อความงานวิจัยที่ขึ้นต้น (title) ว่า แม่ฮ่องสอน ก็จะทำตามขั้นตอนดังรูปที่ 4.13



รูปที่ 4.13 แสดงหน้าจอการค้นหาข้อมูลแม่ฮ่องสอน

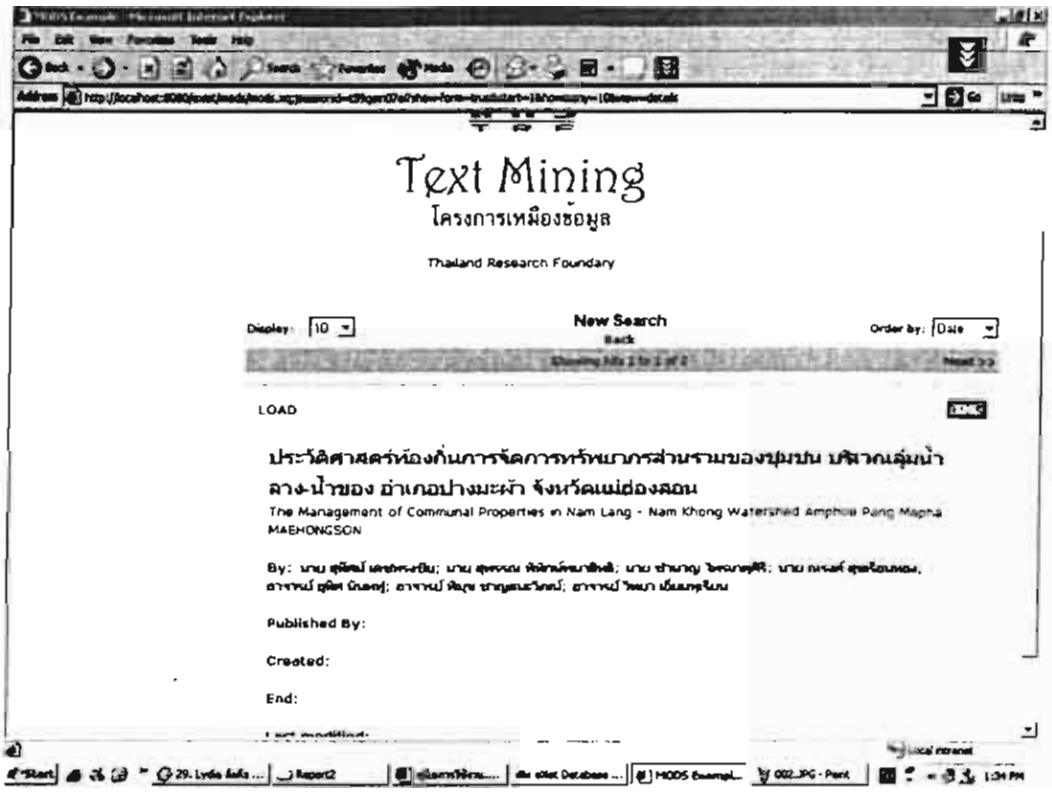
ในหน้าจอการค้นหาข้อมูลเราก็จะพิมพ์คำว่าแม่ฮ่องสอนลงไป หลังจากนั้นก็เลือกรูปแบบการแสดงผล แล้วทำการกดปุ่ม Submit เพื่อให้โปรแกรมทำการประมวลผลข้อมูลให้ ดังรูปที่ 4.14



รูปที่ 4.14 แสดงหน้าจอการค้นหาข้อมูลที่เลือก

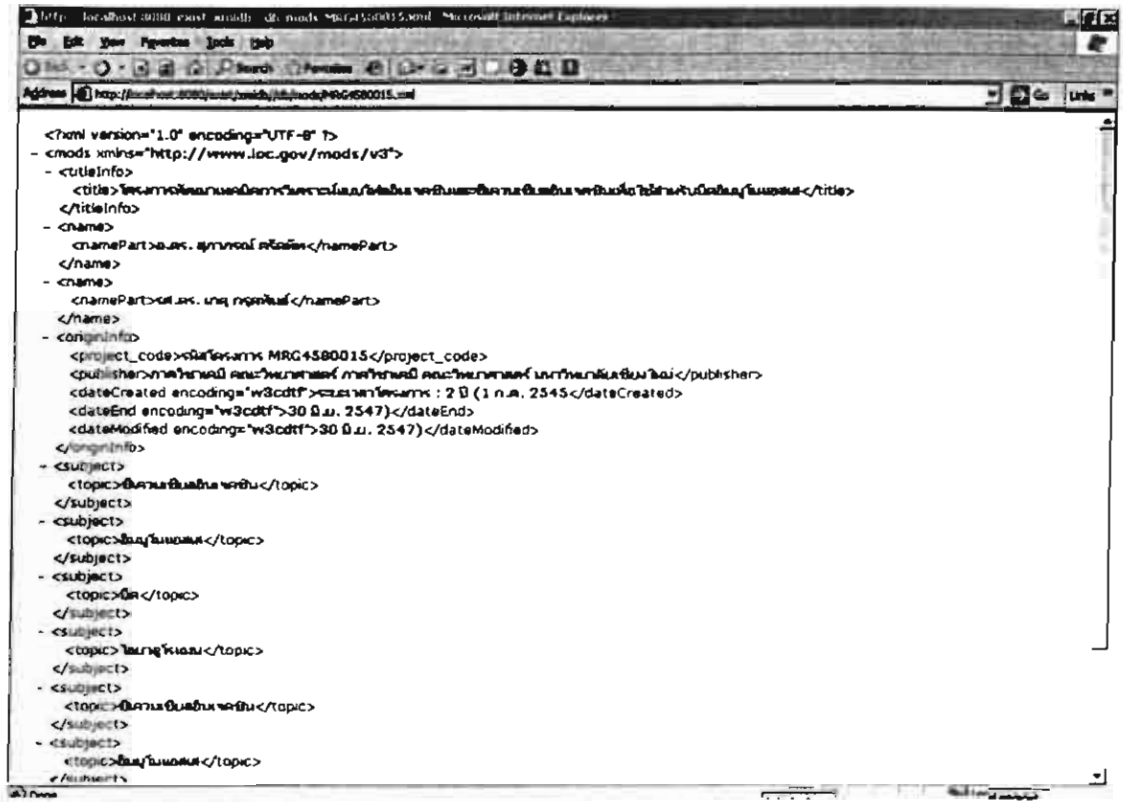
ในส่วนนี้จะเป็นการแสดงผลการเลือกข้อมูลที่ได้กรอกไว้ ซึ่งจะโปรแกรมจะบอกว่ามีข้อมูลทั้งหมด 2 ข้อมูล ในส่วนที่ค้นหา ซึ่งในหน้าจอเราจะพบว่า ในส่วนของ display นั้นเราสามารถกำหนดได้ว่าจะให้หน้าจอเราแสดงผลการค้นหาข้อมูลกี่รายการซึ่งเราจะพบว่าได้กำหนดไว้ที่ 10 รายการเป็นหลัก แต่ถ้าเราต้องการจะเปลี่ยนเป็น 5 รายการก็เพียงแค่นำเมาส์ไปคลิกแล้วเปลี่ยนเป็นเลข 5 ระบบก็จะทำการเปลี่ยน ให้เป็น 5 รายการต่อ 1 หน้าจอ และอีกส่วนหนึ่ง คือ Order by จะเป็นการกำหนดให้ข้อมูลมีการเรียงลำดับแบบไหนบ้าง ซึ่งในหน้าจอเรากำหนดให้เรียงโดยตาม Date , Title , Author ซึ่งผู้ใช้สามารถกำหนดส่วนนี้ได้เองตามความต้องการ

เมื่อทำการกำหนดสิ่งต่าง ๆ ได้แล้วระบบจะทำการประมวลผลและแสดงผลการค้นหาให้ตามที่ผู้ใช้กำหนดไว้ ซึ่งในกรณีที่เราต้องการจะเข้าไปดูรายละเอียดของงานวิจัยที่เราค้นหาในแต่ละรายการนั้นเราก็สามารถคลิกเลือกที่ Details ก็จะปรากฏรายละเอียดของโครงการให้ ดังรูปที่ 4.15



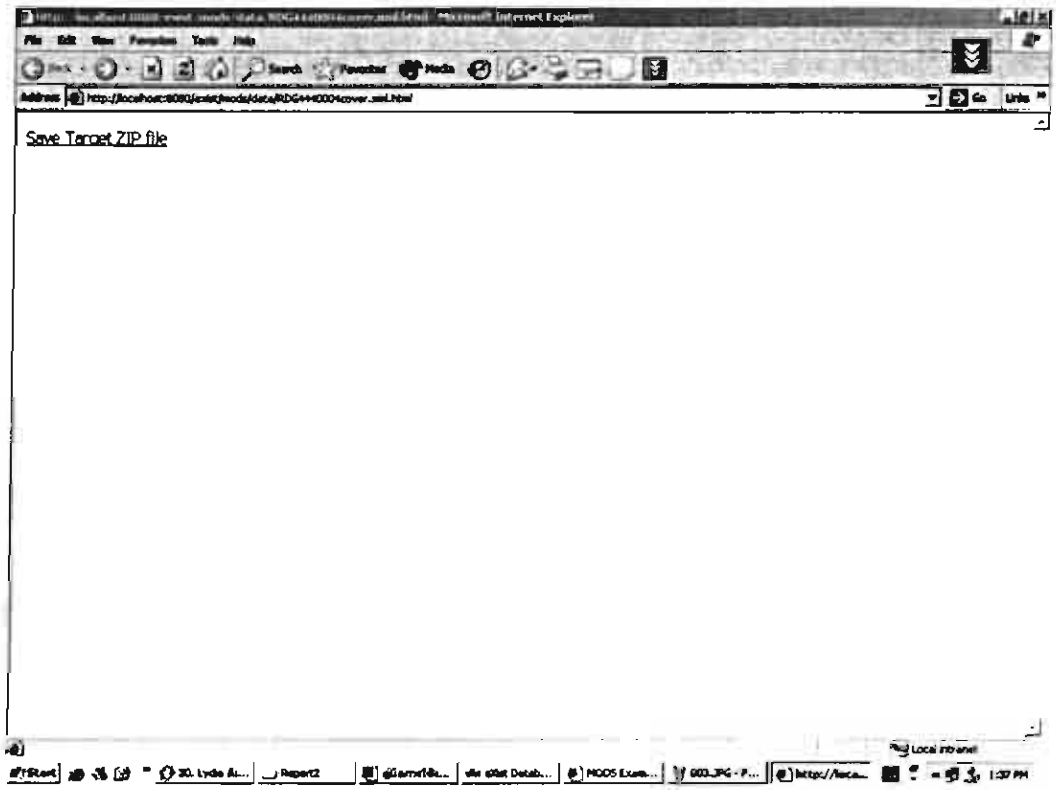
รูปที่ 4.15 แสดงหน้าจอรายละเอียดข้อมูล

ในส่วนนี้จะเป็นการแสดงรายละเอียดของงานวิจัยที่เราเลือกดูข้อมูล ซึ่งจะแสดงรายการต่าง ๆ ให้ว่าใครเป็นผู้ทำวิจัย หัวข้อเรื่องวิจัยอะไร มีรหัสโครงการอะไร ระยะเวลาโครงการ ฯลฯ ซึ่งรายละเอียดในส่วนนี้เราสามารถเข้าไปดูและทำการคัดลอกข้อมูลรายงานการวิจัยได้ในส่วนที่เป็นข้อมูล XML ได้ที่ปุ่ม XML ก็จะแสดงข้อมูลต่าง ๆ ของการ Tag ข้อมูลดังรูปที่ 4.16



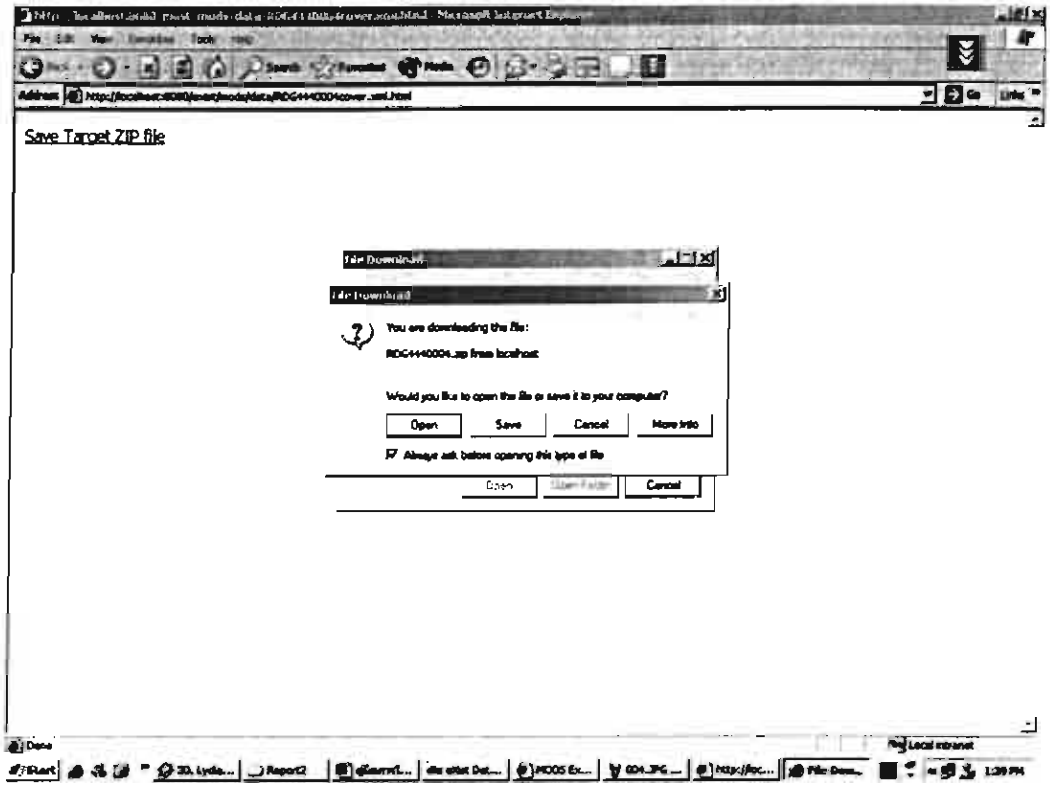
รูปที่ 4.16 แสดงการ Tag ข้อมูล

จะแสดงการ Tag ข้อมูลที่ได้เข้าไปดูรายละเอียดโครงสร้างของข้อมูล แต่ถ้าต้องการที่จะทำการโหลดข้อมูลทั้งงานวิจัยก็จะทำการเลือกที่ปุ่ม Load เพื่อทำการโหลดข้อมูลดังรูปที่ 4.17



รูปที่ 4.17 แสดงหน้าจอการเข้าโหลดข้อมูลงานวิจัย

ในหน้าจอนี้เมื่อเราทำการคลิกที่ Save Target ZIP file จะปรากฏหน้าจอการโหลดข้อมูลดังรูปที่ 4.18



รูปที่ 4.18 แสดงการsave ข้อมูลที่ทำการโหลด

ในหน้าจอนี้จะปรากฏให้เราเลือกว่าเราจะทำการบันทึกข้อมูลหรือว่าจะเปิดข้อมูลขึ้นมาดู ซึ่งถ้าเราต้องการบันทึกข้อมูลเก็บไว้ดูเราก็คลิกที่ปุ่ม Save เพื่อบันทึกข้อมูลได้เลย เมื่อทำการบันทึกเรียบร้อยระบบก็จะเสร็จสิ้นการทำงาน เช่นเดียวกับการโหลดข้อมูลทางอินเทอร์เน็ตทั่วๆ ไป

การพัฒนาโปรแกรมสำหรับการดูข้อมูลจากหมืองข้อมูลงานวิจัย

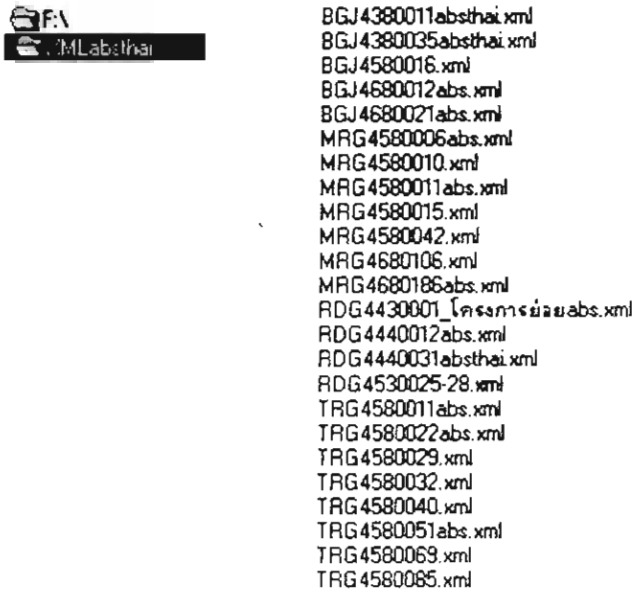
จากการออกแบบการทำงานของโปรแกรม ทั้ง 4 ส่วน สามารถพิจารณาผลลัพธ์ที่ได้ในแต่ละส่วนดังนี้

1. **ส่วนของการนำเข้าข้อมูล (Open File)**

ส่วนของการนำเข้าข้อมูล สามารถทำการเลือก เปิดไฟล์ที่เป็น XML ที่ได้ทำการ Tag ไว้ได้โดยสามารถทำการเลือกไดร์ฟ ที่มีข้อมูลอยู่ (จากตัวอย่างทำการเลือก จาก ไดร์ฟ F

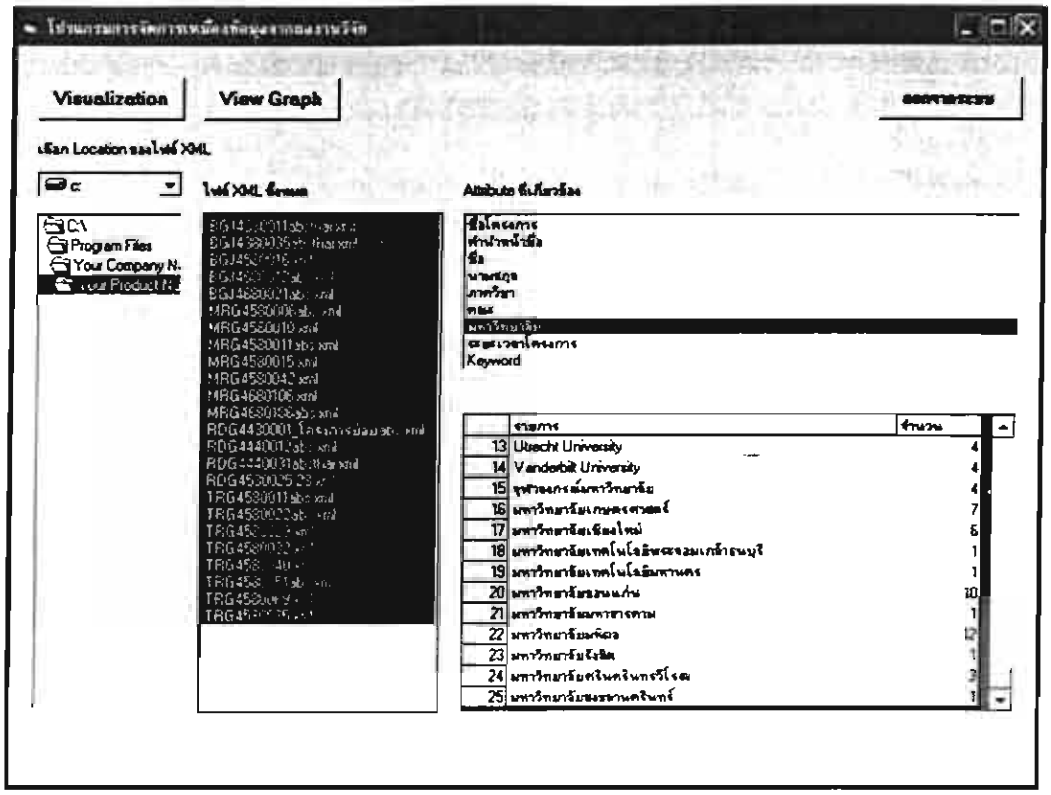
)

โปรแกรมจะแสดงข้อมูลไฟล์ xml ทั้งหมดที่มีอยู่ ดังรูป



รูปที่ 4.19 แสดงไฟล์ xml ที่นำเข้าสู่โปรแกรม

โดยที่เมื่อทำการเปิดไฟล์แล้ว โปรแกรมจะแสดงไฟล์ที่ได้ทำการเลือกพร้อมทั้งแสดงTag ที่สนใจไว้ทางหน้าต่างด้านขวามือ ซึ่งสามารถนำเข้าไฟล์ได้หลายๆไฟล์ โดยจะมีการคำนวณค่าทางสถิติของแต่ละฟิลด์ไว้ด้านล่าง ดังรูป

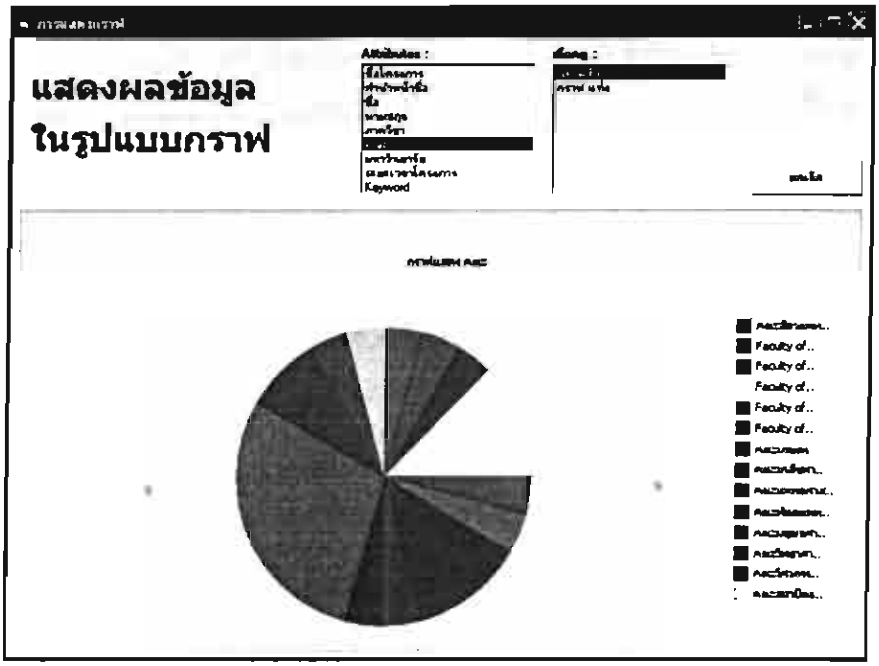


รูปที่ 4.20 แสดงส่วนนำเข้าไฟล์เอกสารงานวิจัยที่อยู่ในรูป xml

ซึ่งจะแสดงส่วนของ tag หรือฟิลด์ต่างๆที่สนใจ เช่นจากรูปเมื่อนำไฟล์เข้าประมวลผลแล้ว สามารถเลือกดูTag ที่สนใจได้ ในที่นี้ เลือกดูชื่อมหาวิทยาลัยที่มีผลงานวิจัย โดยจะมีจำนวนของงานวิจัยปรากฏในส่วนของคอลัมภ์จำนวน

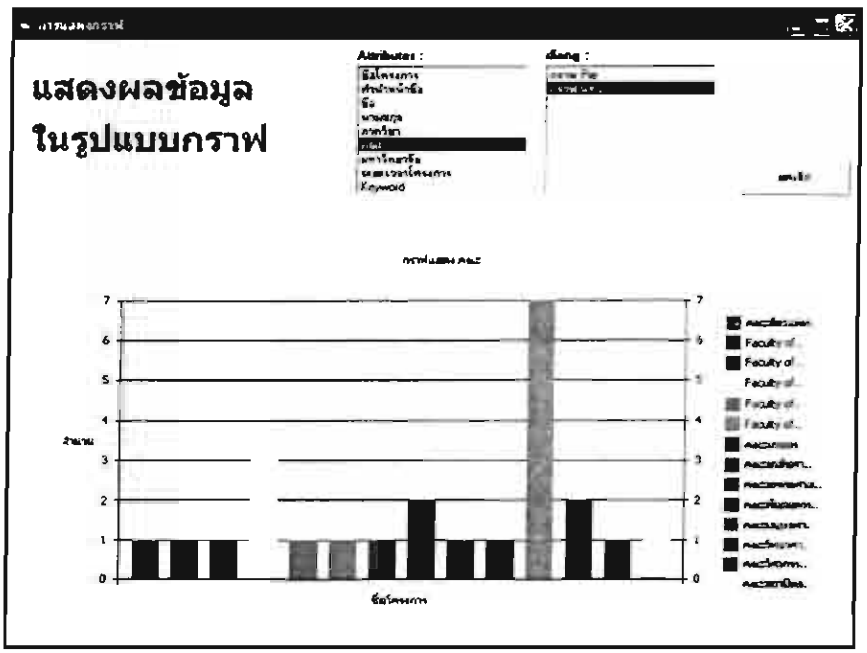
2. ส่วนของการนำเสนอข้อมูลในรูปของกราฟ (View Graph)

เมื่อต้องการดูข้อมูลที่สนใจในรูปแบบของกราฟ สามารถทำการเลือกดูข้อมูลได้ในรูปแบบของกราฟ วงกลมและ กราฟ แท่ง โดยอันดับแรก ต้องทำการเลือก attribute ที่สนใจก่อนแล้วถึงทำการเลือกรูปแบบของกราฟทางด้านขวามือ จะปรากฏข้อมูลดังรูป



รูปที่ 4.21 แสดงกราฟวงกลมแสดงสัดส่วนของข้อมูลจากงานวิจัย

นอกจากนี้ สามารถทำการเลือกมุมมองในรูปแบบของกราฟแท่งได้เช่นเดียวกัน

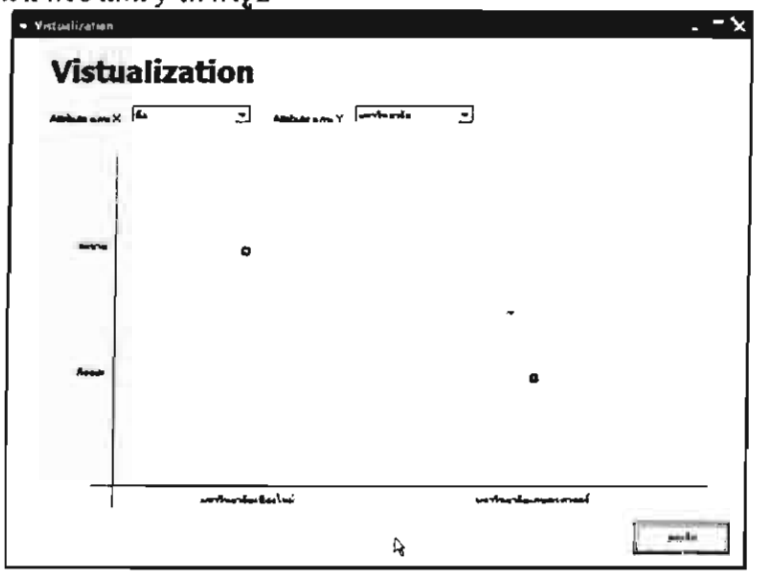


รูปที่ 4.22 แสดงกราฟแท่งแสดงสัดส่วนของข้อมูลจากงานวิจัย

จากตัวอย่างเป็นการดูข้อมูลว่ามีคณะใดบ้างที่มีการทำงานวิจัย มากที่สุดซึ่งโปรแกรมจะทำการรวบรวมข้อมูลจาก Tag <FACULTY> ทั้งหมดจากทุกๆไฟล์เอกสารงานวิจัย โดยสามารถดูข้อมูลได้ เมื่อเอาเมาส์วางจะปรากฏข้อมูล ของแต่ละส่วนของกราฟ โดยจะมีสีแทนความหมายคณะต่างๆ

3. ส่วนของการนำเสนอข้อมูลที่เกี่ยวข้องกัน (Visualization)

เป็นส่วนที่สามารถทำการเลือกดูข้อมูลที่มีความสัมพันธ์กันได้ โดยสามารถกำหนดข้อมูลที่สนใจให้แสดงอยู่บนแกน x หรือ แกน y ได้ ดังรูป

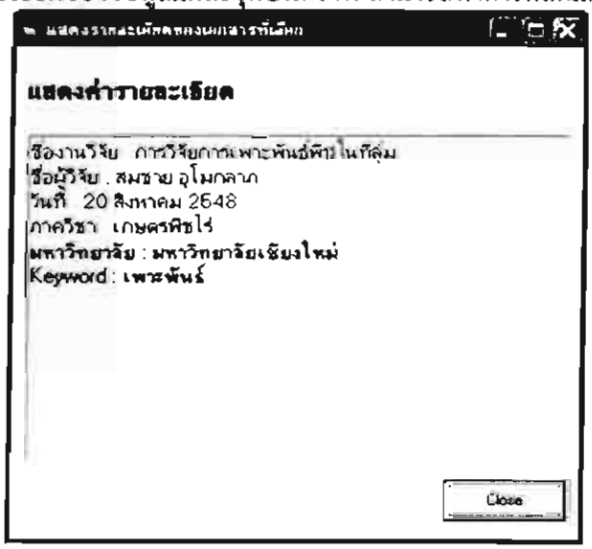


รูปที่ 4.23 แสดงกราฟความสัมพันธ์ระหว่างข้อมูล

จากตัวอย่างเป็นการดูข้อมูล ที่มีความสัมพันธ์กันระหว่าง ชื่อผู้ทำงานวิจัยและมหาวิทยาลัย ซึ่งจะแสดงเป็นจุดบนกราฟ ซึ่งสามารถคลิกดูรายละเอียดของข้อมูลได้ ดังแสดงในส่วนที่ 4

4. ส่วนของการดูรายละเอียดของข้อมูลที่สัมพันธ์กัน (Detail)

หากต้องการที่จะดูรายละเอียดของข้อมูลแต่ละจุดบนกราฟ สามารถทำการคลิกเลือกได้ดังรูป



รูปที่ 4.24 แสดงรายละเอียดของข้อมูลบนกราฟ

เมื่อทำการคลิกเลือกที่จุดบนกราฟก็จะแสดงข้อมูลรายละเอียดทั้งหมดใน Tag XML ทั้งหมด

บทที่ 5

สรุปผลการดำเนินงาน

ปัญหาและอุปสรรคในการดำเนินงาน

ในการดำเนินงานในโครงการพบปัญหาต่าง ๆ ดังต่อไปนี้

1. การแปลงเอกสารข้อมูลต้องอาศัยการค้นหาโปรแกรมต่าง ๆที่เกี่ยวข้องในการทำงานอาทิเช่น โปรแกรมที่ใช้ในการแปลงข้อมูลจาก Microsoft Word ไปเป็น Text ซึ่งจุดนี้พบว่ามีได้ค้นหาโปรแกรมการทำงานที่ยุ่งยากพอสมควรเนื่องจากว่า โปรแกรมส่วนใหญ่ทำมาเพื่อรองรับเอกสารที่เป็นภาษาอังกฤษมากกว่าภาษาไทย จึงทำให้เสียเวลาในการค้นหาโปรแกรมช่วยงานพอสมควร และนอกจากนี้ยังได้มีการสร้างโปรแกรมเพื่อช่วยงานให้สามารถดำเนินการได้เร็วขึ้น อาทิเช่น การตัดคำโดยอาศัยโปรแกรม Swath ซึ่งจะบอกคุณลักษณะของคำมาให้ แต่ก็ยังมีปัญหาตามมาด้วยคือ เมื่อตัดคำเรียบร้อยแล้ว คำบางคำไม่สามารถสื่อความหมายตามงานวิจัยได้ ในการตัดคำ เนื่องจากว่าภาษาไทยเป็นภาษาที่ไม่ตายตัว ทำให้เวลาตัดคำบางครั้งคำที่ตัดออกมาไม่สามารถสื่อใจความได้ มีการตัดคำที่ผิดจากความเป็นจริง ไม่ตรงตามความหมายเกิดขึ้น นับเป็นปัญหาสำคัญในการดำเนินงานในโครงการ ฯ เช่นคำว่า จังหวัด ซึ่งควรจะติดกัน แต่เมื่อทำการตัดคำแล้ว โปรแกรมจะแยกออกให้เป็นคำ ๆ ไป ทำให้เวลาที่ทำการ ค้นหาข้อมูลเกิดความผิดพลาดได้ นี้ก็เป็นอีกหนึ่งปัญหาที่เกิดขึ้นจากการใช้โปรแกรม Swath แต่ก็ได้เตรียมวิธีการแก้ไขปัญหานี้ไว้แล้วคือ การเขียนโปรแกรมขึ้นมาเพื่อรองรับให้คำที่สมควรอยู่ติดกันอยู่ใกล้กัน โดยอาศัยพจนานุกรม เป็นตัวเปรียบเทียบคำอีกครั้งหนึ่งถึงจะได้ออกมา ซึ่งปกติแล้ว คำว่า " โครงการวิจัย " จะต้องแยกออกจากกันเป็น โครง + การ + วิจัย ทำให้การค้นหามีความยุ่งยากในการกำหนดรูปแบบ เลยได้มีการรวมคำเพื่อให้สื่อความหมายมากยิ่งขึ้นในการค้นหาข้อมูล ดังตัวอย่าง โครงการวิจัย@NCMN|"@NCMN|การ@FIXN|ประเมิน@VSTA|สถาน@CMTR|ภาพ@NCMN|องค์@CMTR|ความ@FIXN|รู้@VSTA|จังหวัด@NCMN|ศรี@NCMN|สะ@NCMN|เก@VATT|ษ@VATT|"@NPRP|
บท@CLTV|คัด@VACT|ย่อ@VACT|

|การ@FIXN|วิจัย@VACT| |"@FIXN|การ@FIXN|ประเมิน@VSTA|สถาน@CMTR|ภาพ@NCMN|องค์@CMTR|ความ@FIXN|รู้@VSTA|จังหวัด@NCMN|ศรี@NCMN|สะ@NCMN|เก@VATT|ษ@VATT|"@NPRP| |มี@VSTA|วัตถุประสงค์@NCMN|เพื่อ@JSBR|ประมวล@VACT| |วิเคราะห์@VACT| |ประเมิน@VSTA| |และ@JCRG|สังเคราะห์@VACT|องค์@CMTR|ความ@FIXN|รู้@VSTA|เกี่ยว@VSTA|กับ@RPRE|จังหวัด@NCMN|ศรี@NCMN|สะ@NCMN|เก@VATT|ษ@VATT|อย่าง@FIXV|เป็น@VSTA|ระบบ@NCMN| |เพื่อ@JSBR|นำเสนอ@VACT| |ทิศทาง@NCMN|และ@JCRG|โจทย์@NCMN|การ@FIXN|วิจัย@VACT|ตาม@RPRE|ความ@FIXN|สำคัญ@VATT|

เร่งด่วน@ADVN| |เพื่อ@JSBR|การ@FIXN|แก้@VACT|ปัญหา@NCMN|และ@JCRG|พัฒนา@VACT| จังหวัด@NCMN|ใน@RPRE|อนาคต@NCMN| |กระตุ้น@VACT|ให้@JSBR|เกิด@VSTA| กระบวนการ@NCMN|เรียนรู้@VACT|ของ@RPRE|คน@CNIT|ใน@RPRE|ท้องถิ่น@NCMN| | พัฒนา@VACT|และ@JCRG|เสริมสร้าง@VACT|ความ@FIXN|เข้มแข็ง@ADVN|ทาง@RPRE| วิชาการ@NCMN|และ@JCRG|การ@FIXN|วิจัย@VACT|แก่@RPRE|นักวิจัย@NCMN|ระดับ@NCMN| ท้องถิ่น@NCMN| |เชื่อม@VACT|โยง@VACT|ข้อมูล@NCMN|จาก@RPRE|การ@FIXN|วิจัย@VACT| เข้า@XVAE|กับ@RPRE|วิสัย@NCMN|ทัศน์@NCMN|ใน@RPRE|การ@FIXN|พัฒนา@VACT| จังหวัด@NCMN| |โดย@FIXV|คำนึง@VACT|ถึง@RPRE|พื้นฐาน@VATT|ความ@FIXN|เป็น@VSTA| จริง@VATT|และ@JCRG|ความ@FIXN|ต้องการ@XVMM|ของ@RPRE|คน@CNIT|ใน@RPRE| ท้องถิ่น@NCMN| |โครงการ@CNIT|ดำเนินการ@VACT|วิจัย@VACT|โดย@JSBR|การ@FIXN| รวบรวม@VACT|เอกสาร@NCMN|ต่าง@PDMN| ๆ@NPRP| |ที่@PREL|เกี่ยวข้อง@VSTA|กับ@RPRE| องค์@CMTR|ความ@FIXN|รู้@VSTA|ใน@RPRE|จัง@ADVN|หวัด@NCMN|ศรี@NCMN|สะ@NCMN| เก@VATT|ษ@VATT|ใน@RPRE|ด้าน@NCMN|ต่าง@PDMN| |ๆ@NPRP| |

2. ข้อมูลที่ได้มามีโครงสร้างที่ไม่แน่นอนทำให้เวลาทำการ Tagging จะต้องหาโครงสร้างที่มีความแน่นอนเป็นตัวมาตรฐานในการ Tagging ข้อมูลซึ่งพบว่าในแต่รายงานมีความหลากหลายในการกำหนดหัวข้อดำเนินการที่แตกต่างกันไปอาทิเช่น ในบทคัดย่อที่เป็นภาษไทย จะมีด้วยกัน 2 แบบคือแบบที่ 1 จะมีในส่วนของผู้วิจัยว่าชื่ออะไร รหัสโครงการอะไร ฯลฯ แต่แบบที่ 2 จะไม่มีในส่วนของผู้วิจัยมีเฉพาะเนื้อหาเท่านั้น ดังตัวอย่าง ทั้ง 2 แบบ

แบบที่ 1 บทคัดย่อ

เลขที่สัญญา: RSA/06/2544

ชื่อโครงการ: คาลิก[4]ซารีนคาวิโนนไดเมอร์สำหรับเป็นตัวตรวจจับไอออนของโลหะแอลคาไลและแอลคาไลน์เอิร์ธ

ผู้ดำเนินการวิจัย:

หัวหน้าโครงการ: ผู้ช่วยศาสตราจารย์ ดร. ธวัชชัย ตันฑุลานิ

ภาควิชาเคมี คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

E-mail Address: tthawatc@chula.ac.th

ระยะเวลาดำเนินโครงการ: 1 ธันวาคม 2543 ถึง 30 พฤศจิกายน 2546

วัตถุประสงค์: เพื่อสังเคราะห์คาลิก[4]ซารีนคาวิโนนไดเมอร์และคาลิก[4]ซารีนที่มีเฟอร์โรซีนเอไมด์เป็นองค์ประกอบและเพื่อศึกษาสมบัติการจับกับไอออนของโลหะแอลคาไลและแอนไอออนตามลำดับ จากนั้นวัดความสามารถในการเป็นเซนเซอร์ของสารที่สังเคราะห์ได้ด้วยไซคลิกโวลแทมเมทรีและสแควร์เวฟโวลแทมเมทรี ฯลฯ

แบบที่ 2 บทคัดย่อ

เซลล์สัตว์ชั้นสูงมีกระบวนการตอบสนองต่อภาวะเครียดซึ่งเกิดจากการสะสมของ unfolded protein ใน endoplasmic reticulum (ER) ที่เรียกว่า Unfolded protein response (UPR) กระบวน การนี้จะทำหน้าที่ส่งสัญญาณไปยัง nucleus เพื่อกระตุ้นการสร้างโปรตีนที่ช่วยเร่งกระบวนการม้วนพับของ unfolded protein เหล่านี้มีโครงสร้างที่ถูกต้องและสามารถถูกส่งออกจาก ER อันจะนำไปสู่การลดการภาวะเครียด จากการศึกษากลไกการส่งสัญญาณนี้ ฯลฯ

จากตัวอย่างทั้ง 2 แบบทำให้ต้องค้นหารูปแบบที่มีความสมบูรณ์ที่สุดเป็นตัวกำหนดการทำ Tagging ข้อมูล ซึ่งอาศัยการค้นหามากเช่นกัน ดังตัวอย่างที่ได้แสดงไว้แล้วในรูปที่ 4.1 – 4.4 ซึ่งเป็นโครงสร้างที่ได้มีการกำหนดให้เป็นมาตรฐานเดียวกันในการ Tagging งานเอกสารในการดำเนินโครงการ ฯ ในครั้งนี้

3. โปรแกรม และ ภาษาที่ใช้ในการดำเนินงานเป็นภาษาที่ใหม่ซึ่งต้องอาศัยเวลาในการศึกษาพร้อมปฏิบัติไปพร้อม ๆ กัน ทำให้บางช่วงงานมีความล่าช้า เกิดขึ้น
4. การหาเอกสารอ้างอิงของงานวิจัยที่เคยทำในลักษณะเดียวกันไม่ค่อยมี ทำให้ต้องเสียเวลาในการลองถูกลองผิดในการทำงานจนกระทั่งได้ในสิ่งที่ต้องการเป็นเวลานาน

สรุปผลการดำเนินงาน

จากการดำเนินการโครงการเหมืองข้อมูล จะต้องมีการศึกษาถึงรูปแบบโครงสร้างของภาษาไทย ซึ่งเป็นที่เข้าใจกันแล้วว่าภาษาไทยมีความยืดหยุ่นในตัวเองมากพอสมควร นั่นก็เพราะว่าภาษาไทยเราเป็นภาษาที่สามารถนำไปประกอบกับคำต่าง ๆ แล้วสามารถที่จะแปลความหมายได้หลากหลายความหมายเช่นเดียวกัน เช่นว่า คำว่า ตากลม เราสามารถที่จะแบ่งแยกคำออกเป็น 2 คำคือ ตา + กลม และ ตาก + ลม ซึ่งในแต่ละความหมายนั้นเราจะดูจากรูปประโยคที่อยู่ติดกันไปว่า คำ ๆ เดียวกันนี้เราจะแปลความหมายได้เป็นอะไรนั้นเอง นั่นคือข้อกำหนดอีกอย่างของภาษาไทยที่มีความยืดหยุ่นต่างจากภาษาอังกฤษ ซึ่งมีความตายตัวในความหมายมากกว่าภาษาไทย ซึ่งทำให้เกิดการแยกคำของภาษาไทยมีความยุ่งยากมาก ดังงานวิจัยที่ได้มีการดำเนินการมาในช่วงแรก ซึ่งมีการศึกษาดำเนินการเพื่อที่จะทำให้สามารถแยกคำได้ โดยความหมายของคำไม่เปลี่ยนแปลงมากจนกระทั่งว่าไม่สามารถจะกลับมาเป็นคำเดิมได้ ซึ่งการดำเนินการนี้เราได้ศึกษาถึงภาษาต่าง ๆ เป็นจำนวนมากและโปรแกรมต่าง ๆ ที่เราสามารถจะนำมาเป็นแบบอย่างเพื่อทำการพัฒนาต่อไปเพื่อให้สามารถทำงานกับข้อมูลที่เราได้รับมา ซึ่งตัวของข้อมูลเองนั้นก็มีความหลากหลาย

รูปแบบในการจัดทำ ซึ่งก็เป็นปัญหาอีกอย่างหนึ่งด้วยเช่นกันในการจัดหารูปแบบที่จะใช้เป็นมาตรฐานในการเป็นต้นแบบในการกำหนดโครงสร้างของข้อมูลต่อไปด้วย

และจากการศึกษามาทั้งหมดที่ได้กล่าวมาตั้งแต่ต้นนั้นเราพบว่า การกำหนดโครงสร้างรูปแบบให้กับข้อมูลจะเป็นหนทางที่จะทำให้เราสามารถที่จะกำหนดรูปแบบโครงสร้างของงานวิจัยให้เป็นมาตรฐานเดียวกันเพื่อให้สามารถทำการสืบค้นข้อมูลต่าง ๆ โดยอาศัยโปรแกรมที่ได้จัดทำขึ้นมีความสะดวกรวดเร็วขึ้น ดังนั้นเราจึงได้เลือกใช้ภาษา XML (eXtensible Marked-up Language) เป็นตัวกำหนดโครงสร้างพื้นฐานของข้อมูลต่าง ๆ โดยการกำหนดเครื่องหมาย Tag ต่าง ๆ ให้กับข้อมูลซึ่งมีการปรับเปลี่ยนการ Tag ข้อมูลหลายครั้ง เพื่อที่จะได้โครงสร้างของการ Tag ข้อมูลที่เหมาะสมและนำมาสร้างเป็นโครงสร้างที่ทำงานได้ตามจริงในการสืบค้นข้อมูลงานวิจัย ตามที่ได้แสดงโครงสร้างการ Tag ไว้ในบทที่ 2 และบทที่ 3 ไว้แล้วนั้น

นอกจากนี้ดังที่ได้กล่าวมาแล้วในหัวข้อปัญหาและอุปสรรคที่พบในการดำเนินงาน ยังทำให้การทำงานมีความล่าช้าในการวิเคราะห์โครงสร้างในแต่ละรายงานการวิจัย เนื่องจากการกำหนดโครงสร้างของงานแต่ละงานไม่มีความสอดคล้องกัน จะต้องมีการแยกส่วนการทำงานและมีการกำหนดโครงสร้างให้ด้วย ซึ่งทำให้มีความล่าช้าในส่วนนี้ ซึ่งจะต้องมีการ ตรวจสอบข้อมูลหลายรอบ เพื่อให้ได้ข้อมูลที่มีคุณภาพในการวิเคราะห์ต่อไป

จากการศึกษาและใช้งาน XML ทำให้เราสามารถที่จะทำงานโดยอาศัยโครงสร้างที่สร้างขึ้นมามีความสะดวกรวดเร็วและในเวลาเดียวกัน เราก็สามารถปรับเปลี่ยนโครงสร้างเพื่อให้มีความเหมาะสมกับงานที่ดำเนินการอยู่ได้ง่ายขึ้น แต่การดำเนินงานก็ต้องศึกษาภาษา XML ไปพร้อม ๆ กัน ซึ่งทำให้มีความรู้ความชำนาญในการเปลี่ยนแปลงต่าง ๆ มากยิ่งขึ้น ทำให้การดำเนินงานมีความรวดเร็วขึ้นในภายหลัง

ในอนาคตการที่จะทำให้ข้อมูลต่าง ๆ ที่เกี่ยวกับงานวิจัยมีความสามารถในการสืบค้นข้อมูลได้นั้น ทุกรายงานวิจัยจะต้องมีรูปแบบโครงสร้างของข้อมูลที่เหมือนกัน ดังเช่นงานวิจัยนี้ที่เราได้จัดสร้างโครงสร้างรูปแบบ การกำหนดไว้ แล้วนั้นจะทำให้การนำข้อมูลเข้าสู่โปรแกรมการสืบค้นมีความสะดวกและง่ายต่อการสืบค้นต่อไปในอนาคตได้

สรุปผลการดำเนินงานของการพัฒนาโปรแกรมสำหรับการดูข้อมูลจากหมืองข้อมูลงานวิจัย

โปรแกรมสามารถทำการแสดงผลในรูปแบบของ จำนวนตัวเลขความถี่และ กราฟ แต่ยังมีข้อจำกัดในด้านของปัญหาในการทำ Text Mining ดังปัญหาที่พบ ดังนี้ ปัญหาส่วนใหญ่จะเกิดจากข้อมูลเอกสารที่มีอยู่ ข้อมูลที่เก็บอยู่ใน Document Database จะเป็นข้อมูลที่มีลักษณะโครงสร้างไม่ชัดเจน (Semi-Structured Data) คือโครงสร้างข้อมูลที่ประกอบด้วย Structured data รวมอยู่ด้วยกันทำให้ ไม่ครบถ้วนสมบูรณ์ และความแตกต่างกันไปแต่ละเอกสารงานวิจัย ทำให้จำเป็นต้องมีการจัดการและตรวจสอบเอกสารข้อมูลที่ต้องทำการ tag XML ก่อน ซึ่งจะทำให้เสียเวลาในการตรวจสอบ ทั้งนี้เนื่องมาจากปัญหาทางด้านโครงสร้างของเอกสาร และความแตกต่างกันของเอกสาร

ด้านโครงสร้างของเอกสาร

เอกสารของงานวิจัยแต่ละงานวิจัยจะมีโครงสร้างที่แตกต่างกันไป บางเอกสารอาจจะมี ครบทุกๆ Tag แต่บางเอกสารจะมีเพียงแค่ Tag บางอันเท่านั้น จากตัวอย่างดังนี้

ด้านความแตกต่างกันของข้อมูล

ในบางเอกสาร ข้อมูลใน tag อาจจะมีการเขียนที่แตกต่างกัน เช่น คำว่า นางสาว ในบางงานวิจัยก็จะใช้คำว่า น.ส. ซึ่งทำให้โปรแกรมคำนวณค่าผิด หรือแม้แต่การสะกดคำผิด ก็จะทำให้โปรแกรมไม่สามารถแยกได้ว่าเป็นคำๆเดียวกัน เช่น คณะวิทยาศาสตร์ หากพิมพ์เป็น คณะวิดยาศาสตร์ ก็จะทำให้โปรแกรมคำนวณค่าเป็น คณะที่ต่างกัน

นอกจากปัญหาดังที่กล่าวมาข้างต้นแล้ว ในเรื่องของความหมายของคำก็มีส่วนช่วยทำให้การประมวลผลของคำสามารถทำได้ ถูกต้องมากยิ่งขึ้น กล่าวคือ ถ้าหากมีกระบวนการตัดคำ(Word Segmentation)เข้ามาช่วยในการวิเคราะห์และตรวจสอบคำที่อยู่ใน Tag XML ก็จะทำให้ การสรุปข้อมูลนั้นถูกต้องมากยิ่งขึ้น ทั้งนี้เนื่องมาจากว่าคำในภาษาไทยนั้นมีความซับซ้อน และ บางคำมีลักษณะพ้องรูปพ้องเสียง จึงให้ในบางครั้งถึงแม้ว่า จะได้ผลลัพธ์ออกมา เหมือนกัน แต่ความหมายก็อาจจะแตกต่างกันได้

เนื่องด้วยสาเหตุนี้จึงทำให้ต้องใช้เวลาในการจัดการและตรวจสอบข้อมูลที่มีอยู่ให้มีความถูกต้องก่อนจึงจะสามารถนำข้อมูลไปประมวลผลต่อไป

ข้อเสนอแนะ

โปรแกรมนี้ ยังคงมีข้อจำกัดในด้านของเอกสาร ซึ่งแนวทางในการพัฒนาต่อไป ถ้าหากมีการจัดการกับเอกสารเพื่อให้เอกสารงานวิจัยมีโครงสร้างรูปแบบที่แน่นอนและความถูกต้องมากยิ่งขึ้นก็จะช่วยให้โปรแกรมนี้มีประสิทธิภาพมากยิ่งขึ้น ซึ่งเมื่อข้อมูลมีความถูกต้องมากขึ้นแล้วก็สามารถที่จะนำเทคนิคที่เกี่ยวกับ data mining อื่นๆมาใช้ประมวลผลเพื่อทำให้เกิดองค์ความรู้ใหม่ๆจากเอกสารงานวิจัยที่มีอยู่ได้

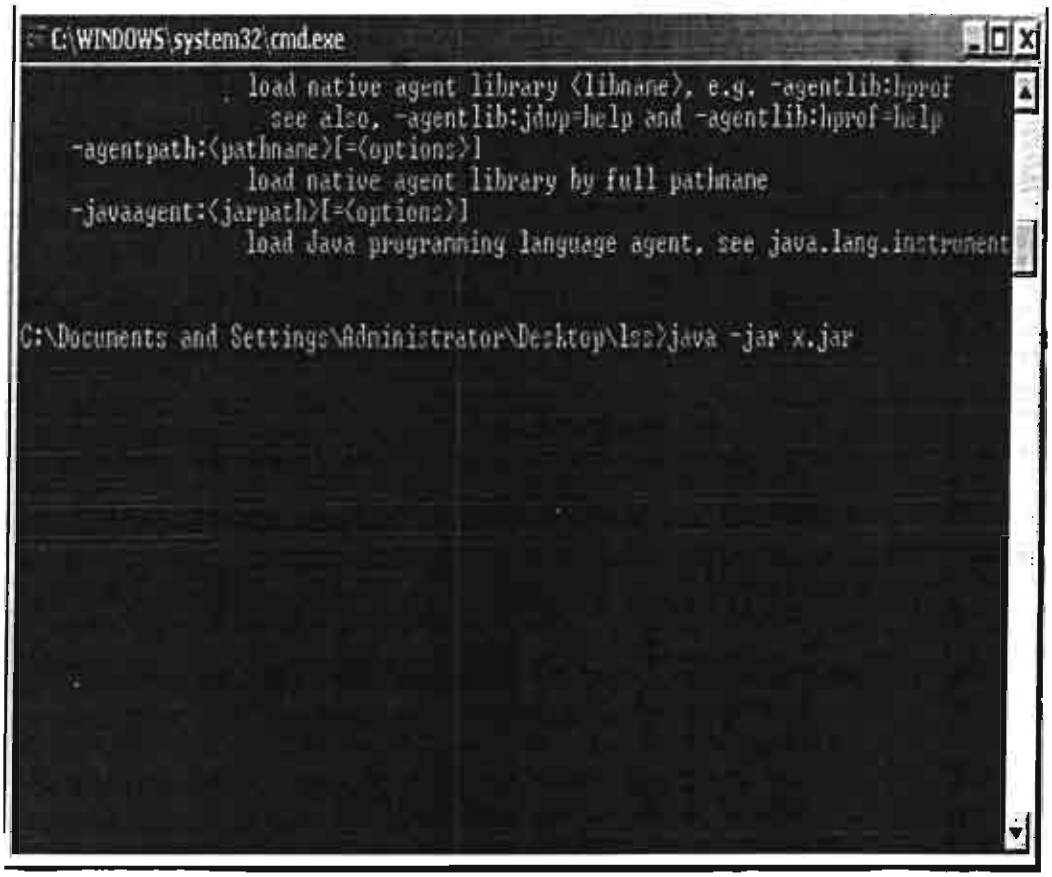
ภาคผนวก

คู่มือการใช้งาน

1 ลงโปรแกรม Java เพื่อใช้เป็นโปรแกรมที่จะแตกไฟล์จากโปรแกรม eXist

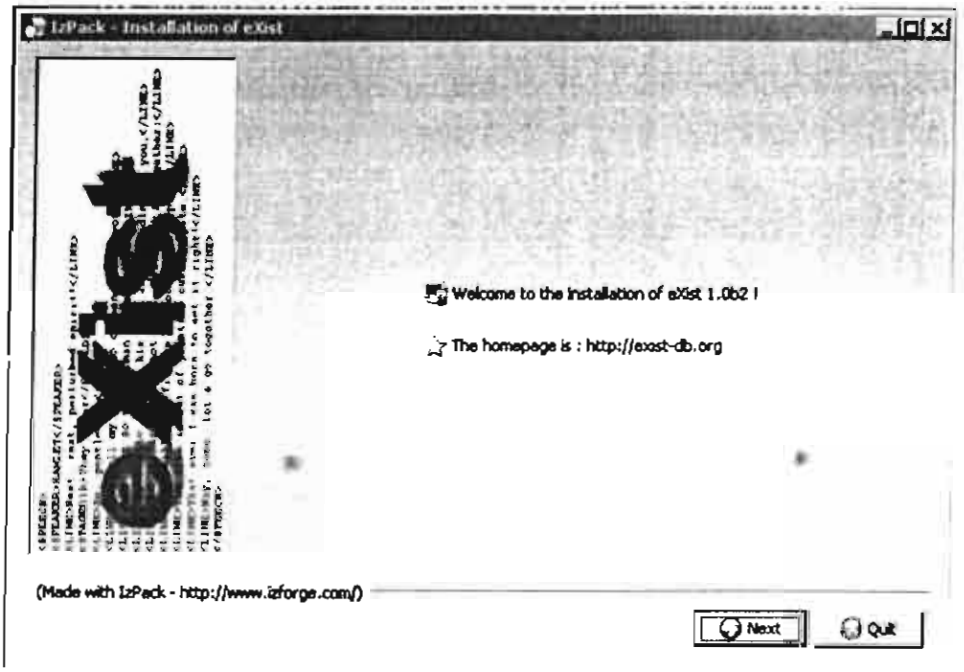
2. ลงโปรแกรม eXist

การลงโปรแกรม eXist เพื่อใช้เป็นโปรแกรมจัดการฐานข้อมูลที่จัดทำขึ้น การลงโปรแกรมมีขั้นตอนดังต่อไปนี้



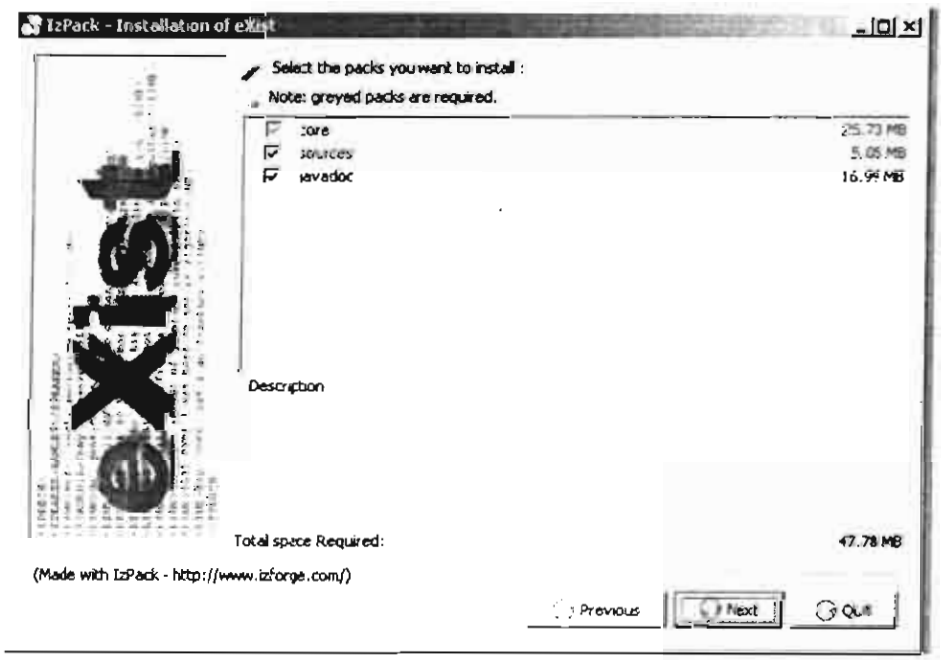
แสดงหน้าจอแสดงการเรียกใช้ Java เพื่อทำการแตกไฟล์ exist

เมื่อทำการติดตั้งโปรแกรม Java เรียบร้อยแล้วก็ทำการ run โปรแกรมเพื่อทำการแตกไฟล์ eXist หลังจากนั้นทำการติดตั้งโปรแกรม exist ดังรูป



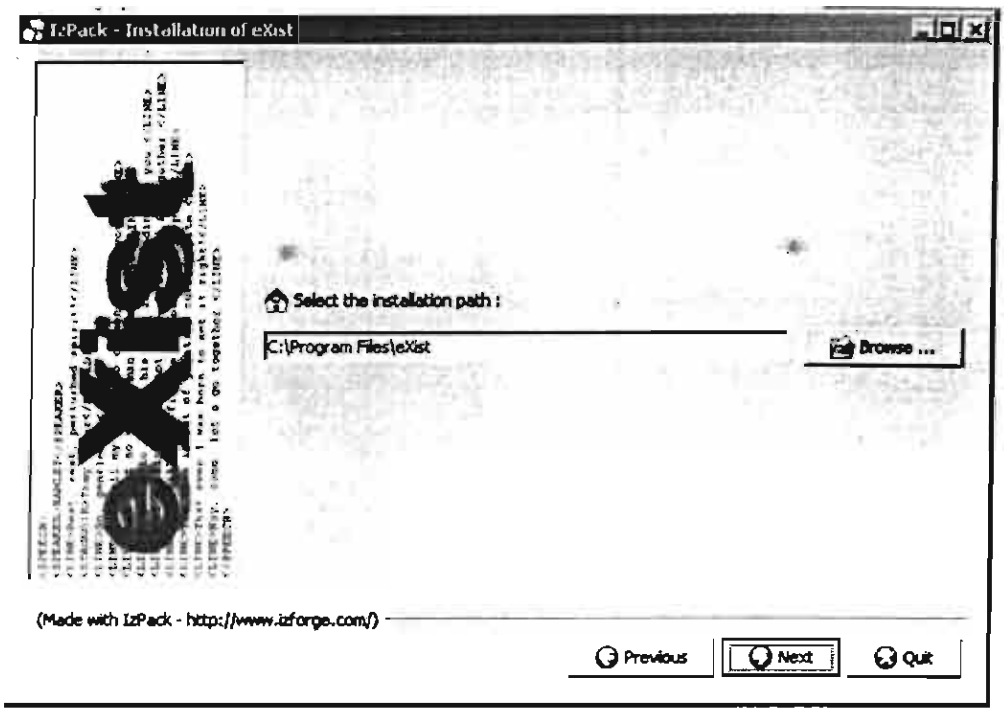
แสดงการแตกไฟล์โปรแกรม exist ในการติดตั้ง

เมื่อขึ้นหน้าจอการติดตั้งให้กดปุ่ม Next เพื่อทำงานต่อ หลังจากนั้นจะปรากฏรายละเอียดการลง ดังรูป



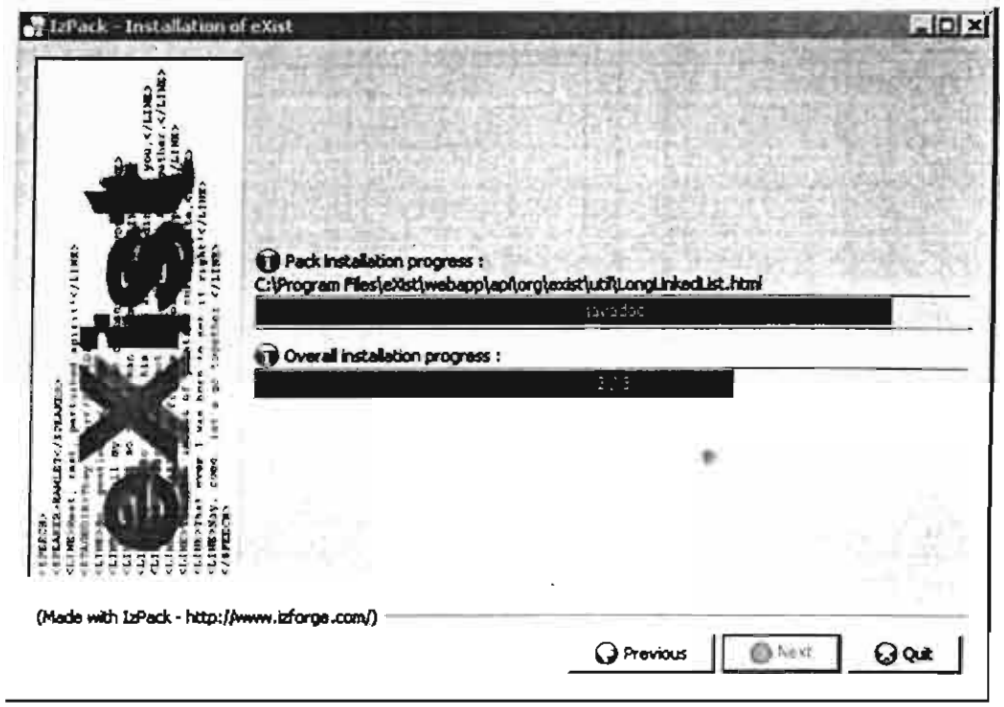
แสดงรายละเอียดการลงโปรแกรม

เมื่อปรากฏหน้าจอรายละเอียดการลง จะถามว่าจะลงอะไรบ้าง ซึ่งแล้วแต่จะเลือกว่าจะเอาอะไรบ้าง ซึ่งโดยมากจะเลือกทั้งหมดเพราะเกี่ยวกับการใช้งานตัวโปรแกรม เสร็จแล้วกดปุ่ม Next เพื่อทำงานต่อไป ซึ่งจะเป็นหน้าจอแสดงการเลือกไดเร็กทรอรี่ ที่จะติดตั้งโปรแกรม ดังรูป



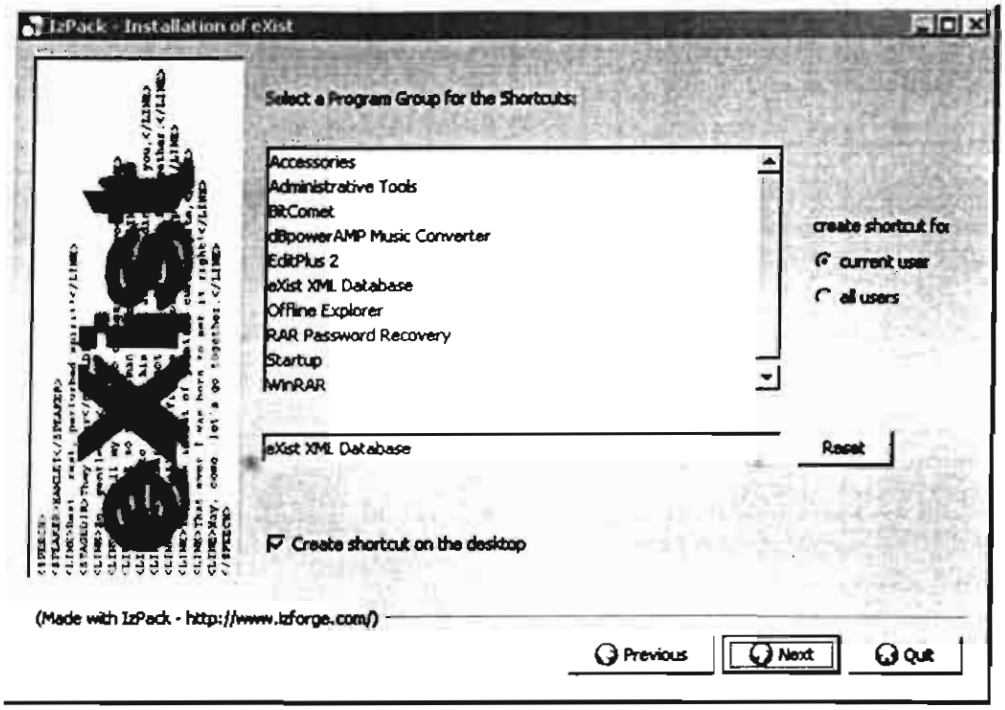
แสดงการเลือกไดเร็กทรอรี่ ที่จะติดตั้ง

การเลือกไดเร็กทรอรี่ ที่จะติดตั้ง โดยปกติแล้วระบบจะเลือกให้ที่ไดร์ C ในตัว Program File และลงใน Exist เป็นหลัก แต่ถ้าต้องการเปลี่ยนก็สามารถเลือกไปที่ปุ่ม Browse... ซึ่งสามารถเลือกไดเร็กทรอรี่ที่ต้องการลงได้ เมื่อทำการเลือกเสร็จก็กดปุ่ม Next เพื่อทำงานต่อไป



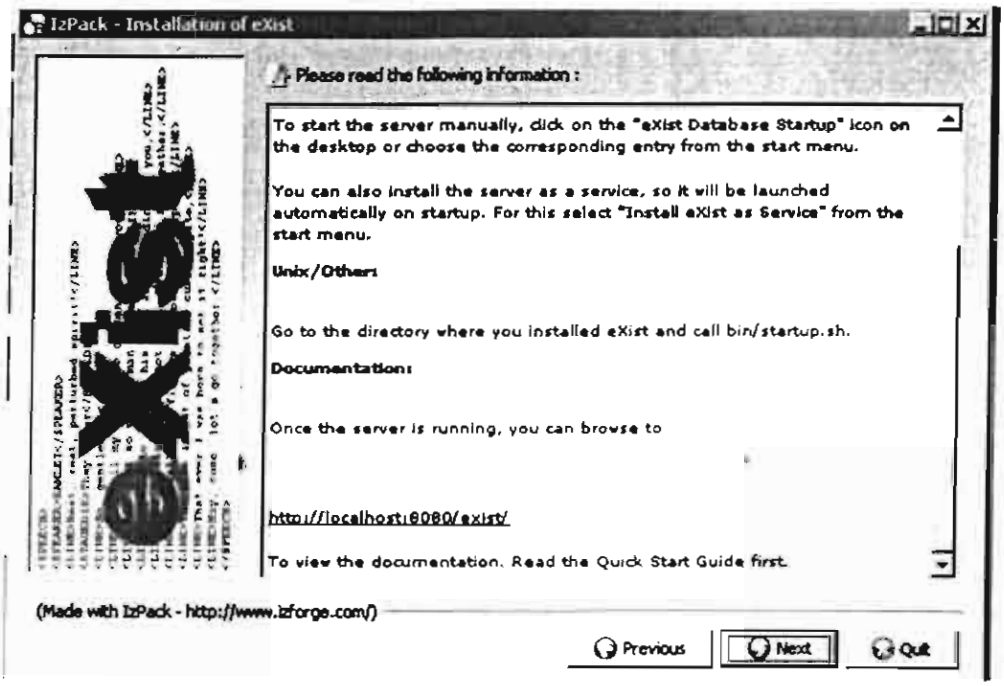
แสดงการติดตั้งระบบ

จะแสดงขั้นตอนการลงโปรแกรมซึ่งระบบจะทำงานโดยอัตโนมัติในการติดตั้งส่วนนี้ แต่ถ้าต้องการหยุดการติดตั้งก็สามารถกดปุ่ม Quit เพื่อทำการยกเลิกการลงโปรแกรม ได้ และเมื่อต้องการย้อนกลับเพื่อไปทำการปรับเปลี่ยนข้อมูลต่าง ๆ ก็สามารถกดปุ่ม Previous ก็จะย้อนกลับไปให้ 1 หน้าจอเพื่อทำการปรับแก้ไขข้อมูล แต่ถ้าไม่ต้องการปรับเปลี่ยนใด ก็รอจนกว่าปุ่ม Next จะปรากฏบนหน้าจอ แสดงว่าระบบได้ลงโปรแกรมให้เรียบร้อยดังรูป



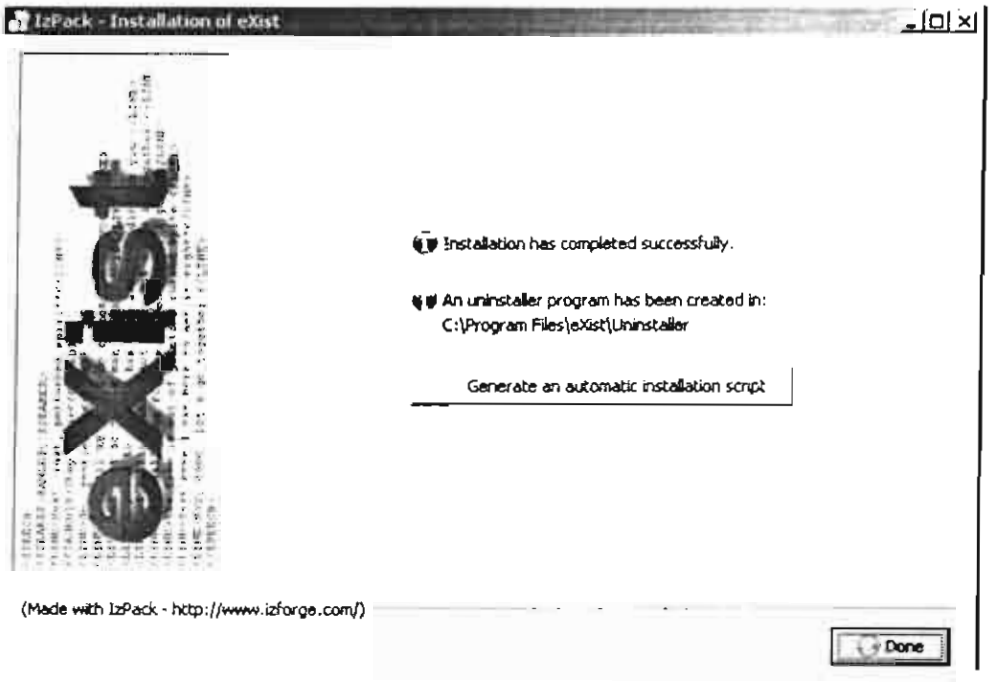
แสดงหน้าจอการทำ Shortcut

เมื่อระบบทำการลงโปรแกรมเรียบร้อยแล้วจะมีหน้าจอถามเพื่อให้ทำ Shortcut หรือไม่ ซึ่งจะปรากฏเครื่องหมาย ถูก ในช่อง Create shortcut on the desktop ถ้าไม่ต้องการทำให้คลิกในช่องสี่เหลี่ยมเพื่อให้เครื่องหมายถูกหายไป ก็จะไม่ปรากฏ shortcut บนหน้าจอคอมพิวเตอร์ เมื่อทำการเลือกเสร็จก็กดปุ่ม Next เพื่อทำงานต่อไป



แสดงรายละเอียดของโปรแกรม

เป็นหน้าจอที่แสดงรายละเอียดต่าง ๆ ของโปรแกรม หน้าจอนี้เราสามารถกดปุ่ม Next เพื่อทำงานต่อไปได้เลย



แสดงหน้าจอการเสร็จสิ้นการติดตั้งโปรแกรม

เป็นหน้าจอที่แสดงว่าได้ทำการติดตั้งโปรแกรม exist เรียบร้อยแล้ว ซึ่งจะต้องกดปุ่ม Done เพื่อออกจากการติดตั้งโปรแกรม

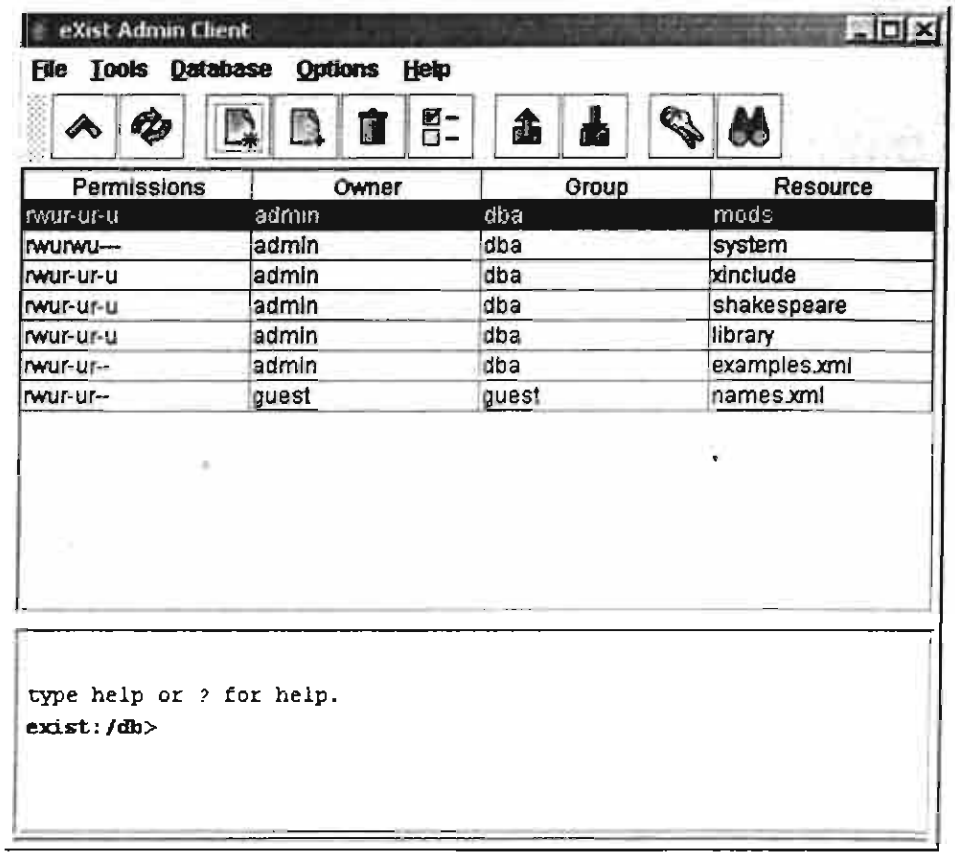
เมื่อออกจากการติดตั้งโปรแกรมแล้วจะปรากฏรูปของ eXist 3 รูปดังรูป

แสดง eXist Database Startup

เมื่อนำเมาส์มาคลิกที่ Shortcut นี้จะเป็นตัวที่เรียกใช้ฐานข้อมูลต่าง ๆ หรือที่เราเรียกว่า start database ในเครื่องเพื่อที่จะทำการนำข้อมูลต่าง ๆ มาเก็บไว้ในฐานข้อมูลต่อไป

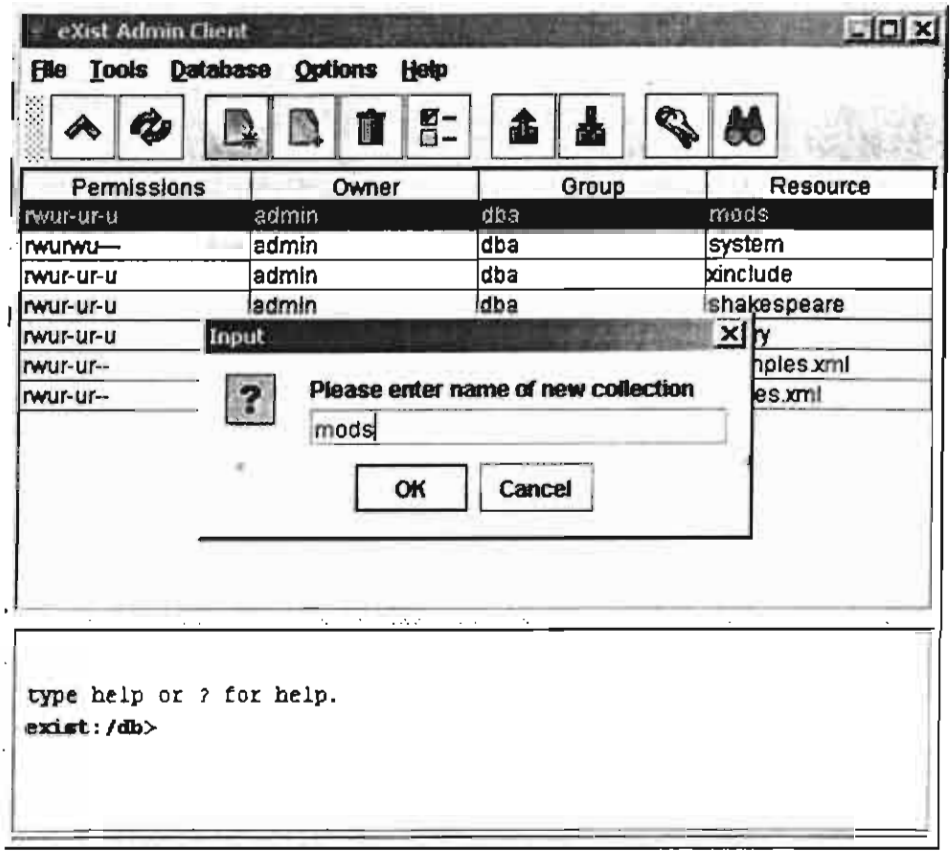
แสดง eXist Client Shell

เมื่อนำเมาส์มาคลิกที่ Shortcut นี้จะเป็นตัวที่จัดการเกี่ยวกับฐานข้อมูลที่จะนำเข้าในตัวโปรแกรม ดังรูป



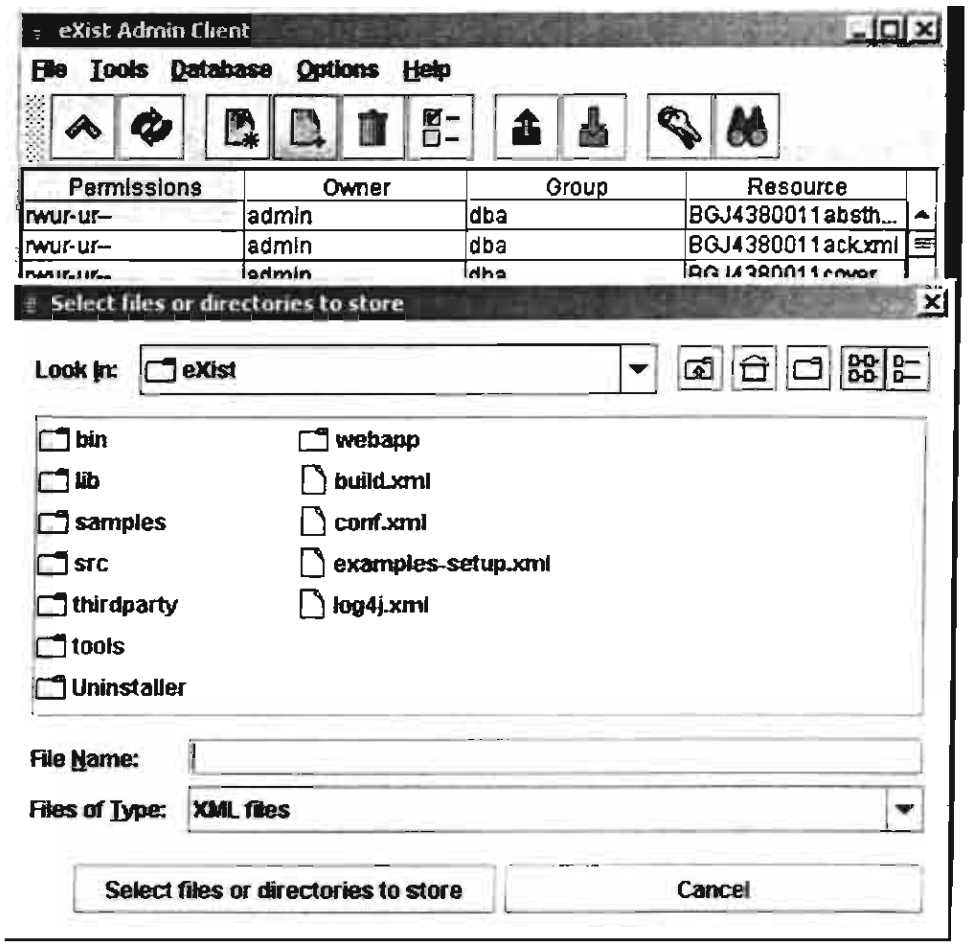
แสดงการนำข้อมูลเข้าสู่ฐานข้อมูล

การนำข้อมูลเข้าสู่ฐานข้อมูลนั้นเราสามารถทำได้โดยการนำเมาส์ไปคลิกที่ปุ่มสัญลักษณ์เป็นกระดาษและเครื่องหมายบวกด้านล่างขวา เมื่อเลือกข้อมูลได้แล้วก็จะปรากฏหน้าจอให้เลือก collection สำหรับเก็บข้อมูลดังรูป



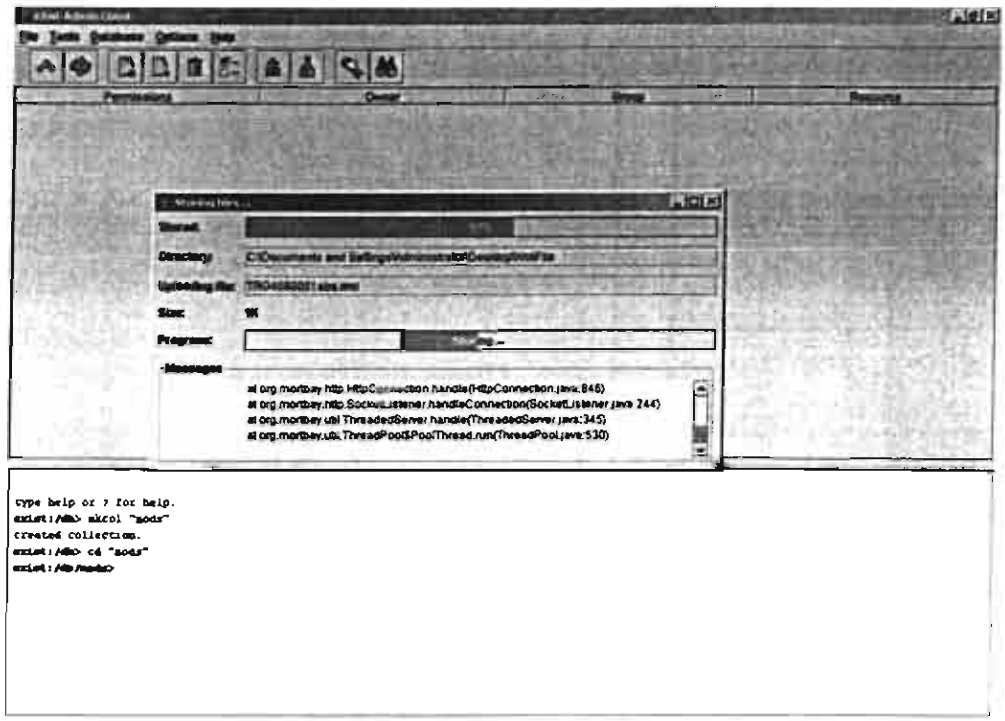
แสดงการเลือก collection ให้กับข้อมูล

ในหน้าจอนี้จะเป็นการเลือก collection ให้กับตัวข้อมูลที่ต้องการจะนำเข้าสู่ฐานข้อมูล เมื่อใส่ชื่อได้แล้วก็ทำการกดปุ่ม OK เพื่อทำงานต่อไป ดังรูป



แสดงการเลือกไฟล์ข้อมูลที่ต้องการจัดเก็บไว้ในฐานข้อมูล

ในหน้าจอนี้จะเป็นการเลือกไฟล์ข้อมูลที่ต้องการจัดเก็บไว้ในฐานข้อมูลโดยเราต้องเลือกที่อยู่ของไฟล์ข้อมูลของเราเอง ในช่อง Look In ซึ่งไฟล์ที่เราเลือกจะต้องเป็นไฟล์นามสกุล XML ที่เราได้สร้างขึ้นมาเท่านั้น หลังจากเลือกไฟล์ได้แล้วก็กดปุ่ม Select files or directories to store เพื่อให้โปรแกรมทำการโหลดข้อมูลในไฟล์ไปเก็บไว้ในฐานข้อมูลให้กับเราใน collection ที่เราได้เลือกไว้ ดังรูป



แสดงการโหลดไฟล์ข้อมูลที่ต้องการจัดเก็บไว้ในฐานข้อมูล

เมื่อทำการโหลดข้อมูลเรียบร้อยแล้วก็จะเข้าสู่หน้าจอ eXist Local Homepage เพื่อทำการเปิดดูข้อมูลที่เราได้ทำการเลือกไฟล์ต่าง ๆ เข้าสู่ฐานข้อมูล โดยจะคลิกที่ shortcut eXist Local Homepage ดังรูป

แสดง eXist Local Homepage

แสดงหน้าจอการเข้าใช้ shortcut eXist Local Homepage

ในหน้าจอนี้เมื่อเราคลิกที่ shortcut eXist Local Homepage จะปรากฏการเข้าใช้งาน eXist ให้คลิกที่คำว่า Main ซึ่งจะปรากฏหน้าจอดังต่อไปนี้

แสดงหน้าจอการค้นหาข้อมูลต่าง ๆ

ในหน้าจอนี้จะเป็นหน้าจอแรกในการใช้งานโปรแกรมที่ได้จัดทำขึ้น จะแสดงครั้งแรกเมื่อเปิดใช้โปรแกรม ซึ่งจะมีการถามโดยให้ผู้ใช้สามารถเลือกว่าจะเลือกดูข้อมูลจากส่วนใด เช่น จาก Title, author , keyword ซึ่งเมื่อเลือกอย่างใดอย่างหนึ่งแล้วก็จะทำการใส่ข้อมูลลงในช่องต่อมาว่าจะใส่ข้อความอะไรเพื่อให้สอดคล้องกับสิ่งที่ต้องการค้นหา เช่นตัวอย่างเป็นการค้นหาข้อมูลจากคำว่า Title ในช่อง Search in และคำสำคัญคือ แม่ฮ่องสอน ในช่อง Search what และในบรรทัดต่อมาจะเป็นการเลือกว่าจะให้แสดงข้อมูลที่พบในลักษณะไหนบ้าง

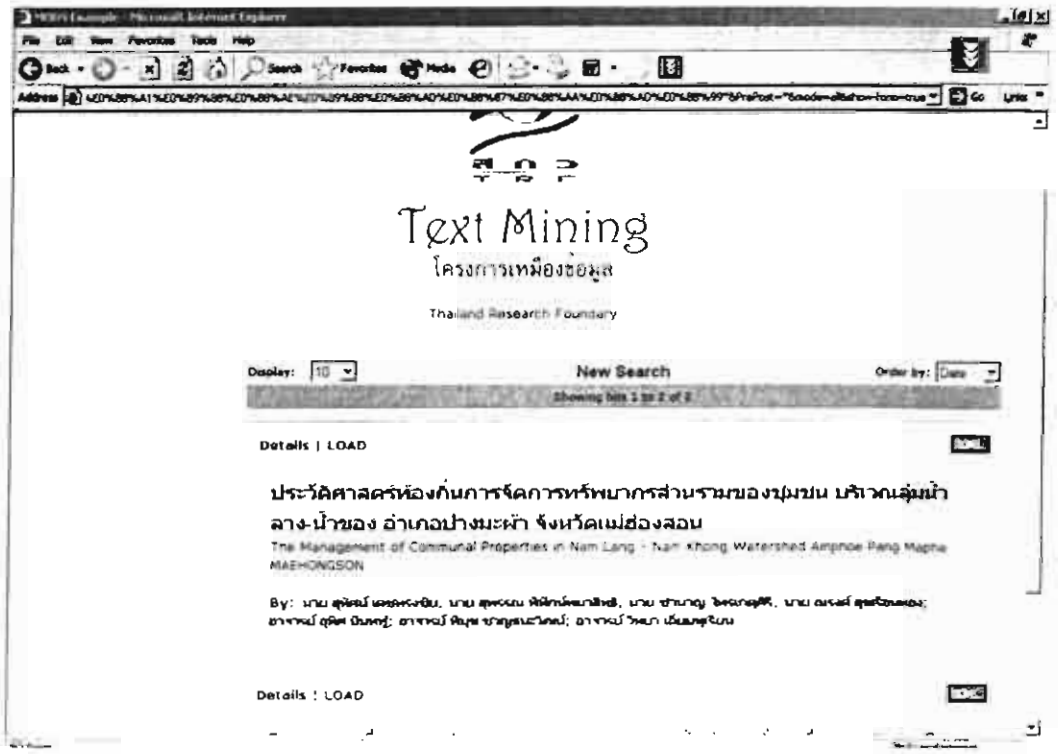
All terms เป็นการแสดงรายการทั้งหมดที่ค้นพบ

Any term เป็นการแสดงเฉพาะที่ค้นหา

All Terms near เป็นการแสดงรายการทั้งหมดและที่ใกล้เคียงกัน ซึ่งเราสามารถเติมคำที่ต้องการค้นหาได้ในช่องที่ 2 ทางขวามือ เช่นเราต้องการค้นหาข้อความงานวิจัยที่ขึ้นต้น (title) ว่า แม่ฮ่องสอน ก็จะทำตามขั้นตอนดังรูป

แสดงหน้าจอการค้นหาข้อมูลแม่ฮ่องสอน

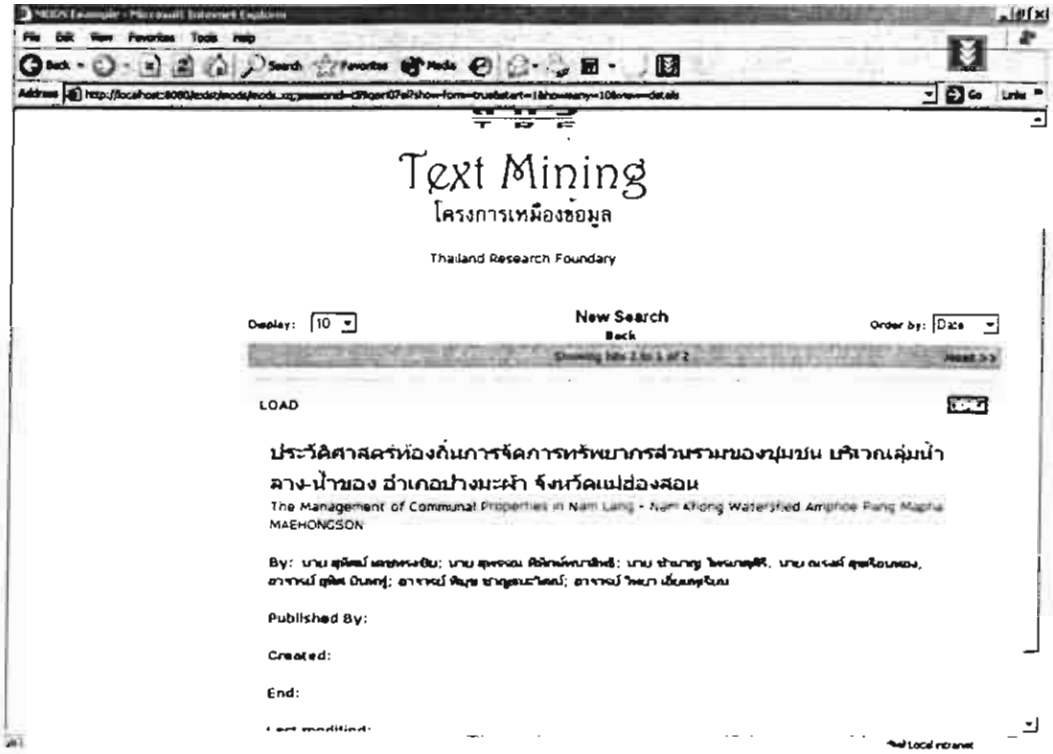
ในหน้าจอการค้นหาข้อมูลเราก็จะพิมพ์คำว่าแม่ฮ่องสอนลงไป หลังจากนั้นก็เลือกรูปแบบการแสดงผล แล้วทำการกดปุ่ม Submit เพื่อให้โปรแกรมทำการประมวลผลข้อมูลให้



แสดงหน้าจอการค้นหาข้อมูลที่เลือก

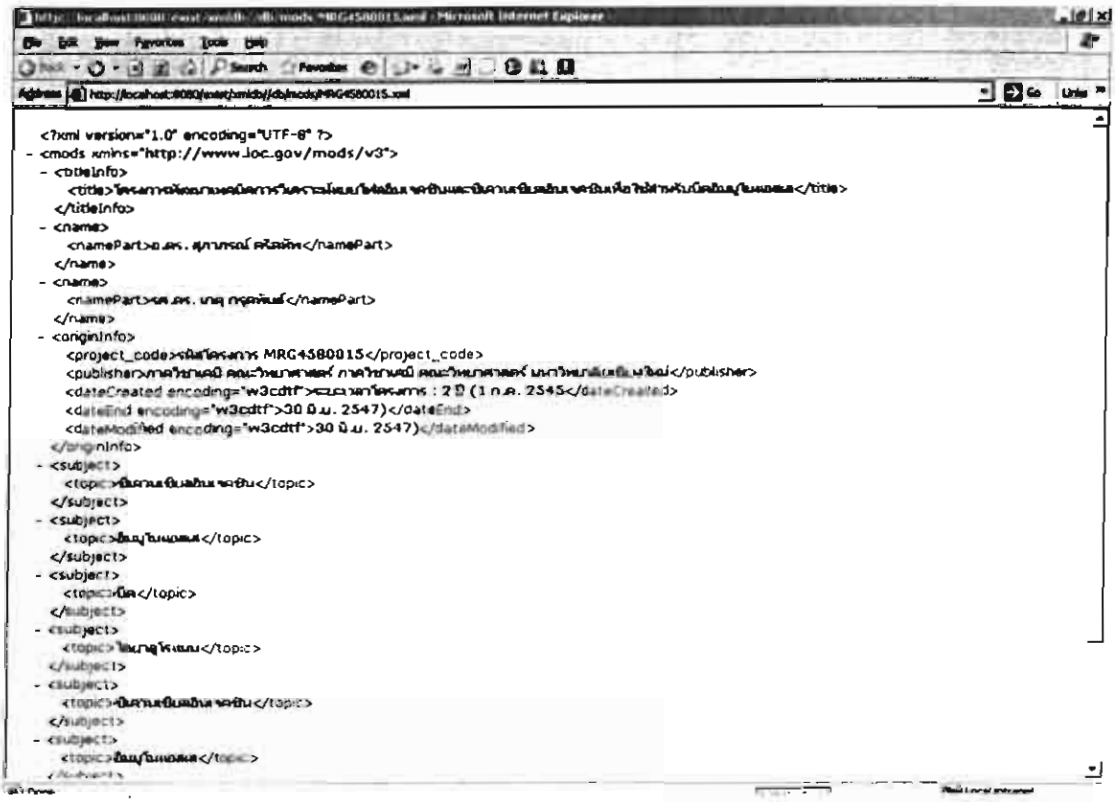
ในส่วนนี้จะเป็นการแสดงผลการเลือกข้อมูลที่ได้กรอกไว้ ซึ่งจะโปรแกรมจะบอกว่ามีข้อมูลทั้งหมด 2 ข้อมูล ในส่วนที่ค้นหา ซึ่งในหน้าจอเราจะพบว่า ในส่วนของ display นั้นเราสามารถกำหนดได้ว่าจะให้หน้าจอเราแสดงผลการค้นหาข้อมูลกี่รายการซึ่งเราจะพบว่าได้กำหนดไว้ที่ 10 รายการเป็นหลัก แต่ถ้าเราต้องการจะเปลี่ยนเป็น 5 รายการก็เพียงแค่นำเมาส์ไปคลิกแล้วเปลี่ยนเป็นเลข 5 ระบบก็จะทำการเปลี่ยน ให้เป็น 5 รายการต่อ 1 หน้าจอ และอีกส่วนหนึ่ง คือ Order by จะเป็นการกำหนดให้ข้อมูลมีการเรียงลำดับแบบไหนบ้าง ซึ่งในหน้าจอเรากำหนดให้เรียงโดยตาม Date , Title , Author ซึ่งผู้ใช้สามารถกำหนดส่วนนี้ได้เองตามความต้องการ

เมื่อทำการกำหนดสิ่งต่าง ๆ ได้แล้วระบบจะทำการประมวลผลและแสดงผลการค้นหาให้ตามที่ผู้ใช้กำหนดไว้ ซึ่งในกรณีที่เราต้องการจะเข้าไปดูรายละเอียดของงานวิจัยที่เราค้นหาในแต่ละรายการนั้นเราก็สามารถคลิกเลือกที่ Details ก็จะปรากฏรายละเอียดของโครงการให้



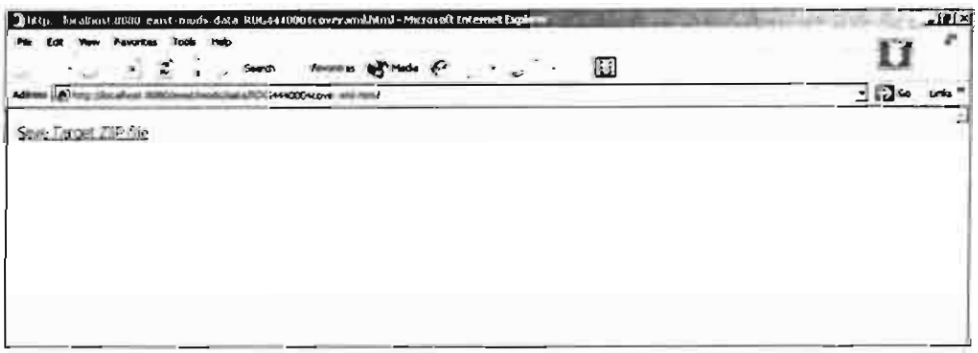
แสดงหน้าจอรายละเอียดข้อมูล

ในส่วนนี้จะเป็นการแสดงรายละเอียดของงานวิจัยที่เราเลือกดูข้อมูล ซึ่งจะแสดงรายการต่าง ๆ ให้ว่าใครเป็นผู้ทำวิจัย หัวข้อเรื่องวิจัยอะไร มีรหัสโครงการอะไร ระยะเวลาโครงการ ฯลฯ ซึ่งรายละเอียดในส่วนนี้เราสามารถเข้าไปดูและทำการคัดลอกข้อมูลรายงานการวิจัยได้ในส่วนที่เป็นข้อมูล XML ได้ที่ปุ่ม XML ก็จะแสดงข้อมูลต่าง ๆ ของการ Tag ข้อมูลดังรูป



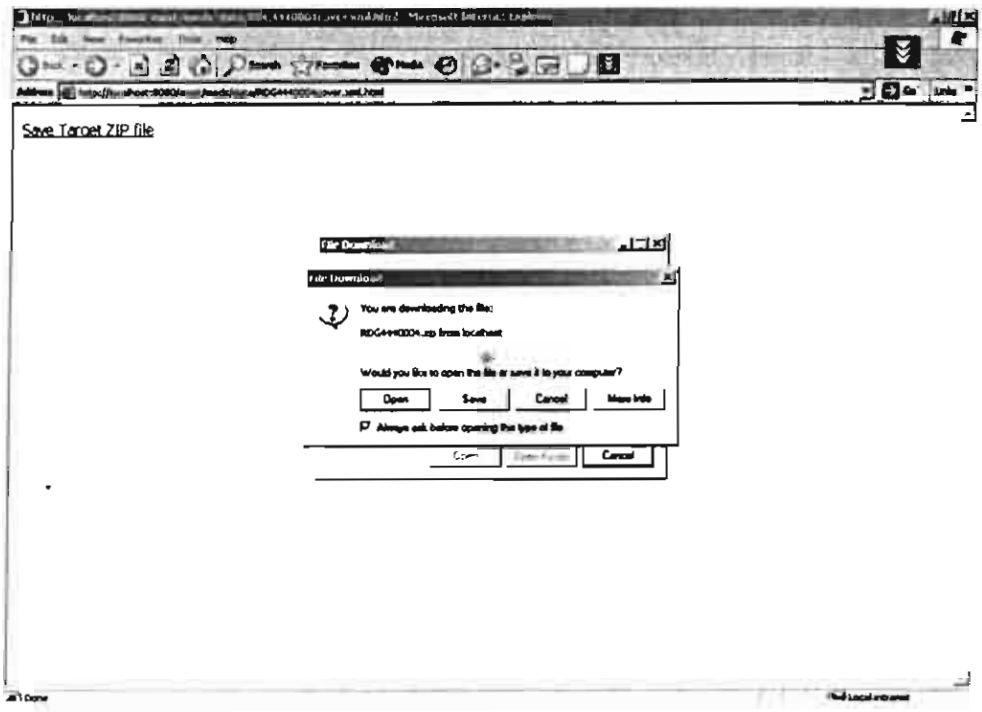
แสดงการ Tag ข้อมูล

จะแสดงการ Tag ข้อมูลที่ได้เข้าไปดูรายละเอียดโครงสร้างของข้อมูล แต่ถ้าต้องการที่จะทำการโหลดข้อมูลทั้งงานวิจัยก็จะทำการเลือกที่ปุ่ม Load เพื่อทำการโหลดข้อมูลดังรูป



แสดงหน้าจอการเข้าโหลดข้อมูลงานวิจัย

ในหน้าจอนี้เมื่อเราทำการคลิกที่ Save Target ZIP file จะปรากฏหน้าจอการโหลดข้อมูลดังรูป



แสดงการsave ข้อมูลที่ทำการเลือกไฟล์

ในหน้าจอนี้จะปรากฏให้เราเลือกว่าเราจะทำการบันทึกข้อมูลหรือว่าจะเปิดข้อมูลขึ้นมาดู ซึ่งถ้าเราต้องการบันทึกข้อมูลเก็บไว้ดูเราก็คลิกที่ปุ่ม Save เพื่อบันทึกข้อมูลได้เลย เมื่อทำการบันทึกเรียบร้อยระบบก็จะเสร็จสิ้นการทำงาน เช่นเดียวกับการโหลดข้อมูลทางอินเทอร์เน็ตทั่วๆ ไป

บรรณานุกรม

[1] นววรรณ พันธุเมธา. ไวยากรณ์ไทย . กรุงเทพฯ 2527.

[2] พระยาอุปกิตศิลปสาร. หลักภาษาไทย.ไทยวัฒนาพานิช ,กรุงเทพฯ 2541 .

[3] เรืองเดช ปันเขื่อนขัติย์. ภาษาศาสตร์ภาษาไทย. สถาบันวิจัยภาษาและวัฒนธรรมเพื่อพัฒนาชนบท , มหาวิทยาลัยมหิดล, กรุงเทพฯ 2540.

[4] เอกพน จิรังสุวรรณ,สมนึก คีรีโต . การวิเคราะห์และเปรียบเทียบภาษาสืบค้นสำหรับ XML ระหว่าง Xquery และ XSLT . มหาวิทยาลัยเกษตรศาสตร์.

[5] A. Deutsch, M. Fernadez, D. Florescu, A. Levy, and D. Suciu, A Query Language for XML. http://www.research.att.com/~mff/files/final.html.

[6] A Tutorial on Clustering Algorithms. http://www.elet.polimi.it/upload/matteucc/Culstering/tutorial_thml/index.html.

[7] Alin Deutsch, Mary Fernandez , Daniela Florescu. A Quer Language for XML. http://research.att.com/~mff/files/final.html.

[8] J.Robie,XQL (XML Query Language), August 1999. http://metalab.unc.edu/xql/xql-proposal.html.

[9] Machomsomboon, T,. and Opasanont, B. 2000. Thai Word Segmentation without a Dictionary by Using Decision Trees. The fourth Symposium on Natural Language Processing 2000.

[10] Navavan Bandhumedha."Noun Phrase Deletion in thai." Unpublished Ph.D. dissertation, University of Washington, 1976.

[11] Priscilla Walmsley. Introduction to Xquery.2004. http://wwwdatypic.com

[12] Queries on XML documents. http://www.brics.dk/~amoeller/XML/querying/queries.html.

[13] Somlertlamvanich, V. 1993. Word Segmentation for Thai in a Machine Translation System. NECTEC. (in Thai).

[14] Tariq Rashid : "Clustering". http://www.cs.bris.ac.uk/home/tr1690/documentation/fuzzy_clustering_initial_re.

[15] Voula Giouli,Srelios Piperidis . Copopra and HLT , Current trends in corpus processing and annotation . http://www.larflast.bas.bg/balric/eng_files/corpora7_1.php .

[16] Wirote Aroonmanakun. Collocation and Thai Word Segmentation. Chulalongkorn University.XML-QL : A Query Language for XML . http://www.w3.org/TR/NOTE-xml-ql/.