



Expertise Ranking for Online Social Network Group Members

การจัดลำดับความเชี่ยวชาญสำหรับสมาชิกกลุ่มเครือข่ายสังคมออนไลน์

25 เมษายน 2013

โชคชัย พุฒตาล

คณะวิศวกรรมคอมพิวเตอร์,
สาขาวิชา วิศวกรรมซอฟต์แวร์



ตัวอย่างของปัญหา

- กนกพร เป็นนักพัฒนาจาวาแอปพลิเคชัน
- เธอประสบปัญหาที่เกี่ยวข้องกับการเชื่อมต่อฐานข้อมูล
- เธอจึงค้นหาคำตอบจากเว็บไซต์ค้นหาเช่น Google และ Bing



ตัวอย่างของปัญหา (ต่อ)

- ผลการค้นหาที่ได้รับก็ไม่สามารถให้คำตอบที่เธอต้องการ
- เธอจึงตัดสินใจว่าจะไปตามในกลุ่มเครือข่ายสังคมออนไลน์
- เธอได้ถามคำถามกับกลุ่มของนักพัฒนาจากวาบบนเฟซบุ๊ก



ปัญหาที่พบ

- มีผู้ให้ความสนใจและตอบคำถามมากมาย
- คำตอบนั้นมีหลากหลายมากมายทั้งดีและไม่ดี
- กนกพรควรเชื่อใคร ?

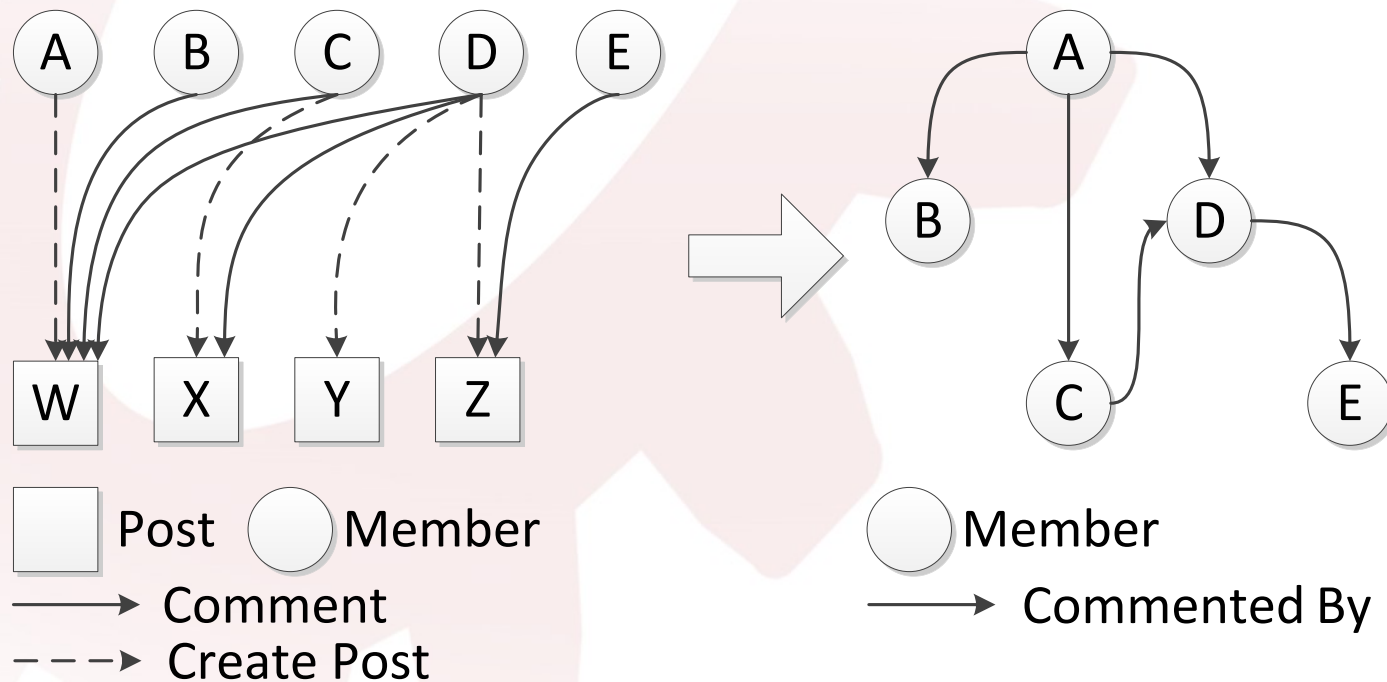


วิธีการแก้ปัญหา

- การระบุตัวผู้เชี่ยวชาญในกลุ่มเฟซบุ๊ก
- การจัดลำดับความเชี่ยวชาญของสมาชิกในกลุ่มเฟซบุ๊ก



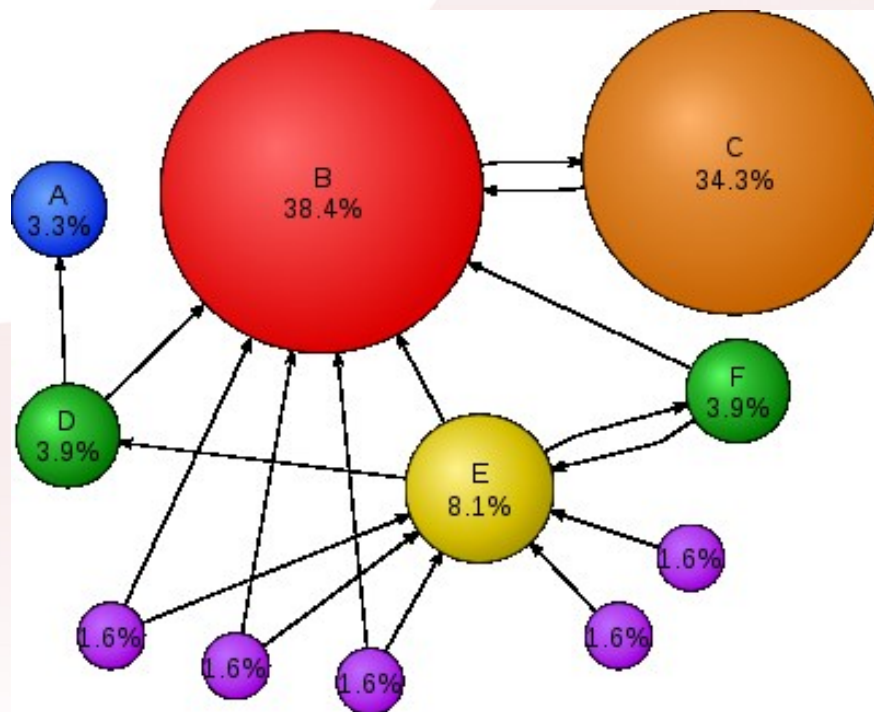
Community Expertise Network





PageRank

ขั้นตอนวิธีที่ใช้คำนวณหาค่าของความสำคัญของแต่ละหน้าเอกสาร





PageRank (ต่อ)

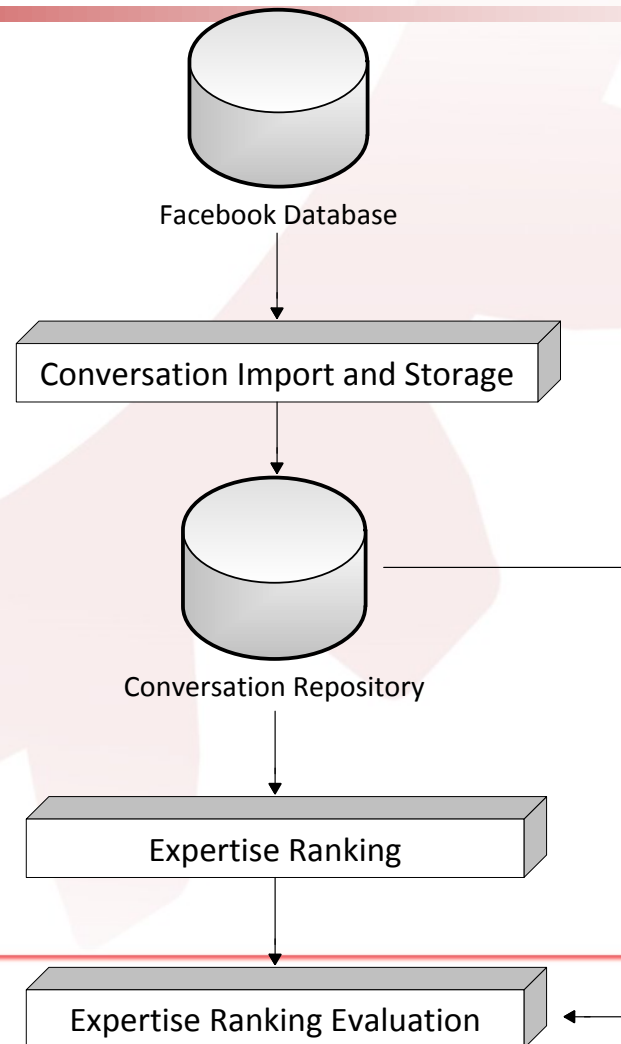
$$r(i) = (1 - d) + d * \sum_{j \in B(i)} \frac{r(j)}{N(j)}$$

โดยที่

$r(i)$	คือ	ค่าเพจเรงค์ของโหนด i
$r(j)$	คือ	ค่าเพจเรงค์ของโหนด j
$N(j)$	คือ	จำนวนของลิงค์ออกจากโหนด j
d	คือ	ค่าถ่วงน้ำหนัก (~0.85)
$B(i)$	คือ	โหนดทั้งหมดที่มีลิงค์ไปยังโหนด i



ภาพรวมของขั้นตอนการดำเนินงานวิจัย





ภาพรวมของขั้นตอนการดำเนินงานวิจัย (ต่อ)

- นำเข้าและจัดเก็บข้อมูลการสนทนา
(Conversation Import and Storage)
- นำข้อมูลที่ได้มาเก็บลงในฐานข้อมูลการสนทนา
(Conversation Repository)
- การจัดลำดับความเชี่ยวชาญ
(Expertise Ranking)
- การประเมินผลการจัดลำดับความเชี่ยวชาญ
(Expertise Ranking Evaluation)



การนำเข้าและจัดเก็บข้อมูลการสนทนา

Conversation Import and Storage

Conversation Import



Conversation Extraction

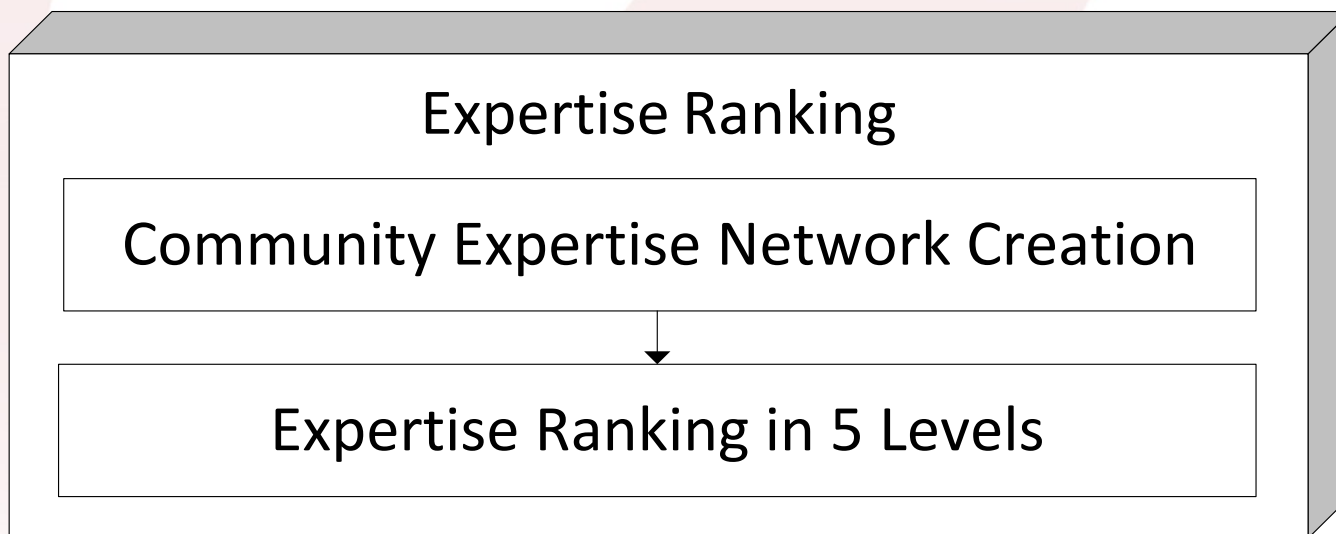


การนำเข้าและจัดเก็บข้อมูลการสนทนา (ต่อ)

- ในการนำเข้าข้อมูลของกลุ่มเฟสบุ๊ก (Conversation Import) จะใช้การนำเข้าข้อมูลผ่านทางเอพีไอเฟสบุ๊ก (Facebook API) โดยการนำข้อมูลออกมานั้นจะสามารถทำได้หลากหลายวิธี ข้อมูลที่ได้จะอยู่ในแบบของเจสัน (JSON)
- จากนั้นสกัดข้อมูลการสนทนา (Conversation Extraction) เพื่อนำไปจัดเก็บลงในฐานข้อมูลของการสนทนา (Conversation Repository)



การจัดลำดับความเชี่ยวชาญ





การจัดลำดับความเชี่ยวชาญ (ต่อ)

- นำเอาข้อมูลการสนทนามาแปลงให้อยู่ในรูปของเครือข่ายความเชี่ยวชาญในชุมชน (Community Expertise Network Creation) เพื่อนำมาใช้ในการคำนวณลำดับความเชี่ยวชาญของสมาชิก (Expertise Ranking in 5 Levels) กำหนดให้สมาชิกที่อยู่ในระดับ 4 และ 5 ถือเป็นผู้เชี่ยวชาญ
- โดยขั้นตอนวิธีการจัดลำดับความเชี่ยวชาญนั้น จะมี 2 รูปแบบ คือ ขั้นตอนวิธีที่ใช้คุณภาพของความเห็นร่วมกับขั้นตอนวิธีแบบเพจเรงก์ และขั้นตอนวิธีที่ใช้เพจเรงก์อย่างเดียว



คำอธิบายของแต่ละระดับความเชี่ยวชาญ

ระดับ	หมวดหมู่	รายละเอียด
5	ผู้เชี่ยวชาญมาก	มีความเชี่ยวชาญในหัวข้อของกลุ่มเป็นอย่างมาก สามารถตอบคำถามที่ยากได้
4	ผู้เชี่ยวชาญ	มีความเชี่ยวชาญในหัวข้อของกลุ่ม สามารถตอบคำถามในกลุ่มได้เกือบทั้งหมด
3	ผู้มีประสบการณ์	มีประสบการณ์เกี่ยวกับหัวข้อของกลุ่ม
2	ผู้กำลังศึกษา	กำลังศึกษาและมีประสบการณ์เล็กน้อยกับหัวข้อของกลุ่ม ไม่สามารถตอบคำถามที่มีความยากได้
1	ผู้เริ่มต้น	เพิ่งเริ่มต้นศึกษา



การประเมินผลการจัดลำดับความเชี่ยวชาญ

Expertise Ranking Evaluation

Expertise Ranking Evaluatuion

Expert Identification Evaluation



การประเมินผลการจัดลำดับความเชี่ยวชาญ (ต่อ)

- การประเมินผลจะทำการจัดลำดับความเชี่ยวชาญสองขั้นตอนวิธี
 1. ขั้นตอนวิธีที่นำเอาคุณภาพของความเห็นมาพิจารณาร่วมกับเพจแรงค์
 2. ขั้นตอนวิธีที่ใช้เพจแรงค์อย่างเดียว
- การประเมินผลแบ่งออกเป็นสองขั้นตอน
 1. การประเมินผลการจัดลำดับความเชี่ยวชาญ (Expertise Ranking Evaluation)
 2. การประเมินผลการระบุตัวผู้เชี่ยวชาญ (Expert Identification Evaluation)

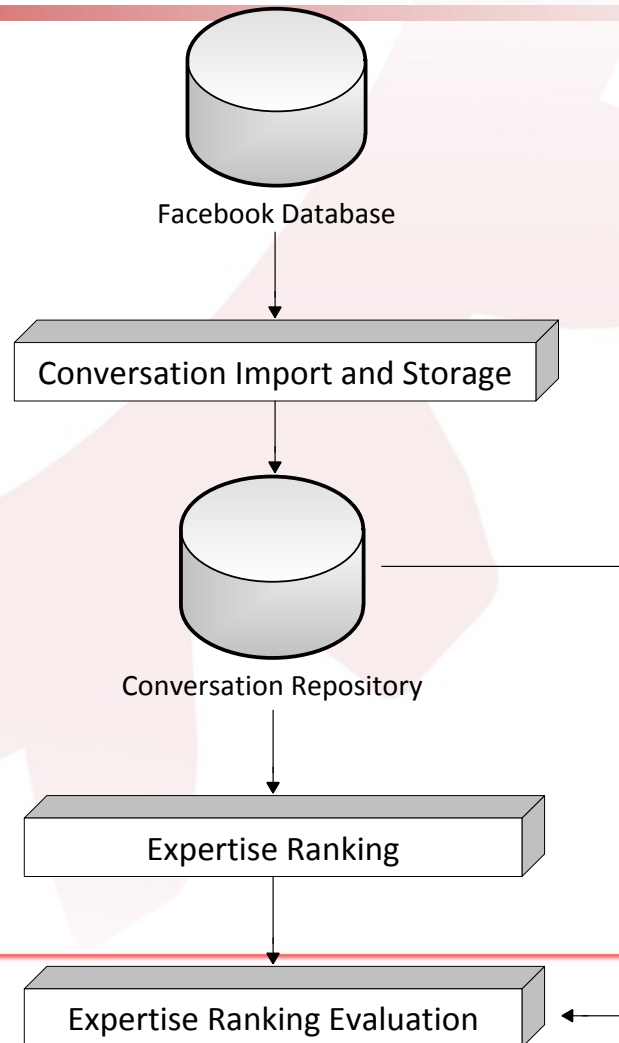


การประเมินผลการจัดลำดับความเชี่ยวชาญ (ต่อ)

- การประเมินจะทำโดยการเปรียบเทียบระหว่างผลที่ได้จากขั้นตอนวิธี และ ผลที่ได้จากการจัดลำดับความเชี่ยวชาญโดยผู้เชี่ยวชาญภาษาจาวา



สรุปภาพรวมของขั้นตอนการดำเนินงานวิจัย





ฟังก์ชันการทำงานของโปรแกรมสำหรับงานวิจัย

- สามารถนำเข้าข้อมูลการสนทนาของกลุ่มตัวอย่างผ่านเฟซบุ๊กเอพีไอ และนำมาจัดเก็บลงฐานข้อมูลในรูปแบบตารางความสัมพันธ์ได้
- สามารถนำข้อมูลบทสนทนามาคำนวณหาค่าความเชี่ยวชาญของสมาชิกแต่ละคนได้ โดยทั้งขั้นตอนวิธีที่ใช้เพจแรงค์ และขั้นตอนวิธีที่ใช้คุณภาพของความเห็นร่วมกับเพจแรงค์
- สามารถนำข้อมูลค่าความเชี่ยวชาญของสมาชิกแต่ละคนมาใช้ในการทำแบบทดสอบที่ใช้สำหรับวัดผลการทดลองได้
- สามารถวัดผลที่ได้จากการคำนวณทั้งสองขั้นตอนวิธี โดยเปรียบเทียบกับผลที่ได้จากผู้เชี่ยวชาญได้



เครื่องมือที่ใช้ในการพัฒนา

- ซับไลม์ 2 (Sublime 2)
- ชุดเครื่องมือพัฒนาโหนดเจเอส 10.1 (NodeJS 10.1)
- โหนดดีบี (Node-DB)
- ฐานข้อมูลมายเอสคิวแอล 5.5.24 (MySQL Database 5.5.24)



กลุ่มเฟซบุ๊กที่เลือกใช้ในการทดลอง

- กลุ่ม Java
- มีสมาชิกกลุ่ม 41,717 คน
- มีจำนวนบทสนทนา 12,269 บทสนทนา
- มีความคิดเห็น 56,013 ความคิดเห็น
- วันที่นำเข้าข้อมูลคือ 8 มกราคม 2556
- <https://facebook.com/groups/Javagroup123/>

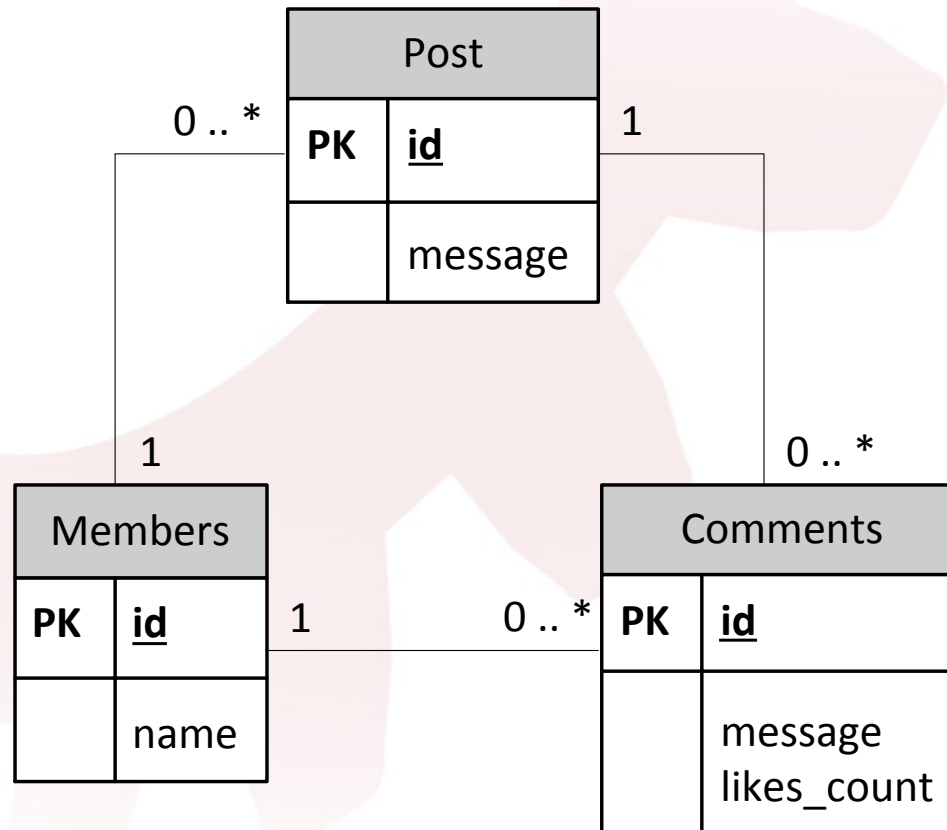


การนำเข้าและจัดเก็บข้อมูลการสนทนา

- ข้อมูลบทสนทนาทั้งหมดจะนำเข้าผ่านเฟซบุ๊กเอพีไอ โดยจะนำเข้าผ่านทางระบบฟีด (Feed) ซึ่งสามารถนำเข้าข้อมูลมาได้ครั้งละ 25 บทสนทนา งานวิจัยนี้ใช้การเขียนโปรแกรมที่พัฒนาจากโหนดเจเอส ในการนำเข้าข้อมูล ซึ่งข้อมูลดิบที่นำเข้าทั้งหมดจะถูกนำมาจัดเก็บลงใน
- ตารางนำเอาข้อมูลดิบที่อยู่ในรูปของเจสันที่ถูกจัดเก็บไว้ในฐานข้อมูลมาจำแนกและจัดเก็บให้อยู่ในรูปแบบตารางความสัมพันธ์ เพื่อความสะดวกในการนำไปใช้จัดลำดับความเชี่ยวชาญ



แผนภาพอีอาร์ (ER Diagram)





การจัดลำดับความเชี่ยวชาญ

- งานวิจัยนี้จะใช้การจัดลำดับความเชี่ยวชาญโดยการสร้างกราฟความสัมพันธ์ของสมาชิกจากบทสนทนา จากนั้นนำมาคำนวณเพื่อหาค่าความเชี่ยวชาญผ่านขั้นตอนวิธีเพจเรงก์เพื่อให้ได้ค่าความเชี่ยวชาญ
- แต่เนื่องจากคุณภาพของความเห็นอาจแสดงถึงความเชี่ยวชาญของผู้แสดงความเห็นได้ จึงนำมาใช้ในการหาค่าความเชี่ยวชาญเบื้องต้น (Preliminary Expertise Score) ก่อนที่จะนำไปใช้ร่วมกับการคำนวณด้วยขั้นตอนวิธีแบบเพจเรงก์ต่อไป



คุณภาพของความเห็น

1. ค่าถูกใจของความเห็น (Likes count)
2. ค่าความยาวของความเห็น (Words count)
3. ค่าความซับซ้อนของความเห็น (Entropy)
4. ค่าความมีสาระประโยชน์ของความเห็น (TF-IDF)



ขั้นตอนวิธีในการคำนวณค่าความเชี่ยวชาญเบื้องต้น

1. ค่าถูกใจของความเห็นของสมาชิกแต่ละคนในหนึ่งบทสนทนา

$$l_{ij} = \frac{(\sum_{j \in i} lc_{ij}) + 1}{pl_i + m_i}$$

โดยที่

lc_{ij} คือ ค่าถูกใจของแต่ละความเห็นของสมาชิก j ในบทสนทนา i

pl_i คือ ค่าถูกใจรวมจากทุกความเห็นในบทสนทนา i

m_i คือ จำนวนสมาชิกที่ร่วมตอบคำถามในบทสนทนา i



ตัวอย่างการคำนวณค่าถูกใจ

คำถาม: Can someone explain the difference between...

สมาชิก A: (likes_count=2) Abstract Class can contain...

สมาชิก B: (likes_count=1) take a look at this...

$$l_{iA} = \frac{(\sum_{A \in i} lc_{iA}) + 1}{pl_i + m_i} = \frac{2 + 1}{(2 + 1) + 2} = 0.60$$



ขั้นตอนวิธีในการคำนวณค่าความเชื่อชาติเบื้องต้น

2. ความยาวของความเห็นของสมาชิกแต่ละคนในหนึ่งบทสนทนา

$$ln_{ij} = \sum_{j \in i} cl_j$$

โดยที่

cl_j คือ ค่าความยาวของแต่ละความเห็นของสมาชิก j ในบทสนทนา i



ขั้นตอนวิธีในการคำนวณค่าความเชื่อวชาญเบื้องต้น

3. ความซับซ้อนของความเห็นของสมาชิกแต่ละคนในหนึ่งบทสนทนา

$$e_{ij} = \sum_{j \in i} \text{entropy}(c_j)$$

โดยที่

$\text{entropy}(c_j)$ คือ ค่าความซับซ้อนของแต่ละความเห็นของสมาชิก j ในบทสนทนา i ซึ่งคำนวณได้จากสมการ



Entropy

$$entropy(c_j) = \frac{1}{\lambda} \sum_{i=1}^n p_i [\log_{10}(\lambda) - \log_{10}(p_i)]$$

โดยที่

c_j คือ ความเห็นที่ j

λ คือ จำนวนของคำทั้งหมดในความเห็นที่ j

p_i คือ ความถี่ของแต่ละคำในความเห็นที่ j

n คือ จำนวนของคำทั้งหมดที่ไม่ซ้ำกันที่อยู่ในความเห็นที่ j



ขั้นตอนวิธีในการคำนวณค่าความเชื่อวชาญเบื้องต้น

4. ความมีเนื้อหาสาระของความเห็นของสมาชิกแต่ละคนในหนึ่งบทสนทนา

$$t_{ij} = \sum_{j \in i} \text{inform}(c_j)$$

โดยที่

$\text{inform}(c_j)$ คือ ค่าความมีเนื้อหาสาระของแต่ละความเห็นของสมาชิก j ในบทสนทนา i ซึ่งคำนวณได้จากสมการ



Informativeness (TF-IDF)

$$\text{inform}(c_j) = \sum_{t_i \in c_j} tf_{ij} \times idf_i$$

$$tf_{ij} = \frac{n_{ij}}{\sum_k n_{kj}}$$

$$idf_i = \log \frac{|C|}{|c:t_i \in c| + 1}$$



TF-IDF (ต่อ)

โดยที่

c_j	คือ	ความเห็นที่ j
tf_{ij}	คือ	Term Frequency ของ Term i ในความเห็นที่ j
n_{ij}	คือ	ความถี่ของ Term i ในความเห็นที่ j
$ C $	คือ	จำนวนของความเห็นทั้งหมดในบทสนทนา
$ c: t_i \in c $	คือ	จำนวนของความเห็นที่มี Term i เป็นส่วนประกอบ



ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ในบทสนทนา i

$$pesc_{ij} = \frac{sl_{ij} + sln_{ij} + se_{ij} + st_{ij}}{4}$$

sl_{ij} คือ ค่าถูกใจของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์

sln_{ij} คือ ค่าความยาวของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์

se_{ij} คือ ค่าความซับซ้อนของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์

st_{ij} คือ ค่าความมีสาระประโยชน์ของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์



ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ในบทสนทนา i

$$sl_{ij} = \frac{l_{ij}}{L_i}$$

โดยที่

sl_{ij}	คือ	ค่าถูกใจของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์
l_{ij}	คือ	ค่าถูกใจของสมาชิก j ในบทสนทนา i ที่ได้จากสมการ
L_i	คือ	ค่าถูกใจที่มีค่าสูงที่สุดในบทสนทนา i



ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ในบทสนทนา i

$$sln_{ij} = \frac{ln_{ij}}{LN_i}$$

โดยที่

sln_{ij} คือ ค่าความยาวของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่าน
การนอร์มัลไลซ์

ln_{ij} คือ ค่าความยาวของความเห็นของสมาชิก j ในบทสนทนา i ที่ได้จาก
สมการ

LN_i คือ ค่าความยาวของความเห็นซึ่งมีค่าสูงที่สุดในบทสนทนา i



ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ในบทสนทนา i

$$se_{ij} = \frac{e_{ij}}{E_i}$$

โดยที่

se_{ij} คือ ค่าความซับซ้อนของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์

e_{ij} คือ ค่าความซับซ้อนของความเห็นของสมาชิก j ในบทสนทนา i ที่ได้จากการ

E_i คือ ค่าความซับซ้อนของความเห็นซึ่งมีค่าสูงที่สุดในบทสนทนา i



ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ในบทสนทนา i

$$st_{ij} = \frac{t_{ij}}{T_i}$$

โดยที่

st_{ij} คือ ค่าความมีสาระประโยชน์ของความเห็นของสมาชิก j ในบทสนทนา i ที่ผ่านการนอร์มัลไลซ์

t_{ij} คือ ค่าความมีสาระประโยชน์ของความเห็นของสมาชิก j ในบทสนทนา i ที่ได้จากการ

T_i คือ ค่าความมีสาระประโยชน์ของความเห็นซึ่งมีค่าสูงที่สุดในบทสนทนา i



ค่าความเชี่ยวชาญเบื้องต้น pes_j ของสมาชิก j

$$pes_j = \sum_{i \in P_j} pesc_{ij}$$

โดยที่

P_j คือ บทสนทนาทั้งหมดที่สมาชิก j ร่วมแสดงความเห็น



ขั้นตอนวิธีในการคำนวณค่าความเชื่อวชาญด้วยเพจแรงค์

- การนำเอาบทสนทนาที่มีความเห็นของสมาชิกมาคำนวณด้วยขั้นตอนวิธีแบบเพจแรงค์เพื่อหาค่าความเชื่อวชาญนั้น จำเป็นต้องแปลงบทสนทนาทั้งหมดให้อยู่ในรูปของกราฟ โดยการนำเอาความสัมพันธ์ระหว่างสมาชิกมาใช้ในการสร้างกราฟเครือข่ายความเชื่อวชาญในชุมชน
- ขั้นตอนต่อมาคือการคำนวณหาค่าความเชื่อวชาญ $ER(j)$ ของสมาชิก j แต่ละคนด้วยขั้นตอนวิธีแบบเพจแรงค์



คำนวณหาความเชี่ยวชาญ ของสมาชิก j แต่
ละคบด้วยปัจจัยบทความวิธีแบบแฟลจแรงค์

$$ER(j) = (1 - d) + d \left(\frac{ER(U_1)}{C(U_1)} + \dots + \frac{ER(U_n)}{C(U_n)} \right)$$

โดยที่

$ER(U_i)$ คือ ค่าความเชี่ยวชาญของสมาชิก U_i ที่สมาชิก j

แสดงความเห็น

$C(U_i)$ คือ จำนวนสมาชิกที่แสดงความเห็นในบทความที่
สมาชิก U_i เป็นผู้เริ่มต้น

d คือ ค่าถ่วงน้ำหนักมีค่า 0.85



คำนวณค่าความเชี่ยวชาญ ของสมาชิก แต่ละคนด้วย ขั้นตอนวิธีแบบเพจแรงค์

การคำนวณค่าความเชี่ยวชาญของสมาชิกจะทำการคำนวณซ้ำหลายรอบ เพื่อให้ค่าที่ได้เข้าใกล้ค่าที่ต้องการ โดยในแต่ละรอบจะนำค่าที่ได้ในรอบก่อนหน้ามาใช้เป็นค่าเริ่มต้น สำหรับรอบแรกนั้นจะกำหนดค่าความเชี่ยวชาญของสมาชิกแต่ละคนดังสมการ

$$ER(U_i)_{initial} = \frac{1}{N}$$

โดยที่

N คือ จำนวนสมาชิกทั้งหมดในเครือข่ายความเชี่ยวชาญในชุมชน



การกำหนดค่าเริ่มต้นพิเศษให้กับเพจแรงค์

- เนื่องจากงานวิจัยต้องการทดลองว่าคุณภาพของความเห็นนั้นส่งผลต่อการจัดลำดับความเชี่ยวชาญเพียงใด จึงกำหนดค่าความเชี่ยวชาญเบื้องต้นของสมาชิกแต่ละคน
- อย่างไรก็ตาม การกำหนดค่าเริ่มต้นพิเศษให้กับขั้นตอนวิธีเพจแรงค์นั้น จะไม่ส่งผลกระทบต่อผลลัพธ์สุดท้าย
- ดังนั้นจึงจำเป็นต้องปรับปรุงสมการเพจแรงค์ใหม่ โดยเปลี่ยนแปลงที่ส่วน $1 - d$ สมการเพจแรงค์ที่ปรับปรุงแล้วเป็นดังนี้



คำนวณหาค่าความเชี่ยวชาญ ของสมาชิกแต่ละคนด้วย
ขั้นตอนวิธีแบบเพจเรงก์ และคุณภาพของความเห็น

$$ER(j) = S(j)(1 - d) + d \left(\frac{ER(U_1)}{C(U_1)} + \dots + \frac{ER(U_n)}{C(U_n)} \right)$$

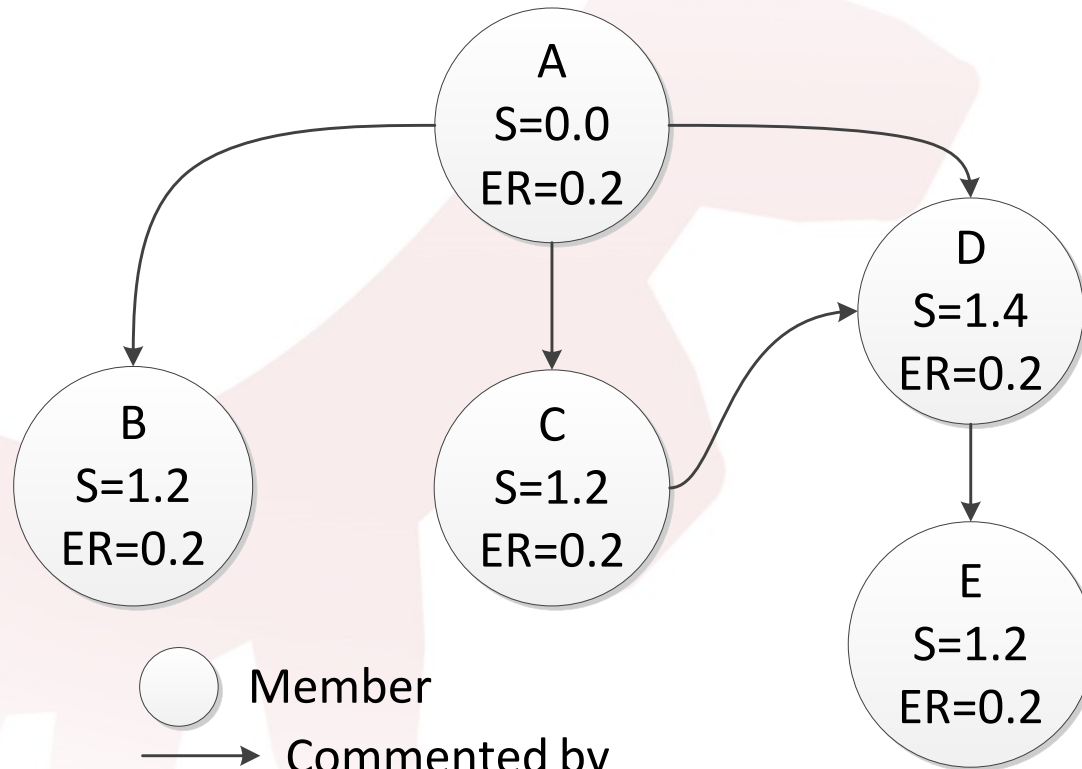
โดยที่

$$S(j) = N \frac{pes_j}{\sum_{m=1}^N pes_m}$$

$S(j)$ คือ ค่าความเชี่ยวชาญเบื้องต้นของสมาชิก j ที่ถูกนอร์มัลไลซ์
และนำไปถ่วงน้ำหนักด้วยจำนวนสมาชิกในเครือข่ายความเชี่ยวชาญในชุมชน



ตัวอย่าง: คำนวณหาค่าความเชี่ยวชาญของสมาชิก
ขั้นตอนวิธีแบบเพจแรงค์ และคุณภาพของความเห็น



○ Member
→ Commented by
ER = Expertise score



ตัวอย่าง: คำนวณหาค่าความเชี่ยวชาญของสมาชิก
ขั้นตอนวิธีแบบเพจแรงค์ และคุณภาพของความเห็น

- โหนดของสมาชิก D แสดงความเห็นในบทสนทนาของสมาชิก 2 คนคือ A และ C มีค่า S คือ 1.4
- โหนดของสมาชิก A มีสมาชิกมาแสดงความเห็นในบทสนทนาที่ตนตั้งทั้งหมด 3 คน มีค่า ER เริ่มต้นคือ 0.2
- โหนดของสมาชิก C มีสมาชิกมาแสดงความเห็นในบทสนทนาที่ตนตั้งทั้งหมด 1 คน มีค่า ER เริ่มต้นคือ 0.2

$$\begin{aligned} ER(D) &= S(D)(1 - 0.85) + 0.85 \left(\frac{ER(U_A)}{C(U_A)} + \frac{ER(U_C)}{C(U_C)} \right) \\ &= 1.4(1 - 0.85) + 0.85 \left(\frac{0.2}{3} + \frac{0.2}{1} \right) \cong 0.44 \end{aligned}$$



การจัดลำดับความเชี่ยวชาญ

- หลังจากที่ยกย่องค่าความเชี่ยวชาญของสมาชิกเสร็จสิ้น ทำการกรองสมาชิกที่ไม่เคยแสดงความเห็นในบทสนทนาใดเลยออก
- จากนั้นจึงจัดลำดับความเชี่ยวชาญของสมาชิกที่เหลือ จำนวน 2,564 คน โดยเรียงลำดับจากมากไปหาน้อย แล้วจึงแบ่งออกเป็น 5 ระดับตามลำดับคะแนนที่ได้ ระดับแสดงถึงความเชี่ยวชาญจากมากไปหาน้อย
- โดยกำหนดให้สมาชิกที่อยู่ในระดับ 5 และ 4 คือผู้เชี่ยวชาญ และสมาชิกในระดับต่ำลงมาก็คือผู้มีส่วนร่วมในกลุ่ม



แนวทางการประเมินและการวัดผล

- งานวิจัยนี้เลือกใช้การเปรียบเทียบระหว่าง ผลลัพธ์จากการทดลองกับ ผลลัพธ์การจัดลำดับความเชี่ยวชาญจากแบบสอบถาม
- ผู้เชี่ยวชาญภาษาจาวา 10 คน ผู้เชี่ยวชาญแต่ละคนจะทำแบบสอบถาม จำนวน 20 ข้อ แต่ละข้อจะประกอบด้วยหัวข้อบทสนทนา และความเห็น ของสมาชิกจำนวน 2-5 ความเห็น โดยให้จัดลำดับความเชี่ยวชาญของ สมาชิกที่แสดงความเห็นแต่ละคน และระบุตัวสมาชิกที่เป็นผู้เชี่ยวชาญ



ตัวอย่างคำตอบแบบสอบถามของผู้เชี่ยวชาญ

คำถาม #	เรียงลำดับผู้เชี่ยวชาญ	ผู้ที่เป็นผู้เชี่ยวชาญ
1	CEBDA	BCDE
2	BCA	BC
3	CBDA	ABCD
4	AB	AB
5	DCBA	ABCD
6	CABD	ABCD
7	BCA	B
8	EBCDA	ABCDE
9	CBEAD	BCDE
10	CAB	ABC

คำถาม #	เรียงลำดับผู้เชี่ยวชาญ	ผู้ที่เป็นผู้เชี่ยวชาญ
11	ABC	AB
12	ACB	AC
13	ADBC	ABCD
14	ABCD	ABCD
15	BCDA	ABCD
16	DCAB	ACD
17	ACB	AC
18	DCAB	ABCD
19	BAC	ABC
20	DACB	ABCD



การวัดผล

1. ความแม่นยำในการจัดลำดับความเชี่ยวชาญ
2. ความแม่นยำในการระบุตัวผู้เชี่ยวชาญ



ความแม่นยำในการจัดลำดับความเชี่ยวชาญ

1. “สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน”

เป็นค่าดัชนีการวัดความสัมพันธ์ระหว่างการจัดลำดับสองชุดบนข้อมูลชุดเดียวกัน โดยจะวัดครั้งละคู่ลำดับ

2. “ระยะห่างลำดับที่ของเคนดัลล์”

เป็นค่าดัชนีการวัดความขัดแย้งระหว่างการจัดลำดับสองชุดบนข้อมูลชุดเดียวกัน โดยจะวัดครั้งละ



สัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน

$$\rho_i = 1 - \frac{6 \sum d_j^2}{n(n^2 - 1)}$$

โดยที่

d_j คือ ค่าผลต่างของลำดับที่ของความเชี่ยวชาญของสมาชิก j

n คือ จำนวนของความเห็นในบทสนทนา (จำนวนของกลุ่มลำดับ)

****** ค่าจะอยู่ในช่วง -1 ถึง 1 โดย -1 หมายถึงการจัดลำดับสองชุดมีลำดับที่ตรงข้ามกัน
และ 1 หมายถึงการจัดลำดับสองชุดมีลำดับที่ตรงกัน



ระยะห่างลำดับที่ของเคนคัลล์

$$k_i = \frac{D}{n(n-1)/2}$$

โดยที่

D คือ จำนวนคู่ของความเห็นที่จัดลำดับแตกต่างกัน

n คือ จำนวนของความเห็นในบทสนทนา (จำนวนของคู่ลำดับ)

****** ผลลัพธ์จะอยู่ในช่วง 0 ถึง 1 โดย 0 หมายถึงการจัดลำดับสองชุดมีลำดับที่ตรงกัน
และ 1 หมายถึงการจัดลำดับสองชุดมีลำดับที่ตรงข้ามกัน



ความแม่นยำในการระบุตัวผู้เชี่ยวชาญ

ค่าความแม่นยำในการระบุตัวผู้ผู้เชี่ยวชาญ คือ สัดส่วนความถูกต้องในการระบุตัวผู้ผู้เชี่ยวชาญในแต่ละบทสนทนา i ต่อการระบุตัวผู้ผู้เชี่ยวชาญทั้งหมด โดยผู้ผู้เชี่ยวชาญนั้นหมายถึงสมาชิกที่มีความเชี่ยวชาญระดับ 4 และ 5

$$EIC_i = \frac{ic}{ic+iic}$$

โดยที่

ic คือ จำนวนการระบุตัวผู้ผู้เชี่ยวชาญที่ถูกต้อง

iic คือ จำนวนการระบุตัวผู้ผู้เชี่ยวชาญที่ไม่ถูกต้อง



ผลการทดลองที่ได้จากการจัดลำดับความเชี่ยวชาญโดยใช้ คุณภาพของความเห็นร่วมกับขั้นตอนวิธีแบบเพจแรงค์

ผู้เชี่ยวชาญ #	ค่าเฉลี่ยของ p	ค่าเฉลี่ยของ k	ค่าเฉลี่ยของ EIC
1	0.26	0.40	0.81
2	0.14	0.44	0.67
3	0.20	0.41	0.66
4	0.02	0.50	0.73
5	0.31	0.38	0.76
6	0.15	0.43	0.85
7	0.19	0.41	0.80
8	0.04	0.51	0.72
9	0.33	0.36	0.80
10	0.07	0.47	0.76
ค่าเฉลี่ย	0.17	0.43	0.76



ผลการทดลองที่ได้จากการจัดลำดับความเชี่ยวชาญ โดยขั้นตอนวิธีแบบเพจแรงค์

ผู้เชี่ยวชาญ #	ค่าเฉลี่ยของ p	ค่าเฉลี่ยของ k	ค่าเฉลี่ยของ EIC
1	0.18	0.43	0.81
2	0.07	0.47	0.67
3	0.12	0.45	0.67
4	0.00	0.50	0.74
5	0.28	0.38	0.78
6	0.08	0.46	0.86
7	0.14	0.45	0.80
8	0.01	0.51	0.72
9	0.31	0.36	0.81
10	-0.03	0.51	0.76
ค่าเฉลี่ย	0.12	0.45	0.76



อธิบายผลการทดลอง

- ค่าความแม่นยำในการระบุตัวผู้เชี่ยวชาญ *EIC* มีค่าเท่ากันในทั้งสองขั้นตอนวิธี
- ค่าสัมประสิทธิ์สหสัมพันธ์ลำดับที่ของสเปียร์แมน ρ ของทั้งสองขั้นตอนวิธีมีค่าเป็นบวก คือ 0.17 และ 0.12 ซึ่งแสดงว่าการจัดลำดับด้วยทั้งสองขั้นตอนวิธีมีความสอดคล้องกับการจัดลำดับด้วยผู้เชี่ยวชาญ แต่เนื่องจากค่าบวกนั้นมีค่าน้อย ทำให้บ่งบอกถึงความสอดคล้องของการจัดลำดับได้ไม่มากนัก



อธิบายผลการทดลอง (ต่อ)

- ค่าระยะห่างลำดับที่ของเคนคัลล์ k ของทั้งสองขั้นตอนวิธีคือ 0.43 และ 0.45 แสดงถึงความสอดคล้องกับการจัดลำดับด้วยผู้เชี่ยวชาญในระดับปานกลาง ค่าที่ได้น้อยกว่า 0.5 จึงเป็นค่าที่ค่อนข้างไปทางดี
- ค่าความแม่นยำสามแบบของขั้นตอนวิธีที่พิจารณาคูณภาพของความเห็นร่วมกับการคำนวณเพจเรงค์ มีค่าไม่น้อยกว่าค่าความแม่นยำในกรณีที่ใช้เพจเรงค์อย่างเดียว จึงเป็นสัญญาณที่ดีว่า การนำคูณภาพของความเห็นมาพิจารณาร่วมด้วยน่าจะเป็นประโยชน์ต่อการจัดลำดับความเชี่ยวชาญ



วิเคราะห์ผลการทดลอง

การที่ผลการทดลองยังบ่งบอกได้ไม่ชัดเจนนักถึงผลของคุณภาพของ
ความเห็นต่อการจัดลำดับความเชี่ยวชาญ อาจเนื่องมาจากปัจจัยหลาย
ประการ โดยแบ่งออกเป็นสองปัจจัยดังนี้

1. ปัจจัยที่เกี่ยวข้องกับการประเมิน
2. ปัจจัยที่เกี่ยวข้องกับลักษณะของข้อมูลความเห็นในบทสนทนา



ปัจจัยที่เกี่ยวข้องกับการประเมิน

1. การระบุตัวผู้เชี่ยวชาญได้ผลที่น่าพอใจมากกว่าการวัดความแม่นยำในการจัดลำดับความเชี่ยวชาญ เนื่องจากการจัดลำดับของสมาชิกที่ให้ความเห็นในบทสนทนาหนึ่ง ๆ เรียงจากมากไปน้อยทำได้ยากกว่าการให้ผู้เชี่ยวชาญระบุว่าสมาชิกคนใดเป็นหรือไม่เป็นผู้เชี่ยวชาญ



ปัจจัยที่เกี่ยวข้องกับการประเมิน (ต่อ)

2. ผู้เชี่ยวชาญจะพิจารณาแต่ละบทสนทนา จากนั้นจึงจัดลำดับความเชี่ยวชาญของสมาชิกที่แสดงความเห็นแต่ละคนและระบุตัวสมาชิกที่เป็นผู้เชี่ยวชาญ ผลการประเมินจึงเป็นการพิจารณาความเชี่ยวชาญของสมาชิกคนหนึ่ง ๆ เมื่อเทียบกับสมาชิกคนอื่นโดยไม่ได้พิจารณาความสามารถของสมาชิกคนนี้ในทุกบทสนทนาที่สมาชิกให้ความเห็น



ปัจจัยที่เกี่ยวข้องกับการประเมิน (ต่อ)

3. การกำหนดลำดับความเชี่ยวชาญและการระบุตัวผู้เชี่ยวชาญโดยผู้เชี่ยวชาญภาษาจาวา เพื่อใช้เป็น “มาตรฐาน” สำหรับเปรียบเทียบกับผลลัพธ์จากขั้นตอนวิธีทั้งสองนั้นทำได้ยาก เนื่องจากต้องหาผู้ที่มีประสบการณ์เชี่ยวชาญในภาษาจาวามาทำการวิเคราะห์บทสนทนาจำนวนมากที่เกิดขึ้นระหว่างสมาชิกจำนวนมากเช่นกัน ในที่นี้ผู้วิจัยเลือกบทสนทนาจำนวนเพียง 20 บทสนทนา จากทั้งหมด 12,269 บทสนทนา เพื่อใช้ในการสร้าง “มาตรฐาน” สำหรับการทดสอบ



ปัจจัยที่เกี่ยวข้องกับลักษณะของข้อมูลความเห็นใน บทสนทนา

1. บางกรณีความแตกต่างระหว่างความเห็นในระดับทั่วไปกับความเห็นจากผู้เชี่ยวชาญนั้นเด่นชัด แต่ในบางกรณีความแตกต่างอาจไม่เด่นชัดนักเนื่องจากสมาชิกที่มาให้ความเห็นที่หลังมักไม่แสดงความเห็นเช่นเดียวกับความเห็นที่สมาชิกอื่นให้ก่อนหน้าไปแล้ว ทำให้การแสดงความเห็นจะเป็นเพียงส่วนที่เพิ่มเติมจากความเห็นที่มีอยู่แล้วเท่านั้น ถึงแม้สมาชิกที่มาแสดงความเห็นจะมีความเชี่ยวชาญสูงก็ตาม
2. ในบางกรณีผู้เชี่ยวชาญอาจให้ความเห็นที่ดีแต่ไม่ได้ให้บ่อย ในสภาพเช่นนี้จึงกระทบต่อคะแนนความเชี่ยวชาญได้



สรุปผลการวิจัย

งานวิจัยนี้นำเสนอขั้นตอนวิธีในการจัดลำดับความเชี่ยวชาญในกลุ่มสังคมออนไลน์แบบถาม-ตอบปัญหา โดยนำเอาคุณภาพของความเห็น มาใช้ร่วมกับขั้นตอนวิธีแบบเพจเรงค์ในการหาค่าความเชี่ยวชาญของสมาชิกเพื่อนำไปใช้ในการจัดลำดับความเชี่ยวชาญ จากนั้นสรุปผลโดยเปรียบเทียบกับผลการจัดลำดับโดยผู้เชี่ยวชาญผ่านแบบสอบถาม ผลการทดลองแม้จะยังไม่บ่งบอกเด่นชัดนักว่าการพิจารณาคุณภาพของความเห็นส่งผลต่อการประเมินความเชี่ยวชาญ แต่มีแนวโน้มในเชิงบวกว่า น่าจะช่วยเสริมให้การประเมินความเชี่ยวชาญดียิ่งขึ้นได้



อุปสรรคและข้อจำกัด

1. ในการคัดเลือกบทสนทนาสำหรับแบบสอบถามนั้นทำได้ยาก
เนื่องจากข้อมูลบทสนทนามีจำนวนมากถึง 12,269 บทสนทนา
 - หากเลือกบทสนทนาที่มีสมาชิกที่เชี่ยวชาญในระดับที่แตกต่างกันมาแสดงความเห็น จะช่วยให้ผู้เชี่ยวชาญสามารถตอบแบบสอบถามได้ง่ายเพราะเห็นความแตกต่างของแต่ละความเห็นชัดเจน
 - หากเลือกบทสนทนาที่มีสมาชิกที่เชี่ยวชาญในระดับที่ใกล้เคียงกันมาแสดงความเห็น จะทำให้โอกาสที่ผู้เชี่ยวชาญจะจัดลำดับแตกต่างจากการจัดลำดับของขั้นตอนวิธีของงานวิจัย จะมีมากขึ้น ซึ่งส่งผลให้สามารถตรวจพบการจัดลำดับที่ผิดของขั้นตอนวิธีของงานวิจัยได้ง่ายกว่า



อุปสรรคและข้อจำกัด (ต่อ)

2. เนื่องจากแบบสอบถามจำเป็นต้องใช้ความเชี่ยวชาญในการตอบคำถาม ทำให้ใช้เวลานานในการค้นหาผู้เชี่ยวชาญที่เหมาะสมมาทำแบบทดสอบ แต่ด้วยข้อจำกัดด้านเวลา จึงจำเป็นต้องให้ผู้เชี่ยวชาญบางคนที่มีประสบการณ์ภาษาจาวาไม่มากนัก แต่มีความชำนาญในภาษาใกล้เคียงเช่น ภาษาซีชาร์ป (**C#**) มาร่วมทำแบบสอบถามด้วย
3. บทสนทนาส่วนใหญ่มีสมาชิกในกลุ่มร่วมแสดงความเห็นจำนวนมาก เนื่องจากความเห็นมีจำนวนมากทำให้ไม่สามารถเลือกความเห็นทั้งหมดในบทสนทนามาเป็นตัวเลือกของคำตอบของแบบสอบถามได้



แนวทางการวิจัยต่อ

1. ปรับปรุงวิธีการวัดผล เนื่องจากจำนวนบทสนทนาในแบบสอบถามที่ใช้ในการวัดผลนั้นยังมีจำนวนน้อยมากเมื่อเทียบทั้งหมด รวมทั้งผู้เชี่ยวชาญที่ร่วมในการทำแบบสอบถามยังมีจำนวนน้อย การเพิ่มจำนวนบทสนทนาและจำนวนผู้เชี่ยวชาญ หรือแม้กระทั่งเปลี่ยนรูปแบบของแบบสอบถามโดยให้ผู้เชี่ยวชาญแต่ละคนทำการพิจารณาความเห็นทั้งหมดในทุกบทสนทนาที่สมาชิกคนหนึ่ง ๆ มีส่วนร่วมน่าจะช่วยเพิ่มประสิทธิภาพของการประเมิน และทำให้เห็นผลการทดลองที่ชัดเจนขึ้น



แนวทางการวิจัยต่อ (ต่อ)

2. เนื่องจากยังมีปัจจัยอื่นที่เกี่ยวข้องกับคุณภาพของความเห็น เช่น ระยะเวลาที่แสดงความเห็น ความเกี่ยวข้องของความเห็นกับหัวข้อของกลุ่ม ค่าความสามารถในการอ่านได้ง่าย (**Readability**) ความยากของคำถาม ประวัติของสมาชิกในกลุ่ม เป็นต้น จึงสามารถนำปัจจัยเหล่านี้มาพิจารณาเพิ่มเติมได้



แนวทางการวิจัยต่อ (ต่อ)

- การปรับปรุงโดยการให้น้ำหนักของความสัมพันธ์กับขั้นตอนวิธีเพจเรงก์ (Weighted Page Rank) โดยเมื่อสมาชิกแสดงความเห็นให้กับสมาชิกคนเดิมหลายครั้งก็จะทำให้มีน้ำหนักของความเชี่ยวชาญสูงขึ้น
- เนื่องจากความเห็นจำนวนมากมีส่วนผสมของรหัสต้นฉบับ (Source code) ทำให้การระบุคุณภาพของความเห็นอาจผิดพลาดได้ จึงควรวิเคราะห์คุณภาพของข้อความ และรหัสต้นฉบับแยกจากกัน



THANK YOU

Q&A

โชคชัย พุฒตาล

คณะวิศวกรรมคอมพิวเตอร์,
สาขาวิชา วิศวกรรมซอฟต์แวร์



เครือข่ายสังคมออนไลน์ (Social Network)

