รหัสโครงการ13P33W002

การวิเคราะห์และจัดประเภททัศนคติเชิงความคิดเห็น
BEST 2011 : การแข่งขันสุดยอดซอฟต์แวร์ประมวลผลภาษาไทย
(Thai Language Processing Software Contest)

รายงานฉบับสมบูรณ์

เสนอต่อ

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ

กระทรวงวิทยาศาสตร์และเทคโนโลยี

ได้รับเงินทุนอุดหนุนโครงการวิจัย พัฒนาและวิศวกรรม

โครงการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ 13

ประจำปีงบประมาณ 2553

โดย

นางสาวณัฐวรรณ สุวรรณจิต

นางสาวพัชรินทร์ อุดมชัยเดช

อาจารย์ที่ปรึกษาโครงการ อาจารย์โอภาส วงษ์ทวีทรัพย์

ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร

กิตติกรรมประกาศ

ปริญญานิพนธ์เรื่อง การวิเคราะห์และจัดประเภททัศนคติเชิงความคิดเห็น สำเร็จลุล่วงไปได้ด้วยดีต้องขอขอบพระคุณ อาจารย์โอภาส วงษ์ทวีทรัพย์ อาจารย์ที่ปรึกษาปริญญานิพนธ์ ที่คอยให้ความรู้ คำแนะนำ คำสั่งสอน และข้อคิดเห็นเป็นอย่างดีมาตลอด และยังให้คำปรึกษาที่เป็นประโยชน์ในการพัฒนาและแก้ไขปัญหาที่เกิดขึ้นระหว่างการจัดทำปริญญานิพนธ์ฉบับนี้

ขอขอบพระคุณบิดา มารดา และสมาชิกในครอบครัวทุกคนที่เป็นห่วง เอาใจใส่ ดูแล และเป็นกำลังใจที่สำคัญ รวมถึงเพื่อนๆ พี่ๆ น้องๆ ที่คอยให้กำลังใจ คำแนะนำ และช่วยเหลือกันเสมอมา

ขอขอบพระคุณโครงการการแข่งขันพัฒนาโปรแกรมคอมพิวเตอร์แห่งประเทศไทย ครั้งที่ 13 จากศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ สำนักงานพัฒนาวิทยาศาสตร์และเทคโนโลยีแห่งชาติ ที่ให้เงินทุนสนับสนุนในการพัฒนางานในครั้งนี้

สุดท้ายนี้ขอขอบคุณคณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร ที่ให้โอกาสข้าพเจ้าได้เข้ามาศึกษาเรียนรู้ประสบการณ์ใหม่ๆ มากมาย ได้พบเพื่อนๆ พี่ๆ น้องๆ ที่น่ารัก คณาจารย์ที่คอยดูแล สั่งสอนเสมอมา ซึ่งถ้าไม่มีทุกๆ ท่านที่กล่าวมา ปริญญานิพนธ์นี้คงไม่เกิดขึ้น และไม่สามารถสำเร็จลุล่วงได้

ณัฐวรรณ สุวรรณจิต
พัชรินทร์ อุดมชัยเดช

บทคัดย่อ

ในปัจจุบันการแสดงความคิดเห็นในประเด็นต่างๆ ถูกแสดงผ่านอินเตอร์เน็ตมากขึ้น ซึ่งข้อคิดเห็นเหล่านี้เป็นประโยชน์อย่างยิ่งต่อบริษัท และองค์กรต่างๆ ในการหาทัศนคติที่มีต่อสินค้าและบริการ เพื่อนำเอาข้อมูลเหล่านี้ไปปรับปรุงคุณภาพของผลิตภัณฑ์ให้ดียิ่งขึ้น แต่ด้วยปริมาณข้อคิดเห็นจำนวนมากที่เกิดขึ้นนั้น ทำให้การอ่านข้อคิดเห็นทั้งหมดเพื่อให้ได้ข้อสรุปทัศนคติของข้อคิดเห็นในหัวข้อที่สนใจออกมานั้นเป็นไปได้ยาก ซึ่งคงดีกว่าหากอ่านเพียงบางข้อคิดเห็นแล้วสามารถสรุปได้ว่าทัศนคติที่มีต่อหัวข้อนั้นเป็นอย่างไร และกลุ่มของผู้ที่แสดงความคิดเห็นที่คล้ายคลึงกันนั้นเป็นบุคคลใด มีข้อมูลที่เกี่ยวข้องอย่างไรบ้าง โดยในโครงการนี้จะสนใจในเรื่องของการตัดสินทัศนคติจากข้อคิดเห็นในเว็บไซต์ และจัดกลุ่มผู้ที่แสดงความคิดเห็นที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกัน

โดยโครงการนี้เป็นการสร้างระบบที่มีความสามารถในการตัดสินทัศนคติของข้อคิดเห็นภาษาไทยที่เกี่ยวกับหัวข้อจากเว็บไซต์ และจัดกลุ่มของผู้ใช้จากทัศนคติที่ผู้ใช้แสดงไว้บนเว็บไซต์ผ่านทางเว็บเบราว์เซอร์ (Web browser)

ทางผู้พัฒนาคาดหวังว่าระบบที่สร้างขึ้นจะเป็นประโยชน์แก่ผู้ใช้ และจะได้รับการพัฒนาต่อเพื่อเพิ่มความสามารถในการใช้งาน และได้รับความสนใจมากขึ้นในอนาคต

Abstract

At present,commenting on various issues vai the internet mostly. Those comments are useful to comapany and and organization to find the attitude towards products and services.The company will bring this information to improve better quality of the product.But the amount of comments are difficult to read all comments to conclude the attitude of comments on interest topic.Thus,it would be better to read some comments and it can conclude the attitude toward the topic,a group of people who post similar comments and information about group.

This project interested in judging the attitude from comments on website and cluster opinion holder who has similar comments.This project creates a system that has the ability to judge the attitude of thai comments on topic from website and cluter opinion holder from his comments via a web browser.

The developers expect this system will be useful to user and will be develop toc enhance the ability to use and receive much attention in the future.

คำสำคัญ: BEST 2010, Opinion Mining, ทัศนคติ (Attitudes), การจัดประเภท (Clustering)

สารบัญ

บทนำ

เนื่องจากความแพร่หลายของการใช้งานอินเตอร์เน็ตในปัจจุบัน การแสดงความคิดเห็นในประเด็นต่างๆ จึงถูกแสดงผ่านอินเตอร์เน็ตมากขึ้น ซึ่งข้อคิดเห็นเหล่านี้เป็นประโยชน์อย่างยิ่งต่อบริษัท และองค์กรต่างๆ ในการหาทัศนคติที่มีต่อสินค้า และบริการ เพื่อนำเอาข้อมูลเหล่านี้ไปปรับปรุงคุณภาพของผลิตภัณฑ์ให้ดียิ่งขึ้น แต่ด้วยปริมาณข้อคิดเห็นจำนวนมากที่เกิดขึ้นนั้น ทำให้การอ่านข้อคิดเห็นทั้งหมดเพื่อให้ได้ข้อสรุปทัศนคติของข้อคิดเห็นในหัวข้อที่สนใจออกมานั้นเป็นไปได้ยาก ซึ่งคงดีกว่าหากอ่านเพียงบางข้อคิดเห็นแล้วสามารถสรุปได้ว่าทัศนคติที่มีต่อหัวข้อนั้นเป็นอย่างไร และกลุ่มของผู้ที่แสดงความคิดเห็นที่คล้ายคลึงกันนั้นเป็นบุคคลใด มีข้อมูลที่เกี่ยวข้องอย่างไรบ้าง เช่น เพศ (Sex) อายุ (Age) ความสนใจ (Interests) เป็นต้น เพื่อนำข้อมูลจากการจัดกลุ่มของผู้แสดงความคิดเห็นไปใช้ในการวางแผนการตลาดให้เกิดประโยชน์ต่อธุรกิจต่อไป

จากการศึกษาการทำเหมืองความคิดเห็น (Opinion Mining) ทำให้ทราบถึงหลักการ และแนวความคิดในการหาทัศนคติ และตัดสินทัศนคติจากข้อคิดเห็นว่ามีทัศนคติเป็นเชิงบวก เชิงลบ หรือเป็นกลาง ซึ่งระบบที่รองรับการวิเคราะห์ทัศนคติจากข้อคิดเห็นที่เป็นภาษาไทยนั้นไม่เป็นที่แพร่หลายมากนัก และยังไม่มีการจัดกลุ่มของผู้แสดงความคิดเห็นเพื่อนำข้อมูลโดยรวมของผู้แสดงความคิดเห็นมาใช้ประโยชน์แต่อย่างใด

ดังนั้นข้าพเจ้าจึงเกิดแนวความคิดในการพัฒนาระบบการวิเคราะห์ และจัดประเภททัศนคติเชิงความคิดเห็นขึ้นมา เพื่อช่วยตัดสินทัศนคติจากข้อคิดเห็นในเว็บไซต์ที่ผู้ใช้สนใจได้อย่างรวดเร็ว โดยใช้หลักในการทำเหมืองความคิดเห็น (Opinion Mining) และแสดงผลในรูปแบบของกราฟ รวมถึงใช้หลักความน่าจะเป็นตรวจสอบความถูกต้องของทัศนคติในแต่ละข้อคิดเห็น โดยให้ผู้ใช้ตัดสินข้อคิดเห็นตอบกลับมา (Relevance Feedback) เพื่อให้ทัศนคติโดยรวมของข้อคิดเห็นนั้นๆ มีความถูกต้องแม่นยำมากขึ้น และทำการจัดกลุ่มผู้ที่แสดงความคิดเห็นจากข้อคิดเห็น โดยแสดงผลลัพธ์ในรูปแบบของแผนภาพ เพื่อให้ผู้ใช้สามารถมองเห็นกลุ่มของผู้แสดงความคิดเห็นที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกัน และข้อมูลที่เกี่ยวข้องของผู้แสดงความคิดเห็นในแต่ละกลุ่มได้ชัดเจนยิ่งขึ้น

3. วัตถุประสงค์และเป้าหมาย

3.1 สร้างระบบที่มีความสามารถในการตัดสินทัศนคติของข้อคิดเห็นภาษาไทยที่เกี่ยวกับหัวข้อจากเว็บไซต์ได้

3.2 เพื่อให้สามารถจัดกลุ่มของผู้ใช้จากทัศนคติที่ผู้ใช้แสดงไว้บนเว็บไซต์ได้

4. รายละเอียดของการพัฒนา

4.1 เนื้อเรื่องย่อ (Story Board)

4.1.1 หลักการทำงานของระบบ

การทำงานของระบบประกอบด้วยกันอยู่ 2 ส่วนคือ ส่วนของการหาทัศนคติจากข้อคิดเห็นที่เกี่ยวข้องกับหัวข้อที่สนใจ และส่วนของการจัดกลุ่มของผู้แสดงความคิดเห็นที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกัน

การทำงานในส่วนของการหาทัศนคติจากข้อคิดเห็น จะรับข้อมูลเข้าเป็น URL จากเว็บไซต์พันทิป โดยระบบจะเก็บข้อความจากหน้าเว็บไซต์นั้นมาทำการตัดคำ และหาคำสำคัญของแต่ละข้อคิดเห็น แล้วตัดสินทัศนคติของข้อคิดเห็นจากคำสำคัญโดยเปรียบเทียบจากทัศนคติของคำที่เก็บไว้ และสรุปข้อมูลอยู่ในรูปแบบของกราฟ และแสดงข้อคิดเห็นที่มีต่อหัวข้อนั้น เพื่อให้ผู้ใช้สามารถปรับปรุงทัศนคติที่มีต่อข้อคิดเห็นนั้น ๆ และระบบจะนำข้อมูลที่ได้ไปปรับปรุงทัศนคติของคำในฐานข้อมูลของระบบ

การทำงานในส่วนของการจัดกลุ่มของผู้แสดงความคิดเห็น ที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกันจะรับข้อมูลเข้าเป็น URLจากเว็บไซต์ www.vanilla.in.th และ www.cosmetic.in โดยระบบจะนำข้อคิดเห็น และข้อมูลส่วนตัวของผู้แสดงความคิดเห็นมาทำการจัดเก็บ และนำมาจัดกลุ่มของผู้แสดงความคิดเห็น ที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกัน และสรุปข้อมูลโดยรวมของรายชื่อของผู้แสดงความคิดเห็นในแต่ละกลุ่ม

4.1.2 การออกแบบส่วนประสานงานกับผู้ใช้ (User Interface Design)

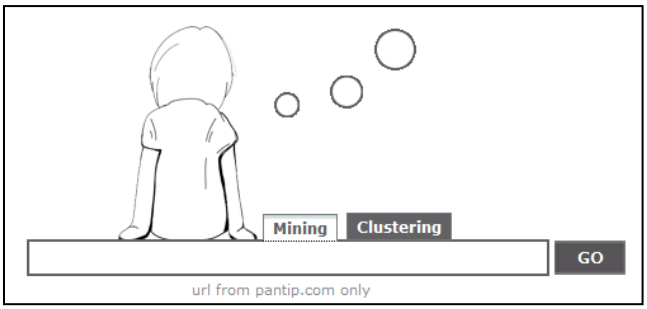
รูปที่ 4.1 ตัวอย่างส่วนประสานงานกับผู้ใช้

ระบบจะประกอบด้วยส่วนหลักๆ อยู่ 3 ส่วน ประกอบด้วยส่วนของการป้อนเว็บไซต์ (Input), ส่วนการแสดงผลการตัดสินทัศนคติ (Opinion Mining) และส่วนการแสดงผลการจัดกลุ่มผู้แสดงความเห็น (Opinion Clustering)

4.1.2.1 ส่วนของการป้อนเว็บไซต์ (Input) ประกอบด้วย

- Mining สำหรับเลือกดูผลลัพธ์ในการตัดสินทัศนคติ

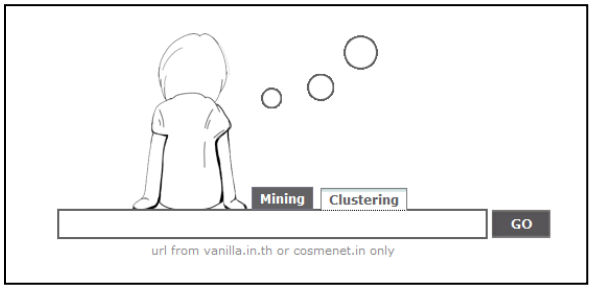
ระบบจะนำเว็บไซต์ที่ผู้ใช้ป้อนเข้ามาไปทำการวิเคราะห์และแสดงผลการตัดสินทัศนคติออกมา



รูปที่ 4.2 ส่วนของการป้อนเว็บไซต์ในการทำ Mining

- Clustering สำหรับเลือกดูผลลัพธ์ในการจัดกลุ่มผู้แสดงความเห็น

ระบบจะนำเว็บไซต์ที่ผู้ใช้ป้อนเข้ามาไปทำการวิเคราะห์และแสดงผลการจัดกลุ่มผู้แสดงความเห็นออกมา

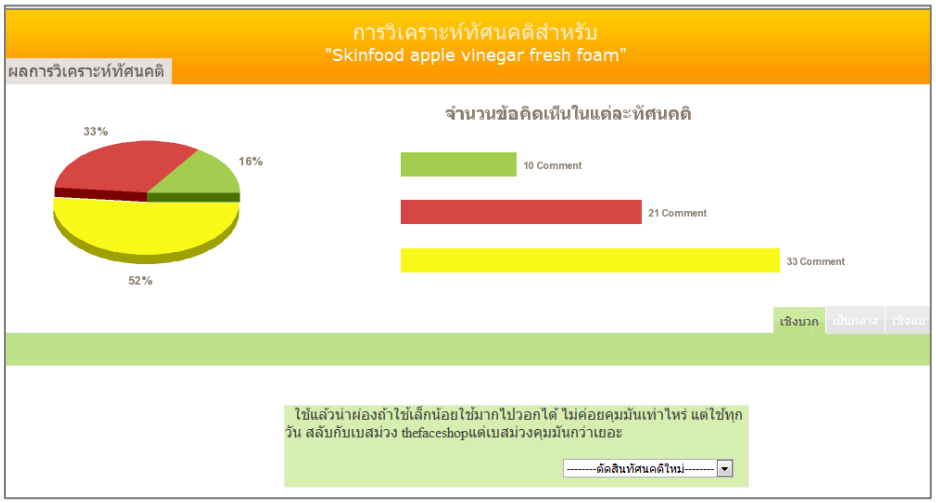


รูปที่ 4.3 ส่วนของการป้อนเว็บไซต์ในการทำ Clustering

4.1.2.2 ส่วนการแสดงผลการตัดสินทัศนคติ (Opinion Mining)

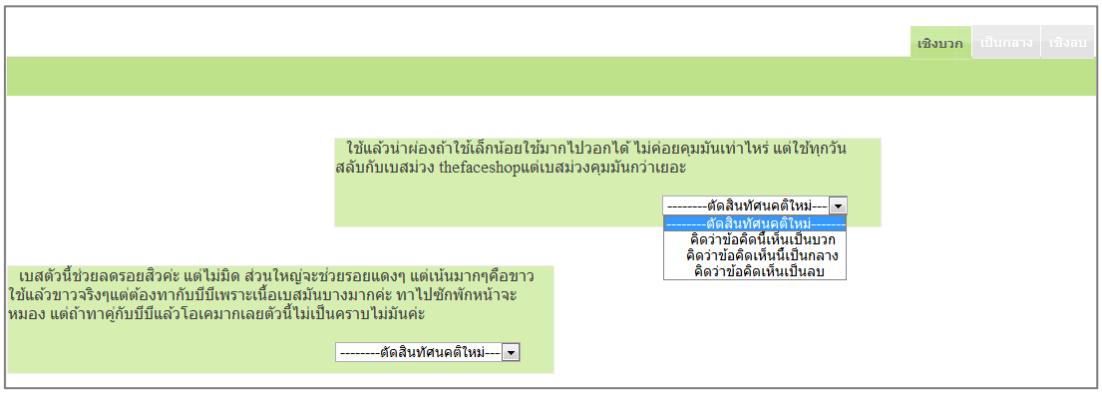
ระบบจะแสดงผลการตัดสินทัศนคติจากเว็บไซต์ที่ผู้ใช้ป้อนเข้ามาในรูปแบบของกราฟวงกลมที่คำนวณอยู่ในรูปแบบของเปอร์เซ็นต์ และกราฟแท่งแนวนอนแสดงจำนวนข้อคิดเห็นในแต่ละทัศนคติ พร้อมทั้งแสดงตัวอย่างข้อคิดเห็นเพื่อให้ผู้ใช้ช่วยตัดสินทัศนคติใหม่ โดยการ

แสดงตัวอย่างข้อคิดเห็นนั้น สามารถเลือกพิจารณาได้ว่าจะพิจารณาในส่วนที่มีทัศนคติเป็นบวก ลบ หรือเป็นกลาง

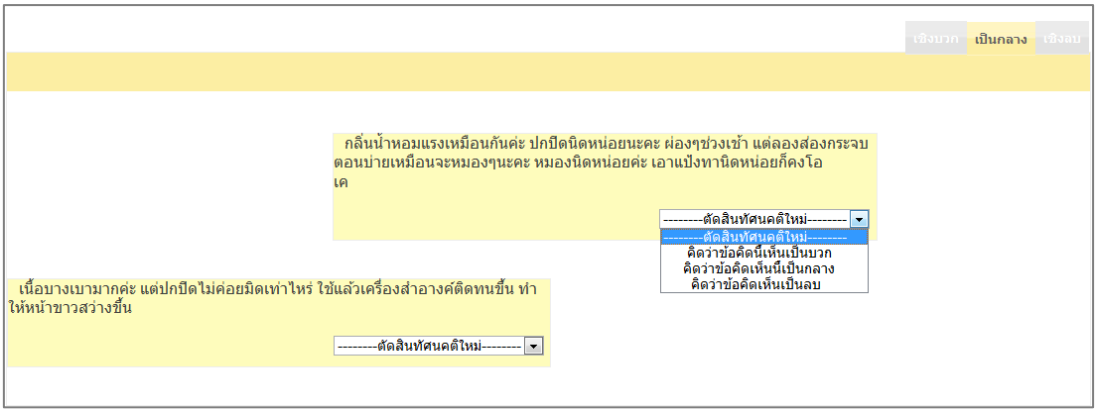


รูปที่ 4.4 ส่วนการแสดงผลการตัดสินทัศนคติ

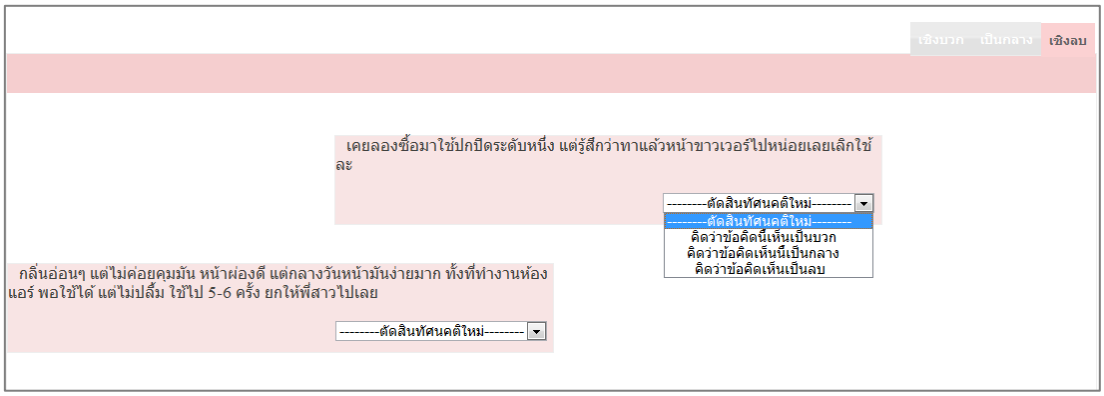
รูปที่ 4.5 กราฟแสดงผลการตัดสินทัศนคติในรูปแบบกราฟวงกลม และกราฟแท่งแนวนอน



รูปที่ 4.6 ส่วนแสดงตัวอย่างข้อคิดเห็นที่มีทัศนคติเป็นบวกเพื่อให้ผู้ใช้ช่วยตัดสินทัศนคติใหม่



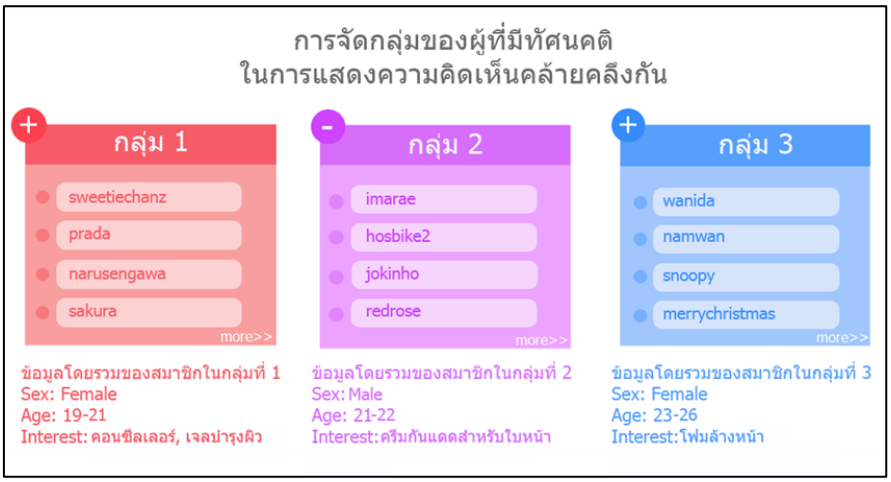
รูปที่ 4.7 ส่วนแสดงตัวอย่างข้อคิดเห็นที่มีทัศนคติเป็นกลางเพื่อให้ผู้ใช้ช่วยตัดสินทัศนคติใหม่



รูปที่ 4.8 ส่วนแสดงตัวอย่างข้อคิดเห็นที่มีทัศนคติเป็นลบเพื่อให้ผู้ใช้ช่วยตัดสินทัศนคติใหม่

4.1.2.3 ส่วนการแสดงผลการจัดกลุ่มผู้แสดงความเห็น (Opinion Clustering)

ระบบจะแสดงผลการจัดกลุ่มผู้แสดงความเห็นจากเว็บไซต์ที่ผู้ใช้ป้อนเข้ามา พร้อมทั้งแสดงข้อมูลโดยสรุปของแต่ละกลุ่ม

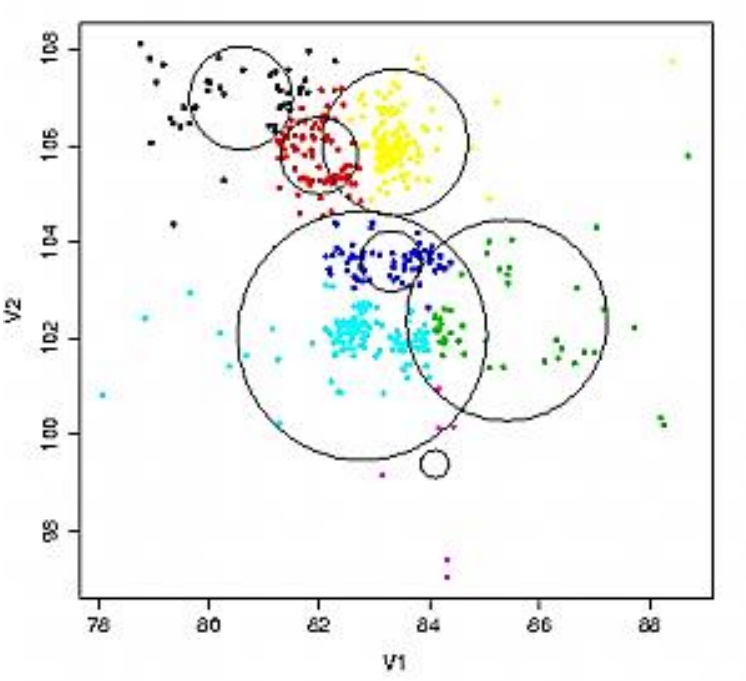


รูปที่ 4.9 ส่วนการแสดงผลการจัดกลุ่มผู้แสดงความเห็น

4.2 ทฤษฎีและความรู้ที่เกี่ยวข้อง

4.2.1 เทคนิคการจำแนกกลุ่มออกเป็น k กลุ่มโดยการพิจารณาจากค่าเฉลี่ย

เทคนิคการจำแนกกลุ่มออกเป็น k กลุ่มโดยการพิจารณาจากค่าเฉลี่ย เป็นวิธีการแบ่งกลุ่มข้อมูลของวัตถุทั้งหมดออกตามจำนวนกลุ่มที่ต้องการ K กลุ่ม และค่า K ต้องน้อยกว่าจำนวนของวัตถุทั้งหมด (N) ซึ่งจำนวนกลุ่มต้องเป็นเลขจำนวนเต็มบวก และการจัดกลุ่มวัตถุต้องอาศัยความเหมือนของวัตถุ โดยวัดจากระยะห่างที่น้อยที่สุดระหว่างวัตถุกับจุดศูนย์กลางของกลุ่มทั้งหมด เพื่อจัดวัตถุเข้าสู่กลุ่มต่างๆ ตามจำนวนกลุ่มที่ต้องการ



รูปที่ 4.10 ภาพแสดงการจัดกลุ่ม โดยวิธีการ K-mean [6]

การใช้เทคนิคการจำแนกกลุ่มออกเป็น k กลุ่มโดยการพิจารณาจากค่าเฉลี่ย จะมีขั้นตอนในการทำงานเบื้องต้นดังต่อไปนี้

1. กำหนดจำนวนกลุ่มตามที่ต้องการ K กลุ่ม
2. สุ่มจุดศูนย์กลางในการเริ่มต้นจัดกลุ่มจำนวน K จุด
3. คำนวณหาระยะห่างระหว่างข้อมูลนั้นกับศูนย์กลางในการจัดกลุ่มที่ทำการสุ่มมาโดยใช้การวัดระยะห่างแบบยูคลิด (Euclidean distance) ซึ่งสมการในการคำนวณหาระยะห่างแบบยูคลิด คือ

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

จัดกลุ่มข้อมูลโดยพิจารณาจากระยะห่างที่น้อยที่สุดระหว่างจุดศูนย์กลางกับข้อมูลนั้น ๆ

4. คำนวณหาค่าเฉลี่ยของระยะห่างระหว่างจุดศูนย์กลางกับข้อมูล และย้ายจุดศูนย์กลางการจัดกลุ่มไปยังจุดที่คำนวณค่าเฉลี่ยมาได้

5. ทำการจัดกลุ่มใหม่อีกครั้งจนกว่าค่าเฉลี่ยของแต่ละกลุ่มจะไม่มีการเปลี่ยนแปลง

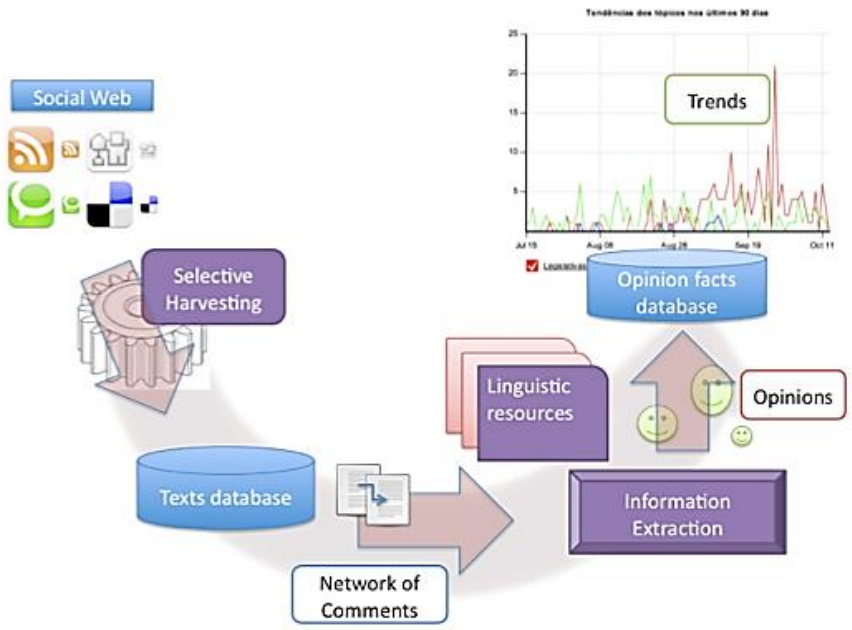
4.2.2 การเรียนรู้ของเครื่อง

การพัฒนาเทคนิควิธี เพื่อให้คอมพิวเตอร์สามารถเรียนรู้ โดยเน้นที่วิธีการเพื่อสร้างโปรแกรมคอมพิวเตอร์จากการวิเคราะห์ชุดข้อมูล การเรียนรู้ของเครื่องจึงเกี่ยวข้องอย่างมากกับสถิติศาสตร์ เนื่องจากทั้งสองสาขาศึกษาการวิเคราะห์ข้อมูลเช่นเดียวกัน ซึ่งอัลกอริทึมการเรียนรู้ของเครื่อง จัดแบ่งได้ตามลักษณะผลลัพธ์ โดยทั่วไปแล้วจะแบ่งเป็น

- การเรียนรู้แบบมีผู้สอน (Supervised Learning) เป็นอัลกอริทึมสร้างฟังก์ชันซึ่งเชื่อมระหว่างข้อมูลเข้ากับผลที่ต้องการ
- การเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เป็นอัลกอริทึมสร้างโมเดลจากชุดข้อมูลเข้า
- การเรียนรู้แบบเสริมกำลัง (Reinforcement Learning) เป็นอัลกอริทึมเรียนแผนซึ่งกำหนดการกระทำของระบบจากสิ่งที่สังเกตได้
- transduction เป็นอัลกอริทึมที่เหมือนกับการเรียนรู้แบบมีผู้สอน แต่ไม่ได้สร้างฟังก์ชันขึ้นมาอย่างชัดเจน โดยเน้นไปที่การพยายามทำนายชุดผลลัพธ์ใหม่ โดยอิงจากชุดข้อมูลเข้าที่เรียน, ชุดผลลัพธ์ที่เรียน, และชุดข้อมูลเข้าใหม่
- การเรียนวิธีการเรียน (Learning to Learn, Meta-Learning) เป็นอัลกอริทึมที่เรียนวิธีการเรียนรู้ของตนเอง โดยปรับปรุง Inductive Bias ที่เป็นข้อสมมติฐานที่อัลกอริทึมใช้ในการเรียนรู้

4.2.3 เทคนิคการทำเหมืองความคิดเห็น

การทำเหมืองความคิดเห็น เป็นกระบวนการอัตโนมัติเพื่อใช้ตรวจสอบทัศนคติของผู้พูดหรือผู้เขียนในหัวข้อเรื่องใดเรื่องหนึ่ง โดยอาศัยการสอนคอมพิวเตอร์ให้พิจารณาอารมณ์ความรู้สึกด้วย Natural Language Processing (NLP) ซึ่งระบบจะใช้หลักการทำเหมืองความคิดเห็น ในการหาทัศนคติจากข้อคิดเห็นของหัวข้อบนเว็บไซต์



รูปที่ 4.11 ภาพแสดงการทำเหมืองความคิดเห็น [7]

โดยหลักการทำเหมืองความคิดเห็นมีดังต่อไปนี้

1. ทำการตัดคำให้กับแต่ละประโยค
2. หาประโยคข้อคิดเห็นโดยนำประโยคที่ทำการตัดคำแล้วมาตรวจสอบว่ามีคำแสดงอารมณ์ (Sentiment word) หรือไม่ ถ้ามีคำแสดงอารมณ์เกิดขึ้นแสดงว่าประโยคนั้นมีการแสดงความคิดเห็น ให้เก็บประโยคนั้นไว้
3. นำคำแสดงอารมณ์ในแต่ละประโยคข้อคิดเห็นมาตรวจสอบว่ามีทัศนคติเป็นอย่างไร (เชิงบวก เชิงลบ หรือเป็นกลาง)
4. สรุปทัศนคติให้กับแต่ละประโยค ซึ่งหากประโยคใดมีคำที่มีทัศนคติเชิงบวกมาก ประโยคนั้นก็จะมีทัศนคติเป็นบวก และในทางกลับกัน หากประโยคใดมีคำที่มีทัศนคติเชิงลบมาก ประโยคนั้นก็จะมีทัศนคติเป็นลบ ส่วนประโยคใดที่มีทัศนคติกำกวมไม่แน่นอน ก็จะให้ประโยคนั้นมีทัศนคติเป็นกลาง
5. สรุปทัศนคติโดยรวมของหัวข้อนั้นว่ามีประโยคที่แสดงทัศนคติเกี่ยวกับหัวข้อที่สนใจเป็นเชิงบวก เชิงลบ หรือเป็นกลางด้วยอัตราส่วนเท่าใด

4.2.4 การตัดคำภาษาไทย (Thai Word Segmentation)

การตัดคำภาษาไทยสามารถแบ่งได้เป็น 3 วิธีหลัก ๆ คือ การตัดคำโดยใช้กฎ (Rules หรือ Grammar), การตัดคำโดยใช้พจนานุกรม และการตัดคำโดยใช้คลังข้อความ (Corpus)

1. การตัดคำโดยใช้กฎ เป็นการพยายามสร้าง Grammar ให้กับภาษาไทยเพื่อที่จะหาวิธีในการสร้างโปรแกรมเพื่อใช้สำหรับตัดคำ

2. การตัดคำโดยใช้พจนานุกรมเป็นวิธีที่ง่ายและรวดเร็วมีความความถูกต้องสูงแต่ขึ้นอยู่กับขนาด และคำในพจนานุกรมที่ใช้ ข้อเสียของวิธีการนี้คือ ต้องใช้ขนาดของหน่วยความจำในการประมวลผลมาก การตัดคำที่ใช้วิธีการนี้ คือ

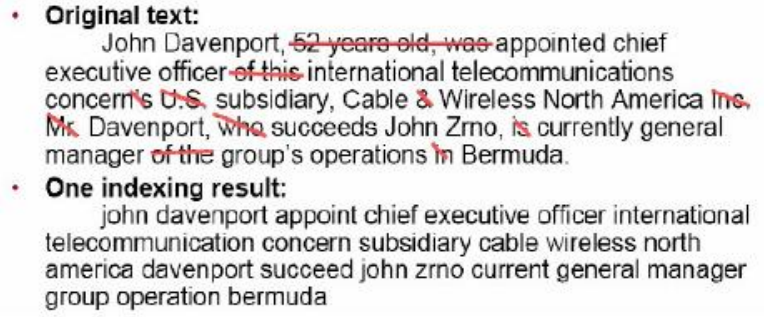
- **วิธีการตัดคำแบบยาวที่สุด (Longest Matching)** จะเลือกคำที่ยาวที่สุดโดยเริ่มจากตัวอักษรซ้ายสุดของข้อความนั้นไปยังตัวอักษรถัดไป จนกว่าจะพบว่าเป็นคำที่มีอยู่ในพจนานุกรม จากนั้นค้นหาคำถัดไปจนกว่าจะครบข้อความ ในกรณีที่พบว่าเป็นคำในพจนานุกรมจากจุดเริ่มต้นเดียวกัน จะเลือกคำที่ยาวที่สุด เช่น “ฉันนั่งตากลมที่หน้าบ้าน” สามารถตัดคำได้ 2 แบบคือ ฉัน-นั่ง-ตา-กลม-ที่-หน้า-บ้าน และ ฉัน-นั่ง-ตาก-ลม-ที่-หน้า-บ้าน จะพบว่าประโยคที่มีคำที่มีความหมายในพจนานุกรมและยาวที่สุดคือ ฉัน-นั่ง-ตาก-ลม-ที่-หน้า-บ้าน
- **วิธีการตัดคำแบบสอดคล้องมากที่สุด (Maximal Matching)** วิธีการตัดคำแบบนี้เป็นการหาวิธีในการตัดคำที่สามารถจะเป็นไปได้ทั้งหมดและเลือกข้อความที่แบ่งแล้วมีจำนวนคำน้อยที่สุด เช่น "ไปหามเหสี" สามารถตัดคำได้ 2 แบบ คือไป-หาม-เห-สี และ ไป-หา-มเหสี ซึ่งแบบที่ 2 มีจำนวนคำที่ตัดได้ 3 คำ แบบที่ 1มี 4คำ จึงเลือกใช้แบบที่ 1 ในการใช้งาน ส่วนในกรณีที่มีจำนวนคำที่เท่ากันจะใช้วิธีการตัดคำแบบยาวที่สุดเข้ามาช่วย

3. หลักการตัดคำโดยใช้คลังข้อมูล (Corpus Based Approach) การตัดคำโดยใช้คลังข้อมูลเป็นการตัดคำโดยนำวิธีการทางสถิติ (Statistical Techniques) เข้ามาใช้ในการประมวลผล โดยใช้คลังข้อมูลทางภาษา (Corpus) เป็นฐานความรู้ในการตัดคำ เพื่อแก้ปัญหาของคำที่ไม่มีในพจนานุกรม เช่น ชื่อเฉพาะ คำที่ยืมมาจากภาษาต่างประเทศ เป็นต้น และความคลุมเครือในการแบ่งขอบเขตของคำได้อย่างมีประสิทธิภาพ แบ่งได้เป็น 2 แนวทาง คือ

- **การตัดคำโดยอาศัยความน่าจะเป็น (Probabilistic Word Segmentation)** วิธีการนี้จะนำเอาค่าทางสถิติการเกิดของคำและลำดับของหน้าที่ของคำ (Part–of-Speech) เข้ามาช่วยในการคำนวณหาความน่าจะเป็น เพื่อที่จะใช้เลือกแบบที่มีโอกาสเกิดมากที่สุด วิธีการนี้สามารถจะตัดคำได้ดีกว่า 2 แบบแรก แต่ข้อจำกัดของวิธีการนี้ คือ จะต้องมีฐานข้อมูลที่มีการตัดคำที่ถูกต้อง และกำหนดหน้าที่ของคำให้เพื่อที่จะได้นำไปใช้ในการสร้างสถิติ
- **การตัดคำโดยอาศัยคุณลักษณะของคำ (Feature Based Word Segmentation)** วิธีนี้การนี้จะพิจารณาจากบริบท (Context) และการเกิดร่วมกันของคำ หรือหน้าที่ของคำ (Collocation) เข้ามาช่วยในการตัด วิธีการตัดคำโดยใช้คลังข้อมูลนี้จำเป็นที่จะต้องมีฐานข้อมูลเป็นจำนวนมาก และจะต้องมีการเรียนรู้การสร้างคำในบริบท หรือการเกิดร่วมกันของคำแต่ละคำ เพื่อให้มีข้อมูลที่จะนำมาใช้ในการตัดคำ

4.2.5 การตัดคำที่ไม่จำเป็นในการค้นคืน (Removing Stop-Word)

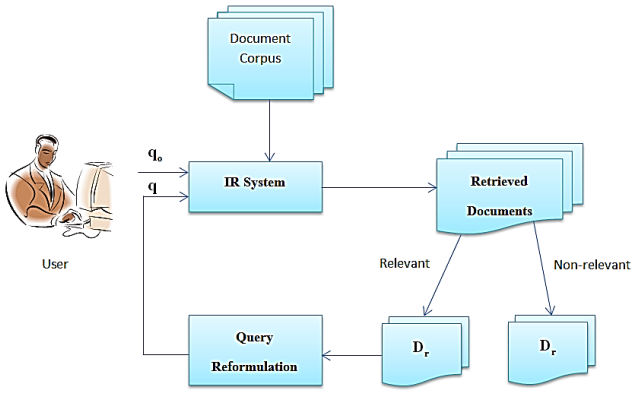
การตัดคำที่ไม่จำเป็นในการค้นคืน เป็นการนำคำที่ไม่มีนัยสำคัญออกโดยที่ไม่ทำให้ความหมายของเอกสารเปลี่ยนแปลงคำที่ไม่มีนัยสำคัญ ในที่นี้หมายถึงคำที่ใช้กันโดยทั่วไป ไม่มีความหมายสำคัญต่อเอกสาร เมื่อตัดออกจากเอกสารแล้วไม่ทำให้ใจความของเอกสารเปลี่ยนแปลง คำหยุดมักเป็นคำที่ปรากฏขึ้นบ่อยครั้งในเอกสารและปรากฏในเอกสารเกือบทุกฉบับ จึงถือได้ว่าคำหยุดเป็นคุณลักษณะที่ไม่เกี่ยวข้องหรือไม่มีประโยชน์ในการค้นคืนหรือการจำแนกหมวดหมู่ ดังนั้นการกำจัดคำหยุดจึงเป็นกระบวนการที่ควรทำก่อนการจัดทำดัชนี เพื่อกำจัดคุณลักษณะที่ไม่เป็นประโยชน์และลดขนาดของดัชนีลง ซึ่งจะช่วยประหยัดทั้งพื้นที่และเวลาในการประมวลผล



รูปที่ 4.12 ภาพแสดงการตัดคำที่ไม่จำเป็นในการค้นคืน

4.2.6 การป้อนความเกี่ยวพันธ์ย้อนกลับ (Relevance Feedback)

การป้อนความเกี่ยวพันธ์ย้อนกลับ เป็นกระบวนการที่ใช้เพื่อปรับปรุงสูตรของการคิวรี่โดยดัดแปลงจากคิวรี่เริ่มต้น โดยที่จะเพิ่มความสำคัญของพจน์ที่เกี่ยวข้องและลดความสำคัญของพจน์ที่ไม่เกี่ยวข้อง



รูปที่ 4.13 วงจรของ Relevance Feedback

ซึ่งระบบที่จะทำขึ้นนั้นเมื่อระบบมีการมีการประมวลผล และจัดทัศนคติเรียบร้อย ระบบสามารถให้ผู้ใช้สามารถปรับปรุงว่าแต่ละข้อคิดเห็นที่ปรากฏอยู่นั้น มีทัศนคติเป็นบวก เป็นลบ หรือเป็นกลาง เพื่อที่ระบบจะนำข้อมูลจากผู้ใช้ มาปรับปรุงฐานข้อมูลของคำที่ทำการจัดเก็บไว้

4.2.7 การวัดระยะห่างแบบยูคลิด (Euclidean distance)

คือระยะทางปกติระหว่างจุดสองจุดในแนวเส้นตรง ซึ่งอาจสามารถวัดได้ด้วยไม้บรรทัด มีที่มาจากทฤษฎีบทพีทาโกรัส ซึ่งสมการในการคำนวณหาระยะห่างแบบยูคลิด คือ

$$d=\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$$

4.3 อุปกรณ์และเครื่องมือที่ใช้

4.3.1 ฮาร์ดแวร์ (Hardware)

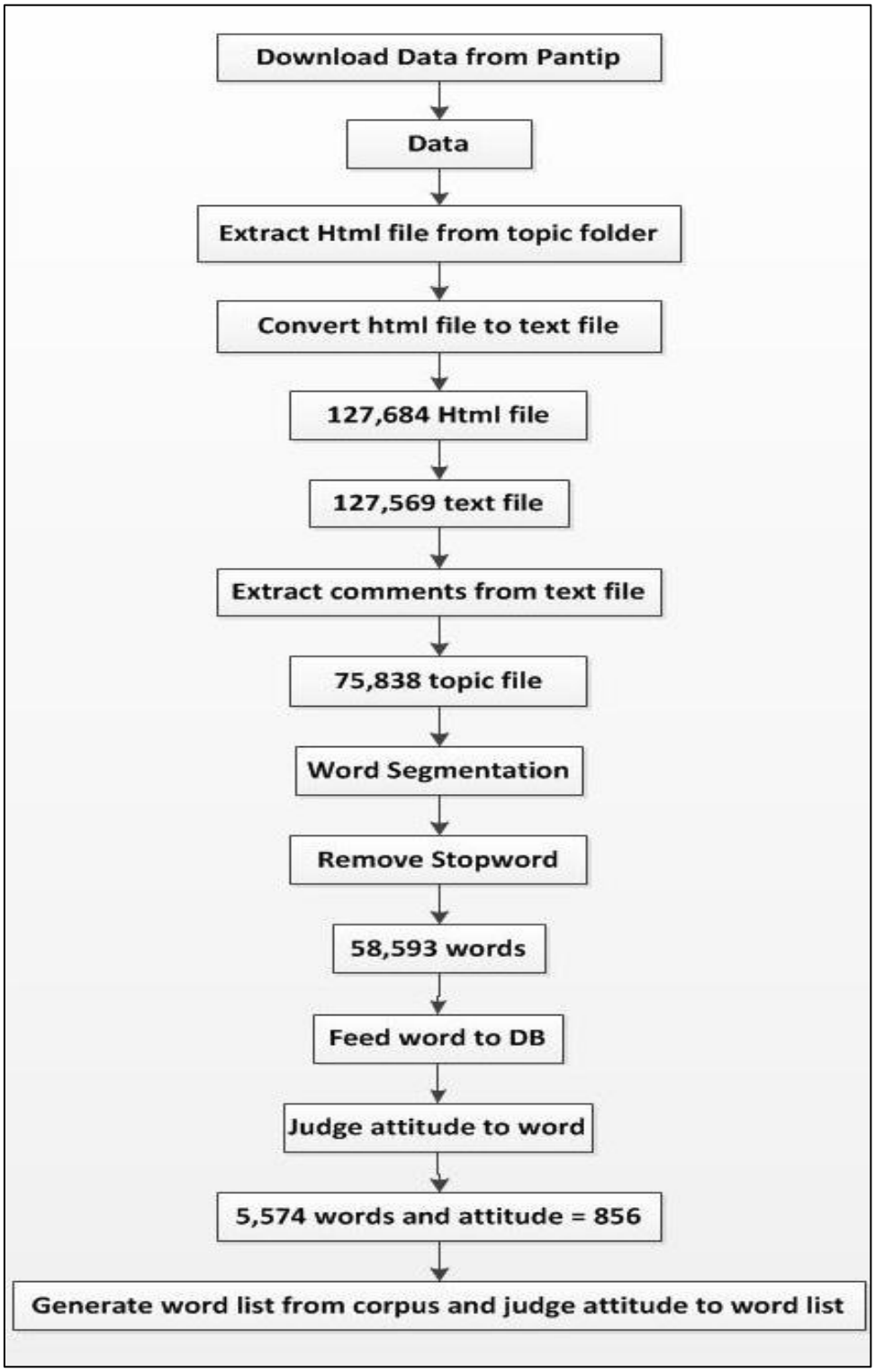
- หน่วยประมวลผลกลาง Intel(R) Core(TM)2 Duo P8400 ความเร็ว 2.26 GHz
- หน่วยความจำ (Memory) ขนาด 3.00 GB
- หน่วยแสดงผล ATI RADEON HD3470,ฮาร์ดดิสก์ (Hard Disk) ขนาด 320GB
- จอภาพ (Monitor),แป้นพิมพ์ (Keyboard), เมาส์ (Mouse)

4.3.2 ซอฟต์แวร์ (Software)

- ระบบปฏิบัติการ Windows 7 Enterprise
- ภาษาหลักที่ใช้ในการพัฒนาโปรแกรม คือ PHP 5.0 และ Python 2.5.1
- โปรแกรม Adobe Photoshop CS3 ใช้ในการตกแต่งภาพ
- Apache Version 2.5 ขึ้นไป ใช้จำลองเป็นเว็บเซิร์ฟเวอร์ (Web Server)
- ระบบจัดการฐานข้อมูล คือ MySQL
- โปรแกรมสำหรับการตัดคำภาษาไทย คือ BEST 2009 [Tlex]
- โปรแกรม WEKA Version 3.6.1 ในการเรียนรู้ด้วยเครื่อง
- โปรแกรม WinHTTrack ในการดาวส์โหลดเว็บไซต์

4.4 รายละเอียดโปรแกรมที่ได้รับการพัฒนาในเชิงเทคนิค

4.4.1 การทำงานในขั้นตอนการเตรียมข้อมูล (Data Preprocessing)



รูปที่ 4.14 แผนภาพแสดงการทำงานในขั้นตอนการเตรียมข้อมูล

4.4.2 ฐานข้อมูลของระบบ (Data Dictionary)

ตารางที่4.1 ฟิลด์ข้อมูลในตารางชื่อ word_table แสดงคำภาษาไทย

ชื่อฟิลด์	ชนิดข้อมูล	ขนาด(ไบต์)	คำอธิบาย
word	varchar		คำภาษาไทย
type	varchar		ประเภทของคำ
frequency	int		ความถี่ของคำ
attitude	int		ทัศนคติของคำ

ตารางที่4.2 ฟิลด์ข้อมูลในตารางชื่อ list_word แสดงคำในแต่ละข้อคิดเห็น

ชื่อฟิลด์	ชนิดข้อมูล	ขนาด(ไบต์)	คำอธิบาย
word_list	Text		คำที่แสดงทัศนคติในแต่ละข้อคิดเห็น
attitudes	int		ทัศนคติของข้อคิดเห็น
code_list	varchar		รหัสสายอักขระของแต่ละข้อคิดเห็น

4.5 ขอบเขตและข้อจำกัดของโครงงาน

4.5.1 ระบบจะนำข้อคิดเห็นในเว็บบอร์ดของเว็บไซต์พันทิปมาวิเคราะห์หาทัศนคติของคำ โดยใช้วิธีการเรียนรู้แบบไม่มีการสอน (Unupervised Learning) และเก็บทัศนคติของคำไว้ในฐานข้อมูล

4.5.2 ระบบจะรับที่อยู่ของเว็บไซต์ (Uniform Resource Locator: URL) ที่ผู้ใช้สนใจผ่านทางเว็บเบราว์เซอร์ (Web Browser) เพื่อหาทัศนคติจากข้อคิดเห็นที่เกี่ยวข้องกับหัวข้อบนเว็บไซต์ที่ผู้ใช้ป้อนเข้ามา ซึ่งระบบจะสามารถตัดสินทัศนคติได้จากเว็บบอร์ดของเว็บไซต์พันทิปเท่านั้น

4.5.3 ระบบจะนำข้อคิดเห็น และข้อมูลส่วนตัวของผู้แสดงความคิดเห็นมาเข้าสู่หลักการจัดเก็บสารสนเทศ (Information Storage) ซึ่งระบบจะเก็บเพียงข้อคิดเห็น และข้อมูลส่วนตัวของผู้แสดงความคิดเห็นจากเว็บไซต์ www.vanilla.in.th และ www.cosmenet.in เท่านั้น

4.5.4 ระบบจะตัดสินออกมาว่าหัวข้อจากเว็บบอร์ดในเว็บไซต์พันทิปนั้น มีทัศนคติจากข้อคิดเห็นภาษาไทยเป็นเชิงบวก เชิงลบ หรือเป็นกลาง ซึ่งระบบจะเก็บข้อคิดเห็นจากหน้าเว็บไซต์นั้นมาทำการตัดคำโดยใช้โปรแกรม BEST 2010 และหาคำที่มีการแสดงทัศนคติโดยการนำมาเปรียบเทียบจากฐานข้อมูล แล้วตัดสินทัศนคติของข้อคิดเห็นเหล่านั้นโดยใช้วิธีการวัดระยะห่างแบบยูคลิด (Euclidean distance)เปรียบเทียบจากทัศนคติของข้อคิดเห็นต่างๆที่เก็บไว้ในฐานข้อมูล

4.5.5 ระบบจะจัดกลุ่มของผู้แสดงความคิดเห็นบนเว็บไซต์ www.vanilla.in.th และ www.cosmenet.in ที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกัน โดยใช้เทคนิคการจำแนกกลุ่มออกเป็น k กลุ่มโดยการพิจารณาจากค่าเฉลี่ย (K-Mean Algorithm) และสรุปข้อมูลโดยรวมของผู้แสดงความคิดเห็นในแต่ละกลุ่ม

4.5.6 ผู้ใช้สามารถมีส่วนร่วมในการตรวจสอบความถูกต้องของทัศนคติในแต่ละข้อคิดเห็นที่ระบบยกตัวอย่างไว้ได้โดยการตัดสินทัศนคติตอบกลับมา (Relevance Feedback)

4.5.7 ระบบจะแสดงผลการตัดสินทัศนคติในรูปแบบของกราฟวงกลม

4.5.8 ระบบจะแสดงการจัดกลุ่มของผู้แสดงความคิดเห็นบนเว็บไซต์ www.vanilla.in.th และ www.cosmenet.in ที่มีทัศนคติในการแสดงความคิดเห็นคล้ายคลึงกันในรูปแบบกลุ่มที่แสดงรายชื่อของผู้แสดงความคิดเห็น และแสดงข้อมูลที่เกี่ยวข้องกันของผู้แสดงความคิดเห็นแต่ละกลุ่ม

5. กลุ่มผู้ใช้โปรแกรม

บุคคลที่มีความสนใจในการตัดสินทัศนคติ และเจ้าของธุรกิจที่มีความสนใจในการหาภาพรวมของสินค้าต่างๆ

6. ผลของการทดสอบโปรแกรม

ระบบสามารถที่จะตัดสินทัศนคติได้ จากการนับคำที่มีทัศนคติโดยเทียบจากฐานข้อมูลที่ทำการสร้างขึ้น โดยผลการทำงานนั้นยังไม่ถูกต้องมากนัก ทำให้ต้องมีการปรับปรุงระบบต่อไปอีก

ระบบสามารถจัดกลุ่มข้อคิดเห็นที่มีทัศนคติคล้ายคลึงกัน โดยเปรียบเทียบจากคำที่มีทัศคติจากฐานข้อมูล แต่ยังไม่สามารถนำข้อมูลส่วนตัวของผู้ที่มาแสดงความคิดเห็นมาทำการสรุปข้อมูลได้

7. ปัญหาที่พบ และแนวอุปสรรค

เนื่องจากในขั้นต้นจะทำการจัดเก็บข้อคิดเห็นทั้งหมด และทำการตัดคำด้วยโปรแกรม BEST และทำการเก็บรายการคำที่ปรากฏในแต่ละข้อคิดเห็นเข้าสู่ฐาข้อมูล จากนั้นจะนำไปข้อมูลเหล่านั้นไปผ่านกระบวนการเรียนรู้แบบไม่มีผู้สอน (Unsupervised Learning) เพื่อที่จะได้ทัศนคติของคำแต่ละคำออกมา แต่เนื่องจากข้อจำกัดของการตัดคำภาษาไทยที่เกิดขึ้น ผลลัพธ์ที่ได้ออกมานั้นมีความผิดพลาด และมีปริมาณคำที่เป็นคำแสลง และไม่มีอยู่ในพจนานุกรม ทำให้การที่จะเก็บทัศนคติขอแต่ละรายการคำที่ปรากฏในแต่ละข้อคิดเห็นเข้าสู่ฐาข้อมูลนั้น จะมีความผิดพลาดเกิดขึ้น

ดังนั้นจึงทำการจัดเก็บคำที่ปรากฏขึ้น และนับความถี่ของคำเพื่อที่จะนำคำที่มีความถี่สูงมาทำการตัดสินทัศนคติให้แต่ละคำ จากนั้นนำคำในแต่ละข้อคิดเห็นที่ทำการเก็บไว้มาเปรียบเทียบกับฐานข้อมูลที่ทำการตัดสินทัศนคติเรียบร้อยแล้ว เพื่อที่จะจัดเก็บรายการคำที่มีทัศนคติในแต่ละข้อคิดเห็น และจัดเก็บลงฐานข้อมูล เพื่อใช้ในงานส่วนถัดไป ซึ่งการแก้ไขดังกล่าวนั้น จะเป็นการทำงานแบบมีผู้สอน (Supervised Learning)

8. แนวทางในการพัฒนาและประยุกต์ใช้ร่วมกับงานอื่นๆ ในขั้นต่อไป

สามารถนำคลังข้อมูลที่มีไปทำงานในอนาคต และอาจปรับปรุงการตัดสินทัศนคติให้มีความถูกต้องแม่นยำ มากยิ่งขึ้น

9. ข้อสรุปและข้อเสนอแนะ

- ระบบยังมีความถูกต้องไม่มากนักเนื่องจากการตัดสินทัศนคติของข้อคิดเห็นนั้นยังไม่มีความสมบูรณ์มากพอ เพราะคำในภาษาไทยนั้นมีความซับซ้อนทางโครงสร้าง และการตัดสินทัศนคติของคำในฐานข้อมูลในขั้นตนนั้น ไม่ได้มาจากผู้เชี่ยวชาญทางด้านภาษา
- ควรมีการตรวจสอบทัศนคติของคำให้มีควาถูกต้องมากยิ่งขึ้นโดยผู้เชี่ยวชาญทางด้านภาษา
- ระบบงานควรรองรับการทำงานสำหรับทุกเว็บไซต์
- ระบบควรรองรับการทำงานกับข้อความที่เป็นภาษาอังกฤษด้วย

10. เอกสารอ้างอิง (Reference)

[1] Yin Luo,ongqi Lin,Yan Fu ,2009, **"Finer Granularity Clustering for Opinion Mining",** Computational Intelligence and Design, 2009. ISCID '09. Second International Symposium on 12-14 Dec. 2009

[2] Xiuguo Chen, Wensheng Yin, Pinghui Tu, Hengxi Zhang ,2009, **"Weighted *k*-Means Algorithm Based Text Clustering"** ,Information Engineering and Electronic Commerce, 2009. IEEC '09.International Symposium on 16-17 May 2009

[3] Alec Go,Richa Bhayani,and Lei Huang,2010 **"Twitter Sentiment"**,[ออนไลน์],เข้าถึงได้:http://twittersentiment.appspot.com/

[4] Joe Lokis,2010, **"twendz™"**,[ออนไลน์],เข้าถึงได้: http://twendz.waggeneredstrom.com/

[5] twitter,2010, **"Tweetfeel"**,[ออนไลน์],เข้าถึงได้: http://www.tweetfeel.com/

[6] Quantitative Archaeology Wiki,2010,[ออนไลน์],เข้าถึงได้: http://wiki.iosa.it/spatial_analysis:k-means

[7] Mário J. Silva,2010 "XLDB",[ออนไลน์],เข้าถึงได้: http://xldb.fc.ul.pt/wiki/Optimism

[8] หนังสือระบบการจัดเก็บและการสืบค้นสารสนเทศด้วยคอมพิวเตอร์ Information Storage and Retieval Systems

11. สถานที่ติดต่อผู้พัฒนา

หัวหน้าโครงการ

ชื่อ นางสาวณัฐวรรณ สุวรรณจิต

วันเกิด 27 ตุลาคม 2531 **ระดับการศึกษา** กำลังศึกษาปริญญาตรี ชั้นปีที่ 4

สถานศึกษา มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

ที่อยู่ตามทะเบียนบ้าน 35/83 หมู่ 3 ตำบลลำผักกูด อำเภอธัญบุรี จังหวัดปทุมธานี 12110

สถานที่ติดต่อ 61/33 ถนนทรงพล ต.พระปฐมเจดีย์ อ.เมืองฯ จ.นครปฐม 73000

โทรศัพท์ 089-6138520 **E-mail** kudo_kyoka@hotmail.com

ผู้ร่วมโครงการ

ชื่อ นางสาวพัชรินทร์ อุดมชัยเดช

วันเกิด 26 พฤษภาคม 2532 **ระดับการศึกษา** กำลังศึกษาปริญญาตรี ชั้นปีที่ 4

สถานศึกษา มหาวิทยาลัยศิลปากร วิทยาเขตพระราชวังสนามจันทร์

ที่อยู่ตามทะเบียนบ้าน 148 หมู่ 2 ต.พญาเย็น อ.ปากช่อง จ.นครราชสีมา 30320

สถานที่ติดต่อ 61/33 ถนนทรงพล ต.พระปฐมเจดีย์ อ.เมืองฯ จ.นครปฐม 73000

โทรศัพท์ 083-8505457 **E-mail** da-i_hotteens@hotmail.com

ภาคผนวก

12. ภาคผนวก

12.1 คู่มือการติดตั้ง

การติดตั้ง AppServ 2.5.10

โปรแกรม AppServ คือโปรแกรมที่รวบรวม Packages ต่างๆ ที่จำเป็นสำหรับการทำ Web Server ไว้ โดย Packages หลักๆ เหล่านั้น ได้แก่

Apache Web Server คือโปรแกรมที่ทำหน้าทีเป็น Web Server

MySQL Database คือโปรแกรมที่ทำหน้าเป็น Database Server

PHP Script Language คือภาษา PHP ที่เอาไว้เขียนโปรแกรมเกี่ยวกับเว็บ

⊙ phpMyAdmin คือตัวควบคุม MySQL Database ผ่านเว็บไซต์

ซึ่งโดยทั่วไปแล้วหากเราต้องการติดตั้ง Apache Web Server และเครื่องคอมพิวเตอร์เราสามารถใช้งาน PHP ได้ และต้องใช้ฐานข้อมูล MySQL ด้วย เมื่อลงโปรแกรมสมบูรณ์แล้วเครื่องคอมพิวเตอร์เราก็เปรียบเสมือน Web Server

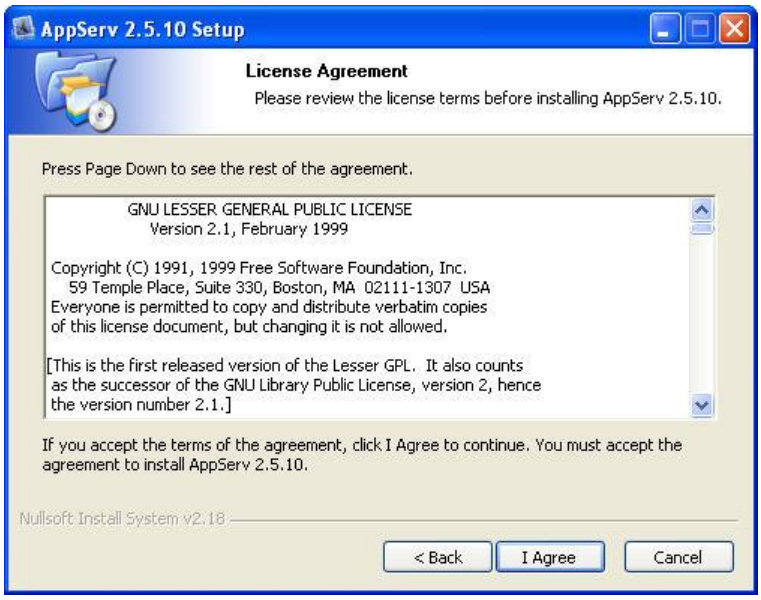
12.1.1 ดาวน์โหลดโปรแกรม AppServ

12.1.2 Double Click ไฟล์ appserv-win32-2.5.10

12.1.3 รอสักครู่จะปรากฏหน้าจอ Welcome ให้กดปุ่ม **Next**

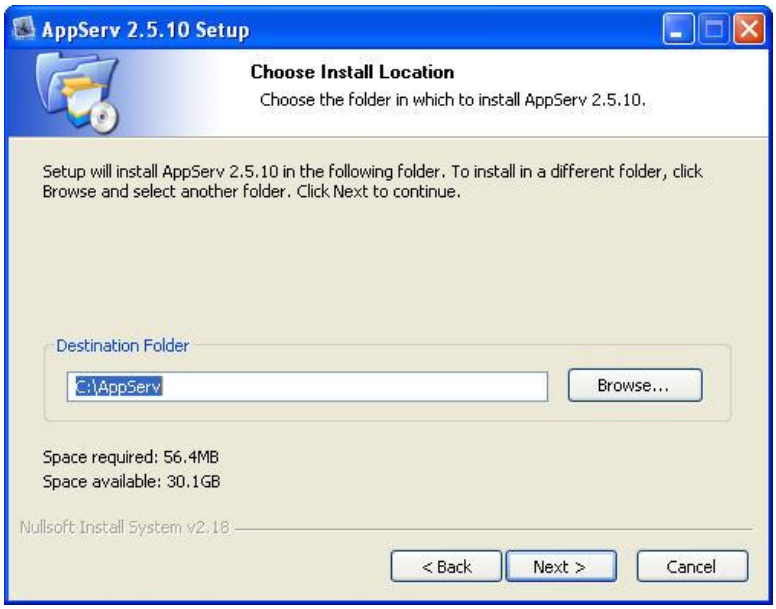
รูปที่ 4.15 หน้าจอ Welcome การติดตั้งโปรแกรมAppServe

12.1.4 กดปุ่ม **I Agree** เพื่อยอมรับข้อตกลงในการใช้ซอฟต์แวร์



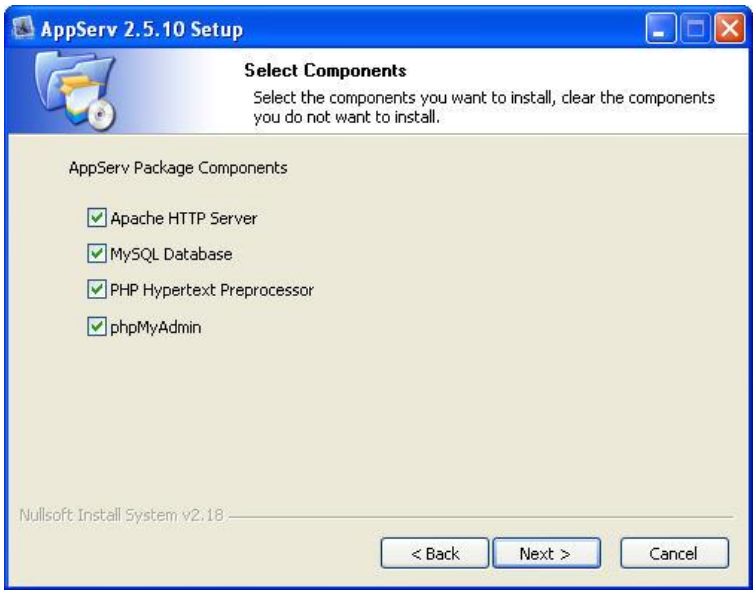
รูปที่ 4.16 หน้าจอตอบรับ ข้อตกลงในการใช้ซอฟต์แวร์

12.1.5 กำหนดโฟลเดอร์สำหรับติดตั้งโปรแกรม AppServ จากนั้นกดปุ่ม **Next**



รูปที่ 4.17 หน้าจอการเลือก Directory ในการติดตั้ง

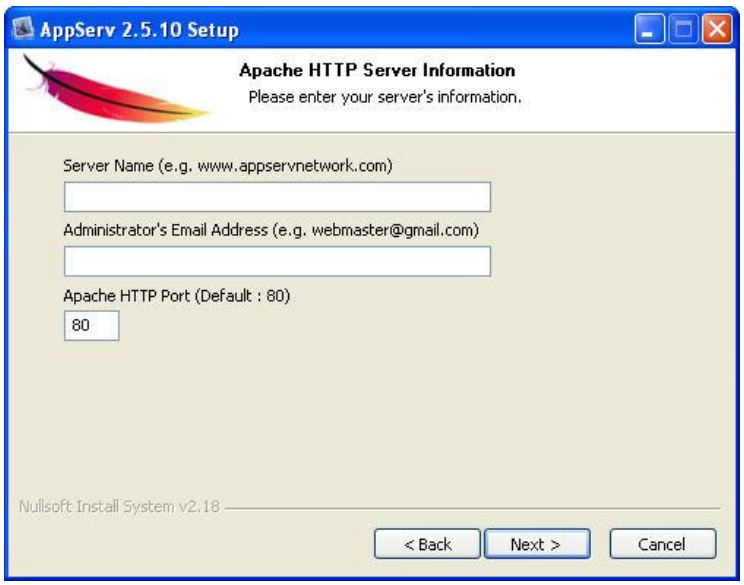
12.1.6 เลือกองค์ประกอบ (Components) สำหรับการติดตั้ง แล้วกดปุ่ม **Next**



รูปที่ 4.18 หน้าจอการเลือกองค์ประกอบต่างๆ ในการติดตั้ง

12.1.7 จะปรากฏหน้าจอสำหรับให้กรอกข้อมูลเซิฟเวอร์ของ (Server Information) ซึ่งประกอบด้วย

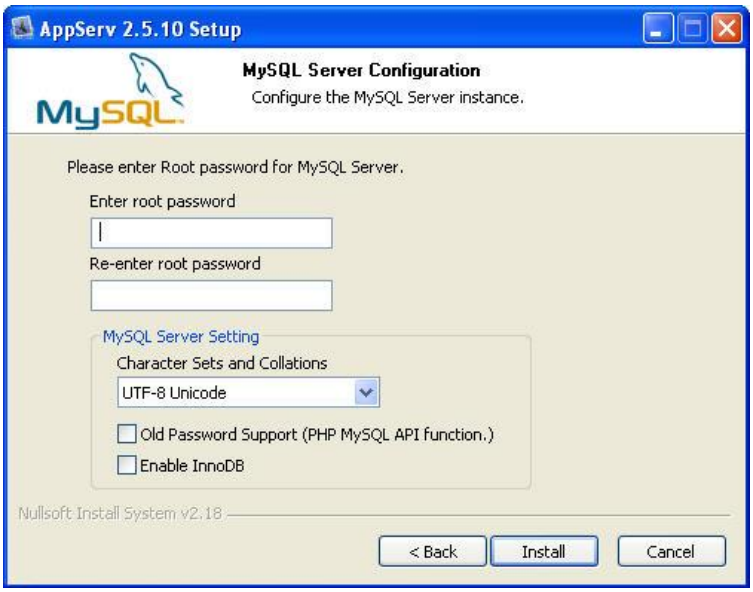
- ชื่อเซิฟเวอร์ หรือ ยูอาร์แอล (URL)
- อีเมล์ของผู้ดูแลเซิฟเวอร์
- พอร์ตสำหรับใช้งาน หรือติดต่อ



รูปที่ 4.18 หน้าจอสำหรับให้กรอกข้อมูลเซิฟเวอร์

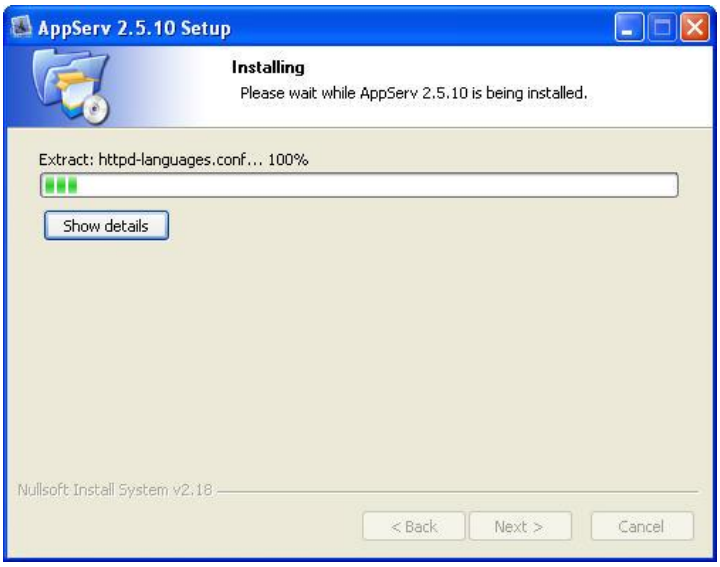
12.1.8 หลังกรอก Server Information แล้ว ขั้นตอนต่อไปคือการกำหนดค่าสำหรับ MySQL Server ซึ่งต้องระบุ

- รหัสผ่าน (Password) สำหรับ root ในที่นี้ใช้ 1234
- ชุดภาษา (Character Sets and Collations) ที่ใช้



รูปที่ 4.19 หน้าจอสำหรับให้กรอกรหัสผ่าน

12.1.10 หลังกำหนดค่าสำหรับ mySQL Server แล้ว ตัวติดตั้งจะดำเนินการติดตั้งองค์ประกอบต่างๆ ลงในระบบ



รูปที่ 4.20 หน้าจอดำเนินการติดตั้งองค์ประกอบต่างๆ ลงในระบบ

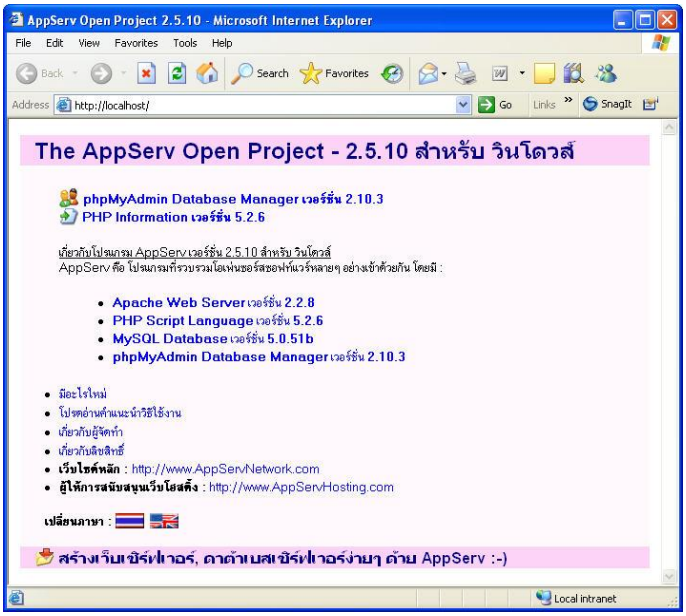
12.1.11 เมื่อการติดตั้งเสร็จสิ้น ให้กดปุ่ม **Finish**

รูปที่ 4.21 หน้าจอดำเนินการติดตั้งเสร็จสิ้น

12.1.12 ระบบจะทำการสตาร์ท (Start) โปรแกรม Apache

รูปที่ 4.22 หน้าจอการสตาร์ท (Start) โปรแกรม Apache

12.1.13 เมื่อสตาร์ท Apache Http Server เสร็จ ให้ทำการเปิดโปรแกรมเว็บบราวเซอร์ และพิมพ์คำว่า http://localhost ลงไปในช่องรับยูอาร์แอล (Address Bar) หากโปรแกรม Apache ทำงานได้เป็นปกติจะปรากฏข้อความในหน้าแรกตามภาพ



รูปที่ 4.23 หน้าจอโปรแกรม Apache ทำงานได้เป็นปกติ

12.1.14 ทำการ copy โฟล์เดอร์ word_opinion,test,php ไปวางที่ C:\AppServ\MySQL\data

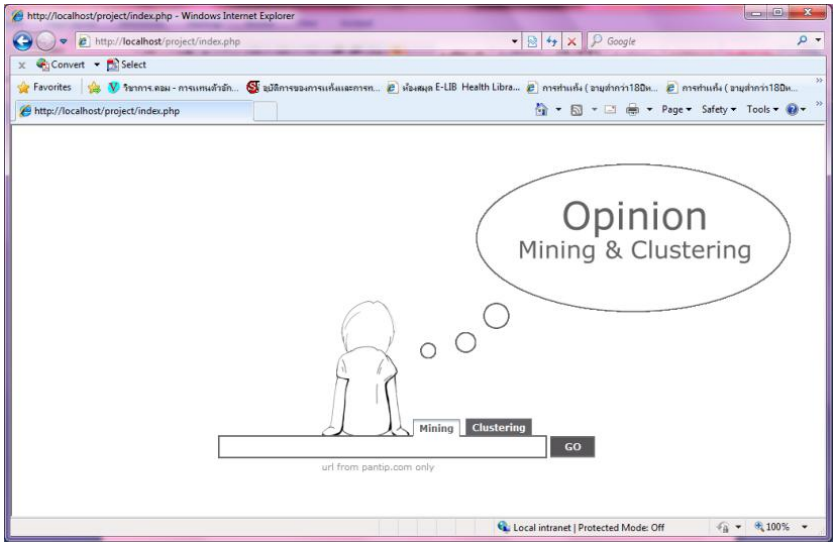
12.1.15 ทำการ copy โฟล์เดอร์ project ไปวางที่ C:\AppServ\www

12.1.16 Install โปรแกรม python 2.6.6

12.1.17 Install โปรแกรม weka 3.6 ที่ C://

12.2 คู่มือการใช้งาน

12.2.1 เปิดเว็บเบาว์เซอร์แล้วพิมพ์ localhost/project/index.php เพื่อที่จะเข้าหน้าหลักที่ต้องการทำงาน

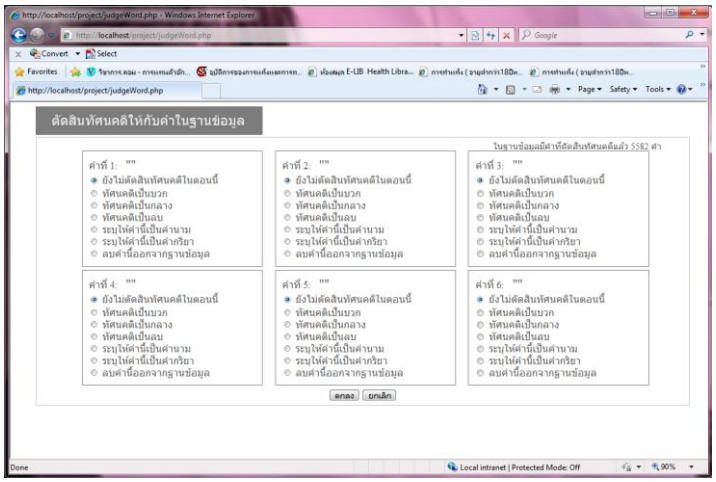


รูปที่ 4.24 ส่วนประสานงานกับผู้ใช้

- Mining สำหรับเลือกดูผลลัพธ์ในการตัดสินทัศนคติ
ระบบจะนำเว็บไซต์ที่ผู้ใช้ป้อนเข้ามาไปทำการวิเคราะห์และแสดงผลการตัดสินทัศนคติออกมา

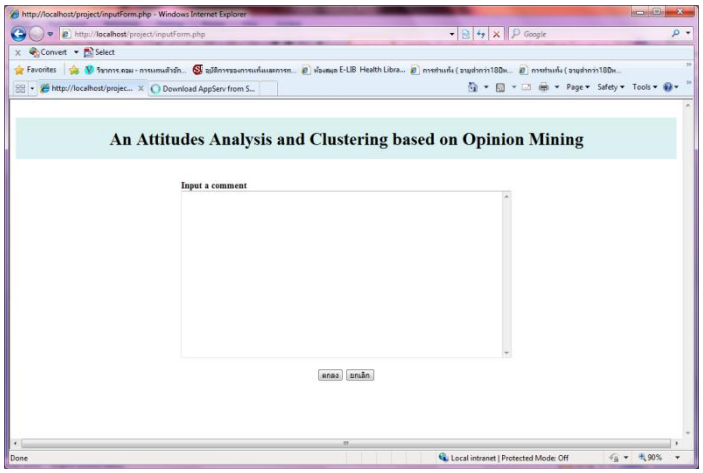
- Clustering สำหรับเลือกดูผลลัพธ์ในการจัดกลุ่มผู้แสดงความเห็น
ระบบจะนำเว็บไซต์ที่ผู้ใช้ป้อนเข้ามาไปทำการวิเคราะห์และแสดงผลการจัดกลุ่มผู้แสดงความเห็นออกมา

12.2.2 หากต้องการเข้าหน้าที่ใช้ในการตัดสินทัศนคติของคำ เปิดเว็บเบาว์เซอร์แล้วพิมพ์ localhost/project/judgeWord.php



รูปที่ 4.25 ภาพส่วนติดต่อผู้ใช้สำหรับการตัดสินทัศนคติในแต่ละคำ

12.2.3 หากต้องการเข้าหน้าที่ใช้ในการตัดสินทัศนคติของข้อคิดเห็นเพียงข้อคิกเห็นเดียว เปิดเว็บเบาว์เซอร์แล้วพิมพ์ localhost/project/inputForm.php



รูปที่ 4.26 ภาพส่วนติดต่อผู้ใช้สำหรับการตัดสินทัศนคติในแต่ละข้อคิดเห็น

12.2.4 code_list.php: เข้ารหัสลิสต์ของคำ เพื่อใช้ในการเปรียบเทียบหาทัศนคติ (เก็บไว้ในตาราง list_word ฟิลด์ code_list)

12.2.5 countWord.php: เพิ่มคำพร้องกับความถี่ของคำใส่เข้าไปในฐานข้อมูล

12.2.6 extractComment.php: ดึงข้อคิดเห็นจากแต่ละกระทู้มาเก็บไว้ใน text file

12.2.7 feedListword.php: เพิ่มลิสต์พร้อมทัศนคติของคำเข้าไปในฐานข้อมูล

12.2.8 wst2.py: ตัดคำภาษาไทย

12.2.9 miningResult.php:แสดงผลลัพธ์ของการตัดสินทัศนคติ

12.2.10 clusterResult.php:แสดงผลลัพธ์ของการจัดกลุ่มผู้แสดงความคิดเห็น