

บทที่ 3

การวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์

วิธีการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์นี้ เป็นการกล่าวถึงปัจจัยที่นำมาใช้ในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์ ขั้นตอนในการเตรียมข้อมูลสถิติยูสเนตนิวส์สำหรับการคัดเลือกกลุ่มข่าว และวิธีการวิเคราะห์ข้อมูลยูสเนตนิวส์ในการคัดเลือกกลุ่มข่าวให้มีความเหมาะสมกับความสามารถในการรับส่งข้อมูลของเครือข่าย

3.1 ปัจจัยที่ใช้ในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์

1) ข้อมูลสถิติความต้องการใช้กลุ่มข่าว

เป็นจำนวนความต้องการใช้กลุ่มข่าวของผู้ใช้บริการภายในเครือข่ายในแต่ละกลุ่มข่าว เนื่องจากเป้าหมายหลักของการให้บริการในระบบเครือข่ายทั่ว ๆ ไป ก็คือ การพยายามตอบสนองความต้องการของผู้ใช้บริการ จึงถือได้ว่าความต้องการของผู้ใช้บริการภายในเครือข่ายเป็นปัจจัยที่สำคัญปัจจัยหนึ่ง

2) ข้อมูลสถิติการเรียกใช้บริการกลุ่มข่าว

เป็นข้อมูลสถิติการเรียกใช้บริการกลุ่มข่าวของผู้ใช้บริการภายในเครือข่ายในแต่ละกลุ่มข่าว ถึงแม้ว่าจะมีการเก็บรวบรวมข้อมูลความสนใจและความต้องการใช้ยูสเนตนิวส์ของผู้ใช้บริการภายในเครือข่ายแล้ว แต่เนื่องจากข้อจำกัดความกว้างแถบความถี่ที่มีอยู่ จึงต้องคำนึงถึงปัจจัยในการใช้บริการยูสเนตนิวส์จริงๆ ว่าผู้ให้บริการภายในเครือข่ายได้ให้บริการกลุ่มข่าวอะไรบ้าง เพื่อเป็นการใช้ความกว้างแถบความถี่ที่มีอยู่ให้เกิดประโยชน์มากที่สุด

3) ข้อมูลสถิติจำนวนใบค์ของแต่ละกลุ่มข่าว

เนื่องจากภายในเครือข่ายต่างๆ ต่างมีข้อจำกัดในด้านความสามารถในการรับส่งข้อมูล ดังนั้นถึงแม้ว่าเครือข่ายจะพยายามตอบสนองความต้องการของผู้ใช้บริการภายในเครือข่ายมากเพียงใด ก็ต้องอยู่ภายในขอบเขตข้อจำกัดดังกล่าว

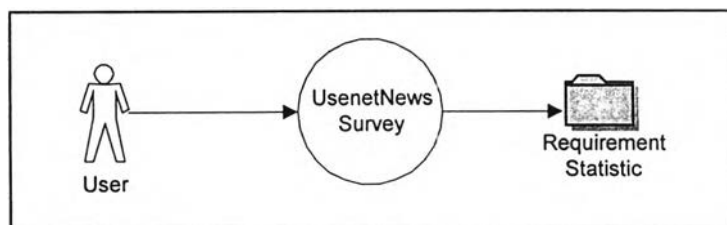
4) ความกว้างแถบความถี่ที่ต้องการในการรับข่าว

เนื่องจากความสามารถในการรับส่งข้อมูลของเครือข่ายมีขีดจำกัด ดังนั้นจำเป็นต้องมีการกำหนดความกว้างแถบความถี่ที่ต้องการสำหรับการบริการยูสเน็ตนิวส์ว่าต้องการที่เปอร์เซ็นต์ของความสามารถในรับ-ส่งข้อมูลของเครือข่ายที่มีอยู่ เพื่อสะดวกในการคัดเลือกกลุ่มข่าวภายในความกว้างแถบความถี่ที่กำหนดไว้

3.2 ขั้นตอนในการเก็บรวบรวมข้อมูลยูสเน็ตนิวส์

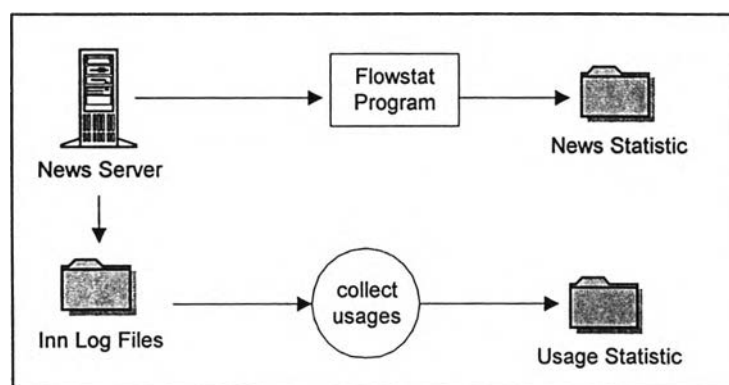
หลังจากได้กำหนดปัจจัยที่ใช้ในการวิเคราะห์ข้อมูลสถิติยูสเน็ตนิวส์แล้ว จะทำการเก็บรวบรวมข้อมูลยูสเน็ตนิวส์สำหรับการคัดเลือกกลุ่มข่าว ดังต่อไปนี้

- 1) เก็บรวบรวมข้อมูลสถิติความต้องการใช้กลุ่มข่าวจากผู้ใช้บริการภายในเครือข่าย โดยให้ผู้ใช้บริการภายในเครือข่ายกรอกข้อมูลแบบสอบถามผ่านทางเว็บเพจที่สร้างขึ้น ดังในรูปที่ 3.1 การเก็บข้อมูลจากผู้ใช้บริการภายในเครือข่าย



รูปที่ 3.1 การเก็บข้อมูลจากผู้ใช้บริการภายในเครือข่าย

2) เก็บรวบรวมข้อมูลสถิติจำนวนไบต์ของแต่ละกลุ่มข่าวและข้อมูลสถิติการเรียกใช้บริการกลุ่มข่าวจากเครื่องบริการข่าวดังในรูปที่ 3.2 การเก็บข้อมูลจากเครื่องบริการข่าวซึ่งข้อมูลสถิติจำนวนไบต์ของแต่ละกลุ่มข่าว เก็บรวบรวมโดยการใส่ซอฟต์แวร์โฟลว์สแตตส์ ส่วนข้อมูลสถิติการเรียกใช้บริการกลุ่มข่าว เก็บรวบรวมโดยการดึงข้อมูลการเรียกใช้บริการกลุ่มข่าวจากแฟ้มข้อมูลล็อกไอเอ็นเอ็น

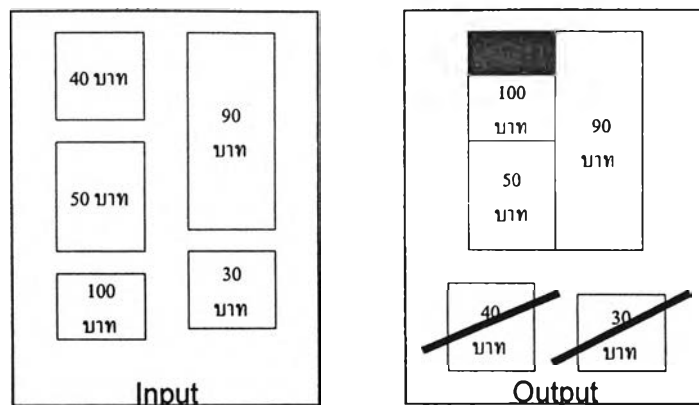


รูปที่ 3.2 การเก็บข้อมูลจากเครื่องบริการข่าว

3.3 วิธีการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์ในการคัดเลือก

ในการทำวิจัยมีเป้าหมายต้องการที่จะทำเครื่องมือทางซอฟต์แวร์ เพื่อช่วยในการบริหาร แต่ในระบบจำเป็นจะต้องทำการอพติไมเซชันในการคัดเลือกกลุ่มข่าว ซึ่งในการวิจัยนี้ได้เลือกหลักการปัญหาเนียบแซค (Knapsack Problem) เนื่องจากมีความอพติไมและใช้ง่าย นอกจากนี้การวิจัยนี้ได้ใช้การทำให้เกิดผลในการแก้ปัญหาเนียบแซคที่หาได้โดยสะดวกจากอินเทอร์เน็ต เพื่อให้เป็นตัวอย่างของส่วนการอพติไม แต่ในการออกแบบได้แยกส่วนนี้ให้สามารถถอดประกอบเปลี่ยนแปลงได้สะดวก เพื่อใส่วิธีการอพติไมอื่น ๆ ให้เหมาะสมตามความต้องการและความสามารถของระบบ

ในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์นี้ได้อาศัยหลักการวิเคราะห์ของ 0/1 Knapsack Problem (Plane and McMillan, 1971: 17-20; Sedgewick, 1992: 596-598) ซึ่งเป็นแนวคิดสำหรับการแก้ปัญหาเรื่องการใส่สิ่งของต่าง ๆ ให้ได้มูลค่ามากที่สุดลงในเป้ที่มีน้ำหนักจำกัด และมีสิ่งของอะไรบางอย่างจะถูกเลือกใส่ลงไปภายในข้อจำกัดดังกล่าว



รูปที่ 3.3 ภาพแนวคิดของปัญหาเน็บแซค (Syslo, Deo, et al 1983)

จากรูปที่ 3.3 ภาพแนวคิดของปัญหาเน็บแซค สามารถอธิบายได้ดังนี้ มีสิ่งของอยู่ 5 ชิ้น แต่ละชิ้นมีมูลค่าแตกต่างกันคือ 30 บาท 40 บาท 50 บาท 90 บาท และ 100 บาท แต่เนื่องจากการใส่สิ่งของมารวมกันมีจำกัด จึงเลือกเฉพาะสิ่งของที่ใส่ลงไปในเรื่องที่ดังกล่าวให้ได้มูลค่ามากที่สุด จากรูปจะเป็นการเลือกเอาสิ่งของที่มีมูลค่า 50 บาท 90 บาท และ 100 บาทเท่านั้น ส่วนสิ่งของที่มีมูลค่า 30 บาท และ 40 บาท จะไม่ถูกเลือก

จากการศึกษาแนวคิดของปัญหาเน็บแซคนี้ พบว่าสามารถนำมาใช้ในการคัดเลือกหากลุ่มข่าวให้เหมาะสมกับความสามารถในการรับส่งข้อมูลของเครือข่ายได้ และการวิจัยในครั้งนี้ได้นำวิธีการแก้ปัญหานเน็บแซคของ Maciej M. Syslo, Narsingh Deo และ Janusz S. Kowalik (Syslo, Deo, et al 1983) มาใช้ โดยมีการกำหนดสมการขั้นต้น ดังนี้

$$\begin{array}{lll} \max & \text{SUM} (P_i * X_i) & i = 1 \text{ to } N \\ \text{subject to} & \text{SUM} (W_i * X_i) \leq V & \text{for } i = 1 \text{ to } N \\ & X_i = 0 \text{ or } 1 & \text{for } i = 1 \text{ to } N \end{array}$$

$$P_1 / W_1 \geq P_2 / W_2 \geq P_3 / W_3 \dots \geq P_n / W_n$$

โดยที่ P_i คือ ผลกำไรของตัวแปร i
 W_i คือ น้ำหนักของตัวแปร i
 N คือ จำนวนสิ่งของทั้งหมดที่จะเลือก
 V คือ น้ำหนักสูงสุดที่เน็บแซคสามารถรับได้

X_i คือ ค่า 0 หรือ 1 โดยสิ่งของที่จะถูกเลือกมีค่า $X_i = 1$ และ สิ่งของที่ไม่ถูกเลือกมีค่า $X_i = 0$

ซึ่งตัวแปร P_i แทนด้วยจำนวนความต้องการที่นำมาจากเพิ่มข้อมูลสถิติความต้องการใช้กลุ่มข่าวรวมกับจำนวนการใช้งานของแต่ละกลุ่มข่าวที่นำมาจากเพิ่มข้อมูลสถิติการเรียกใช้กลุ่มข่าว

ตัวแปร W_i แทนด้วยจำนวนไบต์ของแต่ละกลุ่มข่าว โดยนำมาจากเพิ่มข้อมูลสถิติจำนวนไบต์ของแต่ละกลุ่มข่าว

ตัวแปร N แทนด้วยจำนวนกลุ่มข่าวทั้งหมดที่จะเลือก

ตัวแปร V แทนด้วยความกว้างแถบความถี่ที่ต้องการในการรับข่าว โดยผู้บริหารยูสเนตนิวส์จะเป็นผู้กำหนดความกว้างแถบความถี่ที่ต้องการในการรับข่าวเอง ตามความเหมาะสมของเครือข่าย

ตัวแปร X_i แทนด้วยกลุ่มข่าวที่จะเลือกมีค่า $X_i = 1$ และกลุ่มข่าวที่ไม่ถูกเลือกมีค่า $X_i = 0$

เมื่อผ่านขั้นตอนวิธีการในการแก้ปัญหาเนียบแซคแล้ว จะได้ผลลัพธ์แสดงค่า ดังนี้

- ตรวจสอบการแก้ปัญหาดังกล่าวเป็นการแก้ปัญหาสูงสุด (Optimal Solution) หรือเป็นการแก้ปัญหาได้เพียงบางส่วน (Partial Feasible Solution) ซึ่งจะแสดงค่าเป็นจริงหรือเท็จ

- ผลกำไรรวมที่ได้
- น้ำหนักรวมที่เหลือจากการเลือกสิ่งของแล้ว
- แสดงสิ่งของชิ้นไหนที่ถูกเลือกและสิ่งของชิ้นไหนที่ไม่ถูกเลือก

ขั้นตอนวิธีการแก้ปัญหาเนียบแซคนี้ ใช้เวลาในการประมวลผลเป็น $O(n^2)$ และจากตารางที่ 3.1 การประยุกต์ใช้ขั้นตอนวิธีการแก้ปัญหาเนียบแซคกับการคัดเลือกกลุ่มข่าว เป็นการแสดงการประยุกต์ตัวแปรต่างๆ ที่ใช้ในขั้นตอนวิธีการแก้ปัญหาเนียบแซคกับปัจจัยที่ใช้ในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์ เพื่อทำการคัดเลือกกลุ่มข่าว

ตารางที่ 3.1 การประยุกต์ใช้ขั้นตอนวิธีการแก้ปัญหาแนบแซคกับการคัดเลือกกลุ่มข่าว

ตัวแปรในขั้นตอนวิธีการแก้ปัญหาแนบแซค	ตัวแปรที่มาจากปัจจัยการคัดเลือกกลุ่มข่าว
สิ่งของที่จะเลือก	กลุ่มข่าวที่จะเลือก
จำนวนสิ่งของที่จะเลือก (I)	จำนวนกลุ่มข่าวที่จะเลือก
ผลกำไรของตัวแปร i (P_i)	จำนวนความต้องการและการใช้งานของแต่ละกลุ่มข่าว
น้ำหนักของตัวแปร i (W_i)	จำนวนไบต์ของแต่ละกลุ่มข่าว
น้ำหนักสูงสุดที่รับได้ (V)	ความกว้างแถบความถี่ที่ต้องการในการรับข่าว
สิ่งของที่ถูกเลือก ($X=1$)	ชื่อกลุ่มข่าวที่ถูกเลือก
สิ่งของที่ไม่ถูกเลือก ($X=0$)	ชื่อกลุ่มข่าวที่ไม่ถูกเลือก

ซึ่งขั้นตอนวิธีการแก้ปัญหาแนบแซคที่นำมาประยุกต์ใช้กับการคัดเลือกกลุ่มข่าว จะกล่าวถึงโดยละเอียดในบทที่ 5 การพัฒนาโปรแกรม

จากการเก็บรวบรวมข้อมูลตามปัจจัยที่ใช้ในการวิเคราะห์ข้อมูลสถิติยูสเนตนิวส์ ตลอดจนการใช้แนวความคิดในการแก้ปัญหาแนบแซค และการนำขั้นตอนวิธีการแก้ปัญหาแนบแซคมาประยุกต์ใช้ในการคัดเลือกกลุ่มข่าวแล้ว ในบทต่อไปจะกล่าวถึงการออกแบบระบบ เพื่อนำโปรแกรมในการคัดเลือกกลุ่มข่าวนี้อมาช่วยในการจัดการยูสเนตนิวส์