เทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

An Automatic News Video Feed Filtering Technique for TCP/IP-based Mobile Devices

ชารวีย์ แสงขำ

Charawee Sangkhum

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
มหาวิทยาลัยสงขลานครินทร์

A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Computer Science
Prince of Songkla University
2552

ลิขสิทธิ์ของมหาวิทยาลัยสงขลานครินทร์

ชื่อวิทยานิพนธ์ เทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ผู้เขียน นางสาวชารวีย์ แสงขำ

สาขาวิชา วิทยาการคอมพิวเตอร์

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

…….....

(ดร.ลัดดา ปรีชาวีรกุล)

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

…….....

(ผู้ช่วยศาสตราจารย์ ดร.วิภาดา เวทย์ประสิทธิ์)

คณะกรรมการสอบ

……...ประธานกรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.โอม ศรนิล)

…….....กรรมการ

(ดร.ลัดดา ปรีชาวีรกุล)

…….....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.วิภาดา เวทย์ประสิทธิ์)

…….....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ศิริรัตน์ วณิชโยบล)

บัณฑิตวิทยาลัย มหาวิทยาลัยสงขลานครินทร์ อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษา ตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์

……..…….

(รองศาสตราจารย์ ดร.เกริกชัย ทองหนู)

คณบดีบัณฑิตวิทยาลัย

ชื่อวิทยานิพนธ์ เทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ผู้เขียน นางสาวชารวีย์ แสงขำ

สาขาวิชา วิทยาการคอมพิวเตอร์

ปีการศึกษา 2551

บทคัดย่อ

เทคโนโลยี Really Simple Syndication (RSS) ทำให้ผู้บริโภคสามารถติดตามการนำเสนอของข่าวใหม่ๆได้ นอกจากนี้อุปกรณ์สื่อสารเคลื่อนที่ซึ่งได้รับความนิยมเช่น PDA ในปัจจุบันสามารถเชื่อมต่ออินเตอร์เน็ตเพื่อรวบรวมข่าวต่างๆได้ แต่จะแสดงเป็นข้อความหรือรูปภาพเท่านั้น แม้ว่าเทคโนโลยี Podcast จะถูกพัฒนาขึ้นมาเพื่อตอบสนองให้ผู้ใช้สามารถรวบรวมข้อมูลมัลติมีเดียได้ก็ตาม ผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ก็ยังคงต้องติดตั้งซอฟต์แวร์สำหรับอ่าน Podcast (Podcast Reader) บนอุปกรณ์เคลื่อนที่ของตน งานวิทยานิพนธ์นี้จึงนำเสนอเทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP โดยพัฒนาระบบซึ่งสามารถแสดงเอกสาร RSS (RSS feed) ในรูปแบบวิดีโอข่าวได้ ระบบจะทำการดึงข่าวจากแหล่งข่าวหลายแหล่งมาคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมก่อนที่จะส่งไปแสดงผล อีกทั้งผู้ใช้สามารถรับชมวิดีโอข่าวผ่าน web-based RSS reader ซึ่งไม่ต้องติดตั้งก่อนใช้งานเพื่อลดการประมวลผล

Thesis Title	An Automatic News Video Feed Filtering Technique for TCP/IP-based Mobile Devices
Author	Miss Charawee Sangkhum
Major Program	Computer Science
Academic Year	2008

ABSTRACT

Really Simple Syndication (RSS) technology makes it for users possible to keep track of updated news. Podcasting is the technology based on RSS, aiming to aggregate multimedia files. For desktop users, viewing redundant RSS news video might not be a problem because of the high efficiency of the machine. However, for mobile users, viewing redundant news video in a device with limited resource like PDAs or Smart phones is wasting time, memory-space and battery-power. This thesis proposes an Automatic News Video Feed Filtering Technique for TCP/IP-based Mobile Devices. The system aggregates news video feeds from multiple sources and then filters an appropriate video to be shown on user's mobile device. To reduce the processing on mobile devices, users are able to access news video through a web-based reader.

สารบัญ

หน้า

สารบัญ..... (6)
รายการตาราง..... (8)
รายการภาพประกอบ..... (9)
บทที่
1 บทนำ
1.1 การตรวจเอกสาร 2
1.2 วัตถุประสงค์ของโครงการ..... 3
1.3 ขอบเขตของการดำเนินงาน..... 3
1.4 ขั้นตอนและระยะเวลาการดำเนินงาน 4
1.5 สถานที่และเครื่องมือที่ใช้ 5
1.6 ประโยชน์ที่คาดว่าจะได้รับ..... 6
2 ทฤษฎีที่เกี่ยวข้อง
2.1 สถาปัตยกรรมระบบแบบไคลเอนท์-เซิร์ฟเวอร์ (client-server)..... 7
2.2 ภาษา XML (Extensible Markup Language)..... 9
2.2.1 มาตรฐานต่างๆ ที่เกี่ยวข้องกับ XML..... 10
2.2.2 องค์ประกอบและโครงสร้างของ XML..... 10
2.2.3 การนิยามโครงสร้างเอกสาร XML..... 11
2.2.3.1 Data Type Definition (DTD)..... 12
2.2.3.2 XML Schema..... 14
2.2.4 การแสดงผลของเอกสาร XML..... 15
2.3 XML Parser..... 17
2.3.1 การวิเคราะห์ข้อมูลแบบ Object-based..... 17
2.3.2 การวิเคราะห์ข้อมูลแบบ Event-based..... 18
2.4 เทคโนโลยี Really Simple Syndication (RSS)..... 19
2.5 เทคโนโลยี Podcast..... 23
3. แบบจำลองเทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสาร
เคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

สารบัญ (ต่อ)

หน้า

3.1 แบบจำลองของระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP (An Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices : ANVFFS)..... 29

3.1.1 สถาปัตยกรรมทั่วไป..... 29

3.1.2 ฐานข้อมูลของระบบ..... 28

3.2 กลไกการคัดกรองวิดีโอข่าวที่เหมาะสม..... 32

3.2.1 ขั้นตอนวิธีของ parser..... 34

3.2.2 ขั้นตอนวิธีของการคัดกรองวิดีโอที่เหมาะสม..... 35

4. โปรแกรมคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

4.1 ส่วนของผู้ดูแลระบบ..... 38

4.1.1 การทำงานในส่วนของผู้ดูแลระบบ..... 39

4.1.2 ส่วนแสดงผลของผู้ดูแลระบบ..... 40

4.2 ส่วนของผู้ใช้งานอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP..... 42

4.2.1 การทำงานในส่วนของผู้ใช้..... 42

4.2.2 ส่วนแสดงผลของผู้ใช้..... 45

5. ผลการทดลอง

5.1 ส่วนของผู้ดูแลระบบ..... 47

5.2 การทำงานในส่วนของผู้ใช้ 53

6. บทสรุปและข้อเสนอแนะ

6.1 สรุปผลการวิจัย 58

6.2 ปัญหาและอุปสรรค 59

6.3 ข้อเสนอแนะ 59

บรรณานุกรม 61

ภาคผนวก 64

ก ผลงานตีพิมพ์ในการประชุมวิชาการ JCSSE 2008 65

ข ผลงานตีพิมพ์ในการประชุมวิชาการ ICCEE 2008 74

ประวัติผู้เขียน 80

รายการตาราง

ตาราง		หน้า
1.1	แสดงแผนการดำเนินการวิจัย	5
2.1	ค่าตั้งต้นของการนิยามแอททริบิวต์ใน DTD	13
2.2	การเปรียบเทียบระหว่างวิธีการวิเคราะห์ข้อมูล XML แบบ Object-based และ Event-based	19
2.3	sub-tag ของ <channel>	21
2.4	sub-tag ของ <item>	22
2.5	tag ใน iTunes RSS	25
3.1	แอททริบิวต์ของตาราง Channel	30
3.2	แอททริบิวต์ของตาราง Item	31
3.3	แอททริบิวต์ของตาราง SearchResult	32

รายการภาพประกอบ

ภาพประกอบ		หน้า
2.1	สถาปัตยกรรมแบบไคลเอนท์-เซิร์ฟเวอร์	7
2.2	สถาปัตยกรรมของพร็อกซี่เซิร์ฟเวอร์	8
2.3	โครงสร้างของ XML	11
2.4	การประกาศเอกสาร XML	11
2.5	ตัวอย่างการนิยามโครงสร้างของอิลิเมนต์ใน DTD	12
2.6	ตัวอย่างการนิยามโครงสร้างของแอททริบิวต์ใน DTD	13
2.7	ตัวอย่างการนิยามโครงสร้างของเอนติตี้ใน DTD	13
2.8	ตัวอย่างเอกสาร XML	14
2.9	XML Schema ที่เกี่ยวข้อง	14
2.10	ตัวอย่างการนิยามโครงสร้างของ annotation ใน XML Schema	15
2.11	ลักษณะการทำงานของ XML Parser	17
2.12	แสดงลักษณะการวิเคราะห์ข้อมูลแบบ Object-based	18
2.13	แสดงลักษณะการวิเคราะห์ข้อมูลแบบ Event-based	18
2.14	ตัวอย่างของเอกสาร RSS	23
2.15	รูปแบบโครงสร้างของ RSS enclosure	24
2.16	ตัวอย่าง RSS feed ที่ใช้งาน tag ของ iTunes	26
3.1	สถาปัตยกรรมทั่วไปของระบบ ANVFFS	28
3.2	โครงสร้างฐานข้อมูลระบบ	29
3.3	ลำดับขั้นตอนกลไกการคัดกรองวิดีโอข่าวที่เหมาะสม	33
3.4	ขั้นตอนวิธีของ parser	35
3.5	ขั้นตอนวิธีคัดกรองวิดีโอข่าวที่เหมาะสม	36
4.1	ลักษณะการแสดงผลด้วย AJAX	37
4.2	สถิติการอัพเดทเอกสาร RSS บนเว็บไซต์	38
4.3	แสดงผังการทำงานในส่วนของผู้ดูแลระบบ	39
4.4	แสดงผังการทำงานในส่วนของ parser	40
4.5	หน้าต่างยืนยันตัวตนในส่วนของผู้ดูแลระบบ	41

รายการภาพประกอบ (ต่อ)

ภาพประกอบ		หน้า
4.6	หน้าเว็บในส่วนของผู้ดูแลระบบ	41
4.7	แสดงผังการทำงานในส่วนของผู้ใช้	43
4.8	แสดงผังการทำงานในส่วนของผู้ใช้	44
4.9	หน้าเว็บในส่วนของผู้ใช้แสดงผลบนเครื่องคอมพิวเตอร์ทั่วไป	45
4.10	หน้าเว็บในส่วนของผู้ใช้แสดงผลบนอีมูเลเตอร์อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP	46
5.1	การแสดงสถานะการรวบรวมเอกสาร RSS	48
5.2	ผลการรวบรวมเอกสาร RSS	48
5.3	ข้อมูลของที่สกัดจาก tag <channel> ของเว็บไซต์ต่างๆ	50
5.4	ข้อมูลของที่สกัดจาก tag <item> ของเว็บไซต์ต่างๆ	51
5.5	อัตราของข้อมูลในเอกสาร RSS	52
5.6	ผลการอ่านเอกสาร RSS จากเว็บไซต์ที่ให้บริการ Podcast บนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP	53
5.7	ข้อมูลที่ได้จากการอ่านเอกสาร RSS	54
5.8	หน้าเว็บในส่วนของการรับคำสำคัญและคุณสมบัติการคัดกรองจากผู้ใช้	55
5.9	ข้อมูลที่ได้จากการคัดกรองด้วยคำสำคัญ	55
5.10	ผลการแสดงวิดีโอที่เหมาะสมบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP	56
5.11	การบันทึกวิดีโอข่าวที่คัดกรองได้ลงบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP	57
5.12	การเรียกชมวิดีโอข่าวผ่านโปรแกรมเล่นไฟล์มัลติมีเดีย	57

บทที่ 1

บทนำ

รูปแบบของการรับข้อมูลบนอินเตอร์เน็ตปัจจุบันมีการปรับเปลี่ยนเป็นข้อมูลแบบมัลติมีเดียมากขึ้น เว็บไซต์ผู้ให้บริการข่าวก็เริ่มมีการเผยแพร่ข่าวรูปแบบมัลติมีเดีย เช่น ไฟล์การรายงานข่าวในรูปแบบเสียง และ วิดีโอข่าว และได้มีการจัดทำเป็นรูปแบบของ RSS เรียกว่า Podcast เพื่อให้ผู้ใช้สามารถรับรู้ข่าวสารที่มีการนำเสนอใหม่ได้ ตัวอย่างเว็บไซต์ของแหล่งข่าวที่มีการจัดทำ RSS สำหรับวิดีโอข่าว เช่น ABC (ABC News, 2009) The Washington Post (The Washington Post, 2009) และ Reuters (Reuters, 2009) เมื่อผู้ใช้มีความต้องการที่จะรับข่าว จะต้องสมัครสมาชิก หรือกรอกที่อยู่ของเอกสาร RSS ของแต่ละเว็บไซต์ลงใน RSS reader หรือ RSS aggregator ซึ่งเป็นโปรแกรมประยุกต์สำหรับดึงสรุปข่าว (feed) จากเว็บเซิร์ฟเวอร์ ผู้ใช้ก็จะได้รับเนื้อหาของข่าวที่แต่ละแหล่งข่าวได้จัดสรรไว้ให้ อย่างไรก็ตาม RSS reader เหล่านี้จะทำการแสดงวิดีโอข่าวตามข้อมูลเข้าซึ่งเป็นที่ (URL) ที่อยู่ของเอกสาร RSS ของแต่ละเว็บไซต์และแสดงออกมาเป็นรายการวิดีโอทั้งหมด ผู้ใช้สามารถที่จะเลือกดูข่าวต่างๆได้ด้วยตนเอง สำหรับเครื่องคอมพิวเตอร์ตั้งโต๊ะทั่วไปซึ่งมีประสิทธิภาพสูงนั้น การเข้าชมข่าวในหัวข้อซ้ำซ้อนกันอาจไม่เป็นปัญหา แต่สำหรับผู้ใช้อุปกรณ์ขนาดเล็ก เช่น PDA หรือ Smart phone ซึ่งมีทรัพยากรจำกัด การเข้าชมวิดีโอข่าวที่ซ้ำซ้อนกันทำให้เป็นการสิ้นเปลืองทั้งเวลา หน่วยความจำ และแบตเตอรี่ของอุปกรณ์ งานวิจัยนี้จึงนำเสนอเทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP โดยระบบจะดึงไฟล์รวบรวมหัวข้อของวิดีโอข่าว (feed) จากแหล่งต่างๆ มารวมกัน จากนั้นจึงทำการคัดกรองเพื่อเลือกวิดีโอที่เหมาะสมสำหรับแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP เพื่อเป็นการลดการประมวลผลของอุปกรณ์

นอกจากนี้งานวิจัยทำการพัฒนาระบบเป็น web-based RSS reader ผู้ใช้จึงสามารถดูการนำเสนอข้อมูลข่าวสารใหม่ๆได้เพียงใช้บราวเซอร์บนอุปกรณ์สื่อสารเคลื่อนที่ (mobile browser) โดยผ่านทางเว็บไซต์ ไม่ต้องทำการติดตั้งซอฟต์แวร์บนอุปกรณ์สื่อสารเคลื่อนที่ (mobile reader) เพื่อเป็นการลดการใช้หน่วยความจำที่มีอยู่อย่างจำกัด

1.1 การตรวจเอกสาร

เทคโนโลยี RSS ได้รับความนิยมเพิ่มมากขึ้นเรื่อยๆ โดยเฉพาะเว็บไซต์ที่เผยแพร่ข่าวสาร ซึ่งข้อมูลมีการเปลี่ยนแปลงใหม่อยู่เสมอ ปัจจุบันงานวิจัยที่เกี่ยวข้องกับเทคโนโลยี RSS ทั้งการประยุกต์ใช้ และการพัฒนาคุณสมบัติของ RSS ได้รับการพัฒนาขึ้นอย่างต่อเนื่องดังต่อไปนี้

Aldrich, Bell และ Batzel (2006) ได้นำเสนองานวิจัยที่นำเทคโนโลยี podcast มาใช้ทางด้านการศึกษา โดยได้ทำการออกแบบระบบในการสร้างไฟล์เสียงในรูปแบบ podcast ของการเรียนการสอน และเผยแพร่บนเว็บบล็อก (webblog) โดยอัตโนมัติ ระบบจะสร้างเอกสาร RSS ขึ้นสำหรับแต่ละคาบเรียน จากนั้นจะเริ่ม และหยุดการบันทึกเสียงโดยอัตโนมัติ ตามเวลาที่ระบุแล้วเก็บไฟล์เสียงลงไปในฐานข้อมูลพร้อมทั้งเชื่อมโยงไฟล์กับเอกสาร RSS ที่เกี่ยวข้องโดยอัตโนมัติ ในปีต่อมา Malan (2007) ได้พัฒนาให้สามารถบันทึกการเรียนการสอนได้ทั้งรูปของไฟล์เสียง และไฟล์วิดีโอ ในรูปแบบ podcast โดยอาจารย์จะเป็นผู้เผยแพร่ข้อมูลมัลติมีเดีย ในขณะที่นักเรียนเป็นผู้สมัครขอรับข้อมูล ซึ่งงานวิจัยนี้ได้มีการศึกษาความเป็นไปได้ในการนำมาใช้งานจริง และศึกษาความพึงพอใจของผู้ใช้

Curran และคณะ (2006) ได้นำเสนอ Scheduled RSS Delivery System ซึ่งเป็นระบบที่ใช้ช่วงเวลาที่ไม่ดำเนินงาน (idle) ในการรับข้อมูล RSS เพื่อดาวน์โหลดข้อมูลเสียง หรือวิดีโอในรูปแบบ podcast aggregator ของระบบนี้สามารถตรวจหารายการใหม่ๆ ที่เพิ่มขึ้นมาในเอกสาร RSS นั้น และเพิ่มเข้าไปในรายการเพื่อทำการดาวน์โหลดไฟล์มัลติมีเดียที่เกี่ยวข้องจากฐานข้อมูลของ web-server มายังเครื่องของผู้ใช้

Blekas และคณะ (2006) ได้เสนอแนวทางการปรับปรุงการแสดงเนื้อหาบนอุปกรณ์สื่อสารเคลื่อนที่ โดยนำเทคนิค RSS มาใช้ เนื่องจากข้อมูลรูปแบบนี้ เป็นข้อมูลโดยย่อของข้อมูลทั้งหมด เหมาะสำหรับแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่มีทรัพยากรจำกัด ผู้วิจัยได้สร้าง proxy server ซึ่งเป็นระบบออนไลน์มาใช้ในเทคนิคการปรับปรุงเนื้อหา proxy server จะรับการร้องขอดูหน้าเว็บจากอุปกรณ์สื่อสารเคลื่อนที่ ทำการโหลดเนื้อหาหน้าเว็บที่เกี่ยวข้อง จากนั้นจึงปรับเปลี่ยนเพื่อให้แสดงผลได้เหมาะสมกับอุปกรณ์ วิธีนี้ proxy server ทำงานด้วยเซิร์ฟเวอร์ประสิทธิภาพสูง ทำให้ลดการประมวลผลของอุปกรณ์สื่อสารเคลื่อนที่ลงได้อย่างมาก อย่างไรก็ตามวิธีนี้ยังมีข้อจำกัดคือแสดงเอกสาร RSS ได้เฉพาะข้อความเท่านั้น รูปภาพ และตาราง จะถูกตัดออกจากเนื้อหา และไม่สามารถใช้งานหน้าเว็บที่มี การกรอกแบบฟอร์มของภาษา HTML ได้

Ananthanarayanan (2006) และคณะ ได้นำเสนอแนวคิดของ OWeB หรือ Offline Web Browsing ในการแก้ปัญหาการเข้าถึงเว็บไซต์ บนเครือข่ายที่มีประสิทธิภาพต่ำ โดยคิดค้นขั้นตอนวิธีเพื่อสร้างรูปแบบของหน้าเว็บอัตโนมัติแล้วดึงข้อมูลจากเอกสาร RSS ของเว็บไซต์มาใช้เป็นข้อมูลของหน้าเว็บ ทำให้สามารถเข้าถึงเว็บไซต์ได้โดยไม่ขึ้นกับประสิทธิภาพของเครือข่าย OWeB จะมีส่วนประกอบที่ทำหน้าที่เป็น RSS Reader เพื่อดึงข้อมูลมาเก็บไว้บนเครื่องของผู้ใช้ทำให้สามารถเข้าถึงแบบออฟไลน์ได้

จากที่กล่าวมาจะเห็นว่า ได้มีการพัฒนาและประยุกต์ใช้เทคโนโลยี RSS ในหลายๆ ด้าน อย่างไรก็ตามยังมีงานวิจัยจำนวนน้อยที่กล่าวถึงการรวบรวมวิดีโอข่าวที่ซ้ำซ้อนกันจากหลายแหล่งข่าวมาคัดกรองเพื่อเลือกวิดีโอที่เหมาะสมสำหรับแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP เพื่อเป็นการลดการประมวลผลของอุปกรณ์เหล่านี้ซึ่งมีทรัพยากรจำกัด งานวิจัยนี้จึงได้นำเสนอระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP ซึ่งเป็นระบบที่ทำให้ผู้ใช้สามารถรับชมเอกสาร RSS ในรูปแบบวิดีโอที่มีเพิ่มมากขึ้นเรื่อยๆ โดยสนใจที่วิดีโอข่าวซึ่งมีการนำเสนอใหม่อยู่ตลอดเวลา พร้อมทั้งใช้เทคนิคการคัดกรองเพื่อให้ได้วิดีโอข่าวที่เหมาะสมสำหรับแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP ซึ่งมีขีดจำกัดทั้งด้านหน่วยประมวลผล หน่วยความจำ และประสิทธิภาพของการส่งผ่านข้อมูลผ่านเครือข่ายไร้สาย

1.2 วัตถุประสงค์ของโครงการ

พัฒนาเทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติ ซึ่งสามารถเลือกวิดีโอที่เหมาะสมเพื่อส่งไปแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP ที่มีทรัพยากรจำกัด เช่น ความเร็วของหน่วยประมวลผล เนื้อที่หน่วยความจำ แบตเตอรี่ เป็นต้น

1.3 ขอบเขตของการดำเนินงาน

1.3.1 พัฒนาระบบที่ใช้เทคนิคการคัดกรองวิดีโอข่าวจากหลายแหล่งข่าวเพื่อเลือกวิดีโอข่าวที่เหมาะสมสำหรับแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

1.3.2 ระบบทำการดึงวิดีโอข่าวจาก feed วิดีโอของเว็บไซต์แหล่งข่าวต่างๆที่กำหนดโดยผู้พัฒนา

1.3.3 อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP ครอบคลุมถึงอุปกรณ์ที่สามารถเข้าถึงเครือข่ายอินเตอร์เน็ตแบบไร้สายได้ เช่น PDA Smart phone Laptop

1.3.4 ผู้ใช้สามารถเข้าใช้ระบบซึ่งพัฒนาเป็น web-based RSS reader โดยเข้าถึงผ่านทางอินเตอร์เน็ตบราวเซอร์

1.4 ขั้นตอนและระยะเวลาการดำเนินงาน

1.4.1 ขั้นตอนการดำเนินงาน

1) ศึกษางานวิจัยและเอกสารที่เกี่ยวข้องและกำหนดขอบเขตของปัญหา

2) วิเคราะห์และออกแบบระบบคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

3) ศึกษาและเลือกใช้เทคโนโลยีและเครื่องมือสำหรับการพัฒนาระบบในส่วนต่างๆ

4) พัฒนาระบบคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP บนโปรแกรมจำลองการทำงานระบบปฏิบัติการบน PDA (PDA emulator)

5) เขียนบทความเพื่อตีพิมพ์วารสารทางวิชาการ

6) ทดสอบและปรับปรุงการทำงานของระบบบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

7) จัดทำเอกสารประกอบการวิจัย

1.4.2 ระยะเวลาดำเนินการ

ตุลาคม 2550 – มีนาคม 2552

1.4.3 แผนการดำเนินการวิจัย

แผนการดำเนินงานวิจัย แสดงได้ดังตารางที่ 1.1

ตารางที่ 1.1 แสดงแผนการดำเนินการวิจัย

กิจกรรมและขั้นตอนการดำเนินงาน	ปี พ.ศ. 2550			ปี พ.ศ. 2551												ปี พ.ศ. 2552		
	ตค.	พย.	ธค.	มค.	กพ.	มีค.	เมย.	พค.	มิย.	กค.	สค.	กย.	ตค.	พย.	ธค.	มค.	กพ.	มีค.
1.ศึกษางานวิจัยและเอกสารที่เกี่ยวข้องและกำหนดขอบเขตของปัญหา																		
2.วิเคราะห์และออกแบบระบบคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่																		
3.ศึกษาและเลือกใช้เทคโนโลยีและเครื่องมือสำหรับการพัฒนาระบบในส่วนต่างๆ																		
4.พัฒนาระบบคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่บน ซิมมูเลเตอร์																		
5.เขียนบทความเพื่อตีพิมพ์วารสารทางวิชาการ																		
6.ทดสอบและปรับปรุงการทำงานของระบบบนอุปกรณ์สื่อสารเคลื่อนที่																		
7.จัดทำเอกสารประกอบการวิจัย																		

1.5 สถานที่และเครื่องมือที่ใช้

1.5.1 สถานที่

ห้องปฏิบัติการคอมพิวเตอร์ CS207 (*i*STAR Lab) ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยสงขลานครินทร์ วิทยาเขตหาดใหญ่

1.5.2 เครื่องมือที่ใช้

1) ด้านฮาร์ดแวร์

เครื่องคอมพิวเตอร์ส่วนบุคคล สำหรับออกแบบและพัฒนาระบบคัดกรองวิดีโออัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

1.1) หน่วยประมวลผลกลาง : Intel(R) Pentium (R)

1.2) หน่วยความจำ : 1 GB

1.3) ฮาร์ดดิสก์ : 80 GB

2) ด้านซอฟต์แวร์

2.1) ระบบปฏิบัติการ Microsoft Windows XP

2.2) ระบบจัดการฐานข้อมูล MySQL

2.3) เว็บเซิร์ฟเวอร์

2.4) ภาษา AJAX, PHP, JAVASCRIPT

2.5) โปรแกรมจำลองการทำงานระบบปฏิบัติการบน PDA (Personal Digital Assistant) Windows mobile 5

2.6) โปรแกรม Multimedia Player สำหรับอุปกรณ์สื่อสารเคลื่อนที่

1.6 ประโยชน์ที่คาดว่าจะได้รับ

เทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติ ที่สามารถคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมก่อนที่จะส่งไปแสดงผลบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP เพื่อเป็นการลดการประมวลผลบนอุปกรณ์ของผู้ใช้

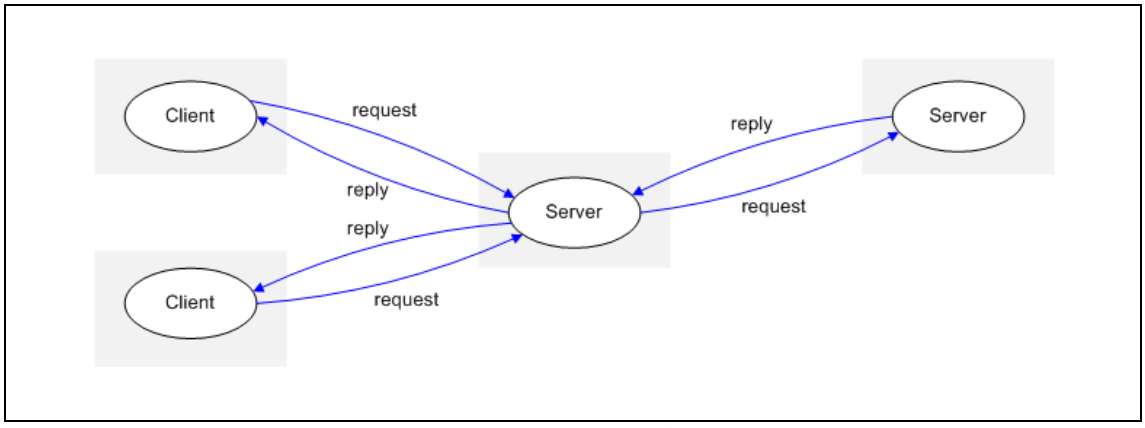
บทที่ 2

ทฤษฎีที่เกี่ยวข้อง

ทฤษฎีต่างๆที่เกี่ยวข้องกับการพัฒนาเทคนิคการคัดกรองวิดีโออัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP ประกอบด้วย สถาปัตยกรรมระบบแบบ client-server ระบบปฏิบัติการสำหรับอุปกรณ์สื่อสารเคลื่อนที่ ภาษา Extensible Markup Language (XML) XML Parser เทคโนโลยี Really Simple Syndication (RSS) เทคโนโลยี podcast และโครงสร้างของวิดีโอ

2.1 สถาปัตยกรรมระบบแบบไคลเอนท์-เซิร์ฟเวอร์ (client-server)

ในระบบที่มีการทำงานร่วมกันระหว่างหลายคอมโพเนนท์ การออกแบบสถาปัตยกรรมเป็นสิ่งสำคัญที่ส่งผลต่อประสิทธิภาพ ความมั่นคง และความปลอดภัยของระบบ สถาปัตยกรรมที่สำคัญคือสถาปัตยกรรมแบบไคลเอนท์-เซิร์ฟเวอร์ (client-server) แบบจำลองของระบบแสดงดังภาพประกอบ 2.1 (Coulouris, Dollimore and Kindberg, 2001)

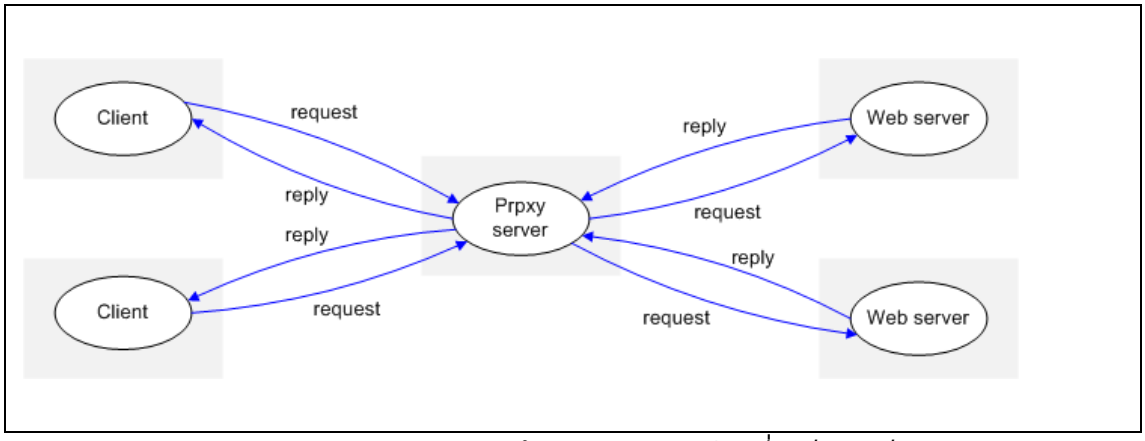


ภาพประกอบ 2.1 สถาปัตยกรรมแบบไคลเอนท์-เซิร์ฟเวอร์

สถาปัตยกรรมดังภาพประกอบ 2.1 เป็นสถาปัตยกรรมพื้นฐาน ที่ถูกนำไปใช้งานอย่างกว้างขวาง โดยโปรเซสของไคลเอนท์ติดต่อกับโปรเซสภายในของเซิร์ฟเวอร์ที่แบ่งกันทำงานหลายเครื่อง แบ่งการจัดการทรัพยากร เซิร์ฟเวอร์หนึ่งอาจจะประพฤติตัวเป็น ไคลเอนท์ของอีกเซิร์ฟเวอร์หนึ่ง เว็บเซิร์ฟเวอร์และบริการบนอินเตอร์เน็ตส่วนใหญ่เป็นไคล-เอนท์ของ DNS (Domain Name System) ที่แปลงระหว่าง Domain name กับ network address อีก

ตัวอย่างหนึ่งบนเว็บ ได้แก่ search engine ที่ให้ผู้ใช้สามารถค้นหาสรุปข้อมูล (contents summary) ของหน้าเว็บบนเครือข่ายอินเตอร์เน็ต โดยสรุปข้อมูลดังกล่าวสร้างขึ้นจากโปรแกรมที่เรียกว่า web- crawler ซึ่งทำงานอยู่เบื้องหลังเว็บ search engine โดยใช้การร้องขอผ่าน HTTP (Hypertext Transfer Protocol) เพื่อเข้าถึงเว็บเซิร์ฟเวอร์ผ่านเครือข่ายอินเตอร์เน็ต search engine จะประพฤติตัวเป็นทั้งไคลเอนท์ และ เซิร์ฟเวอร์ โดยตอบสนองการร้องขอจากบราวเซอร์ของผู้ใช้ แล้วเรียก web crawler ซึ่งเป็นไคลเอนท์ของเว็บเซิร์ฟเวอร์อื่นๆขึ้นมาทำงาน ตัวอย่างดังกล่าวการทำงานของทั้งสองส่วนจะเป็นอิสระต่อกัน

นอกจากนี้ยังมีการพัฒนาสถาปัตยกรรมระบบที่เพิ่มประสิทธิภาพการทำงานของเซิร์ฟเวอร์ได้แก่พร็อกซี่ (proxy) และแคช (cache) แคชคือที่เก็บข้อมูลของออปเจ็กต์ (object) ที่มีการเรียกใช้ล่าสุด ออปเจ็กต์ใหม่ที่ถูกดึงมายังเครื่องจะถูกเพิ่มเข้าไปในที่เก็บข้อมูล cache และอาจแทนที่บางออปเจ็กต์ที่มีอยู่ เมื่อโปรเซสของไคลเอนท์ต้องการเรียกออปเจ็กต์ กระบวนการแคชจะตรวจสอบข้อมูลแคชและดึงออปเจ็กต์ที่เก็บอยู่หากออปเจ็กต์ดังกล่าวอัพเดท แต่หากไม่ออปเจ็กต์ที่อัพเดทใหม่ก็จะถูกดึงเข้ามาแทนที่ ข้อมูลแคชอาจถูกเก็บไว้ที่ไคลเอนท์หรือเก็บไว้ที่พร็อกซี่เซิร์ฟเวอร์ซึ่งสามารถใช้งานได้โดยไคลเอนท์หลายไคลเอนท์



ภาพประกอบ 2.2 สถาปัตยกรรมของพร็อกซี่เซิร์ฟเวอร์

ปัจจุบันมีการใช้งานแคชอย่างกว้างขวาง เว็บบราวเซอร์เองก็มีการเก็บแคชของหน้าเว็บที่ถูกเรียกดูล่าสุด และข้อมูลจากเว็บอื่นๆที่ถูกเก็บไว้ในเครื่องผู้ใช้ โดยใช้การร้องขอผ่าน HTTP ไปยังเว็บเซิร์ฟเวอร์ว่าข้อมูลหน้าเว็บนั้นยังอัพเดทอยู่หรือไม่ก่อนที่จะแสดง จากภาพประกอบ 2.2 พร็อกซี่เซิร์ฟเวอร์จะเก็บแคชจากเว็บเซิร์ฟเวอร์โดยมีเป้าหมาย คือ เพื่อเพิ่มความสามารถการเข้าใช้ และประสิทธิภาพของบริการต่างๆ เพื่อลดภาระงานบนเครือข่ายและเว็บเซิร์ฟเวอร์ พร็อกซี่เซิร์ฟเวอร์ยังสามารถนำมาใช้งานในลักษณะอื่นเช่น ใช้เพื่อเข้าถึงเว็บเซิร์ฟเวอร์อื่นโดยผ่าน firewall

2.2 ภาษา XML (Extensible Markup Language)

ภาษา XML (Extensible Markup Language) เป็นชุดภาษาย่อยของ SGML (Standard Generalized Markup Language) ซึ่ง SGML เป็นภาษาที่ใช้ในการแสดงโครงสร้างและเนื้อหาของเอกสารอิเล็กทรอนิกส์ตามมาตรฐาน ISO 8879:1986 (นิศาชล จำนงศรี, 2551) ในส่วนของ XML ได้รับการพัฒนาขึ้นมาในปีค.ศ. 1998 โดย W3C (World Wide Web Consortium) มีวัตถุประสงค์เพื่อให้มีการแลกเปลี่ยนข้อมูลที่มีโครงสร้าง เช่น Metadata บนอินเตอร์เน็ตทำได้ง่ายขึ้น และมีความชัดเจนพอที่มนุษย์จะสามารถอ่านได้ในกรณีที่ไม่มี XML Browser ซึ่งต่างจาก HTML (Hypertext Markup Language) ที่เป็นแอพพลิเคชันหนึ่งของ SGML โดย HTML เป็นภาษาที่มีการใช้อย่างแพร่หลายในการสร้างเอกสารเว็บ แต่ไม่สามารถรองรับความต้องการในการแสดงเอกสารที่มีโครงสร้างได้ เนื่องจากเป็นแอพพลิเคชันที่พัฒนาขึ้นมาเพื่อนำเสนอเอกสาร เช่น รูปแบบตัวอักษร สีตัวอักษร สีพื้น เป็นต้น และสามารถอธิบายได้เฉพาะโครงสร้างหลัก ๆของเอกสารเท่านั้น เช่น หัวข้อ <Heading> ชื่อเรื่อง <Title> เนื้อหา <Body> ซึ่งไม่สามารถบอกได้ว่ารายละเอียดที่อยู่ในส่วนของเนื้อหาประกอบไปด้วยรายละเอียดส่วนใดบ้าง

ทั้ง HTML และ XML มีลักษณะเป็น Markup Language คือ มีการกำหนด tag เพื่อสื่อความหมายให้แก่ข้อมูลหรือเพื่อระบุถึงการแสดงผลว่าข้อมูลจะถูกแสดงออกมาเช่นไร โดย tag ต่าง ๆ ที่อธิบายถึงข้อมูลจะไม่แสดงผลทางหน้าจอ ซึ่งสามารถแบ่งภาษาที่มีลักษณะเป็น Markup Language ออกเป็น 2 ประเภทใหญ่ ๆ (Layout Galaxy, 2003; Dykes and Tittel, 2005) คือ

1) Specific Markup Language เป็นภาษาที่มีชุดคำสั่งเจาะจงชัดเจนลงไปเลยว่าคำสั่งใด หรือ tag ใดเอาไว้ทำอะไร เป็นภาษาที่ถูกสร้างมาเพื่อการประยุกต์ใช้งานเช่น HTML เนื่องจากแต่ละ tag จะมีลักษณะเฉพาะที่ชัดเจนหรือมีการนิยามเป็นมาตรฐาน เช่น <Table> <Tr> <Td> </Td> </Tr> </Table>เป็นชุดคำสั่งในการสร้างตาราง

2) Generalized Markup Language เป็นภาษาที่เน้นการอธิบายความหมายของเอกสารมากกว่าการแสดงผล มีชุดคำสั่งของตัวเอง สามารถสร้างชุดคำสั่งขึ้นมาเอง หรือนำไปสร้างเป็นภาษาใหม่ขึ้นมาได้ อีกทั้งยังทำงานกับโปรแกรมประยุกต์ได้หลากหลาย เช่น XML โดยมีชุดคำสั่งที่กำหนดขึ้นเองได้เป็นลักษณะของ Meta Language หรือ เป็นภาษาที่สามารถนิยามภาษาอื่นได้

2.2.1 มาตรฐานต่าง ๆ ที่เกี่ยวข้องกับ XML

มีการเสนอเครื่องมือใหม่ ๆ เรื่อยมาตั้งแต่มีการพัฒนา XML 1.0 เพื่อทำงานกับ XML ในรูปของ parsers และเสริมศักยภาพการทำงานของ XML

มาตรฐานต่าง ๆ ที่เกี่ยวข้องกับ XML (Stylus Studio, 2009) ได้ถูกเสนอให้ W3C รับรองเป็นจำนวนมาก ในที่นี้ จะขอยกตัวอย่างแต่พอสังเขปดังนี้

1) Namespaces ช่วยในการกำหนดขอบเขตให้กับชื่อของ element และ attributes เพื่อหลีกเลี่ยงการถูกเรียกใช้ซ้ำกันระหว่างเอกสาร XML จากต่างแหล่งกำเนิดกัน โดยที่ namespace แต่ละชุดเป็นการรวมกันของชื่อต่างๆ ของ XML element และ attributes ที่ถูกระบุพร้อมกับ URI (Uniform Resource Identifier) ที่เป็นแหล่งกำเนิดของเอกสาร XML นั้น

2) XSL (Extensible Stylesheet Language) เป็นภาษาในการกำหนด style sheets ให้กับ XML data สำหรับการนำเสนอในเว็บบราวเซอร์ หรือสื่อนำเสนออื่นๆ มีศักยภาพสูงกว่า CSS (Cascading style Sheet) ที่ใช้เป็น style sheets สำหรับ HTML มาก

3) XLink เป็นการปรับปรุงการเชื่อมโยง (link) ใน HTML โดยสนับสนุนการเชื่อมโยงแบบสองทาง (bi-direction) แบบหนึ่งต่อหลายแห่ง (one-to-many) และแบบแยกตามประเภท มีพื้นฐานการทำงานจาก HyTime (Hypermedia/Time-based Structuring Language)

4) Xpointer เป็นตัวชี้ให้สามารถชี้ไปที่จุดหรือตำแหน่งใดก็ได้ในเอกสารเป้าหมายโดยไม่ต้องกำหนดการเชื่อมโยงแบบ HTML link ไว้ล่วงหน้า มีพื้นฐานการทำงานจาก Text Encoding Initiative (TEI)

5) RDF (Resource Description Framework) เป็น Metadata สำหรับเอกสาร XML คล้ายกับ Metadata tags ของ HTML (เช่น ผู้แต่ง ลิขสิทธิ์ และวันที่จัดพิมพ์ เป็นต้น)

6) XQL (XML Query Language) เป็นส่วนขยายเพิ่มเติมของ XSL สำหรับใช้อ้างถึงและกลั่นกรองส่วนที่เป็น element และ text ของเอกสาร XML ช่วยกำหนดรูปแบบที่ชัดเจน เข้าใจได้ง่าย ในการเข้าถึง element ที่เฉพาะเจาะจงและสำหรับการค้นหาโหนด (node) ที่มีคุณลักษณะพิเศษต่าง ๆ

นอกจากนี้ยังมีมาตรฐานในการกำหนดโครงสร้างของเอกสาร XML ซึ่งจะอธิบายในหัวข้อ

2.2.2 องค์ประกอบและโครงสร้างของ XML

เอกสาร XML เป็นเอกสารที่ประกอบกันขึ้นจากชุดของเอนติตี้ (entity) (นิศาชล จำนงศรี, 2551) ลักษณะเป็นออปเจ็กต์ (object) แต่ละเอนติตี้จะประกอบด้วยอิลิเมนต์ (element) ตั้งแต่ 1 อิลิเมนต์ขึ้นไป ในแต่ละอิลิเมนต์จะมีแอททริบิวต์ (attribute) ที่บ่งบอกลักษณะเฉพาะ และ tag คือโครงสร้างระดับรองลงมา เป็นตัวกำหนดขอบเขตของอิลิเมนต์ดังแสดงในภาพประกอบ 2.3

```
<root>
   <element attribute=”...”>
      <tag>...</tag>
      <tag>...</tag>
        .
        .
        .
   </element>
      .
      .
      .
</root>
```

ภาพประกอบ 2.3 โครงสร้างของ XML

อิลิเมนต์และแอททริบิวต์ของ XML จะเขียนไว้ใน <....> ส่วนค่าแอททริบิวต์จะแสดงไว้ในเครื่องหมายคำพูด "...." ในแต่ละอิลิเมนต์ของ XML จะต้องประกอบด้วย tag เปิด <...> และ tag ปิด </...> tag อยู่นอกสุดจะเป็นอิลิเมนต์รากของเอกสาร (Root Element) และ tag ตัวถัดมาจะเป็นลูกของรากนั้น (Child Elements) ลักษณะดังกล่าวเรียกว่าเป็น nested tags นอกจากนี้ tag ของ XML มีลักษณะเป็น Case-sensitive language ตัวพิมพ์ใหญ่และตัวพิมพ์เล็กในการกำหนดชื่ออิลิเมนต์ให้ผลที่ต่างกัน

XML declaration XML Declaration เป็นการระบุ Version ของ XML ที่ใช้ในการสร้างเอกสาร จะประกาศไว้ที่ ด้านบนสุดของเอกสาร แม้ว่าในส่วนนี้จะเป็นตัวเลือก แต่ผู้สร้างเอกสารควรประกาศไว้เสมอ เพื่อให้ผู้นำเอกสารไปใช้ทราบ Version ของ XML การประกาศทำได้ดังนี้

```
<?xml version="1.0" encoding=UTF-8" standalone="yes">
```

ภาพประกอบ 2.4 การประกาศเอกสาร XML

2.2.3 การนิยามโครงสร้างเอกสาร XML

XML เป็นเอกสารที่ใช้แลกเปลี่ยนข้อมูลระหว่างองค์กรหรือ หน่วยงาน ดังนั้นฝ่ายที่รับเอกสารต้องมั่นใจว่าเอกสารนั้นมีความ ถูกต้อง และนำไปใช้งานได้ ดังนั้นเอกสารจึงต้องมีการกำหนดมาตรฐานของรูปแบบ มีสองลักษณะคือ Well-formed และ Valid

เอกสาร Well-formed XML นั้นจะต้องเป็นไปตามกฎเกณฑ์และวิธีการของ XMLอย่างเคร่งครัด กฎเกณฑ์เบื้องต้นของ Well-formed

1) ต้องมี tag ที่เป็นรากเพียงหนึ่ง tag เท่านั้น
2) ใน tag อื่นที่ไม่ใช่รากจะต้องเป็นลูกหรือหลานของรากเท่านั้น
3) อิลิเมนต์ทุกตัวจะครบคู่กล่าวคือ เมื่อเปิด tag จะต้องเป็น tag เสมอ
4) ชื่อ tag ที่เปิดจะต้องเหมือนกับ tag ปิด ทั้งตัวเล็กตัวใหญ่คือเป็น Case-sensitive
5) ชื่อแอททริบิวต์จะต้องถูกใช้เพียงหนึ่งครั้งเท่านั้นในelementเดียวกัน

เอกสาร Valid XML คือเอกสารที่ถูกต้องสมบูรณ์ในการนำไปใช้งานจริง เอกสาร XML จำเป็นต้องถูกตรวจสอบความถูกต้อง ดังนั้นเอกสารนั้นจะต้องอยู่ในรูปแบบ well-formed และเอกสารนั้นจะต้องถูกต้องตามข้อกำหนดการนิยามโครงสร้างของเอกสาร XML การนิยามโครงสร้างนั้นสามารถทำได้โดยใช้ DTD (Data Type Definition) หรือ XML Schema (Moller, A., Schwartzbach, M. 2003) ซึ่งมีรายละเอียดดังนี้

2.2.3.1 Data Type Definition (DTD)

DTD มีองค์ประกอบหลักดังนี้

1) PCDATA: ระบุส่วนข้อมูลที่ XML Parser สามารถวิเคราะห์ได้
2) CDATA: ระบุข้อมูลประเภทตัวอักษร ซึ่ง Parser จะไม่ทำการวิเคราะห์ข้อมูล
3) Elements: ระบุ tag ที่ใช้ในการอธิบายข้อมูล ตัวอย่างของการนิยามโครงสร้างอิลิเมนต์และเอกสาร XML ตามโครงสร้างดังกล่าวแสดงดังภาพประกอบ 2.5 จากภาพโครงสร้างเอกสาร XML ถูกนิยามไว้ใน note.dtd โดยอิลิเมนต์ชื่อ <note> มี tag ลูกคือ <to> <from> <heading> และ <body> ซึ่งบรรจุข้อมูลที่เป็นข้อความ

note.dtd

```
<!ELEMENT note
(to,from,heading,body)>
<!ELEMENT to (#PCDATA)>
<!ELEMENT from (#PCDATA)>
<!ELEMENT heading (#PCDATA)>
<!ELEMENT body (#PCDATA)>
```

```
<?xml version="1.0"?>
<!DOCTYPE note SYSTEM "note.dtd">
<note>
  <to>Tove</to>
  <from>Jani</from>
  <heading>Reminder</heading>
  <body>Don't forget me this
  weekend</body>
</note>
```

ภาพประกอบ 2.5 ตัวอย่างการนิยามโครงสร้างของอิลิเมนต์ใน DTD

นอกจากนี้ยังมีการกำหนดจำนวนของอิลิเมนต์ลูกต่าง ๆคือ “*” ระบุจำนวน ตั้งแต่ 0 ตัวขึ้นไป “+” ระบุจำนวนตั้งแต่ 1 ตัวขึ้นไป “?” ระบุจำนวน 0 หรือ 1 ตัว เช่น <!ELEMENT [element-name] ([child-name]*)>

4) Attributes: กำหนดข้อมูลเพิ่มเติมให้กับอิลิเมนต์การนิยามโครงสร้างของแอททริบิวต์จะมีค่าตั้งต้นต่าง ๆที่สามารถระบุได้ แสดงดังตารางที่ 2.1

ตารางที่ 2.1 ค่าตั้งต้นของการนิยามแอททริบิวต์ใน DTD

Attribute	การใช้งาน
#REQUIRED	เป็นแอททริบิวต์บังคับ ที่ต้องระบุค่า
#IMPLIED	ไม่จำเป็นต้องระบุแอททริบิวต์ดังกล่าว
#FIXED [fixedvalue]	ค่าของแอททริบิวต์นี้ มีการกำหนดไว้แน่นอน
Default	ค่าตั้งต้นของแอททริบิวต์ซึ่งจะถูกดึงมาใช้หากไม่มีการกำหนดไว้เป็นค่าอื่น

ตัวอย่างการนิยามโครงสร้างแอททริบิวต์แสดงดังภาพประกอบ 2.6 อิลิเมนต์ ชื่อ Author มีสองแอททริบิวต์คือ a_id และ prefix ซึ่งเป็นแอททริบิวต์บังคับ

```
<!ATTLIST Author a_id #REQUIRED
prefix(Mr|Ms|Mrs) #REQUIRED>
```

```
Valid XML:
<Author a_id=“w002” prefix=“Ms”>
<name>Charawee</name>
</Author>
```

ภาพประกอบ 2.6 ตัวอย่างการนิยามโครงสร้างของแอททริบิวต์ใน DTD

5) Entities: ระบุอักขระพิเศษ ที่จะเรียกใช้ในเอกสาร XML การอ้างถึงจะขึ้นต้นด้วย “&” และลงท้ายด้วย “;” ตัวอย่างแสดงดังภาพประกอบ 2.7

```
<!ENTITY writer "Donald Duck.">
<!ENTITY copyright "Copyright
W3Schools.">
```

```
Valid XML:
<author>&writer;&copyright;
</author>
```

ภาพประกอบ 2.7 ตัวอย่างการนิยามโครงสร้างของเอนติตี้ใน DTD

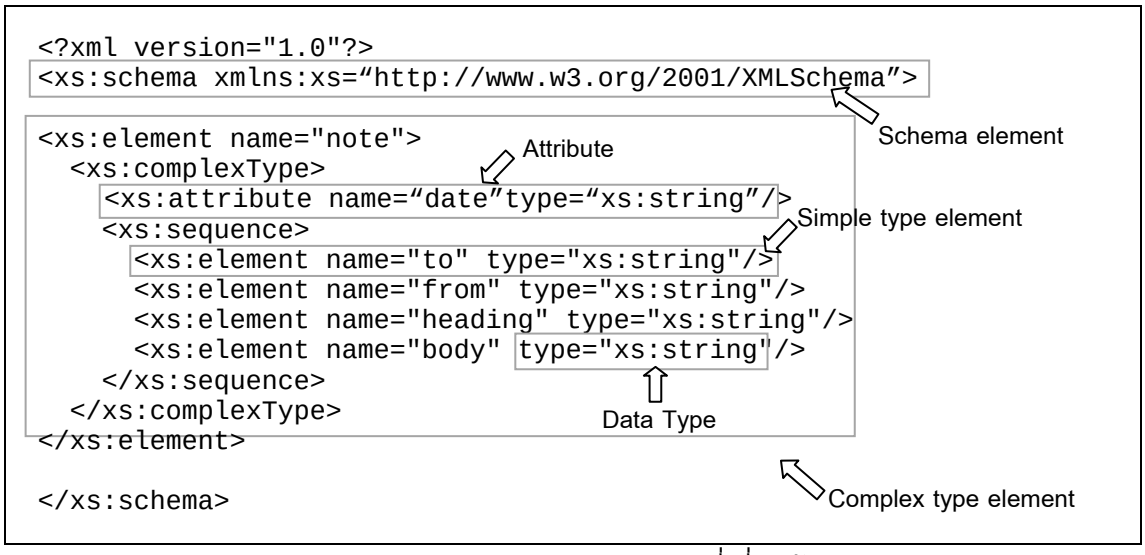
2.2.3.2 XML Schema

เนื่องจากบ่อยครั้งที่มีเอกสารซึ่งประกอบด้วยเนื้อหาที่มีชนิดต่างกัน และต้องการตรวจสอบ Data Type แต่ DTD ไม่ได้ถูกออกแบบมาเพื่อตรวจสอบ Data Type เหล่านี้ ดังนั้น XML Schema จึงถูกสร้างขึ้นมาเพื่อการแก้ปัญหานี้ XML Schema ต่างจาก DTD ตรงที่ถูกเขียนขึ้นในรูปแบบไวยากรณ์ของ XML นอกจากการจัดสร้างข้อมูลที่ DTD นำเสนอแล้ว XML Schema ยังช่วยกำหนด Data Type มีการใช้ Namespace และสามารถกำหนดช่วงค่าของแอททริบิวต์และอิลิเมนต์ได้

ตัวอย่างของเอกสาร XML และ XML Schema ที่เกี่ยวข้องแสดงดังภาพประกอบ 2.8 และ 2.9 ตามลำดับ

```
<?xml version="1.0"?>
<note date=”Jan 1,2008”>
  <to>George</to>
  <from>Laura</from>
  <heading>Reminder</heading>
  <body>Don't forget COKE!</body>
</note>
```

ภาพประกอบ 2.8 ตัวอย่างเอกสาร XML



ภาพประกอบ 2.9 XML Schema ที่เกี่ยวข้อง

จากภาพประกอบ 2.9 อธิบายโครงสร้างต่างๆของ XML Schema ได้ดังต่อไปนี้

1) Schema element: เป็นส่วนที่บอกว่าไฟล์เป็น XML Schema XML Namespace คือ URL ที่มีรายละเอียดของคำศัพท์ XML ต่างๆ ในที่นี้คือคำศัพท์ของ XML Schema ในที่นี้กำหนดการอ้างถึง namespace นี้โดยใช้ prefix “xs”

2) Element: มีสองลักษณะคือ simple type คืออิลิเมนต์ที่มีเพียงข้อมูลเท่านั้น ไม่มีแอททริบิวต์หรืออิลิเมนต์ลูก และ complex type คืออิลิเมนต์ที่บรรจุอิลิเมนต์อื่น และ/หรือแอททริบิวต์

3) Attribute: จะต้องมีชื่อ (name) และอิลิเมนต์ชนิด complex เท่านั้นที่สามารถมีแอททริบิวต์ได้

4) Annotation: ใช้สำหรับอธิบายรายละเอียดเพิ่มเติมที่เป็นประโยชน์ต่อผู้เขียนโปรแกรม หรือแอพพลิเคชัน

```
<xsd:annotation>
  <xsd:documentation xml:lang = “en”>
  here is the documentation text for our schema
  </xsd:documentation>
</xsd:annotation>
```

ภาพประกอบ 2.10 ตัวอย่างการนิยามโครงสร้างของ annotation ใน XML Schema

5) Datatype: ชนิดของข้อมูลในการนิยามโครงสร้างใน XML Schema ตามมาตรฐานของ W3C มีทั้งหมด 44 ชนิด

2.2.4 การแสดงผลของเอกสาร XML

การแสดงผลของเอกสาร XML ต้องอาศัยภาษา หรือวิธีการอื่น ๆ มาช่วยในการนำข้อมูลออกมาแสดงผลตามที่ผู้ใช้เอกสารต้องการ เนื่องจาก XML นั้น เป็นเพียงส่วนของการจัดการข้อมูลไม่สามารถแสดงผลได้ด้วยตนเอง ซึ่งการนำเสนอเอกสาร XML บนเว็บบราวเซอร์นั้นสามารถทำได้ 3 วิธี คือ 1) นำเสนอทาง HTML (Hypertext Markup Language) 2) การเขียนโค้ด CSS (Cascading Style Sheet) และ 3) XSL (Extensible Stylesheet Language) ปัจจุบันมีเว็บบราวเซอร์ที่สนับสนุนการใช้เอกสาร XML เช่น Internet Explorer Version 5.0 ขึ้นไป โดยมี Parser ของเอกสาร XML อยู่ด้วย นั่นคือ หากเขียนชุดคำสั่งถูกต้องบราวเซอร์จะสามารถเรียกเอกสาร XML ขึ้นมาแสดงผลได้ทันที ผลลัพธ์ที่ออกมาจะเหมือนกับชุดคำสั่งทุกประการ

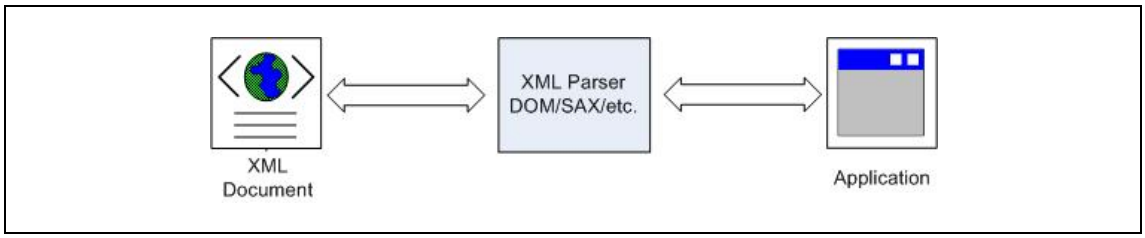
1) การแสดงผลด้วย HTML (Hypertext Markup Language) โดยการสร้างไฟล์ HTML เพื่อกำหนดรูปแบบที่ต้องการแสดงผล เช่น กำหนดให้แสดงผลแบบตาราง แล้วเขียนคำสั่งให้โปรแกรมไปดึงข้อมูลที่ต้องการแสดงผลมาจากไฟล์ XML มาแสดงผลในตารางของ HTML โดยใช้ Attribute

2) การแสดงผลด้วยการเขียนชุดคำสั่ง CSS (Cascading Style Sheet) แล้วนำมาแทรกในเอกสาร XML เพื่อให้เอกสารแสดงผลตาม CSS ที่ออกแบบไว้ การแสดงผลจะเรียกที่ไฟล์ XML อย่างไรก็ตามการใช้ CSS เพื่อแสดงผลเอกสาร XML เหมาะกับการนำเสนอข้อมูลลักษณะง่าย ๆ ที่ไม่ซับซ้อนเช่น การจัดการเรื่องสี และขนาดตัวอักษรเนื่องจาก CSS ยังคงใช้ความสามารถของ HTML

3) การแสดงผลด้วย XSL (Extensible Stylesheet Language) XSL เป็นมาตรฐานใหม่ที่พัฒนาขึ้นมาเพื่อกำหนด Style sheets ในการแสดงผลเอกสาร XML บนบราวเซอร์ หรือสื่อนำเสนออื่นๆ โดยในการแสดงผล สามารถจัดเรียงลำดับข้อมูลได้ สามารถเลือกได้ว่าต้องการให้แสดงข้อมูลส่วนใดบ้าง และสามารถเรียงลำดับข้อมูลจากค่าน้อยไปหาค่ามาก หรือจากค่ามากไปหาค่าน้อยได้ โดยการกำหนดเงื่อนไขการแสดงผลไว้ และนอกจากจะทำหน้าที่ในการกำหนดรูปแบบ (Formatting) ให้กับเอกสาร XML แล้ว XSL ยังมีความสามารถในการแปลงรูปแบบข้อมูล (Transforming) ได้ด้วย เช่น แปลงจาก XML เป็น HTML หรือ PDF เป็นต้น จากภาพรวมทั้งหมดจะเห็นว่า XML สามารถช่วยขยายความสามารถให้กับการแลกเปลี่ยนข้อมูลผ่านทางเครือข่ายรวมทั้งอินเตอร์เน็ตได้มาก โดยไม่ต้องกังวลว่าระบบอื่นจะไม่รู้จัก XML เนื่องจาก XML จะเน้นที่การกำหนดโครงสร้างเนื้อหาของเอกสาร เพื่อรองรับการแลกเปลี่ยนข้อมูลระหว่างแอพพลิเคชั่น โดยไม่ขึ้นอยู่กับแพลตฟอร์มหรือผู้ผลิตรายใดรายหนึ่ง และประเด็นสำคัญคือ XML จะแยกส่วนของข้อมูลออกจากการแสดงผลอย่างชัดเจน ดังนั้น การเปลี่ยนแปลงของแต่ละส่วนจะไม่ส่งผลถึงกันและกัน ซึ่งเป็นประเด็นที่พัฒนาขึ้นมาเพื่อแก้ไขข้อบกพร่องของ HTML ที่ไม่สามารถแยกแสดงโครงสร้างด้านเนื้อหาของเอกสารได้ และยังมีการกำหนด tag แบบตายตัว จึงไม่สามารถกำหนด tag ให้สอดคล้องกับเนื้อหาของเอกสารแต่ละประเภทได้ แต่มีจุดเน้นที่การแสดงรูปร่างหน้าตาของเอกสาร จึงไม่สนับสนุนการแลกเปลี่ยนข้อมูลระหว่างแอพพลิเคชันซึ่งเป็นความต้องการของระบบสารสนเทศยุคใหม่ แนวโน้มในอนาคตคือ จะใช้ XML เป็นรูปแบบกลางในการสร้างเอกสารเพื่อนำไปใช้ในแอพพลิเคชันที่ต่างกันได้ โดยไม่ต้องสร้างข้อมูลขึ้นมาใหม่ หรืออาจกล่าวได้ว่า XML จะกลายมาเป็นภาษาสากลที่แอพพลิเคชันต่าง ๆ จะนำมาพูดคุยให้สามารถสื่อสารเข้าใจกันได้

2.3 XML Parser

การดึงข้อมูลจากเอกสาร XML มาใช้งานนั้นจะต้องอาศัย XML Parser เป็นโปรแกรมซึ่งทำหน้าที่วิเคราะห์และแยกแยะโครงสร้างของเอกสาร XML ฟังก์ชั่นการทำงานหลักของ XML Parser คือ 1) ตรวจสอบไวยากรณ์ของภาษา XML (syntax) 2) ตรวจสอบความถูกต้องของโครงสร้างของเอกสาร เช่นโครงสร้างที่กำหนดด้วย DTD หรือ XML Schema และ 3) สกัดข้อมูลของเอกสารเพื่อให้สามารถเข้าถึงได้โดยโปรแกรมประยุกต์ต่างๆ ลักษณะการทำงานของ XML Parser แสดงดังภาพประกอบ 2.11

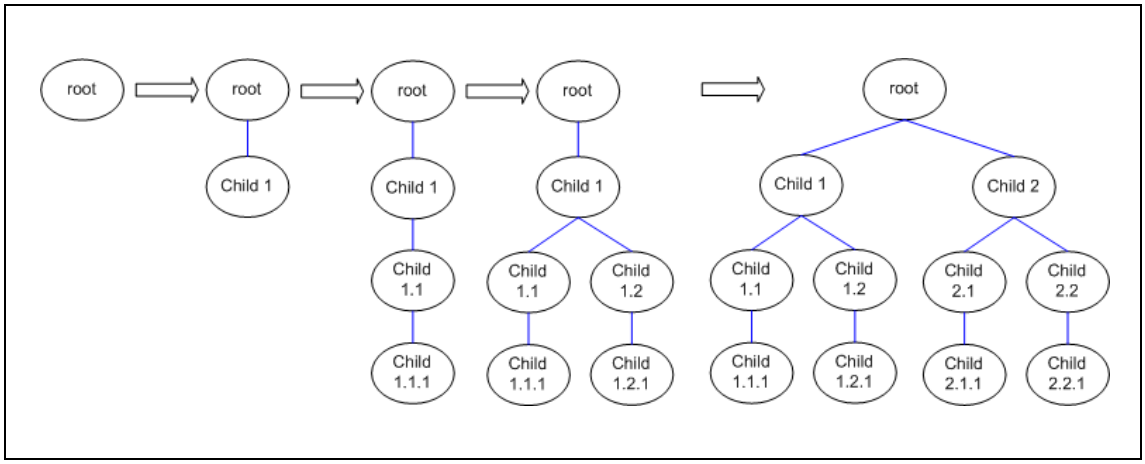


ภาพประกอบ 2.11 ลักษณะการทำงานของ XML Parser

ลักษณะที่ XML Parser ใช้ในการวิเคราะห์ข้อมูลในเอกสาร XML นั้นจะมีอยู่ 2 แบบหลักๆคือ Object-based หรือ Tree-based และ Event-based

2.3.1 การวิเคราะห์ข้อมูลแบบ Object-based

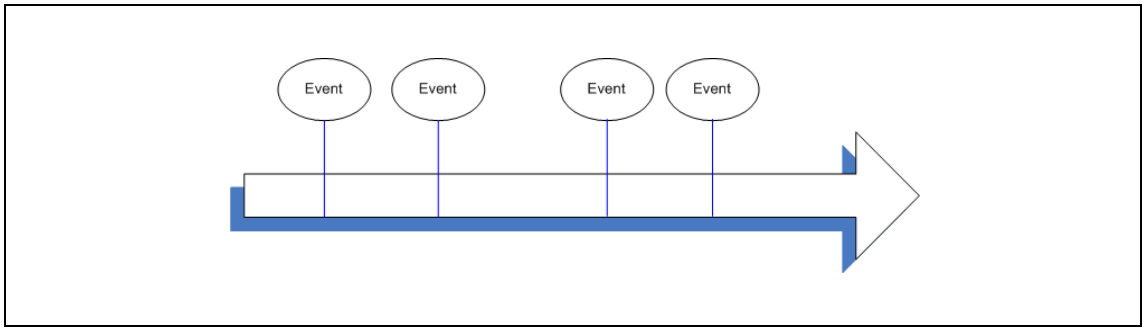
วิธีการแบบ Object-based จะเป็นการอ่านข้อมูลจากแฟ้ม XML ขึ้นมาทั้งหมดแล้วจัดองค์ประกอบต่างๆ ของ XML ให้อยู่ในรูปของโครงสร้างต้นไม้ ดังภาพประกอบ 2.12 เก็บไว้ในหน่วยความจำหลักของคอมพิวเตอร์ เมื่อต้องการนำข้อมูลที่อยู่ในแฟ้ม XML มาใช้งานก็สามารถอ่านข้อมูลที่เก็บอยู่ในโครงสร้างต้นไม้ มาใช้งานได้ทันที โดยไม่ต้องเสียเวลาอ่านแฟ้ม XML ซ้ำอีก API ที่มีการอ่านข้อมูลในลักษณะนี้ได้แก่ DOM (Document Object Model)



ภาพประกอบ 2.12 แสดงลักษณะการวิเคราะห์ข้อมูลแบบ Object-based

2.3.2 การวิเคราะห์ข้อมูลแบบ Event-based

วิธีการแบบ Event-based นั้น จะเป็นการอ่านข้อมูลจากแฟ้ม XML ขึ้นมาทีละชุด โดยอ่านตั้งแต่เริ่มต้นไปเรื่อยๆ แล้วประมวลผลตามเหตุการณ์ต่างๆที่พบ เช่น เปิดอิลิเมนต์ ปิดอิลิเมนต์ ดังภาพประกอบ 2.13 และเลือกเอาเฉพาะสิ่งที่เราต้องการมาใช้งาน และหากต้องการข้อมูลอีกครั้ง จะต้องอ่านเอกสาร XML ซ้ำอีกครั้งหนึ่ง API ที่มีการอ่านข้อมูลในลักษณะนี้เช่น SAX (Simple API for XML)



ภาพประกอบ 2.13 แสดงลักษณะการวิเคราะห์ข้อมูลแบบ Event-based

ตารางที่ 2.2 แสดงการเปรียบเทียบลักษณะการวิเคราะห์ข้อมูลแบบ Object-based และแบบ Event-based

ตารางที่ 2.2 การเปรียบเทียบระหว่างวิธีการวิเคราะห์ข้อมูล XML แบบ Object-based และ Event-based

	Object-based	Event-based
วิธีการเข้าถึงข้อมูล	เข้าถึงข้อมูลแบบสุ่ม (random)	เข้าถึงข้อมูลตามลำดับ (sequence)
วิธีการประมวลผลข้อมูล	อ่านข้อมูลครั้งเดียว แล้วสร้างโครงสร้างข้อมูลแบบ ต้นไม้ไว้ในบัฟเฟอร์	อ่านข้อมูลใหม่ทุกครั้งที่ต้องการประมวลผล
การจัดการข้อมูล	แก้ไขข้อมูลได้	ไม่สามารถแก้ไขข้อมูลได้
การพัฒนาโปรแกรม	การพัฒนาไม่ซับซ้อน	การพัฒนาซับซ้อน
ตัวอย่าง API	DOM	SAX

ข้อดีและข้อเสียของแต่ละวิธีเกี่ยวกับความเร็วในการเข้าถึงข้อมูลและหน่วยความจำที่ใช้งานในแต่ละกรณีนั้นมีความแตกต่างกัน กรณีข้อมูลมีปริมาณน้อยการใช้หน่วยความจำของทั้งสองวิธีจะไม่แตกต่างกันมาก แต่ความเร็วในการเข้าถึงข้อมูลแบบ Event-based จะมีความเร็วมากกว่าแบบ Object-based กรณีข้อมูลมีปริมาณมาก การใช้หน่วยความจำของ Object-based จะมีการใช้หน่วยความจำที่มากกว่า Event-based ซึ่งส่งผลให้ความเร็วในเข้าถึงข้อมูล ช้าลงไปด้วย

2.4 เทคโนโลยี Really Simple Syndication (RSS)

เทคโนโลยีในการรวบรวมข่าวสารนั้นแต่เดิมเริ่มต้นมาจากเทคโนโลยี push และ pull เทคโนโลยี pull (Franklin and Zdonik, 1998; Wikipedia, 2009) จะต้องมีการร้องขอจากผู้ใช้เพื่อรับข้อมูลที่ต้องการ ในขณะที่เทคโนโลยี push นั้นสามารถส่งข้อมูลมายังผู้ใช้ได้โดยไม่ต้องมีการร้องขอ เซิร์ฟเวอร์จะเป็นผู้ควบคุมและกำหนดเนื้อหาและช่วงเวลาที่จะส่งข้อมูลไปให้ผู้ใช้ ต่อมาในปีค.ศ. 1997 จึงได้มีการพัฒนาเทคโนโลยี RSS ขึ้นเพื่อรวบรวมข้อมูลตามแนวคิดของเทคโนโลยี push นั่นคือ เมื่อแหล่งข่าวมีการนำเสนอข่าวใหม่ๆ จะสามารถแจ้งไปให้ผู้ใช้ทราบได้โดยไม่ต้องมีการร้องขอ อย่างไรก็ตามเทคโนโลยี RSS ก็ยังไม่เป็นเทคโนโลยี push แบบเต็มรูปแบบเนื่องจากเมื่อมีข้อมูลใหม่ผู้ใช้จะสามารถทราบได้โดยการตรวจสอบทุกๆ ช่วงเวลา (polling)

เทคโนโลยี RSS มาจากคำว่า Really Simple Syndication (Ayers and Watt, 2005) RSS คือข้อมูลที่อยู่ในรูปของภาษา XML (Extensible Mark-up Language) โดยเวอร์ชัน 1.0 พัฒนาขึ้นเพื่อใช้ในการรวบรวมข้อมูลข่าวสารจากเว็บไซต์ซึ่งมีการเปลี่ยนแปลงอยู่เสมอ เช่น เว็บข่าว หรือ เว็บบล็อก (weblog) เทคโนโลยี RSS ทำให้ผู้ใช้สามารถรับข่าวสารจากเว็บไซต์ที่มีการให้บริการ RSS ซึ่งจะต้องมีการจัดเตรียมเอกสาร XML เอาไว้ เรียกว่า RSS feed หรือเอกสาร RSS ผู้ใช้จะสามารถเรียกดูเนื้อหาในเอกสาร RSS ได้โดยใช้โปรแกรมที่เรียกว่า Reader หรือ Aggregator เมื่อผู้ใช้ต้องการดูข้อมูลจะต้องพิมพ์ที่อยู่ (URL) ของเอกสาร RSS จากนั้น Reader จะดึงข้อมูลมาแสดงบนเครื่องของผู้ใช้ พร้อมทั้งมีการตรวจสอบการเปลี่ยนแปลงข้อมูลใหม่โดยอัตโนมัติ แต่ข้อมูลที่ได้จะเป็นเพียงโดยย่อในรูปแบบของข้อความ เช่น หัวข้อข่าว เท่านั้น ส่วนเนื้อหาหลักจะมีลิงค์เชื่อมโยงไปอีกต่อหนึ่ง เอกสาร RSS จะประกอบด้วย tag (w3schools, 2009) ซึ่งจะอธิบายข้อมูลส่วนต่างๆ แต่ละเอกสาร RSS จะประกอบด้วย tag <channel> ซึ่งเป็นที่มาของ feed และภายในจะมี sub-tag <item> เป็นรายการข้อมูลย่อยภายใน channel ทั้ง <channel> และ <item> ยังประกอบด้วย tag ข้อมูลต่างๆ แสดงดังตารางที่ 2.3 และ ตารางที่ 2.4

ตารางที่ 2.3 sub-tag ของ <channel>

XML sub-tag ของ <channel>	**คำอธิบาย**	**ประเภท** [O (optional): tag ตัวเลือก R (required): tag บังคับ]
<category>	กำหนดหมวดหมู่ของ feed กำหนดได้มากกว่า 1 หมวดหมู่	O
<cloud>	ขั้นตอนการลงทะเบียนเมื่อผู้เผยแพร่มีการอัพเดทข้อมูลใหม่	O
<copyright>	ข้อมูลลิขสิทธิ์ของ feed	O
<description>	คำอธิบายโดยย่อของข้อมูล	R
<docs>	ระบุ URL ของเอกสารกำหนดรูปแบบที่ใช้กับ feed	O
<generator>	ระบุโปรแกรมที่ใช้แสดง feed	O
<image>	รูปภาพอธิบาย feed	O
<language>	ภาษาที่ใช้ในการเขียน feed	O
<lastBuildDate>	วันเวลาล่าสุดที่ทำการปรับปรุงข้อมูล	O
<link>	ลิงค์ไปยังข้อมูลหลัก	R
<managingEditor>	ที่อยู่อีเมล์ไปยังบรรณาธิการผู้ดูแลเนื้อหาของ feed	O
<pubDate>	วันเวลาที่ทำการเผยแพร่ข้อมูล	O
<rating>	ระบุการจัดอันดับความนิยมของ feed	O
<skipDays>	วันที่ที่ RSS Reader ไม่ต้องตรวจสอบการอัพเดทของเนื้อหา	O
<skipHours>	ชั่วโมงที่ RSS Reader ไม่ต้องตรวจสอบการอัพเดทของเนื้อหา	O
<textInput>	แถบข้อความเข้าที่แสดงใน feed	O
<title>	หัวข้อของ chanel	R
<ttl>	time to live ช่วงระยะเวลาที่ข้อมูลยังคงใช้ได้ ก่อนจะมีการเปลี่ยนแปลงใหม่	O
<webMaster>	ระบุที่อยู่อีเมล์ของผู้พัฒนาเว็บไซต์ feed	O

ตารางที่ 2.4 sub-tag ของ <item>

XML sub-tag ของ <item>	**คำอธิบาย**	**ประเภท** [O (optional): tag ตัวเลือก R (required): tag บังคับ]
<author>	ระบุที่อยู่อีเมล์ของผู้เขียน item	O
<category>	กำหนดหมวดหมู่ของ item กำหนดได้มากกว่า 1 หมวดหมู่	O
<comments>	ลิงค์ไปยังคำวิจารณ์เกี่ยวกับ item	O
<description>	คำอธิบาย item	R
<guid>	globally unique identifier ลิงค์สำหรับระบุแต่ละ item	O
<link>	ลิงค์ของเนื้อหาทั้งหมดของ item	R
<pubDate>	วันเวลาที่ทำการเผยแพร่ข้อมูล	O
<source>	แหล่งที่มาของ item	O
<title>	หัวข้อของ item	R
<enclosure>	ไฟล์มัลติมีเดียที่แนบ (เทคโนโลยี Podcast)	O

ในทุกๆไฟล์จะต้องประกอบไปด้วย tag ใหญ่สุดคือ <rss> ซึ่งสามารถระบุเวอร์ชันของเทคโนโลยี RSS ได้ ภายใน <rss> จะมี tag ลูกเพียง tag เดียวคือ <channel> ตามด้วย tag ย่อย ซึ่งเป็น tag บังคับ ได้แก่ <title> <link> และ <description> และ tag ที่เป็นตัวเลือก จะใส่หรือไม่ก็ได้ ภาพประกอบ 2.14 แสดงตัวอย่างส่วนของเอกสาร RSS

ปัจจุบันได้มีการพัฒนา RSS reader ขึ้นหลายรูปแบบ เพื่อรองรับความต้องการของผู้ใช้ ได้แก่ reader ประเภทที่ต้องทำการติดตั้งก่อนใช้งาน เรียกว่า Software reader โดยมีทั้งที่เป็น Desktop reader ซึ่งทำงานบนเครื่องคอมพิวเตอร์ตั้งโต๊ะ เช่น FeedReader (i-Systems Inc, 2009) หรือที่เป็น Mobile RSS reader สำหรับอุปกรณ์สื่อสารเคลื่อนที่ เช่น LiteFeeds (LiteFeeds, 2008) อีกประเภทคือ Web-based RSS reader ซึ่งเป็นโปรแกรมอยู่บนฝั่งเซิร์ฟเวอร์ เข้าถึงได้ผ่านทางเว็บ ไม่ต้องมีการติดตั้งใดๆบนเครื่องของผู้ใช้ เช่น NewsGator Online (NewsGator ® Technologies Inc, 2009) นอกจากนี้บราวเซอร์ยังได้มีการพัฒนา RSS reader ที่ฝังเข้ากับตัวบราวเซอร์เองอีกด้วย เช่น Mozilla Firefox (Mozilla, 2009)

```
<rss version="2.0">
    <channel>
        <title>Liftoff News</title>
        <link>http://liftoff.msfc.nasa.gov/</link>
        <description>Liftoff to Space Exploration.</description>
        <language>en-us</language>
        <pubDate>Tue, 10 Jun 2003 04:00:00 GMT</pubDate>
        <lastBuildDate>Tue, 10 Jun 2003 09:41:01 GMT</lastBuildDate>
        <docs>http://blogs.law.harvard.edu/tech/rss</docs>
        <generator>Weblog Editor 2.0</generator>

        <item>
            <title>Star City</title>
            <link>http://liftoff.msfc.nasa.gov/news
            /2003/news-starcity.asp</link>
            <description>How do Americans get ready
            to work with Russians aboard the
            International Space Station? They take a
            crash course in culture, language
            and protocol at Russia's Star City.
            </description>
            <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
        </item>
          .
          .
          .
    </channel>
</rss>
```

ภาพประกอบ 2.14 ตัวอย่างของเอกสาร RSS

2.5 เทคโนโลยี Podcast

ปัจจุบันผู้ให้บริการข่าวสารได้พัฒนารูปแบบการเผยแพร่ข่าว ที่แต่เดิมเป็นเพียงข้อความข่าว หรือรูปภาพ มาเป็นข้อมูลประเภทมัลติมีเดีย เช่น ไฟล์การรายงานข่าวในรูปแบบเสียง และวิดีโอข่าว Podcasting เป็นเทคโนโลยีอีกรูปแบบหนึ่งที่ทำให้ผู้ชมสามารถดูข้อมูลมัลติมีเดียได้ ไฟล์มัลติมีเดียเหล่านี้เรียกว่า podcast โดย podcast ที่เป็นข้อมูลวิดีโอเรียกเฉพาะได้ว่า vodcast (video on demand broadcast) ทั้งสองเทคโนโลยีนี้ผู้ใช้สามารถเข้าถึงได้ผ่านเทคโนโลยี RSS ข้อมูลมัลติมีเดียสามารถแทรกไว้ในเอกสาร RSS ได้โดยการใส่ URL ของไฟล์มัลติมีเดียดังกล่าวไว้ใน RSS enclosure

RSS เวอร์ชัน 2.0 นั้นได้เพิ่มความสามารถให้เพิ่ม tag ชื่อว่า ‘enclosure’ เข้าไปใน <item> ซึ่ง <enclosure> มีแอททริบิวต์ดังนี้ 1) “url” บอกที่อยู่ของไฟล์มัลติมีเดีย 2) “length” บอกขนาดของไฟล์เป็นไบต์ และ 3) “type” บอกชนิดของไฟล์มัลติมีเดีย เช่น m4a

mp3 mov mp4 m4v และ pdf รูปแบบโครงสร้าง (syntax) ของ enclosure แสดงดังภาพประกอบ 2.16

```
<enclosure url="http://domain.com/file.mp4" length="123456789"
type="video/mp4">
```

ภาพประกอบ 2.15 รูปแบบโครงสร้างของ RSS enclosure

ด้วยเทคโนโลยีดังกล่าวทำให้ RSS reader สามารถรวบรวมไฟล์มัลติมีเดียที่มีการเผยแพร่บนเว็บไซต์ข่าวต่างๆได้ โปรแกรมรวบรวม Podcast หรือ Podcast Reader ที่ได้รับความนิยมอย่างมากโปรแกรมหนึ่งก็คือ “iTunes” จากค่าย Apple ผู้พัฒนาได้กำหนดเอกสารนิยามโครงสร้างข้อมูล XML หรือ Document Type Definition (DTD) สำหรับ iTunes ขึ้นมา ดังนั้นผู้เผยแพร่ Podcast จึงมีการเพิ่ม tag ข้อมูลตาม DTD ที่กำหนดเพื่อให้ Podcast มีความแพร่หลายมากขึ้นในหมู่ผู้ใช้ซอฟต์แวร์ iTunes (Feed For All. 2009)

เมื่อต้องการใช้งาน tag สำหรับ iTunes จะต้องประกาศ XML namespace ซึ่งชี้ไปยังเอกสาร DTD ที่กำหนด tag ของ iTunes จากภายนอก ดังภาพประกอบ 2.17

ข้อมูลใน tag จะต้องเป็นอักขระธรรมดา ไม่เป็นภาษา markup เช่น HTML และมีความยาวจำกัดคือ 255 อักขระ ยกเว้น <itunes:summary> ซึ่งมีความยาวได้ถึง 4,000 อักขระ ตารางที่ 2.5 แสดง tag ของ iTunes RSS ภาพประกอบ 2.17 แสดงตัวอย่างเอกสาร RSS ของ Podcast ที่มีการใช้งาน tag ของ iTunes

ตารางที่ 2.5 tag ใน iTunes RSS

XML tag ใน iTunes RSS	<channel>	<item>	คำอธิบาย
<link>	มี	ไม่มี	ลิงค์ของเว็บไซต์ที่เผยแพร่
<copyright>	มี	ไม่มี	ข้อมูลลิขสิทธิ์ของ Podcast
<itunes:author>	มี	มี	ชื่อผู้เผยแพร่ Podcast
<itunes:block>	มี	มี	การป้องกันไม่ให้รายการปรากฏบน Podcast Reader
<itunes:category>	มี	ไม่มี	หมวดหมู่ของ channel ซึ่งจะกำหนดโดย iTune แบ่งเป็นหมวดหมู่ และหมวดหมู่ย่อย
<itunes:image>	มี	ไม่มี	ภาพประกอบของ Podcast
<itunes:duration>	ไม่มี	มี	ความยาวของไฟล์ Podcast
<itunes:explicit>	มี	มี	คำอธิบาย หรือคอมโพเนนท์อื่นที่เกี่ยวข้อง
<itunes:keywords>	มี	มี	คำสำคัญสำหรับการค้นหาของผู้ใช้
<itunes:owner>	มี	ไม่มี	ชื่อผู้อ้างสิทธิ์ความเป็นเจ้าของ Podcast
<itunes:subtitle>	มี	มี	คำอธิบายโดยย่อของ channel หรือ item
<itunes:summary>	มี	มี	คำอธิบายโดยสรุปของ channel หรือ item

```
<?xml version="1.0" encoding="UTF-8"?>
<rss xmlns:itunes="http://www.itunes.com/DTDs/Podcast-1.0.dtd"
version="2.0">
   <channel>
      <title>RSS Feed Podcast</title>
      <description>Podcast explaining...</description>
      <link>http://www.rss-specifications.com/rss-articles.htm</link>
      <lastBuildDate>Mon, 25 Jul 2005 16:37:32 -0400</lastBuildDate>
      <pubDate>Mon, 25 Jul 2005 09:00:00 -0400</pubDate>
      <generator>FeedForAll Mac v1.0 (1.0.1.0) unlicensed
      version</generator>
      <itunes:category text="Technology">
         <itunes:category text="Information Technology"/>
      </itunes:category>
      <itunes:subtitle>Podcasting made easy with...</itunes:subtitle>
      <itunes:summary>Podcasting... </itunes:summary>
      <itunes:author>Mac OS Team</itunes:author>
      <itunes:owner>
         <itunes:name>Mac OS Team</itunes:name>
         <itunes:email>macsupport@feedforall.com</itunes:email>
      </itunes:owner>
      <itunes:image>
         <url>http://www.feedforall.com/logo.jpg</url>
         <title>FeedForAll Logo</title>
         <link>http://www.feedforall.com</link>
      </itunes:image>
      <itunes:link rel="image" type="video/jpeg" href="...">Podcasting
      Tutorial</itunes:link>

      <item>
         <title>Tips to Creating a Blog</title>
         <description>Tips for creating a blog...</description>
         <link>http://www.rss-specifications.com </link>
         <enclosure url="http://www.rss-
         specifications.com/podcasts/tips.mp3" length="1157907"
         type="audio/mpeg"></enclosure>
         <pubDate>Mon, 25 Jul 2005 09:00:00 -0400</pubDate>
         <itunes:category text="Technology">
            <itunes:category text="Information Technology"/>
         </itunes:category>
         <itunes:explicit>no</itunes:explicit>
         <itunes:duration>4:49</itunes:duration>
         <itunes:keywords>BLOG create howto</itunes:keywords>
      </item>
        .
        .
        .
   </channel>
</rss>
```

XML namespace ระบุที่อยู่ของเอกสาร DTD ของ iTunes

ภาพประกอบ 2.16 ตัวอย่าง RSS feed ที่ใช้งาน tag ของ iTunes

บทที่ 3

แบบจำลองเทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

วิทยานิพนธ์นี้ได้ทำการออกแบบเทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP โดยจะแบ่งออกเป็น 2 ส่วนคือระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP หรือ Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices (ANVFFS) เพื่อการรวบรวมไฟล์สรุปหัวข้อของวิดีโอข่าว (video feed) จากแหล่งต่างๆ และส่วนของเทคนิคการคัดกรองวิดีโอข่าวที่เหมาะสม ก่อนที่จะถูกส่งไปแสดงผลบนเครื่องของผู้ใช้

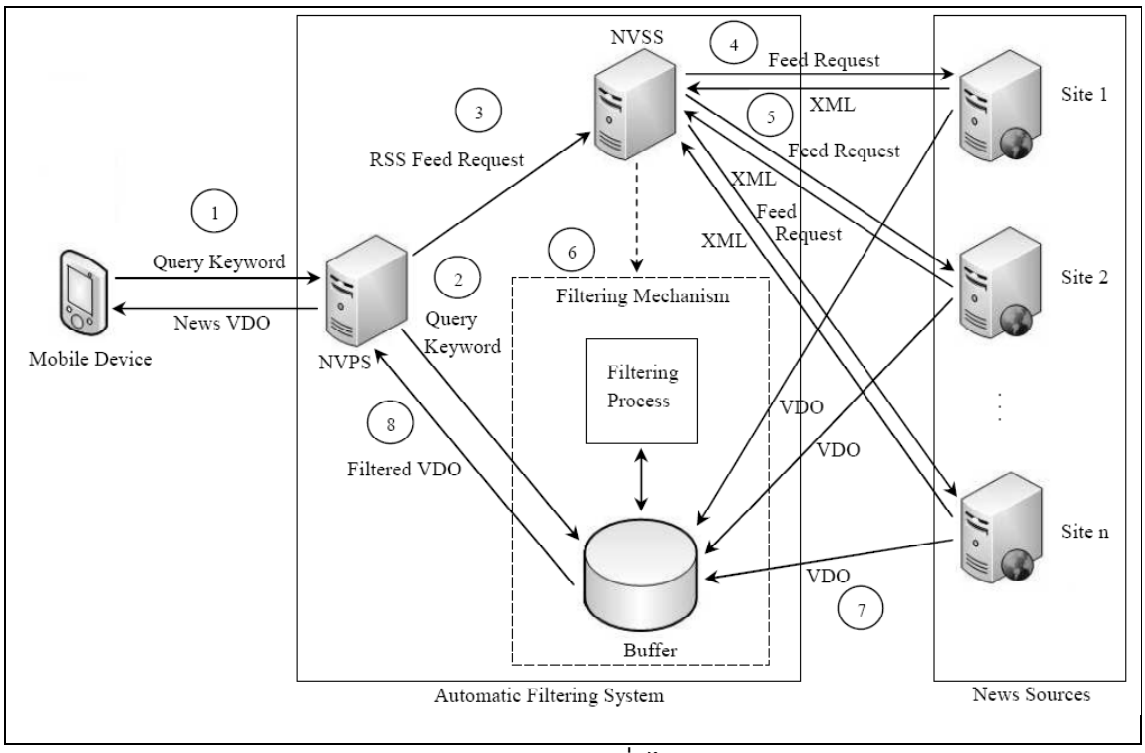
3.1 แบบจำลองของระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP (Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices : ANVFFS)

ในส่วนของ ANVFFS มีการออกแบบในส่วนของสถาปัตยกรรมทั่วไป และฐานข้อมูล ดังนี้

3.1.1 สถาปัตยกรรมทั่วไป

ระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่แบ่งการทำงานออกเป็นส่วนประกอบหลักๆคือ

- News Video Proxy Server (NVPS) เป็นเซิร์ฟเวอร์ซึ่งตรวจสอบการร้องขอของผู้ใช้ว่ามีการประมวลผลไปแล้วหรือไม่ หากใช่จะส่งข้อมูลวิดีโอที่ได้คัดกรองไว้แล้วไปได้ทันที
- News Video System Server (NVSS) เป็นเซิร์ฟเวอร์หลักซึ่งจะมีส่วนของการให้บริการหน้าเว็บในการรับการร้องขอจากผู้ใช้ และส่วนประมวลผลเพื่อดึงเอกสาร RSS และไฟล์วิดีโอมาทำการคัดกรองวิดีโอที่เหมาะสม
- Data Repository ฐานข้อมูลที่ใช้เก็บข้อมูลที่สกัดออกมาจากเอกสาร RSS และจากไฟล์วิดีโอ



ภาพประกอบ 3.1 สถาปัตยกรรมทั่วไปของระบบ ANVFFS

ส่วนประกอบต่างๆทำงานร่วมกันเป็นขั้นตอนดังแสดงในแบบจำลองดังภาพประกอบ 3.1 อธิบายได้ดังนี้

1) ผู้ใช้บริการร้องขอจากหน้าเว็บไซต์ โดยกรอกคำสำคัญ (keyword) ของข่าวที่ต้องการรับชม

2) NVPS จะทำการตรวจสอบว่ามีการร้องขอดังกล่าวมาแล้วก่อนหน้านี้หรือไม่ ถ้ามีอยู่แล้วแสดงว่าได้มีการดึงวิดีโอข่าวมาทำการคัดกรองแล้ว จึงข้ามไปขั้นตอนที่ 8 ส่งวิดีโอไปแสดงผลบนหน้าจอของผู้ใช้ได้เลย

3) หากเป็นการร้องขอใหม่ NVPS จะส่งผ่านการร้องขอไปยัง NVSS

4) NVSS ทำหน้าที่ในการส่งการร้องขอเอกสาร RSS ซึ่งกำหนดโดยผู้ดูแลระบบไปยังเว็บไซต์ผู้ให้บริการข่าวต่างๆ โดยจะทำหน้าที่ในลักษณะของ Back-end คอยอัพเดทข่าวทุกๆช่วงเวลาที่กำหนด

5) แต่ละเซิร์ฟเวอร์ของผู้ให้บริการข่าวจะส่งเอกสาร RSS ในรูปแบบของเอกสาร XML กลับมายัง NVSS

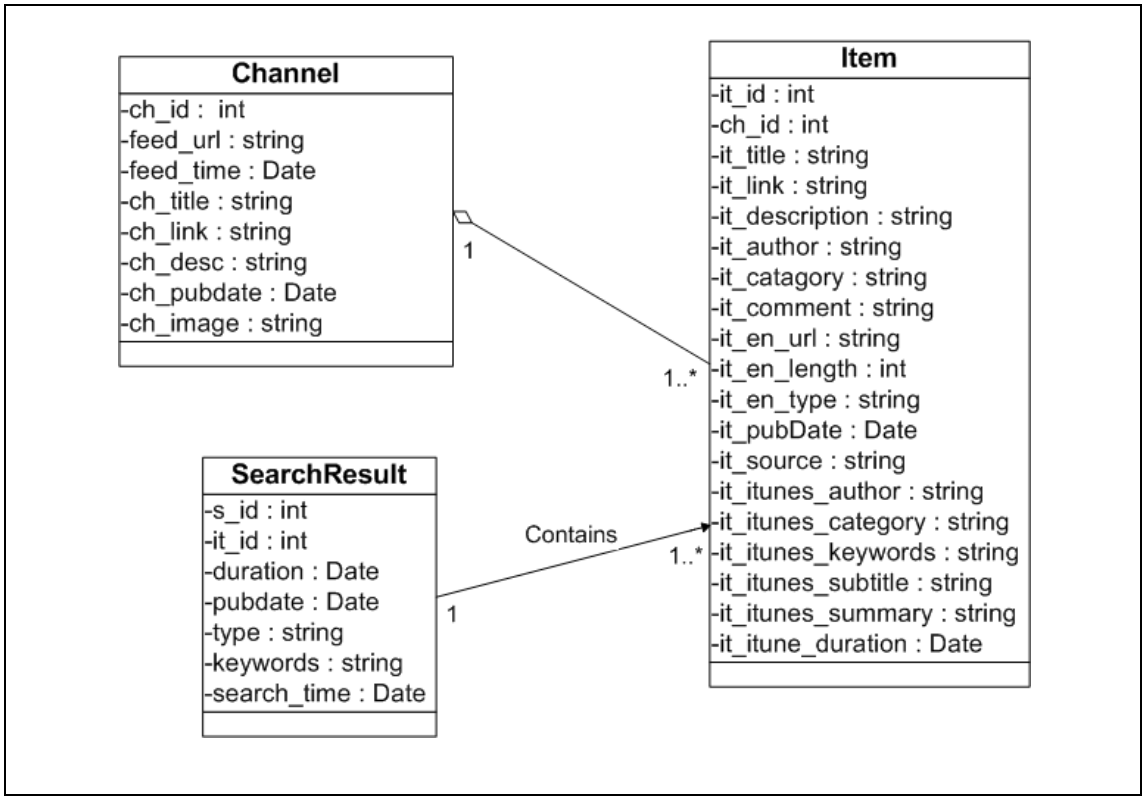
6) NVSS จะสกัดเอกสาร RSS ซึ่งอยู่ในรูปของ XML เพื่อนำข้อมูลต่างๆ เช่น หัวข้อข่าว (<title>) คำอธิบายข่าว (<description>) และที่อยู่ของไฟล์วิดีโอ (<enclosure>) มาเก็บไว้ในฐานข้อมูลของระบบ

7) วิดีโอข่าวที่เกี่ยวข้องกับคำสำคัญ ถูกดึงมาผ่านกลไกการคัดกรอง (filtering process) เพื่อให้ได้วิดีโอที่เหมาะสม

8) วิดีโอที่ได้ทำการคัดกรองออกมาแล้วจะถูกส่งไปยัง NVPS เพื่อแสดงผ่านทางบราวเซอร์บนอุปกรณ์ของผู้ใช้ โดยการคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมสามารถทำได้โดยเลือกจากคุณสมบัติต่างๆ เช่น เลือกจากความยาวของวิดีโอข่าว (duration) ขนาดของวิดีโอ (size) ชนิดของวิดีโอ (format) หรือตามเนื้อหาของวิดีโอ (content)

3.1.2 ฐานข้อมูลของระบบ

ฐานข้อมูลของระบบ ANVFFS เป็นฐานข้อมูลชั่วคราว (buffer) ซึ่งจะเก็บข้อมูลที่ได้จากการสกัดเอกสาร RSS ที่จำเป็นสำหรับขั้นตอนการคัดกรองเอกสาร RSS แต่ละไฟล์จะมีข้อมูล 1 channel และในแต่ละ channel จะประกอบด้วย item ตั้งแต่ 1 item ขึ้นไป ภาพประกอบ 3.2 แสดงโครงสร้างของฐานข้อมูลที่ทำการออกแบบ



ภาพประกอบ 3.2 โครงสร้างฐานข้อมูลระบบ

จากภาพ แบ่งตารางข้อมูลออกเป็น 3 ตาราง

1) ตาราง Channel เปรียบเสมือนช่องรายการเผยแพร่ข่าว มีแอททริบิวต์ต่างๆ ดังตารางที่ 3.1

ตารางที่ 3.1 แอททริบิวต์ของตาราง Channel

Attribute	คำอธิบาย
ch_id	รหัส channel
feed_url	ที่อยู่ของเอกสาร RSS ที่รวบรวมมา
feed_time	วันเวลาที่รวบรวมเอกสาร RSS
ch_title	หัวข้อ channel
ch_link	ลิงค์ของ channel
ch_desc	คำอธิบาย channel
ch_pubdate	วันเวลาที่เผยแพร่เอกสาร RSS
ch_image	รูปภาพที่เกี่ยวข้องของ channel

2) ตาราง Item ด้วยความนิยมของโปรแกรมรวบรวมมัลติมีเดีย RSS เช่น iTunes ทำให้ผู้เผยแพร่ข่าวเพิ่ม tag ของ iTunes เข้าไปเพื่อให้การทำงานผ่านเทคโนโลยี RSS นั้นสามารถใช้งานได้อย่างกว้างขวาง tag ของ iTunes ที่เพิ่มเข้ามามีข้อมูลที่เป็นประโยชน์สำหรับการคัดกรองวิดีโอข่าว หรือแม้กระทั่งผู้เผยแพร่บางรายได้ตัด tag มาตรฐานออกจากเอกสาร RSS แล้วแทนที่ด้วย tag ของ iTunes ตารางนี้จึงได้ถูกออกแบบให้สามารถจัดเก็บข้อมูลสำหรับ iTunes ร่วมกับข้อมูลของเอกสาร RSS ทั่วไป แอททริบิวต์ต่างๆของตาราง Item ดังแสดงในตารางที่ 3.2

ตารางที่ 3.2 แอททริบิวต์ของตาราง Item

Attribute	การใช้งาน
it_id	รหัส item
ch_id	รหัส channel ที่บรรจุ item นั้นๆ
it_title	หัวข้อ item
it_link	ลิงค์ของ item
it_description	คำอธิบาย item
it_author	ผู้เขียน item
it_catagory	ประเภทของ item
it_comment	ลิงค์ไปยังคำวิจารณ์เกี่ยวกับ item
it_en_url	ที่อยู่ไฟล์มัลติมีเดีย
it_en_length	ขนาดของไฟล์มัลติมีเดียเป็นไบต์
it_en_type	ชนิดของไฟล์มัลติมีเดีย
it_pubDate	วันเวลาที่เผยแพร่ item
it_source	แหล่งข้อมูลของ item
it_itunes_author	iTunes: ผู้เขียน item
it_itunes_category	iTunes: ประเภทของ item
it_itunes_keywords	iTunes: คำสำคัญที่เกี่ยวข้องกับ item
it_itunes_subtitle	iTunes: หัวข้อย่อยของ item
it_itunes_summary	iTunes: คำอธิบายโดยสรุปของ item
it_itune_duration	iTunes: ความยาวของไฟล์มัลติมีเดียที่แนบ

อย่างไรก็ตามยังมี tag ที่มีการกำหนดโครงสร้างสำหรับโปรแกรมเฉพาะอื่นๆ แต่ไม่เป็นที่นิยมเท่ากับ tag ของ iTunes

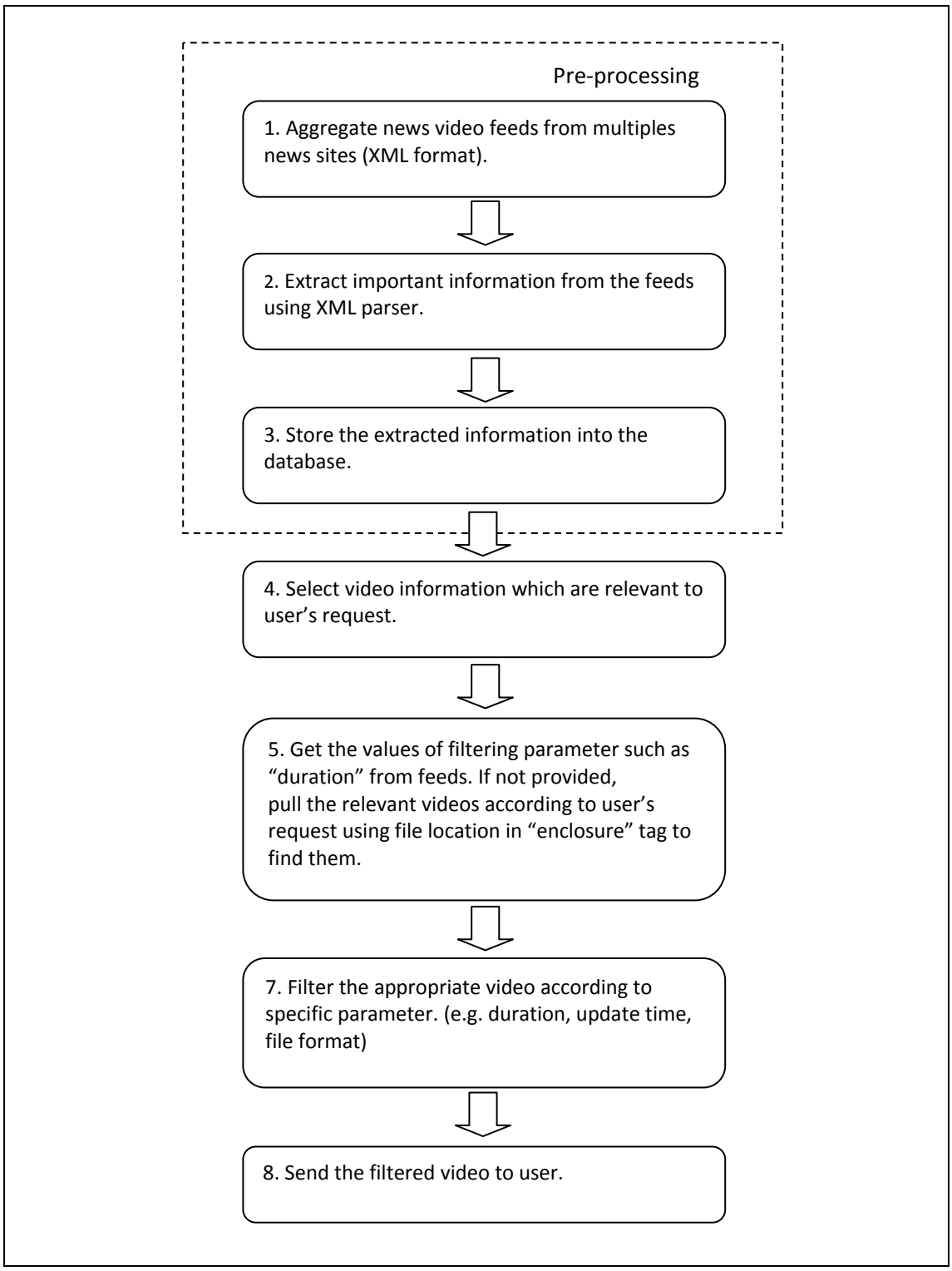
3) ตาราง SearchResult เป็นตารางที่เก็บข้อมูลประวัติการค้นหาของผู้ใช้ซึ่งจะใช้เมื่อมีการร้องขอมายัง NVPS แล้วตรวจสอบว่าซ้ำกับการร้องขอก่อนหน้าหรือไม่ เพื่อให้สามารถดึงผลจากการร้องขอก่อนหน้าไปใช้ได้ทันที และเก็บค่าของคุณสมบัติที่จะใช้ในขั้นตอนการคัดกรองวิดีโอข่าวที่เหมาะสม แอททริบิวต์ต่างๆของตาราง SearchResult แสดงดังตารางที่ 3.3

ตารางที่ 3.3 แอททริบิวต์ของตาราง SearchResult

Attribute	การใช้งาน
s_id	รหัสการค้นหา
it_id	รหัส item ที่ตรงกับคำสำคัญที่ผู้ใช้ร้องขอ
duration	ความยาวของไฟล์มัลติมีเดียที่ตรงกับการค้นหา
pubdate	วันเวลาเผยแพร่ item ที่ตรงกับการค้นหา
type	ชนิดของไฟล์มัลติมีเดียที่ตรงกับการค้นหา
keywords	คำสำคัญที่ผู้ใช้ค้นหา
search_time	วันเวลาที่ผู้ใช้ค้นหา

3.2 เทคนิคการคัดกรองวิดีโอข่าวที่เหมาะสม

ในขั้นตอนของกลไกการคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสม จะต้องมีส่วนของ Pre-processing ประมวลผลเพื่อให้ได้มาซึ่งข้อมูลต่างๆที่สกัดจากเอกสาร RSSในส่วนนี้จะทำงานโดย parser จากนั้นข้อมูลที่ได้จึงนำไปทำการคัดกรองต่อไป ลำดับขั้นตอนต่างๆของการคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมแสดงในภาพประกอบ 3.3



ภาพประกอบ 3.3 ลำดับขั้นตอนกลไกการคัดกรองวิดีโอข่าวที่เหมาะสม

จากภาพอธิบายลำดับขั้นตอนได้ดังนี้

1) ทำการรวบรวมเอกสาร RSS จากเว็บไซต์ผู้ให้บริการข่าวซึ่งจะอยู่ในรูปของเอกสาร XML ในขั้นตอนนี้ระบบจะทำงานเป็น Back-end คอยรวบรวม feed ใหม่ทุกช่วงเวลาเพื่อให้ได้ข่าวล่าสุด

2) สกัดข้อมูลที่จำเป็นในการคัดกรองเช่น title, description, enclosure ด้วย XML parser

3) จัดเก็บข้อมูลดังกล่าวในฐานข้อมูล

4) ทำการเปรียบเทียบการร้องขอจากผู้ใช้ว่ามีวิดีโอข่าวเรื่องใดที่เกี่ยวข้องกับคำสำคัญของผู้ใช้บ้าง ซึ่งในส่วนนี้สามารถที่จะเปรียบเทียบคำสำคัญกับ title และ description ที่ได้จัดเก็บไว้ในฐานข้อมูลแล้ว จะได้ออกมาเป็นกลุ่ม ข้อมูลของวิดีโอ

5) หลังจากได้ข้อมูลของกลุ่มวิดีโอที่เกี่ยวข้องแล้วตรวจสอบเพื่อหาคุณสมบัติที่ต้องการนำมาคัดกรอง เช่นใช้ความยาวของไฟล์ (duration) หากไม่มีข้อมูลดังกล่าวจากการสกัดจากเอกสาร RSS ให้ทำการดึงวิดีโอข่าวเหล่านั้นตามที่อยู่ของไฟล์ใน enclosure เพื่อนำมาหาค่าของคุณสมบัติที่ต้องการ

6) จากค่าของคุณสมบัติที่ได้ นำไปคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่โดยสามารถพิจารณาจากตัวแปรต่างๆเช่น ระยะเวลาของวิดีโอ ขนาด เวลาที่เผยแพร่เอกสาร RSS หรือชนิดของไฟล์

7) ส่งวิดีโอที่ได้คัดกรองแล้วไปยังผู้ใช้

รายละเอียดการทำงานของขั้นตอนต่างๆเขียนเป็นขั้นตอนวิธีได้ดังนี้

3.2.1 ขั้นตอนวิธีของ parser

การออกแบบ parser ได้ออกแบบการทำงานในลักษณะของ Event-based ข้อมูลของเอกสาร RSS ที่รวบรวมจากแหล่งต่างๆจะต้องมีการอัพเดทอยู่ตลอดเวลา ในการรวบรวมแต่ละครั้งจะมีการเก็บไว้ในฐานข้อมูลชั่วคราว จึงสามารถใช้วิธีแบบ Event-Based อ่านเอกสาร RSS ไปตามลำดับเพียงครั้งเดียว การเรียกใช้ข้อมูลครั้งต่อไปสามารถเรียกจากฐานข้อมูลโดยตรงไม่ต้องอ่านเอกสาร XML ใหม่ ซึ่งจะเร็วกว่าแบบ Object-Based ที่ต้องมีการสร้างโครงสร้างต้นไม้ (Tree) ทุกๆครั้งที่ต้องรวบรวมข่าวใหม่ และเก็บไว้ในหน่วยความจำตลอดเวลา

ถึงแม้จะมี API (Application Programmable Interface) เช่น SAX ในการพัฒนา parser ลักษณะ Event-based แต่การใช้งานมีความซับซ้อน และเนื่องจากเอกสาร RSS นั้นยังไม่มีโครงสร้างที่เป็นมาตรฐานที่แน่นอน จึงได้ออกแบบ parser ด้วยขั้นตอนวิธีใหม่เพื่อให้สามารถเก็บข้อมูลเฉพาะ tag ที่ต้องการลงในฐานข้อมูล และสามารถเรียกใช้ข้อมูลได้โดย

ไม่ต้องผ่าน parser ใหม่ทุกครั้ง แต่จะใช้เฉพาะเมื่อต้องการรวบรวมข่าวที่มีการแก้ไข, เปลี่ยนแปลงใหม่เท่านั้น ขั้นตอนวิธีโดยรวมของ parser ที่ได้ออกแบบแสดงดังภาพประกอบ 3.4

```
1  Method parseRSS(feedURLs)
2    for each feedURL
3      if(invalid URL)
4        download feed content from feedURL
5        get data from <channel> tag
6        insert into “Channel” table in database
7        for each  item in the feed
8          get data in <item> tag
9          insert into “Item” table in database
10       end for
11     else
12       report invalidation
13     end if
14   end for
15 end method
```

ภาพประกอบ 3.4 ขั้นตอนวิธีของ parser

parser จะรับ URL ของแต่ละเว็บไซต์ผู้ให้บริการข่าวมาเพื่อดึงเอกสาร RSS จากเซิร์ฟเวอร์ต่างๆ จากนั้นจะทำการอ่านเนื้อหาในเอกสาร XML ที่ได้ ตามลำดับ เมื่อพบ tag <channel> จะดึงข้อมูลภายในมาเก็บลงในฐานข้อมูล และในแต่ละ item ของ channel ก็จะดึงข้อมูลภายใน รวมถึงและ tag ของ iTunes มาเก็บลงในฐานข้อมูล เพื่อเรียกใช้ในขั้นตอนการคัดกรองต่อไป

3.2.2 ขั้นตอนวิธีของการคัดกรองวิดีโอที่เหมาะสม

ในขั้นตอนการคัดกรองเพื่อเลือกวิดีโอที่เหมาะสมได้ออกแบบโดยเลือกใช้คุณสมบัติ ความยาวของวิดีโอ (duration) เนื่องจากมีความสัมพันธ์กับเวลาในการดาวน์โหลดและขนาดของไฟล์ ถึงแม้ว่าภายใน tag <enclosure> จะมีแอททริบิวต์ “length” หรือขนาดของไฟล์ซึ่งมีหน่วยเป็นไบต์ แต่ก็ไม่ได้เป็นแอททริบิวต์บังคับ โดยส่วนมากจะไม่ระบุค่าของแอททริ-บิวต์นี้ หรือระบุเป็น 0 ซึ่งไม่ใช่ค่าจริง ส่วน duration นั้นจะระบุไว้ใน tag <itune:duration> ในเอกสาร RSS ส่วนใหญ่ที่สนับสนุน iTunes RSS Reader นอกจากนี้ยังมีคุณสมบัติ filetype ที่สามารถคัดกรองวิดีโอออกเป็นกลุ่มๆ และ pubDate ที่บอกความอัพเดทของวิดีโอข่าวได้ แต่ไม่สามารถนำมาพิจารณาในประเด็นการใช้ทรัพยากรและเวลาที่เหมาะสมกับอุปกรณ์สื่อสารเคลื่อนที่ รวมทั้งได้มีการออกแบบในส่วนของ NVSS ให้ทำงานเป็น Back-end รวบรวมข่าว

ทุกๆช่วงเวลาเป็นการอัพเดทในขั้นหนึ่งแล้ว ขั้นตอนวิธีของการคัดกรองวิดีโอข่าวที่เหมาะสม แสดงดังภาพประกอบ 3.5

```
1   Method filterByDuration(keywords)
2     If(duplicate request has been made)
3       send the filtered video to user
4     else
5       separate keywords to tokens of keyword
6     for each keyword
7       find the relevant video using information such as <description> <itune:summary>
8       for each relevant video
9         get filtering property, in this case <duration>
10        insert into "SearchResult" table in database
11      end for
12    end if
13    for relevant video with no duration provided
14      get video
15      get duration from video object and store in the database
16    end for
17    select video with least duration
18    show to user
19  end method
```

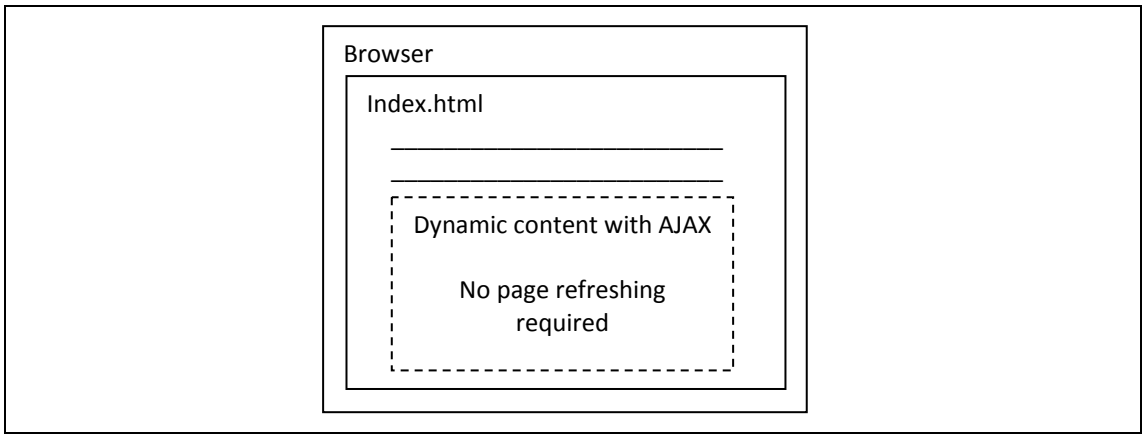
ภาพประกอบ 3.4 ขั้นตอนวิธีของการคัดกรองวิดีโอข่าวที่เหมาะสม

เมื่อรับคำสำคัญ (keyword) มาจากผู้ใช้ในรูปแบบสายอักขระ (string) แล้วนำไปตรวจสอบในฐานข้อมูลประวัติการค้นหา (ตาราง searchResult) ว่าตรงกับการร้องขอที่เคยร้องขอมาแล้วหรือไม่ หากใช่ สามารถส่งวิดีโอที่ได้คัดกรองไว้แล้วไปให้ผู้ใช้ได้ทันที หากยังไม่เคยร้องขอ นำคำสำคัญมาตัดแบ่งเป็นคำสำคัญย่อย แล้วเปรียบเทียบหาวิดีโอที่เกี่ยวข้องกับคำสำคัญนั้น จากกลุ่มของวิดีโอที่เกี่ยวข้องดึงค่าของคุณสมบัติที่ต้องการนำมาคัดกรอง ในที่นี้คือ duration ไปเก็บไว้ในฐานข้อมูล วิดีโอที่ไม่มีค่าของ duration ที่อ่านได้จากเอกสาร RSS ให้ทำการดาวน์-โหลดไฟล์วิดีโอมาเพื่อหาค่า เมื่อได้ค่า duration ของแต่ละวิดีโอแล้ว เลือกวิดีโอที่มีค่า duration น้อยที่สุด ส่งไปยังผู้ใช้

บทที่ 4

โปรแกรมคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

วิทยานิพนธ์นี้ได้ทำการออกแบบเทคนิคและแบบจำลองของระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP (Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices : ANVFFS) และได้ทำการพัฒนาระบบตามแบบจำลองดังกล่าว ซึ่งใช้งานได้กับ Windows Mobile 5 Emulator บราวเซอร์บนอุปกรณ์สื่อสารเคลื่อนที่ที่ใช้ในการทดสอบระบบคือ Opera for Windows Mobile ส่วนที่ติดต่อกับผู้ใช้ได้พัฒนาเป็น web-based สามารถเข้าถึงได้ผ่านเครือข่ายอินเตอร์เน็ต และได้มีการนำเทคโนโลยี AJAX (Asynchronous JavaScript and XML) เข้ามาใช้ ซึ่งเป็นเทคโนโลยีที่มีการทำงานร่วมกัน ระหว่าง JavaScript กับ XML เป็นการลดทรัพยากรในการแสดงผลเนื่องจากจะเปลี่ยนแปลงเนื้อหาเฉพาะส่วนที่จะเป็นเท่านั้น ลักษณะการแสดงผลด้วย AJAX แสดงดังภาพประกอบ 4.1

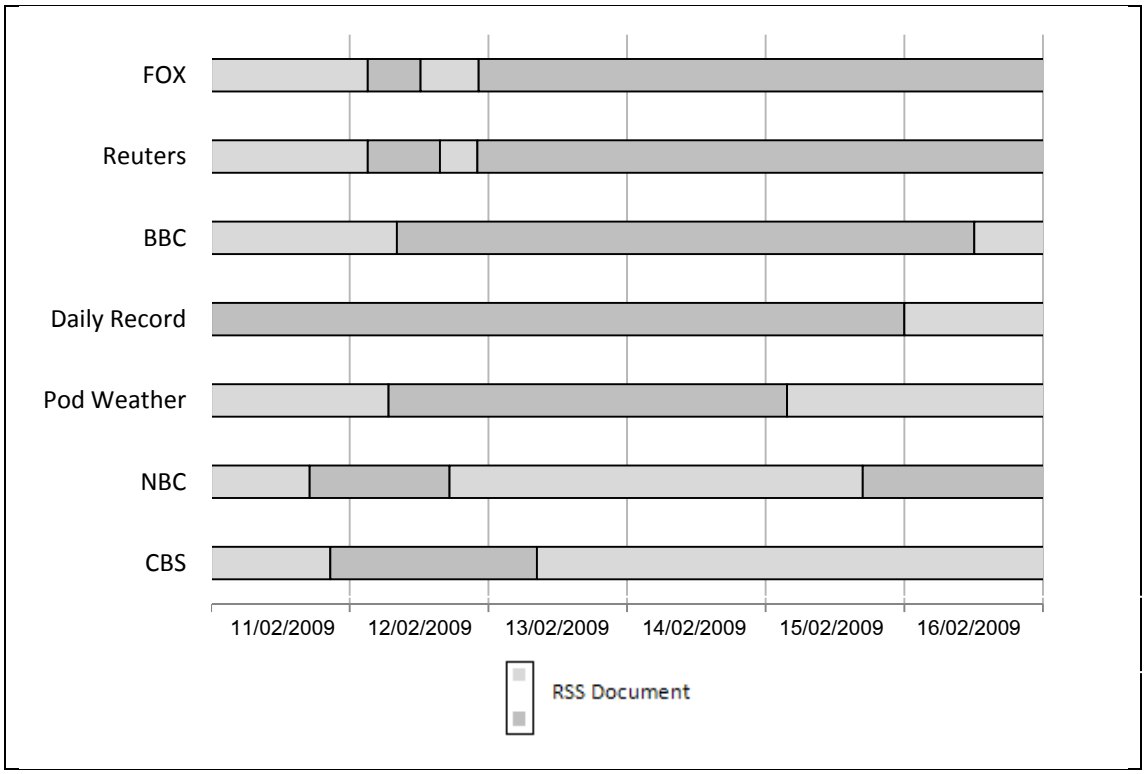


ภาพประกอบ 4.1 ลักษณะการแสดงผลด้วย AJAX

การพัฒนาแบ่งออกเป็นส่วนของผู้ดูแลระบบ และส่วนของผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

4.1 ส่วนของผู้ดูแลระบบ

ในส่วนของหน้าเว็บสำหรับผู้ดูแลระบบนี้ ผู้ดูแลระบบสามารถสั่งเริ่มต้นการทำงานในลักษณะเป็น Back-end โดยจะทำการรวบรวมเอกสาร RSS ของวิดีโอข่าว หรือ podcast มาจากเซิร์ฟเวอร์ของผู้ให้บริการต่างๆมาเก็บไว้ในฐานข้อมูลของระบบ ANVFFS เพื่อสกัดข้อมูลที่สำคัญ และจะทำการรวมรวมใหม่ทุกๆช่วงเวลาที่กำหนด (interval) ซึ่งกำหนดเพื่อให้สัมพันธ์กับช่วงระยะเวลาการอัพเดทเอกสาร RSS บนเว็บไซต์ผู้ให้บริการข่าว ช่วงเวลาของการอัพเดทสามารถศึกษาได้จาก tag <pubDate> ซึ่งแสดงวันเวลาในการเผยแพร่ และในบางเว็บไซต์อาจไม่ระบุ pubDate แต่จะระบุ tag <ttl> คือช่วงเวลาเป็นนาทีที่เนื้อหานั้นทันต่อเหตุการณ์ก่อนที่แหล่งข่าวจะอัพเดทใหม่ (time to live) จากการศึกษาพบว่าเว็บไซต์ผู้ให้บริการข่าวต่างๆ มีช่วงเวลาของการอัพเดทเอกสาร RSS แสดงดังภาพประกอบ 4.1 เก็บข้อมูลตั้งแต่วันที่ 13 กุมภาพันธ์ ถึง 16 กุมภาพันธ์ 2552



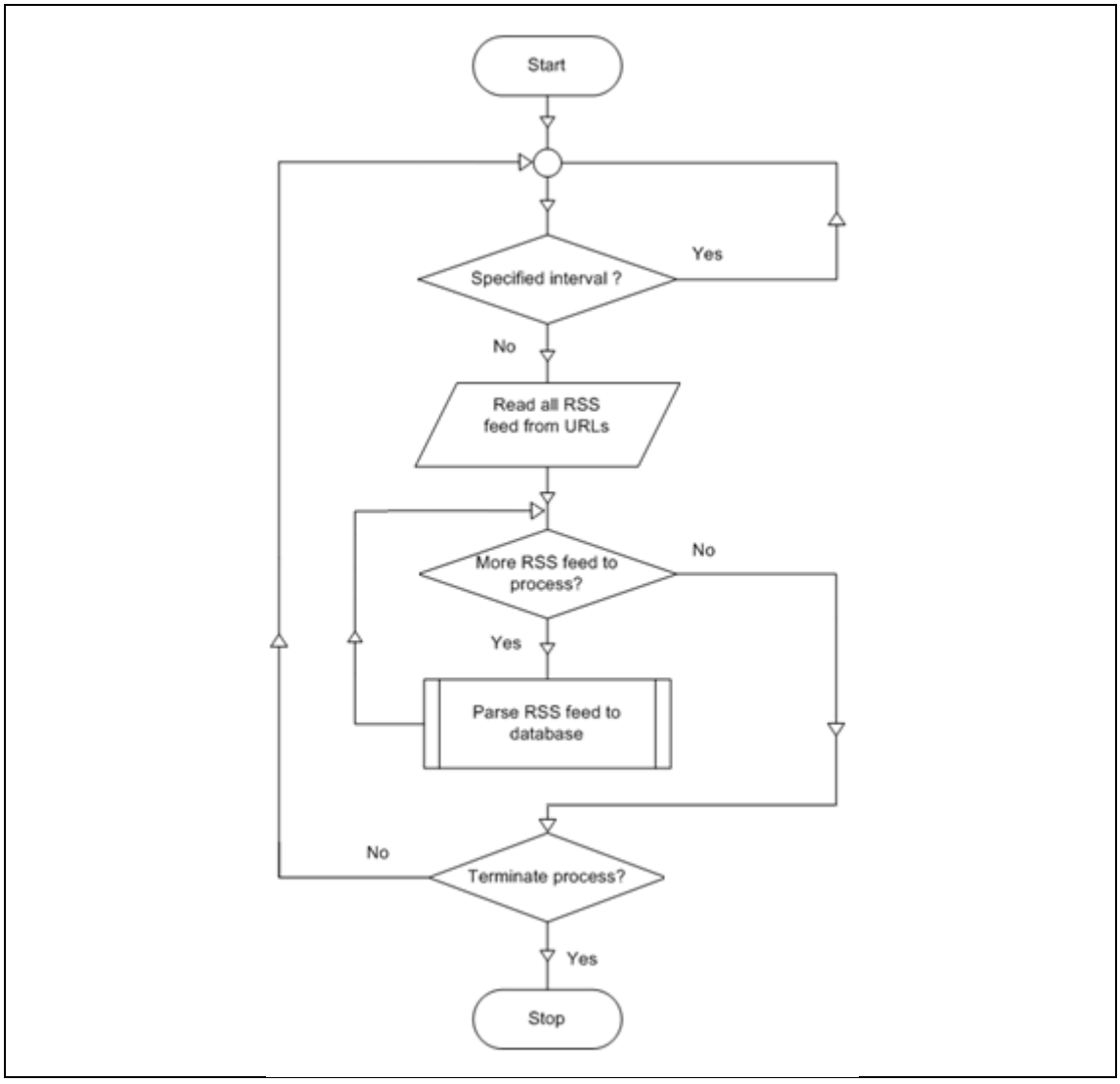
ภาพประกอบ 4.1 สถิติการอัพเดทเอกสาร RSS บนเว็บไซต์

นอกจากนี้ยังมีเว็บไซต์ที่กำหนดช่วงระยะเวลาที่ข่าวคงความทันต่อเหตุการณ์ (ttl) ได้แก่ Pod Weather = 24 ชม. Blogdigger = 1 ชม. ABC = 1 ชม. CNN = 4 ชม. และ

The Washington Post = 1 ชม. งานวิจัยนี้จึงได้กำหนดช่วงเวลาในการรวบรวมข่าวทุกๆ 1 ชม. เพื่อให้ข่าวทันต่อเหตุการณ์

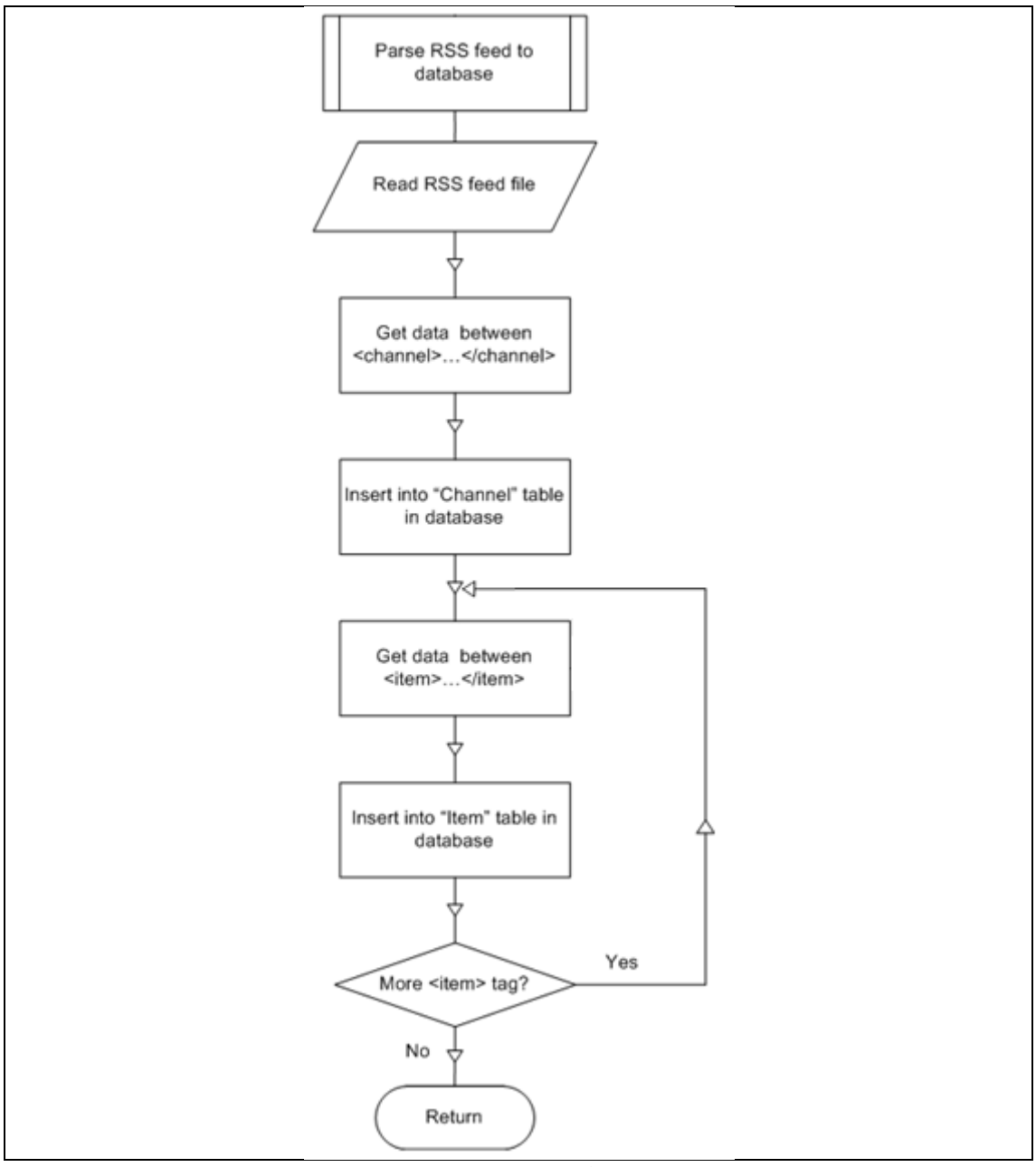
4.1.1 การทำงานในส่วนของผู้ดูแลระบบ

ผังการทำงานในส่วนของผู้ดูแลระบบแสดงดังภาพประกอบ 4.2



ภาพประกอบ 4.2 ผังการทำงานในส่วนของผู้ดูแลระบบ

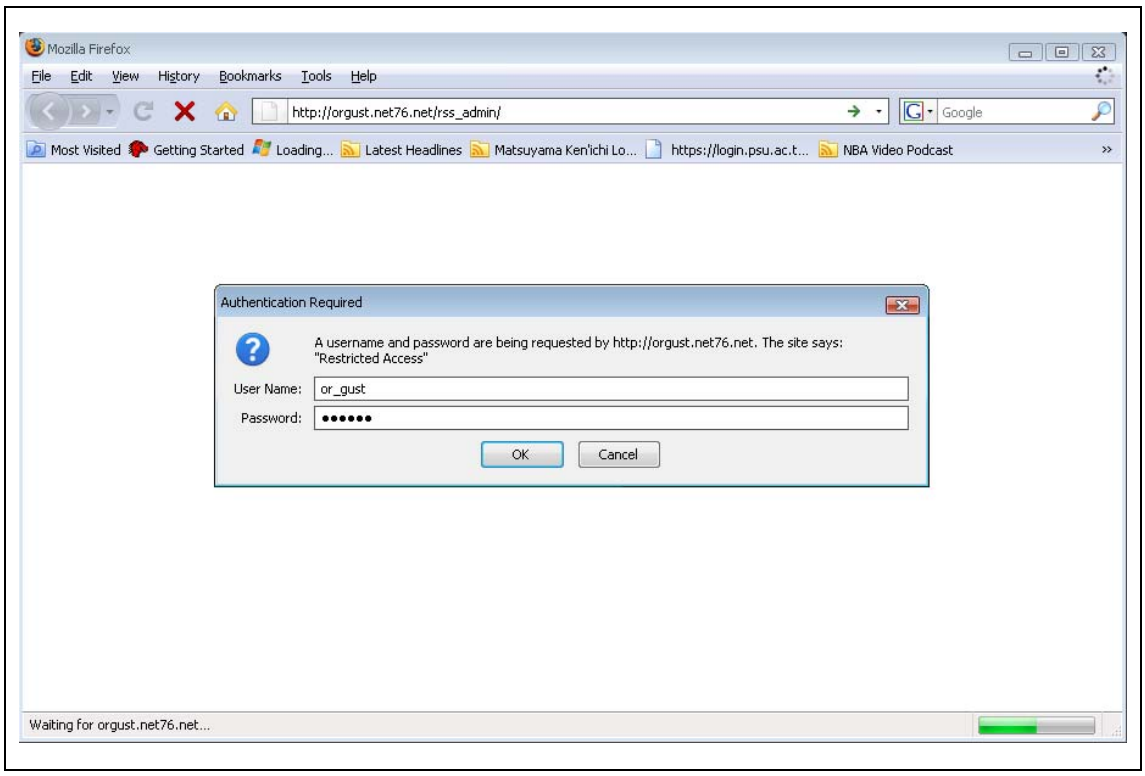
ผังการทำงานย่อยในส่วนของ parser แสดงดังภาพประกอบ 4.3



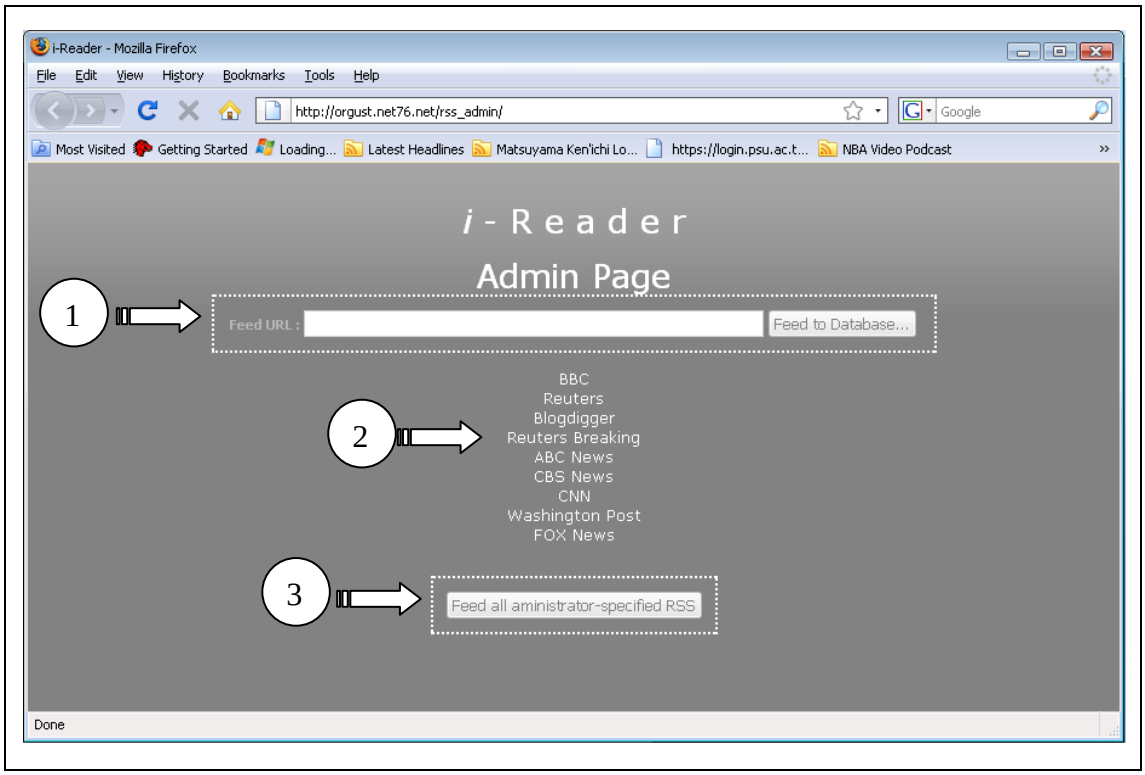
ภาพประกอบ 4.3 ผังการทำงานในส่วนของ parser

4.1.2 ส่วนแสดงผลของผู้ดูแลระบบ

เมื่อเข้าถึงหน้าเว็บในส่วนของผู้ดูแลระบบจะปรากฏหน้าต่างเพื่อยืนยันตัวตน ดังภาพประกอบ 4.4 จากนั้นจะเข้าสู่หน้าเว็บของผู้ดูแลระบบแสดงดังภาพประกอบ 4.5



ภาพประกอบ 4.4 หน้าต่างยืนยันตัวตนในส่วนของผู้ดูแลระบบ



ภาพประกอบ 4.5 หน้าเว็บในส่วนของผู้ดูแลระบบ

หมายเลข 1: เป็นส่วนของช่องรับ URL เพื่อให้ดึงเอกสาร RSS ตามที่อยู่ที่ระบุจากเซิร์ฟเวอร์ของผู้ให้บริการข่าว และทำการสกัดข้อมูล XML ภายในเอกสาร RSS มาเก็บลงในฐานข้อมูล

หมายเลข 2: ส่วนแสดงเว็บไซต์ของผู้ให้บริการข่าวต่างๆ ที่กำหนดโดยผู้พัฒนาระบบ ที่ต้องการรวบรวม

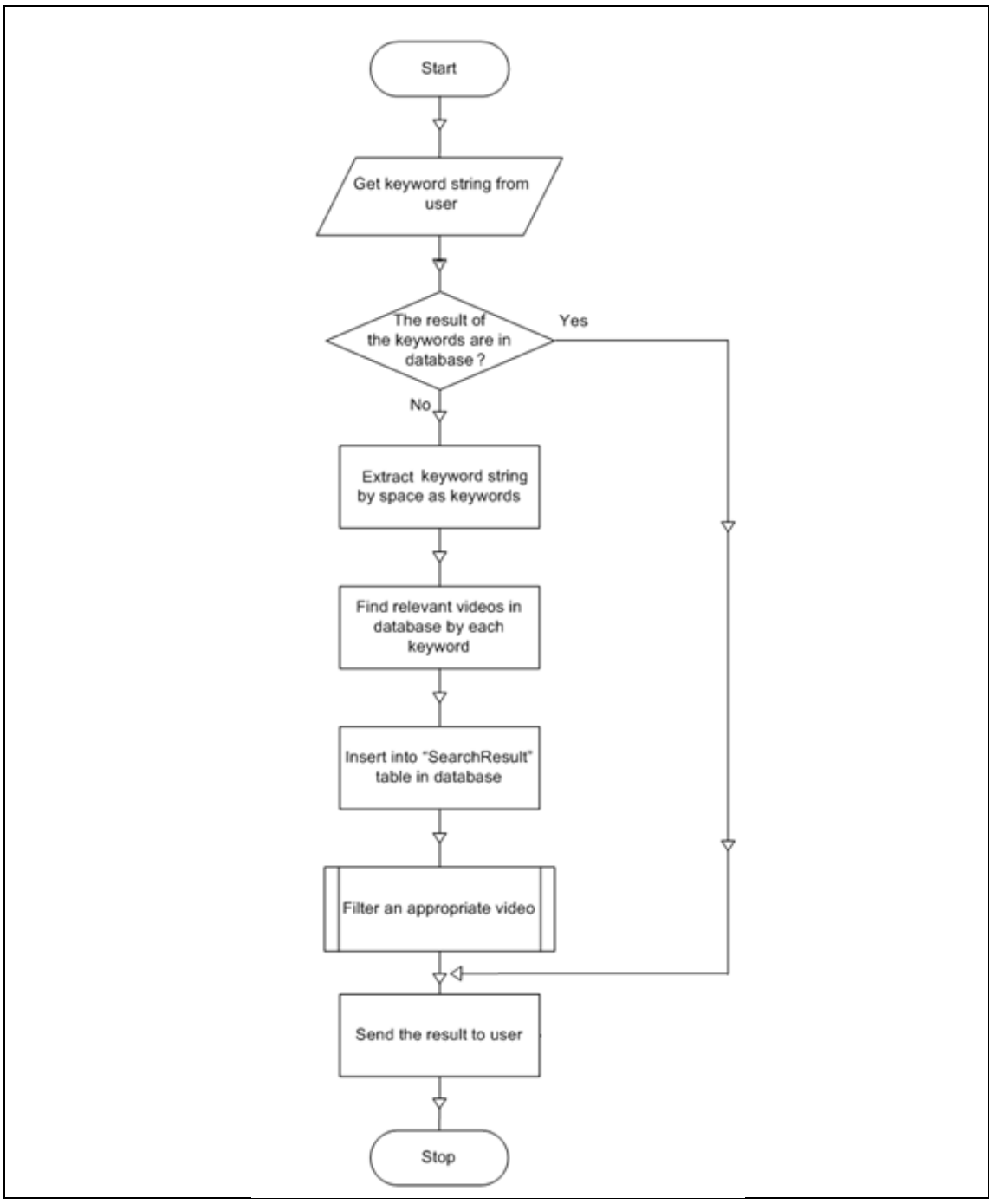
หมายเลข 3: ส่วนของการสั่งงานเริ่มต้นการรวบรวมเอกสาร RSS ทั้งหมดที่ระบุในหมายเลข 2 และสกัดข้อมูล XML ภายในเอกสาร RSS มาเก็บลงในฐานข้อมูล และทำการวนซ้ำรวบรวมใหม่ทุกๆช่วงเวลาตามที่กำหนดไว้

4.2 ส่วนของผู้ใช้งานอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ในส่วนนี้ผู้ใช้สามารถเข้าถึงได้ผ่านเครือข่ายอินเตอร์เน็ตเช่นเดียวกัน ระบบจะรับข้อมูลเข้ามาเป็นคำสำคัญของวิดีโอที่ผู้ใช้ต้องการเรียกดู และคุณสมบัติที่ต้องการ ผ่านกระบวนการคัดกรองเพื่อให้ได้วิดีโอข่าวที่เหมาะสม

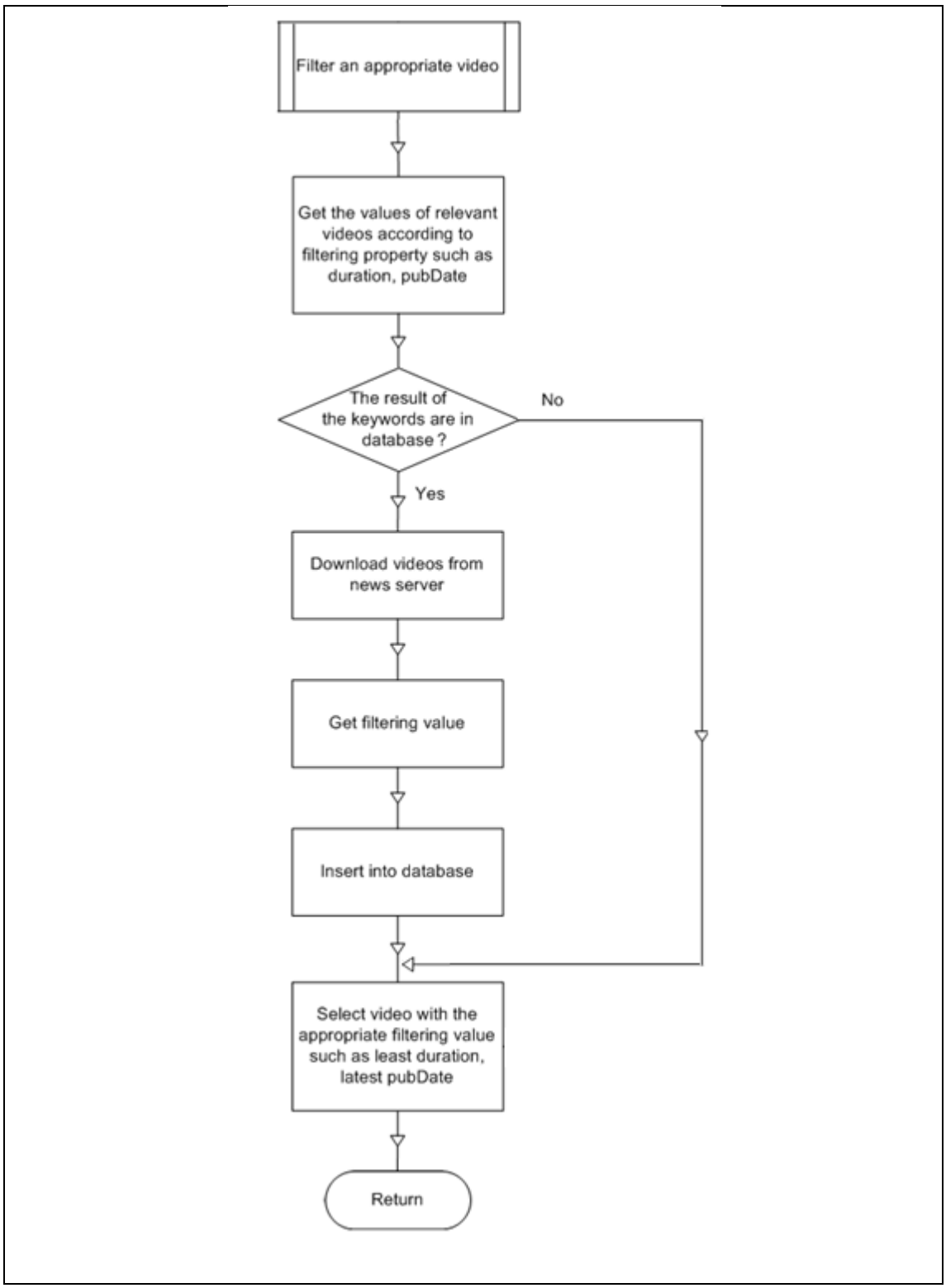
4.2.1 การทำงานในส่วนของผู้ใช้

ผังการทำงานของระบบในส่วนของผู้ใช้แสดงดังภาพประกอบ 4.6



ภาพประกอบ 4.6 ผังการทำงานในส่วนของผู้ใช้

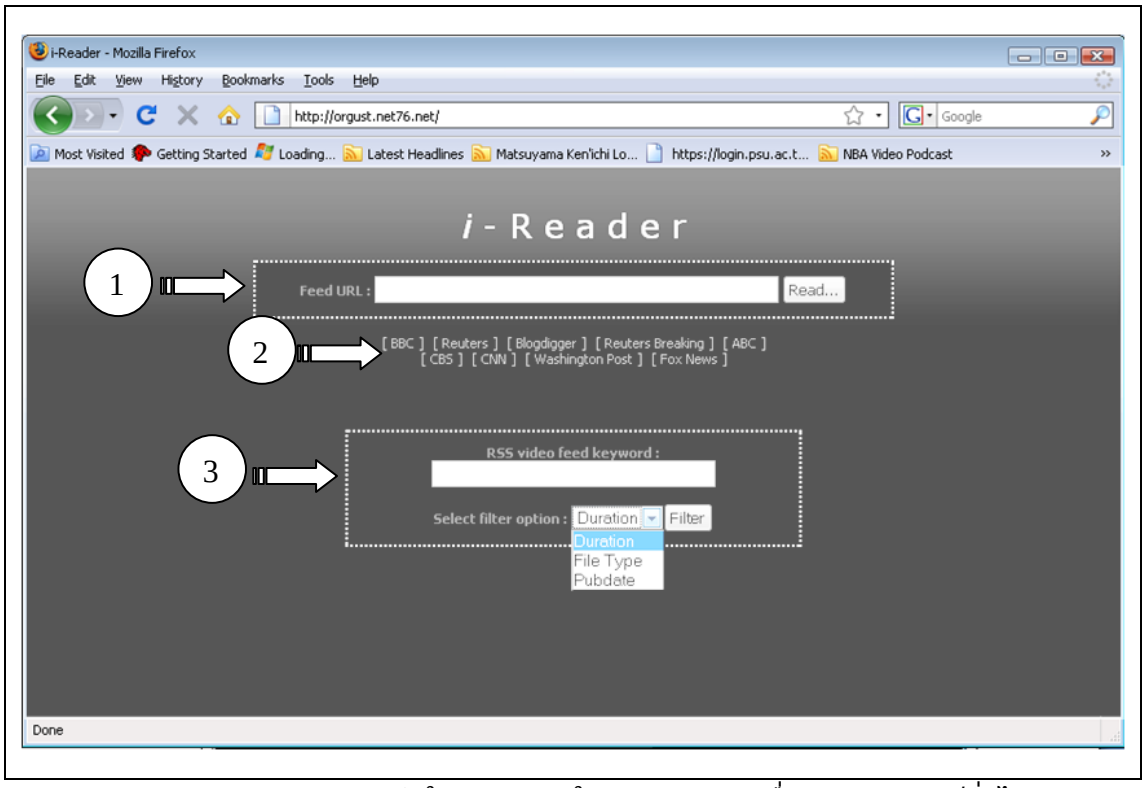
ผังการทำงานย่อยในส่วนของการคัดกรองแสดงดังภาพประกอบ 4.7



ภาพประกอบ 4.7 ผังการทำงานในส่วนของการคัดกรอง

4.2.2 ส่วนแสดงผลของผู้ใช้

หน้าเว็บพัฒนาเพื่อให้สามารถแสดงผลได้ทั้งบนหน้าจอของคอมพิวเตอร์ทั่วไป และบนอุปกรณ์สื่อสารเคลื่อนที่ซึ่งในการพัฒนาทดสอบบนโปรแกรมอีมูเลเตอร์ (emulator) ดังภาพประกอบ 4.8 และ 4.9 ตามลำดับ



ภาพประกอบ 4.8 หน้าเว็บในส่วนของผู้ใช้แสดงผลบนเครื่องคอมพิวเตอร์ทั่วไป

ภาพประกอบ 4.9 หน้าเว็บในส่วนของผู้ใช้แสดงผลบนอีมูเลเตอร์อุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

จากภาพประกอบ 4.9 อธิบายส่วนต่างๆของหน้าเว็บได้ดังนี้

หมายเลข 1: เป็นส่วนของช่องรับ URL เพื่อให้ดึงเอกสาร RSS ตามที่อยู่ที่ระบุจากเซิร์ฟเวอร์ของผู้ให้บริการข่าว ต่างจากส่วนของผู้ดูแลระบบคือไม่ต้องทำการสกัดข้อมูล XML ภายในเอกสารมาเก็บลงในฐานข้อมูล แต่จะแสดงบนหน้าจอของผู้ใช้

หมายเลข 2: ลิงค์ไปยังเว็บไซต์ผู้ให้บริการข่าวที่สำคัญๆบางส่วน

หมายเลข 3: ส่วนของการรับข้อมูลเข้าเป็นคำสำคัญจากผู้ใช้เพื่อนำไปคัดกรองเลือกวิดีโอที่เหมาะสม และส่วนของตัวเลือกระบุคุณสมบัติที่ต้องการในการคัดกรอง

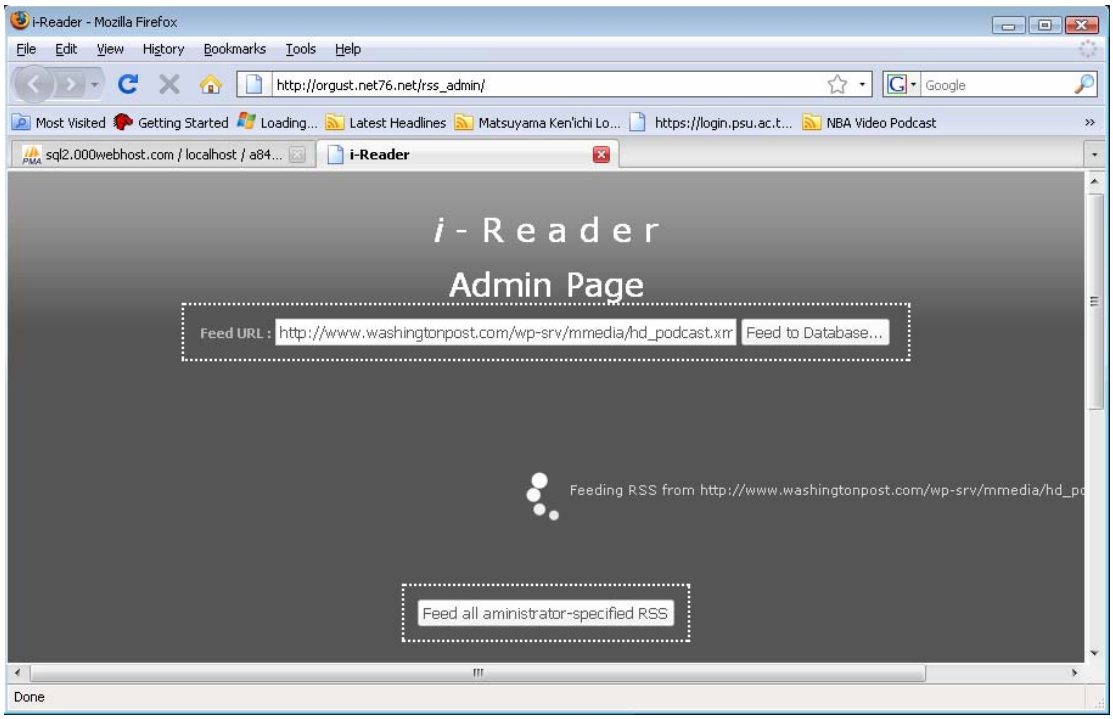
บทที่ 5

ผลการศึกษา

วิทยานิพนธ์ฉบับนี้ได้ทำการทดสอบการทำงานของเทคนิคและระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP (ANVFFS) โดยทดสอบกับเว็บไซต์เผยแพร่วิดีโอข่าวต่างๆได้แก่ CBS News (CBS News, 2009) NBC (NBC, 2009) Pod Weather (Pod Weather, 2009) Daily Record (Daily Record, 2009) Reuters (Reuters, 2009) Blogdigger (Blogdigger, 2009) ABC News (ABC News, 2009) CNN (CNN, 2009) The Washington Post (The Washington Post, 2009) และ Fox News (Fox News, 2009) ผลการทดลองอธิบายแยกเป็นส่วนของผู้ดูแลระบบ และส่วนผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่

5.1 ส่วนของผู้ดูแลระบบ

เมื่อเข้าสู่หน้าเว็บในส่วนของผู้ดูแลระบบ จะปรากฏหน้าต่างให้กรอกชื่อผู้ใช้และรหัสผ่านเพื่อยืนยันตัวตน เมื่อยืนยันตัวตนแล้วจะเข้าสู่หน้าจอหลัก ผู้ดูแลระบบสามารถกดปุ่ม “Feed all administrator-specified RSS” เพื่อเริ่มทำการรวบรวมเอกสาร RSS จากเว็บไซต์ที่เผยแพร่ Podcast ต่างๆ ที่กำหนดโดยผู้พัฒนาระบบ เมื่อกดปุ่มแล้วจะแสดงสถานะการรวบรวมจากแต่ละเว็บไซต์ดังภาพประกอบ 5.1



ภาพประกอบ 5.1 การแสดงสถานะการรวบรวมเอกสาร RSS

เมื่อรวบรวมเอกสาร RSS จากเว็บไซต์ที่เผยแพร่ Podcast ต่างๆแล้ว ระบบ ANVFFS จะทำการสกัดข้อมูลภายในเอกสารด้วย parser แล้วเก็บลงในฐานข้อมูลของระบบ ดังภาพประกอบ 5.2 ซึ่งทำการรวบรวมเอกสาร RSS จาก 11 เว็บไซต์และสกัด item จากเว็บไซต์ทั้งหมดได้รวม 218 item

	Table	Action	Records	Type	Collation
☐	channel		11	MyISAM	latin1_general_ci
☐	item		218	MyISAM	latin1_general_ci
☐	searchResult		0	MyISAM	latin1_general_ci
	3 table(s)	Sum	229	MyISAM	latin1_general_ci

ภาพประกอบ 5.2 ผลการรวบรวมเอกสาร RSS

ตัวอย่างของข้อมูลที่สกัดจากเอกสาร RSS ภายใน tag <channel> ของเว็บไซต์ และ tag <item> ของ channel นั้นๆ แสดงดังภาพประกอบ 5.3 และ ภาพประกอบ 5.4 ตามลำดับ

ch_id	feed_url	feed_time	ch_title	ch_link	ch_desc	ch_pubdate	ch_image
1	http://feeds.cbsnews.com /podcast_eveningnews_video...	2009-02-12 12:12:21	Video: CBS Evening News with Katie Couric	http://www.cbsnews.com	The day's news presented by Katie Couric - Evening...	2009-02-10 20:44:21	
2	http://www.blogdigger.com/media /rss.jsp?q=news&sor...	2009-02-12 12:12:31	Blogdigger Media search for news	http://www.blogdigger.com/media /search.jsp?q=news	Blogdigger Media search for news	2009-02-11 20:36:17	http://www.blogdigger.com /images/blogd_logo01a.gif
3	http://feeds.reuters.com/reuters /USVideoTopNews	2009-02-12 12:12:51	Reuters Video: Top News	http://www.reuters.com /news/video?videoChannel=1	Reuters.com is your source for breaking news, busi...	2009-02-12 08:47:55	
4	http://feeds.reuters.com/reuters /video/breakingnew...	2009-02-12 12:13:22	Breaking News: Europe	http://mobile.reuters.com	A daily in-depth look at a top business or geopoli...	0000-00-00 00:00:00	http://mobile.reuters.com/mobile /images/small/logo...
5	http://rss.cnn.com/services/podcasting /nitn/rss.xm...	2009-02-12 12:13:22	CNN Now in the News (video)	http://www.cnn.com/services /podcasting?eref=nitn	Your quick news update and top stories of the day.	2009-02-12 06:33:51	
6	http://www.washingtonpost.com /wp-srv/mmedia/hd_pod...	2009-02-12 12:13:22	HD Podcast | washingtonpost.com	http://www.washingtonpost.com		2009-01-23 13:15:15	
7	http://www.foxnews.com/xmlfeed /videopodcasting/0,4...	2009-02-12 12:13:31	FOX News Flash	http://www.foxnews.com/fnctv /foxnewsflash	Get ready for 90 seconds of news and personality o...	2009-02-12 11:44:03	
8	http://feeds.cbsnews.com /podcast_icast_1?format=xm...	2009-02-12 12:13:41	CBS Radio World News Roundup	http://www.cbsnews.com	The odd, the interesting and the important: CBS Ne...	2009-02-12 08:24:28	
9	http://podcast.msnbc.com/audio /podcast/MSNBC-NN-NE...	2009-02-12 12:13:51	NBC Nightly News (video)	http://www.nightly.msnbc.com	Watch "NBC Nightly News with Brian Williams," prov...	2009-02-11 20:03:25	http://msnbc.vo.llnwd.net /e1/audio/podcast/images/...
10	http://podweather.com/podcast /?feed=rss2	2009-02-12 12:14:01		http://podweather.com/podcast	Podcasting the Weather 365 days a year.We speciali...	2009-02-12 01:56:06	http://podweather.com/podcast /wp-content/plugins/p...
11	http://www.dailyrecord.co.uk/videos-pics/news/news...	2009-02-12 12:14:11	The Daily Record - News - News Videos	http://www.dailyrecord.co.uk/videos-pics/news/news...	Watch the latest news videos from the Daily Record...	2009-02-10 19:00:00	

ภาพประกอบ 5.3 ข้อมูลที่สกัดจาก tag <channel> ของเว็บไซต์ต่างๆ

it_id	ch_id	it_title	it_link	it_description	it_author	it_catagory	it_comment	it_en_url	it_en_length
183	8	FOX News Flash: FOX & Friends Edition		<p><a href="http://feedads.googleadservices.com/~a...				http://podcasts.foxnews.com/011409/011409_fox_flas...	0
184	8	FOX News Flash: FOX & Friends Edition		<p><a href="http://feedads.googleadservices.com/~a...				http://podcasts.foxnews.com/011309/011309_fox_flas...	0
185	9	Roundup: Passing Stimulus Compromise	http://feeds.cbsnews.com/~r/podcast_icast_1/~3/qzC...	Congress poised to pass the stimulus compromise wi...		News		http://feeds.cbsnews.com/~r/podcast_icast_1/~5/uwg...	0
186	9	Roundup: Tornadoes in OK & TX Kill 8	http://feeds.cbsnews.com/~r/podcast_icast_1/~3/qzC...	Eight dead from tornadoes that struck Oklahoma and...		News		http://feeds.cbsnews.com/~r/podcast_icast_1/~5/hMi...	0

...

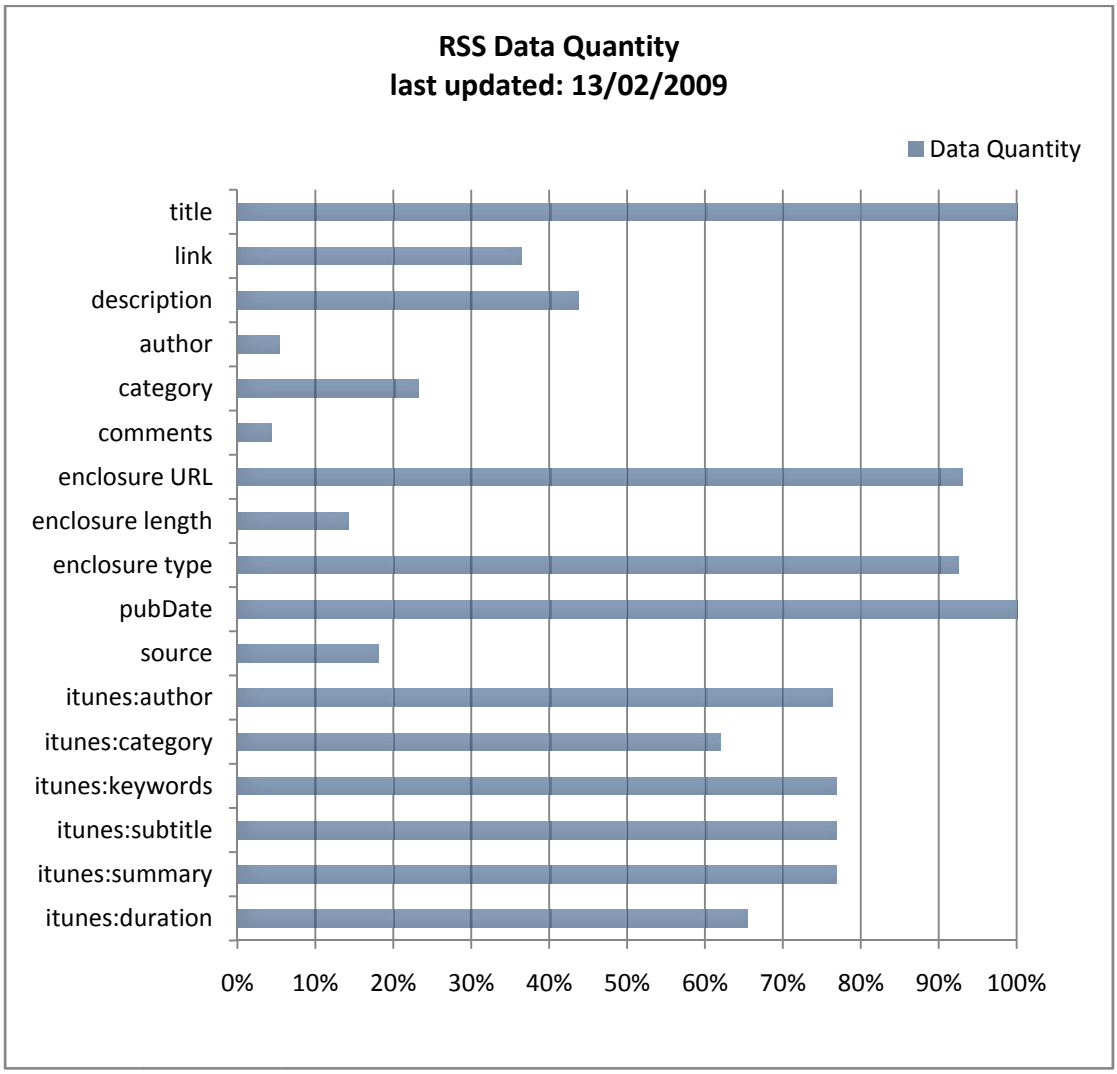
it_en_type	it_pubDate	it_source	it_itunes_author	it_itunes_category	it_itunes_keywords	it_itunes_subtitle	it_itunes_summary	it_itune_duration
video/x-m4v	2009-01-14 18:22:15		FOX News Channel		DMX Steve Doocy unhappy pink rapper jail clothes F...	Rapper DMX complains about wearing pink while in j...	Rapper DMX complains about wearing pink while in j...	03:31:00
video/x-m4v	2009-01-13 10:22:15		FOX News Channel		FOX_Friends Steve Doocy Harvard Gretchen Carlson m...	Britney Spears' searches Harvard University for ne...	Britney Spears' searches Harvard University for ne...	01:44:00
audio/mpeg	2009-02-12 08:24:28		CBS	News	CBS News, couric, 60 minutes, cbs, katie, magid, i...	Congress poised to pass the stimulus compromise wi...	Congress poised to pass the stimulus compromise wi...	00:00:00
audio/mpeg	2009-02-11 08:24:20		CBS	News	CBS News, couric, 60 minutes, cbs, katie, magid, i...	Eight dead from tornadoes that struck Oklahoma and...	Eight dead from tornadoes that struck Oklahoma and...	00:00:00

ภาพประกอบ 5.4 ข้อมูลของที่สกัดจาก tag <item> ของเว็บไซต์ต่างๆ

tag บาง tag ในเอกสาร RSS ที่เป็น tag ตัวเลือกหากไม่มีข้อมูลจะเก็บเป็นค่าว่าง (null) ในฐานข้อมูล อย่างไรก็ตามเนื่องจากเทคโนโลยี RSS นั้นยังไม่มีมาตรฐานที่แน่นอน จึงอาจมีบางเว็บไซต์ละเว้นการใส่ tag บังคับบาง tag ข้อมูลสำคัญที่สามารถนำมาคัดกรองจึงมีปริมาณลดลง

นอกจากนั้นระบบจะทำงานในลักษณะของ Back-end ลบข้อมูลเดิมที่ได้รวบรวมไปแล้วและทำการรวบรวมใหม่ทุกๆช่วงเวลาที่กำหนด เพื่อเป็นการลดการประมวลผลและทรัพยากรของผู้ใช้

จากการเก็บข้อมูล ณ วันที่ 13 ก.พ. 2552 ข้อมูลเอกสาร RSS จากเว็บไซต์ต่างๆที่ได้ทำการทดสอบ พบว่ามีเปอร์เซ็นต์ของอัตราข้อมูลต่างๆใน tag <item> ดังภาพประกอบ 5.5

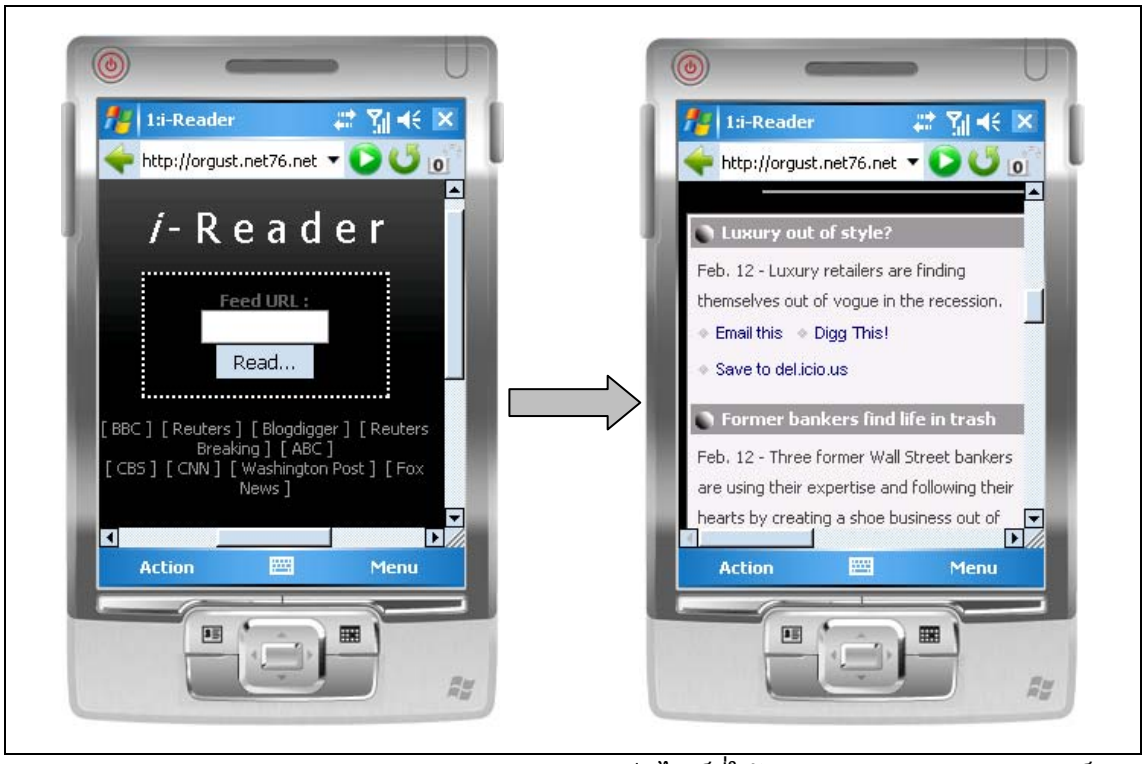


ภาพประกอบ 5.5 อัตราของข้อมูลในเอกสาร RSS

จากภาพประกอบ 5.5 พบว่าในเอกสาร RSS มีการระบุข้อมูลที่อธิบายแต่ละ item ดังนี้ description = 43.6% itunes:subtitle = 76.9% และ itunes:summary = 76.9% ข้อมูล description และ itunes:subtitle นั้นมีปริมาณสูง จึงนำมาใช้เป็นข้อมูลสำคัญในการคัดกรองเพื่อเลือกวิดีโอที่เกี่ยวข้องกับคำสำคัญของผู้ใช้ ถึงแม้ว่า itunes:summary จะมีปริมาณมากเช่นกันแต่พบว่าเกือบ 100% เป็นข้อมูลที่ซ้ำกับ itunes:subtitle จึงไม่จำเป็นต้องนำมาใช้

5.2 การทำงานในส่วนของผู้ใช้

เมื่อเข้าสู่หน้าเว็บไซต์สำหรับผู้ใช้ ผู้ใช้สามารถกรอก URL ที่อยู่ของเอกสาร RSS หรือเลือกจากลิงค์ของเว็บไซต์ที่ใส่ไว้ ระบบจะดึงไฟล์เอกสาร RSS มาทำการสกัดข้อมูลแล้วแสดงผลบนหน้าจอของผู้ใช้ ผลจากการอ่านเอกสาร RSS จากเว็บของ Reuters แสดงดังภาพประกอบ 5.6



ภาพประกอบ 5.6 ผลการอ่านเอกสาร RSS จากเว็บไซต์ที่ให้บริการ Podcast บนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ภาพประกอบ 5.6 (ทางขวามือ) เนื่องจากหน้าจอของอุปกรณ์สื่อสารเคลื่อนที่มีพื้นที่จำกัดจึงไม่สามารถแสดงผลได้หมดในหน้าจอเดียว โดยผลจากการอ่านเอกสาร RSS ทั้งหมดแสดงให้เห็นได้ดังภาพประกอบ 5.7

ภาพประกอบ 5.7 ข้อมูลที่ได้จากการอ่านเอกสาร RSS

ในส่วนของการคัดกรองวิดีโอข่าวอัตโนมัติ ผู้ใช้สามารถกรอกคำสำคัญของวิดีโอข่าวที่ต้องการรับชม ในช่อง “RSS video feed keyword” พร้อมทั้งเลือกคุณสมบัติที่ต้องการให้ระบบทำการคัดกรองคือ duration แสดงดังภาพประกอบ 5.8 ผู้ใช้สามารถกรอกคำสำคัญได้มากกว่า 1 คำโดยเว้นวรรคระหว่างคำ จำนวนของคำสำคัญที่กรอกมีผลต่อความเกี่ยวข้องของวิดีโอที่คัดกรองได้

ภาพประกอบ 5.8 หน้าเว็บในส่วนของการรับคำสำคัญและคุณสมบัติการคัดกรองจากผู้ใช้

เมื่อผู้ใช้กรอกคำสำคัญ ตัวอย่างเช่นคำว่า “found daughter” ระบบจะทำการตรวจเช็คกับประวัติการร้องขอในตาราง SearchResult ในฐานข้อมูล ในที่นี้ไม่เคยมีการร้องขอมาก่อน ระบบก็จะทำงานในขั้นตอนถัดไปคือคัดกรองเอาวิดีโอที่เกี่ยวข้องกับคำสำคัญดังกล่าวแล้วดึงข้อมูลที่สำคัญมาเก็บไว้ในตาราง SearchResult ผลการค้นหาวิดีโอที่เกี่ยวข้องจากวิดีโอทั้งหมดจากหลายเว็บไซต์ ได้ผลลัพธ์เป็นข้อมูลวิดีโอ 2 ระเบียน (record) ดังภาพประกอบ 5.9

«T»	s_id	it_id	duration	pubdate	type	keywords	search_time
	1	82	00:05:00	2008-07-16 07:32:49	video/mp4	found daughter	2009-02-13 09:31:04
	2	173	01:32:00	2009-02-04 11:52:17	video/x-m4v	found daughter	2009-02-13 09:31:04

ภาพประกอบ 5.9 ข้อมูลที่ได้จากการคัดกรองด้วยคำสำคัญ

จากนั้นระบบจะเลือกวิดีโอที่มีระยะเวลาสั้นที่สุดส่งไปยังผู้ใช้ จากตัวอย่างวิดีโอที่สั้นที่สุดคือแถวแรก s_id = 1 ซึ่งมี duration = 5.00 นาที ส่งไปแสดงผลยังอุปกรณ์สื่อสารเคลื่อนที่ของผู้ใช้ดังภาพประกอบ 5.10

บทที่ 6

บทสรุปและข้อเสนอแนะ

งานวิจัยนี้บรรลุตามวัตถุประสงค์คือได้ออกแบบและพัฒนาเทคนิคและแบบจำลองของระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP โดยมีชื่อว่า ANVFFS ซึ่งสามารถแสดงเอกสาร RSS ในรูปแบบวิดีโอข่าวได้ ระบบจะทำการดึงข่าวจากแหล่งข่าวหลายแหล่งมาคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมที่สุดก่อนที่จะส่งไปแสดงผลบนอุปกรณ์ของผู้ใช้

งานวิจัยนี้ได้รับการตีพิมพ์ผลงานวิจัยในระดับประเทศคือ เรื่อง ระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ ซึ่งตีพิมพ์ใน The 5th Joint Conference on Computer Science and Software Engineering (JCSSE 2008) ที่จังหวัดกาญจนบุรี ประเทศไทย ระหว่างวันที่ 7-9 พฤษภาคม 2551 ดังแสดงในภาพผนวก ก. และตีพิมพ์ในระดับนานาชาติ เรื่อง An Automatic Filtering Approach for News Video Feeds on TCP/IP-Based Mobile Devices ซึ่งตีพิมพ์ใน 2008 International Conference on Computer and Electrical Engineering (ICCEE 2008) ที่จังหวัดภูเก็ต ประเทศไทย ระหว่างวันที่ 20-22 ธันวาคม 2551 ดังแสดงในภาคผนวก ข.

6.1 สรุปผลการวิจัย

ระบบ ANVFFS ที่พัฒนาขึ้นโดยใช้เทคนิคการคัดกรองวิดีโอข่าวอัตโนมัติสำหรับอุปกรณ์สื่อสารเคลื่อนที่รองรับโปรโตคอล TCP/IP ที่ได้ทำการออกแบบ สามารถลดการประมวลผลของผู้ใช้ ในสามขั้นตอนดังนี้

- *ขั้นตอนของ NVSS*

จะทำหน้าที่เป็น Back-end รวบรวมข่าวต่างๆจากเว็บไซต์ผู้เผยแพร่ Podcast ตามที่กำหนดไว้ ทำให้ผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่ไม่ต้องอ่านเอกสาร RSS ทั้งหมดจากทุกเว็บไซต์เพื่อหาวิดีโอข่าวที่ต้องการชม และยังทำการรวบรวมใหม่ทุกๆช่วงเวลาที่กำหนดเพื่อให้ได้ข้อมูลของข่าวที่ทันต่อเหตุการณ์

- *ขั้นตอนของ NVPS*

ซึ่งจะทำการตรวจสอบการร้องขอโดยการกรอกคำสำคัญของผู้ใช้ว่ามีการร้องขอเหมือนกันก่อนหน้านั้นหรือไม่ หากไม่มี ระบบก็จะทำการประมวลผลเพื่อหาวิดีโอข่าวที่เหมาะสม หากมีการร้องขอที่เหมือนกันก่อนหน้านี้ระบบจะดึงเอาผลลัพธ์คือวิดีโอข่าวที่

เหมาะสมที่คัดกรองมาแล้วส่งไปแสดงผลบนอุปกรณ์ของผู้ใช้ จึงลดเวลาการประมวลผล เนื่องจากไม่ต้องประมวลผลใหม่ทุกครั้ง ประวัติของการร้องขอจะเก็บไว้ในช่วงเวลาหนึ่งเท่านั้น เมื่อ NVSS ทำการรวบรวมข่าวใหม่ ก็จะล้างข้อมูลประวัติการร้องขอใหม่เพื่อให้ผลลัพธ์การคัดกรองวิดีโอข่าวที่เหมาะสมทันต่อเหตุการณ์

- *ขั้นตอนของการคัดกรองวิดีโอข่าวที่เหมาะสมสำหรับอุปกรณ์สื่อสารเคลื่อนที่*

เนื่องจากทรัพยากรของอุปกรณ์สื่อสารเคลื่อนที่นั้นมีจำกัดการเปิดดูวิดีโอข่าวในเรื่องที่ต้องการจากหลายๆเว็บไซต์จะสูญเสียทั้งเวลา หน่วยความจำ รวมถึงกำลังไฟฟ้าของแบตเตอรี่เป็นจำนวนมาก จึงได้พัฒนาระบบเพื่อให้คัดกรองข่าวที่เกี่ยวข้องกับการร้องขอของผู้ใช้จากเว็บไซต์ต่างๆที่กำหนด แล้วเลือกวิดีโอข่าวที่เหมาะสมนั่นคือวิดีโอข่าวที่มีระยะเวลาสั้นที่สุดส่งไปแสดงผลบนเครื่องของผู้ใช้ จึงเป็นการลดการใช้ทรัพยากรเนื่องจากผู้ใช้รับชมวิดีโอข่าวที่เหมาะสมเพียงวิดีโอเดียว

6.2 ปัญหาและอุปสรรค

6.2.1 เนื่องจากเทคโนโลยี RSS ยังไม่เป็นมาตรฐาน ดังนั้นข้อมูลต่างๆในเอกสาร XML จึงไม่เป็นรูปแบบที่แน่นอน ทำให้การพัฒนาในส่วนของ parser เพื่อสกัดข้อมูลจากเอกสาร RSS นั้นเป็นไปได้ยาก

6.2.2 Podcast นั้นสามารถรวมไฟล์มัลติมีเดียได้ทุกรูปแบบ การรับชมจะขึ้นกับโปรแกรมประยุกต์ที่ใช้ จึงเป็นอุปสรรคกับอุปกรณ์สื่อสารเคลื่อนที่ซึ่งทั้งบราวเซอร์ที่ใช้นั้นส่วนมากจะไม่รองรับการเล่นไฟล์วิดีโอ จึงต้องบันทึกไฟล์ลงบนหน่วยความจำของเครื่องเพื่อเรียกใช้โปรแกรมเล่นไฟล์มัลติมีเดีย (media player) ขึ้นมาเล่นวิดีโอข่าว แต่อย่างไรก็ตามโปรแกรมเล่นไฟล์มัลติมีเดียก็ยังสนับสนุนรูปแบบไฟล์ได้ไม่มากและแตกต่างกัน

6.2.3 เพื่อให้การพัฒนาเป็นไปอย่างมีประสิทธิภาพจึงต้องเสียเวลาในการศึกษาเครื่องมือ (tool) ต่างๆ เช่น AJAX JMF CURL ฯลฯ เป็นเวลานาน เพื่อเลือกใช้ให้เหมาะสม

6.2.4 ข้อมูลเข้าของระบบคือเอกสาร RSS มีการเปลี่ยนแปลงตลอดเวลา ทำให้การทดสอบระบบ และแก้ไขข้อผิดพลาดเป็นไปได้ยาก

6.3 ข้อเสนอแนะ

6.3.1 ในการเลือกคุณสมบัติสำหรับการคัดกรองเพื่อเลือกวิดีโอข่าวที่เหมาะสมโดยใช้ความยาวของวิดีโอ (duration) นั้น หากไม่มีการระบุข้อมูลดังกล่าวไว้ในเอกสาร RSS ระบบจะต้องดาวน์โหลดไฟล์วิดีโอมาเก็บที่เซิร์ฟเวอร์ก่อนนำไปวิเคราะห์หาค่า duration ซึ่งทำ

ให้ภาระงานของระบบเพิ่มมากขึ้นตามจำนวนวิดีโอ จึงอาจมีการพัฒนาเทคนิคโดยการใช้คุณสมบัติอื่นเช่นขนาดของวิดีโอ (file size) ซึ่งสามารถดึงจากส่วนหัวของไฟล์ (file header) แทนที่จะต้องดึงไฟล์วิดีโอทั้งไฟล์

6.3.2 นอกจากการเลือกวิดีโอข่าวที่เหมาะสมจากคุณสมบัติต่างๆแล้ว อาจมีการปรับปรุงเนื้อหาของวิดีโอ (video content) เช่นการทำวิดีโอให้เป็น slide show เพื่อให้สามารถลดการใช้ทรัพยากรของอุปกรณ์สื่อสารเคลื่อนที่ลงได้อีก

6.3.3 พัฒนาขั้นตอนวิธีในการรวบรวมวิดีโอข่าวได้จากเว็บไซต์ที่เผยแพร่ข่าวในรูปแบบอื่นๆ เช่นไฟล์ flash ไม่จำกัดเฉพาะ Podcast

ภาคผนวก ก.

ผลงานตีพิมพ์

เรื่อง	ระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่
Conference	The 5th International Joint Conference on Computer Science and Software Engineering (JCSSE 2008)
สถานที่	จังหวัดกาญจนบุรี ประเทศไทย
วันที่	7-9 พฤษภาคม 2551

ภาพประกอบ 5.10 ผลการแสดงวิดีโอที่เหมาะสมบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ผู้ใช้สามารถคลิกที่หัวข้อของ item ซึ่งจะลิงค์ไปยังที่อยู่ (URL) ของไฟล์วิดีโอ จากนั้นจะปรากฏหน้าต่างให้ผู้ใช้เลือกบันทึกไฟล์วิดีโอนั้นลงบนเครื่อง เนื่องจากดังภาพประกอบ 5.11 และสามารถเรียกดูไฟล์วิดีโอข่าวได้ผ่านโปรแกรมเล่นไฟล์มัลติมีเดียต่างๆ ภาพประกอบ 5.12 แสดงการเล่นไฟล์วิดีโอข่าวในรูปแบบ mpg ผ่านโปรแกรมเล่นไฟล์มัลติมีเดีย TCPMP

ภาพประกอบ 5.11 การบันทึกวิดีโอข่าวที่คัดกรองได้ลงบนอุปกรณ์สื่อสารเคลื่อนที่ที่รองรับโปรโตคอล TCP/IP

ภาพประกอบ 5.12 การเรียกชมวิดีโอข่าวผ่านโปรแกรมเล่นไฟล์มัลติมีเดีย

ภาคผนวก ข.

ผลงานตีพิมพ์

เรื่อง	An Automatic Filtering Approach for News Video Feeds on TCP/IP-Based Mobile Devices
Conference	2008 International Conference on Computer and Electrical Engineering (ICCEE 2008)
สถานที่	จังหวัดภูเก็ต ประเทศไทย
วันที่	20-22 ธันวาคม 2551

2008 International Conference on Computer and Electrical Engineering

An Automatic Filtering Approach for News Video Feeds on TCP/IP-based Mobile Devices

Charawee Sangkhum, Ladda Preechaveerakul, Wiphada Wettayaprasit
Computer Science Department,
Faculty of Science, Prince of Songkla University
Hat Yai, Songkhla, Thailand
Email: {s4910220028, ladda.p, wiphada.w}@psu.ac.th

***Abstract*—Really Simple Syndication (RSS) technology makes it for users possible to keep track of updated news. Podcasting is the technology based on RSS aims to aggregate multimedia files. For desktop users, viewing redundant RSS news video might not be a problem because of the high efficiency of the machine. But for mobile users, viewing redundant news video in a device with limited resource like PDAs or Smart phones is wasting time, memory-space and battery-power. This paper proposes an Automatic Filtering for News Video Feeds for use on TCP/IP-based Mobile Devices. The system aggregates news video feeds from multiple sources and then filters an appropriate video to be shown on user's mobile device. To reduce the processing on mobile devices, users are able to access news video through a web-based reader.**

Keywords-Really Simple Syndication, RSS, RSS reader, news video, podcast, mobile device

I. INTRODUCTION

Recently, information on the Internet has been changing into multimedia format. News sources such as ABC [2], The Washington Post [14] and Reuters [11] have been publishing multimedia news as audio and video files. They also made these files to support RSS technology called Podcast such that the users can keep track of updated news. The users need to subscribe to the publisher or provide the URL of RSS feed location to RSS reader or RSS aggregator which is the application to feed news from the server. Each user will then get the result as news list that the news source provided. However, the RSS reader will show all video feeds related to the topic for which a user selects the video he/she wants to see. In the group of news from multiple sources, there might be more than one news-bit in the same topic or redundancy news. For desktop users, viewing redundant news might not be a problem because of the high efficiency of the machine. But for mobile users, viewing redundant news video in a device with limited resource like PDAs or Smart phones is wasting time, memory-space and battery-power. Our research thus proposes an Automatic News Video Feed Filtering Technique for TCP/IP-based Mobile Devices. Our system will aggregate news video feeds from multiple sources then, filter an appropriate video from user-required topic to be shown on TCP/IP-based mobile devices, so that processing costs will be reduced.

Moreover, we developed the system as web-based RSS reader to allow the users to retrieve updated content through the website by only using a mobile web browser. They need not to install a mobile reader on their device in order to save resources.

The paper is organized as follows: Section 2 and 3 present RSS technology and Podcast Technology, respectively. Section 4 describes related work on RSS technology and Podcast technology. Section 5 proposes a model for an Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices. Section 6 presents our experimental results. Finally, we concludes the major finding and future works in Section 7.

II. REALLY SIMPLE SYNDICATION (RSS)

Technology to aggregate contents on the web originated from "push" and "pull" technology [9]. For pull technology, the initial request for data comes from the client, while push technology requests for a given transaction created by a publisher or central server. The server will specify what and when the content will be shown. In 1997, RSS technology [3] was developed to aggregate updated contents according to push technology. In particular, when the resources provide new contents, these will be sent to users without making any request. However, RSS technology is not completely push technology because when new content is published, the user will know by polling every specific period.

RSS providers such as news sites or weblogs, need to provide the contents as XML document called RSS feed which can be displayed using a reader or aggregator. When the users provide RSS feed URL to the reader, which will feed the contents to be displayed on a user's machine, and will automatically check for updated contents. The contents are only concise textual contents like news topics or news summaries. The whole content can be viewed by clicking on the relevant link.

RSS document consists of tags with different functionality. Table 1 summarizes the functionality of the main RSS tags.

Every RSS file has the root tag <rss> where the version of the RSS file is defined. The only child of the <rss> tag is <channel>. In the tag <channel> there are required tags which are <title>, <link> and <description> and another optional tags. Fig. 1 gives an example of an RSS document.

978-0-7695-3504-3/08 $25.00 © 2008 IEEE
DOI 10.1109/ICCEE.2008.116

IEEE computer society

Authorized licensed use limited to: Prince of Songkla University. Downloaded on February 17, 2009 at 01:42 from IEEE Xplore. Restrictions apply.

TABLE I. RSS FEED TAGS

RSS Tag	Description
<rss>	Start RSS information
<channel>	Channel of the feed publisher
<item>	Chunks of summarized information
<title>	The title of the information chunk
<link>	The url to the html website
<description>	The phrases or sentences describing the information.
<ttl>	Time to live indicates the amount of time (in minutes) that indicates how long a channel should be cached before refreshing from the source.
<pubDate>	The publication date for the content in the channel.
<enclosure>	A media-file to be included with an item(Podcast technology).

```
<rss version="2.0">
 <channel>
  <title>…</title>
  <link>…</link>
  <description>…</description>
  <language>…</language>
  <pubDate>…</pubDate>
  <lastBuildDate>…</lastBuildDate>
  <docs>…</docs>
  <generator>…</generator>
  <item>
    <title>…</title>
    <link>…</link>
    <description>…</description>
    <pubDate>…</pubDate>
  </item>
 </channel>
</rss>
```

Figure 1. Example of RSS feed.

There are many types of RSS readers. Software readers need to be installed first such as desktop readers and mobile readers. Web-based readers locate on the server and can be accessed through websites and browser-embedded readers.

III. PODCAST TECHNOLOGY

Currently, news providers are developing and improving methods to publish news. Previously, there were only texts and pictures, but now, multimedia files such as audio and video files are popular. Podcasting is the technology based on RSS aims to aggregate multimedia files. These multimedia files are called Podcast. RSS version 2.0 added a new tag called "enclosure" as the sub-element of <item> tag. The location of each multimedia file is inserted into the <enclosure> tag of RSS feed. The <enclosure> tag has 3 relevant attributes; 'url' says where the multimedia file is located, 'length' determines the file length in bytes, and 'type' describes type of multimedia file. The syntax of <enclosure> tag is shown in Fig. 2. With this technology, users can now keep track on updated multimedia contents.

```
<enclosure url=http://domain.com/file.mp4
     length="123456789" type="video/mp4" />
```

Figure 2. RSS enclosure syntax.

IV. RELATED WORK

RSS technology becomes more prevalent, as there are an increasing number of researches related to RSS and Podcast technology.

Aldrich, Bell and Batzel [4] developed a system that provides a solution which automatically generates the podcasts of class lectures and publishes them on online class blog-space. The system requires class information to record the audio at a specific time, and creates the RSS feed. A link to a finished audio file will then be added to the feed. Subsequently, Malan [5] improved the system so that the class lectures can be recorded as podcast in both audio and video format.

Curran et al. [8] proposed a Scheduled RSS Delivery System, which utilizes idle computer time to subscribe to media RSS channels for downloading podcasts. The aggregator in the system will scan for newly added items and place those on the download list when the computer is idle.

Blekas et al. [1] proposed a method to adapt the web content before it can be accessed by a mobile browser due to its technical limitations. They build on a proxy server solution that is online system using RSS feeds for an adaptation of the web content for mobile phones. The existence of RSS feeds significantly improves the content presented to the mobile users, decreases both size and access cost because RSS feeds contain metadata that summarize information. The proxy server, which usually runs on powerful servers, unleashes mobile devices from computational needs. However, the limitations of this method are that images and tables will be eliminated from the content and HTML forms can not be used.

Ananthayarayanan et al. [7] proposed and implemented the OWeB system to improve the Internet browsing experience over slow and intermittent networks. The aggregator of the system needs to download the RSS feed to construct the homepage. Hence, the results in websites are up-to-date and available offline, and also help in bandwidth savings.

As mentioned above, we found that there are researches applying RSS technology for many purposes. However, as to our knowledge, only a small number of researches mention aggregation of video feeds from multiple sources and filtering the appropriate video to be shown on TCP/IP-based mobile devices. Our research proposes an Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices, on which users can consume news videos that keep increasing and get updated all the time. The system uses a Filtering Technique to select the appropriate video for TCP/IP-based Mobile Devices, which have limited resources as to speed of processing, memory, battery, and the ability to transfer data through wireless networks.

Authorized licensed use limited to: Prince of Songkla University. Downloaded on February 17, 2009 at 01:42 from IEEE Xplore. Restrictions apply.

V. AN AUTOMATIC NEWS VIDEO FEED FILTERING SYSTEM FOR TCP/IP-BASED MOBILE DEVICES (ANVFFS)

An Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices (ANVFFS) aims to aggregate video feeds (podcasts) from the websites and filter an appropriate video to be shown on TCP/IP-based Mobile Devices.

A. General architecture

Three core elements of the system are described: 1) A News Video Proxy Server (NVPS) checks if the request from the user is similar to the previous request. If so, the filtered video can be sent to the user's device. 2) A News Video System Server (NVSS) provides a webpage of a web-based RSS reader that retrieves the request from users and processes it to get the appropriate video. 3) Data Repository (buffer) of extracted information of RSS feeds and video files.

The procedure of the system is shown in Fig. 3.

1) User requests the video by filling out the keyword on the web-based RSS reader.

2) NVPS checks whether the user has sent the same request earlier. If so, NVPS will send the filtered video to user in step 8.

3) If new request, NVPS will pass the request to NVSS.

4) NVSS sends the RSS feed request to News Sources.

5) The server of News Sources send RSS feed in XML format back to NVSS

6) NVSS extracts the information in RSS feed (such as title, description and enclosure) and stores it in the database.

7) Relevant videos are retrieved and pass the filtering process in order to get the appropriate video.

8) Filtered video will be sent to NVPS to be displayed on mobile browser on user's device.

B. An Automatic News Video Feed Filtering Technique

After RSS feeds from multiple sources have been retrieved, the information in XML tags can be extracted to be stored in the data repository, for example, <title>, <description>. The keyword from the user will be matched with this information to find the related items. The system then pulls the relevant video properties and stores them in the buffer. From the group of video feeds that related to the same topic, the system filters the appropriate video according to specific parameter (e.g. duration, size, update time, file format). The algorithm for filtering video feeds by duration is shown in Fig. 4.

```
1  Method filterByDuration(keyword)
2      for each feedURL
3        download feed content from feedURL
4        for each item in the feed
5          get data in <title>, <description> and
6          <enclosure> section
7          if(search keyword in data from <title>
8          and <description> is found)
9              get duration from feed data
10               if(duration not provided)
11                 get video from the location in <enclosure>
12                 extract duration from the file
13             end if
14         else
15         report no results
16         end if
17       end for
18     end for
19     select item with least duration
20     show to user
21 end method
```

Figure 4. News videos filtering algorithm

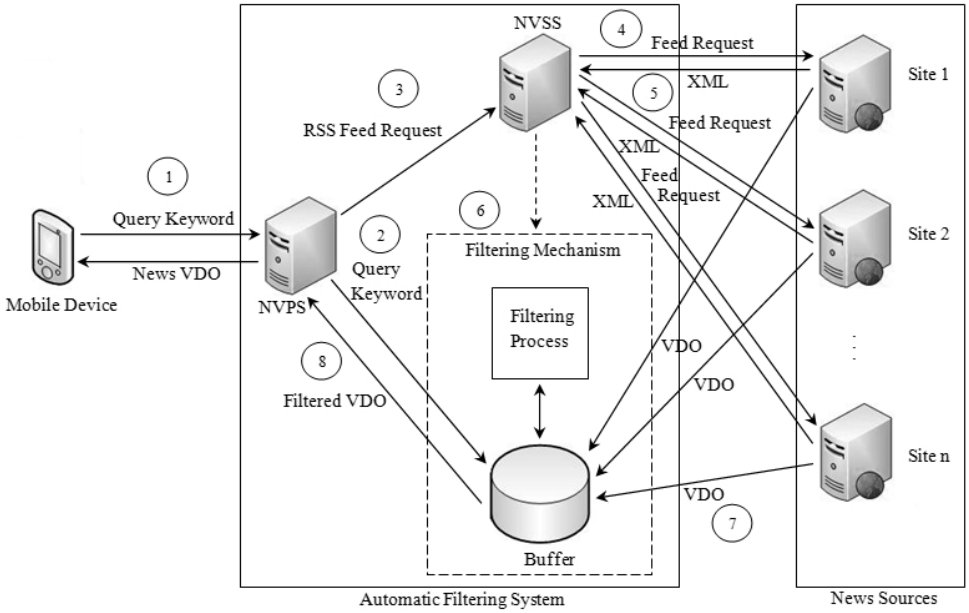


Figure 3. Model of An Automatic News Video Feed Filtering System for TCP/IP-based

Authorized licensed use limited to: Prince of Songkla University. Downloaded on February 17, 2009 at 01:42 from IEEE Xplore. Restrictions apply.

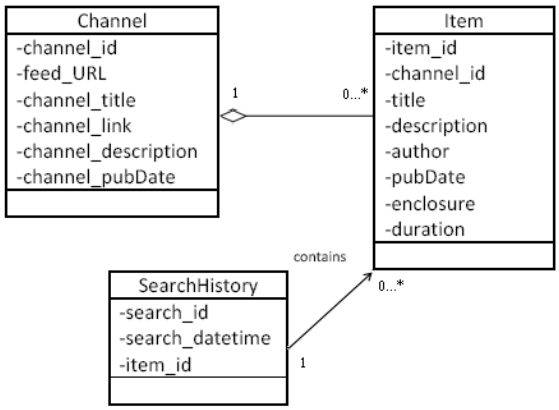


Figure 5. Data repository structure

C. *Data repository*

The data repository contains the information from RSS feeds that are necessary for filtering. We also need to keep the search history for the NVPS to check whether the request is similar to the previous requests.

Fig. 5 shows the design of the database. Each RSS feed has one channel, each of which contains any number of items.

In addition, there is a buffer to temporarily store the retrieved video file for advanced filtering.

VI. EXPERIMENTAL RESULTS

We designed and implemented the system to automatically filter news video feeds. We used Windows Mobile 5 Emulator and Opera for Windows Mobile browser to test accessing via the mobile devices. We developed a web-based reader using AJAX and PHP. The implementation is divided into 2 parts.

1) RSS feed retrieval to extract important information from specific news providers, and to pull data of each video item that is related to a user's keyword to store in the buffer.

2) Video Filtering process to select the appropriate video according to a user's request to be shown on that user's mobile device.

The implementation of the first part is shown in Fig. 6. Fig. 6-a presents the webpage for entering the feed location. Fig. 6-b shows the results of RSS feeds. When the user submits the keyword of the news video they need, the system will search for the matching items in the database of the RSS feeds that had been aggregated from multiple sources and then filter the item with least duration as in Fig. 6-c. When the user selects the item, the relevant video file is downloaded into the user's device and can be played using media player as shown in Fig. 6-d.

VII. CONCLUSIONS AND FUTURE WORK

This paper proposed an Automatic News Video Feed Filtering System for TCP/IP-based Mobile Devices

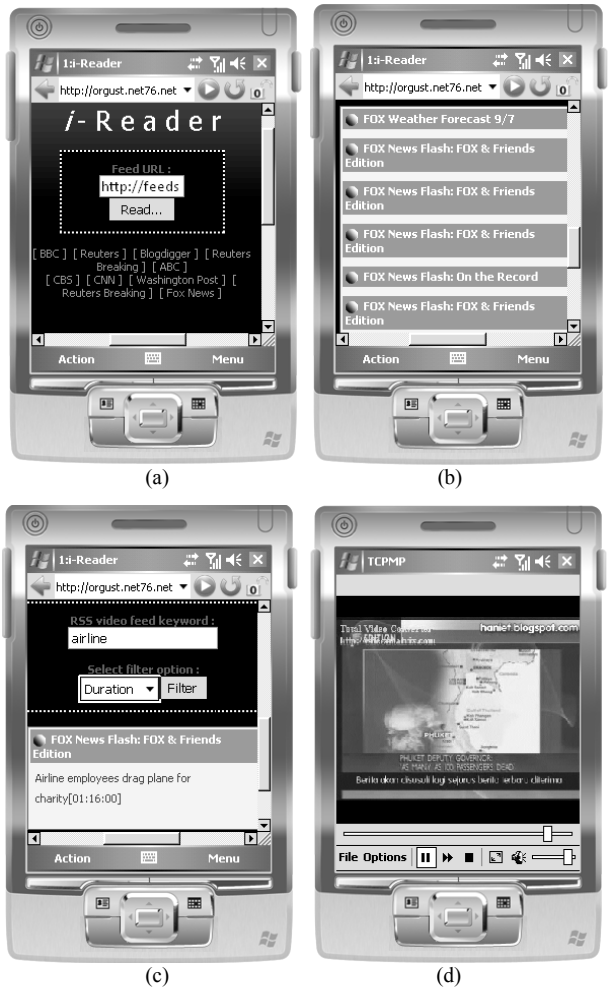


Figure 6. Example of usage: (a) web-based reader (b) RSS feed results (c) filtered RSS item (d) Playing video file

(ANVFFS) that aggregates podcast from multiple websites to be filtered an appropriate video for TCP/IP-based Mobile Devices. The system reduces processing of the mobile devices by using NVPS to check whether the request is similar to the previous if so, the filtered video by NVSS could be sent immediately. Because of the limitation of the mobile devices, the system is designed to relieve the process of downloading video files by filtering only the appropriate video. Additionally, the reader is located on server side and the user doesn't need to install the software. Currently, we are already developing the algorithm for filtering, this type of using file duration and another, more advanced, options.

ACKNOWLEDGMENT

The authors are grateful to Prof. Ir. Sikke A. Hempenius, Faculty of Science, Prince of Songkla University, for reading and correcting the manuscript, including valuable suggestion.

Authorized licensed use limited to: Prince of Songkla University. Downloaded on February 17, 2009 at 01:42 from IEEE Xplore. Restrictions apply.

REFERENCES

[1] A. Blekas, V. Garofalakis, and V. Stefanid, "Use of RSS feeds for content adaptation in mobile web browsing". Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A), New York, USA, 2006, pp. 79-85.

[2] ABC News. Available from: http://abcnews.go.com/ xmldata [2008, Feb 15].

[3] A. Watt and D. Ayers, "Beginning RSS and ATOM Programming", Wiley Publishing, Inc., Indianapolis, Indiana, 2005.

[4] D. Aldrich, B. Bell, and T. Batzel, "Automated podcasting solution expands the boundaries of the classroom". Proceedings of the 34th annual ACM SIGUCCS conference on User services, New York, USA, 2006, pp. 1-4.

[5] D. J. Malan, "Podcasting computer science E-1". 38th SIGCSE technical symposium on Computer science education, New York, USA, 2007, pp. 389-393.

[6] Extensible Markup Language (XML). Available from: http://www.w3.org/ XML/ [2008, May 26].

[7] G. Ananthanarayanan, S. Blagsvedt, and K. Toyama, "OWeB: a framework for offline web browsing". Web Congress LA-Web Fourth Latin American, 2006, pp. 15-24.

[8] K. Curran, S. Mc Kinney, F. Burns, and G. Meredith, "Scheduled RSS feeds for streaming multimedia to the desktop using RSS enclosures". Consumer Communications and Networking Conference, 2006, CCNC. 3rd IEEE, pp. 267-271.

[9] M. Franklin, and S. Zdonik, " 'Data in your face': push technology in perspective". Proceedings of ACM SIGMOD international conference on Management of data, New York, USA, 1998, pp. 516-519.

[10] M. Stephane, "Content-based video retrieval: an overview". Universite De Geneva.

[11] Reuters. Available from: http://www.reuters.com/tools /podcasts /video [2008, Feb 15].

[12] RSS. Available from: http://www.rss-tools.com/rss-tutorials.htm [2008, Feb 15].

[13] T. Lehtonen, et al., "Towards user-friendly mobile browsing". Proceedings of the 2nd international workshop on Advanced architectures and algorithms for internet delivery and applications, New York, USA, 2006.

[14] The Washington Post. Available from: http://www.wash ingtonpost.com/wp-srv/mmedia/podcastfront.htm [2008, Feb 15].

Authorized licensed use limited to: Prince of Songkla University. Downloaded on February 17, 2009 at 01:42 from IEEE Xplore. Restrictions apply.

ประวัติผู้เขียน

ชื่อ สกุล นางสาวชารวีย์ แสงขำ

รหัสประจำตัวนักศึกษา 4910220028

วุฒิการศึกษา

วุฒิ	ชื่อสถาบัน	ปีที่สำเร็จการศึกษา
วศ.บ. (วิศวกรรมคอมพิวเตอร์)	มหาวิทยาลัยสงขลานครินทร์	2549

ทุนการศึกษา

1. ทุนมูลนิธิเพื่อการศึกษาคอมพิวเตอร์และการสื่อสาร (Computer & Communication Education Foundation) ปี 2550
2. ทุนวิจัยภายใต้โครงการแลกเปลี่ยนนักศึกษากับ University of Novi Sad ประเทศเซอร์เบีย ปี 2551

การตีพิมพ์เผยแพร่ผลงาน

1. ชารวีย์ แสงขำ ลัดดา ปรีชาวีรกุล และ วิภาดา เวทย์ประสิทธิ์. 2008. ระบบคัดกรองวิดีโอข่าวอัตโนมัติสำหรับผู้ใช้อุปกรณ์สื่อสารเคลื่อนที่. The 5th Joint Conference on Computer Science and Software Engineering (JCSSE 2008), จังหวัดกาญจนบุรี, ประเทศไทย, หน้า 210-217.
2. Sangkhum, C., Preechaveerakul, L., and Wettayaprasit, W. 2008. An Automatic Filtering Approach for News Video Feeds on TCP/IP-Based Mobile Devices. 2008 International Conference on Computer and Electrical Engineering (ICCEE 2008), Phuket, Thailand, pp. 521-525.