

บทที่ 2

ระบบข่าวและการจัดเก็บข้อมูลสถิติยูสเน็ตนิวส์

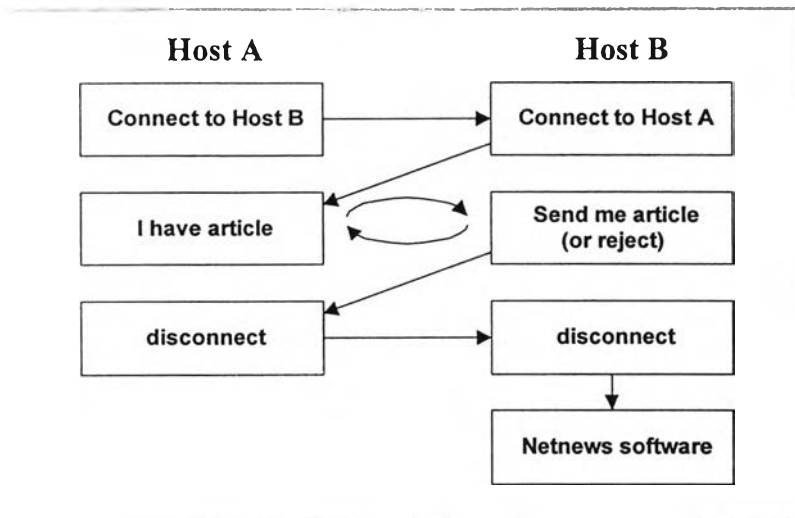
ในบทนี้จะกล่าวถึงระบบข่าว ในส่วนของการแลกเปลี่ยนยูสเน็ตนิวส์ระหว่างเครื่องบริการข่าวหนึ่งไปยังอีกเครื่องบริการข่าวหนึ่ง และการทำงานของไอเอ็นเอ็น (InterNetNews-INN) ซึ่งเป็นซอฟต์แวร์ที่นิยมใช้กันอย่างแพร่หลายบนเครื่องบริการข่าวนอกจากนั้นจะกล่าวถึงการจัดการยูสเน็ตนิวส์ที่ผ่านมาได้ใช้เครื่องมืออะไรในการจัดเก็บข้อมูลยูสเน็ตนิวส์ และมีลักษณะในการจัดเก็บข้อมูลยูสเน็ตนิวส์เป็นอย่างไรบ้าง

2.1. ระบบข่าว

ระบบข่าวสามารถกระจายยูสเน็ตนิวส์หรือแลกเปลี่ยนยูสเน็ตนิวส์จากเครื่องบริการข่าวของเครือข่ายหนึ่งไปยังเครื่องบริการข่าวของเครือข่ายอื่น ๆ ที่เชื่อมโยงถึงกันทั่วโลกได้โดยใช้โพรโทคอลเอ็นเอ็นทีพี

2.1.1. เกณฑ์วิธีขนส่งข่าวของกลุ่มข่าว

เอ็นเอ็นทีพี (Network News Transfer Protocol –NNTP) (O'Reilly and Todino, 1996) เป็นโพรโทคอลประยุกต์ (Application Protocol) ที่ใช้บริการจาก TCP (Transmission Control Protocol) เพื่อให้การรับและส่งข่าวสารมีความถูกต้องสมบูรณ์และเชื่อถือได้ เอ็นเอ็นทีพี จะใช้ในการติดต่อแลกเปลี่ยนข่าวสารระหว่างเครื่องบริการข่าว (News Server) จากเครือข่ายหนึ่งไปยังเครื่องบริการข่าวในเครือข่ายอื่นๆ และระหว่างเครื่องบริการข่าวกับแม่ข่าย (Host) ภายในเครือข่ายเอง

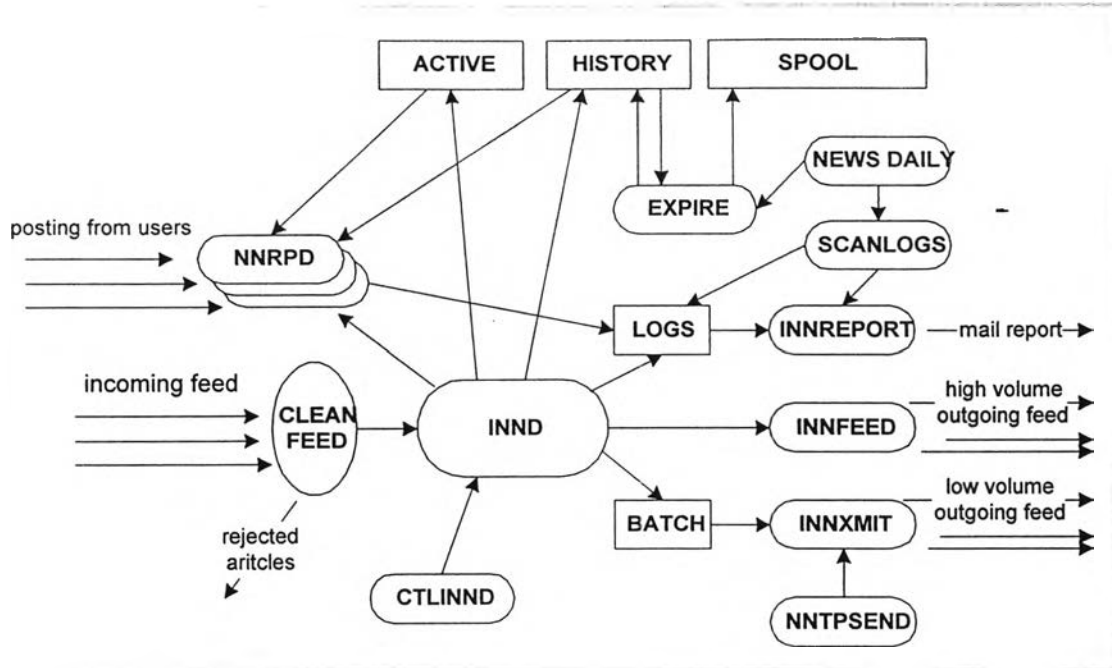


รูปที่ 2.1 รูปแบบจำลองในการรับและส่งข้อมูลของเอ็นเอ็นทีพี (Kantor and Lapsley, 1986)

จากรูปที่ 2.1 รูปแบบจำลองในการรับและส่งข้อมูลของเอ็นเอ็นทีพี ได้แสดงลักษณะพื้นฐานของการสื่อสารของเอ็นเอ็นทีพี คือ เมื่อแม่ข่าย A ทำการติดต่อกับแม่ข่าย B ได้แล้ว แม่ข่าย A จะทำการส่งคำสั่งบอกแม่ข่าย B ว่า มีตัวระบุข้อความ (message-id) ของข่าวหรือไม่ ถ้าแม่ข่าย B มีตัวระบุข้อความของข่าวนี้อยู่ก็จะไม่รับข่าวดังกล่าว และจะเลือกรับเฉพาะกลุ่มข่าวที่ตนเองต้องการและยังไม่มีเท่านั้น โดยแม่ข่าย A จะทำการถามตัวระบุข้อความของข่าวอื่น ๆ ที่มีต่อไปเรื่อย ๆ จนหมด เมื่อเลิกการติดต่อแม่ข่าย A จะส่งคำสั่งเลิกการติดต่อกับทางแม่ข่าย B

2.1.2 ระบบการทำงานของไอเอ็นเอ็น

ไอเอ็นเอ็น (InterNetNews Software – INN) (McDermott and John 1997) สร้างขึ้นโดย Rich Salz ในปี 1992 ถือเป็นซอฟต์แวร์ที่ใช้ในระบบการขนส่งยูสเน็ตนิวส์ตามข้อกำหนด RFC 977 (Kantor and Lapsley, 1986) ซึ่งเป็นที่นิยมใช้บนเครื่องบริการข่าวอย่างกว้างขวาง



รูปที่ 2.2 โครงสร้างของไอเอ็นเอ็น (Samsonova, 1998)

จากรูปที่ 2.2 โครงสร้างการทำงานของไอเอ็นเอ็น ได้แสดงลักษณะการทำงานหลักๆ ของไอเอ็นเอ็น ดังต่อไปนี้

- **INND** เมื่อเครื่องบริการข่าวจากที่อื่นป้อนข่าวเข้ามา INND จะทำหน้าที่ในการรับข่าวดังกล่าว โดยตรวจสอบจากแฟ้มข้อมูลแอคทีฟ (Active file) ซึ่งเป็นแฟ้มข้อมูลสำหรับเก็บชื่อกลุ่มข่าวที่ต้องการจะรับก่อนว่ามีกลุ่มข่าวดังกล่าวหรือไม่ จากนั้นจะทำการเก็บข้อมูลลงในแฟ้มข้อมูลแอคทีฟ แฟ้มข้อมูลประวัติ (History file) และที่เก็บพัก (Spool)

- **NNRPD** เมื่อผู้ใช้บริการภายในเครือข่ายเรียกใช้บริการยูสเน็ตนิวส์จากเครื่องบริการข่าวโดยผ่านโปรแกรมอ่านข่าว (News Client หรือ News Reader) นั้น INND จะสร้างโปรเซส NNRPD เพื่อให้บริการกับโปรแกรมอ่านข่าวแทน

- **CTLINND** ทำหน้าที่ในการส่งคำสั่งคอยควบคุมการทำงานของ INND อาทิ เช่น ตรวจสอบการติดต่อระหว่างเครื่องบริการข่าว การเพิ่มหรือลดกลุ่มข่าว เปลี่ยนโครงสร้างไอเอ็นเอ็น (INN Configuration) สั่งให้ INND หยุดการทำงานชั่วคราวและเริ่มการทำงานอีกครั้ง เป็นต้น

- **INNXMIT** ทำหน้าที่คอยส่งข่าวออกจากเครื่องบริการข่าวไปยังเครื่องบริการข่าวอื่นๆ โดยมี NNTPSEND คอยควบคุมการทำงานของ INNXMIT โดย INNXMIT นี้เหมาะสมสำหรับเครือข่ายที่มีความกว้างแถบความถี่น้อย ส่วนเครือข่ายที่มีความกว้างแถบความถี่มากจะใช้ INNFEED ในการส่งข่าวแทน

2.2 การจัดเก็บข้อมูลสถิติยูสเน็ตนิวส์ที่เกี่ยวข้อง

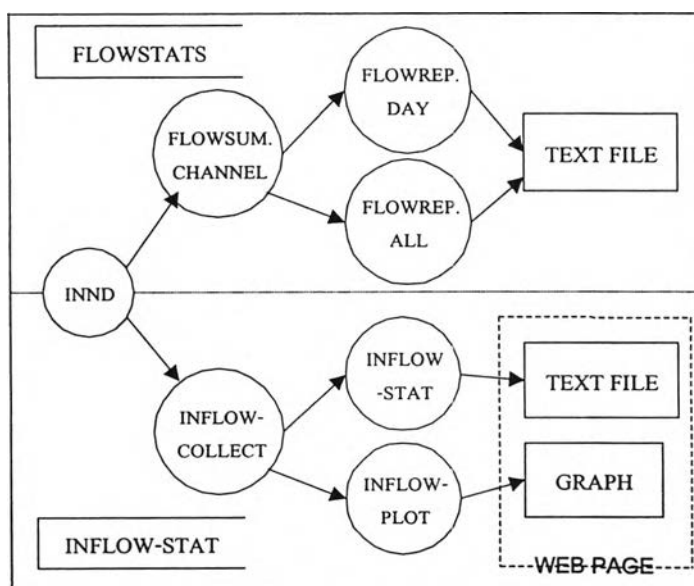
การจัดเก็บข้อมูลสถิติยูสเน็ตนิวส์ที่ผ่านมามีซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูลสถิติยูสเน็ตนิวส์ที่ใช้กันมี 3 ชนิด ดังนี้

- **โฟลว์สแตตส์ (Flowstats)** (Line, 1996)

เป็นซอฟต์แวร์ที่เขียนโดย J.M.Line ใช้ในการเก็บจำนวนหัวเรื่องย่อยและปริมาณของยูสเน็ตนิวส์ที่ป้อนเข้ามาให้กับเครื่องบริการข่าว

- **อินโฟลว์สแตต (Inflow-stat)** (Kugler, 1997)

เป็นซอฟต์แวร์ที่เขียนโดย Felix Kugler ใช้ในการเก็บข้อมูลยูสเน็ตนิวส์ที่ส่งเข้าและออกจากเครื่องบริการข่าวที่ป้อนยูสเน็ตนิวส์ให้ ข้อมูลยูสเน็ตนิวส์ที่เก็บรวบรวมได้แก่ จำนวนและขนาดหัวเรื่องย่อยของกลุ่มข่าวและความหน่วง ในการรับส่งยูสเน็ตนิวส์ของแต่ละเครื่องบริการข่าวและกลุ่มข่าวใหญ่ ๆ 20 กลุ่ม

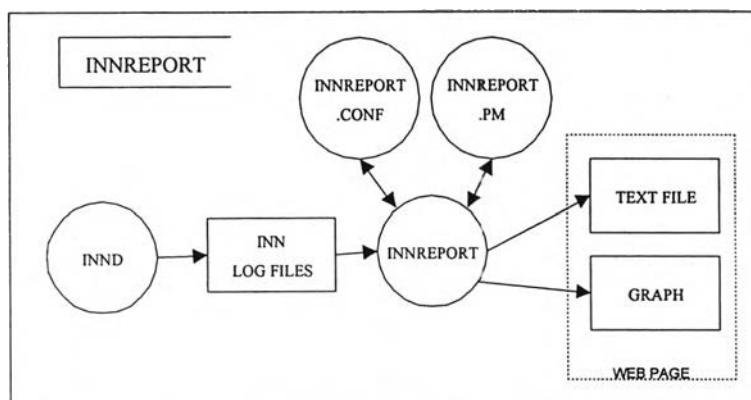


รูปที่ 2.3 รูปแบบการทำงานของโฟลว์สแตตส์และอินโฟลว์สแตต

จากรูปที่ 2.3 รูปแบบการทำงานของโฟลว์สแตตส์และอินโฟลว์สแตต มีลักษณะการทำงานคล้ายๆ กันดังนี้ คือ ทำการควบคุมให้ INN.D ส่งให้แหล่งป้อนข่าว (Newsfeed) ของไอเอ็นเอ็น ส่งข้อมูลยูสเนตนิวส์ที่รับมา ทำการบันทึกข้อมูลยูสเนตนิวส์เก็บลงในกลุ่มแฟ้มข้อมูลและทำการรายงานผลของข้อมูลสถิติยูสเนตนิวส์ออกมา

- **ไอเอ็นเอ็นรีพอร์ต (Innreport) (Tassin, 1996)**

เป็นซอฟต์แวร์ที่เขียนโดย Fabien Tassin ใช้ในการดึงข้อมูลรายละเอียดต่างๆ ที่เก็บอยู่ในแฟ้มข้อมูลล็อก (Log Files) ของไอเอ็นเอ็น มาแสดงผลการทำงานของเครื่องบริการข่าว นั้น ๆ



รูปที่ 2.4 รูปแบบการทำงานของไอเอ็นเอ็นรีพอร์ต

จากรูปที่ 2.4 รูปแบบการทำงานของไอเอ็นเอ็นรีพอร์ต จะทำการอ่านข้อมูลรายละเอียดต่างๆ ที่เก็บอยู่ในแฟ้มข้อมูลล็อกที่สร้างจากไอเอ็นเอ็น แล้วนำมาสรุปผลการทำงานของเครื่องบริการข่าวนั้น ๆ ผ่านออกมาทางเว็บเพจในลักษณะของตารางและกราฟ

ซึ่งรูปแบบแสดงรายงานการจัดเก็บข้อมูลสถิติยูสเนตนิวส์ของซอฟต์แวร์ทั้งสาม ดูได้จากภาคผนวก ข

2.3 เปรียบเทียบซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูลยูสเนตนิวส์

ตารางที่ 2.1 เปรียบเทียบซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูลยูสเนตนิวส์

คุณสมบัติ	โฟลว์สแตตส์	อินโฟลว์ สแตต	ไอเอ็นเอ็น รีพอร์ต
1. เก็บข้อมูลสถิติข่าวหลัก (Primary Newsgroup)	มี	มี	มี
2. เก็บข้อมูลสถิติข่าวรอง (Secondary Newsgroup)	มี	ไม่มี	ไม่มี
3. เก็บข้อมูลการทำงานจากแฟ้มข้อมูลล็อก	ไม่มี	ไม่มี	มี
4. เก็บข้อมูลการหน่วงของข่าว	ไม่มี	มี	ไม่มี
5. ระบุแหล่งป้อนข่าว	ไม่มี	มี	ไม่มี
6. แสดงผลสถิติเป็นแฟ้มข้อความ	มี	ไม่มี	ไม่มี
7. แสดงผลสถิติเป็นกราฟผ่านทางเว็บเพจ	ไม่มี	มี	มี

จากตารางที่ 2.1 เปรียบเทียบซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูลยูสเนตนิวส์ แสดงให้เห็นว่าโฟลว์สแตตส์จะเก็บรายละเอียดชนิดของข่าวหลักและข่าวรอง ทำให้สามารถทราบถึงจำนวนข้อมูลที่แท้จริงของข่าวนั้น ๆ ส่วนอินโฟลว์สแตตเหมาะสำหรับเครือข่ายที่มีการติดต่อกับหลาย ๆ เครื่องบริการข่าวเพื่อวัดประสิทธิภาพการส่งและการหน่วงของข่าวจากเครื่องบริการข่าวต่าง ๆ ส่วนไอเอ็นเอ็นรีพอร์ตเน้นการแสดงผลข้อมูลสถิติของแฟ้มข้อมูลล็อก จึงเหมาะสำหรับแสดงประสิทธิภาพการทำงานของไอเอ็นเอ็นมากกว่า

ถึงแม้ว่าจะมีซอฟต์แวร์ที่ใช้ในการจัดเก็บข้อมูลสถิติแล้ว แต่ซอฟต์แวร์ที่กล่าวข้างต้นยังไม่สามารถใช้ในการคัดเลือกกลุ่มข่าวได้ ดังนั้นในบทต่อไปนี้จะเป็นการกล่าวถึงวิธีการในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์ เพื่อทำการคัดเลือกกลุ่มข่าวให้มีความเหมาะสมกับความสามารถในการรับส่งข้อมูลยูสเนตนิวส์