



PHP4: CONDITIONAL STRUCTURE

WRITTEN BY THAPANAPONG RUKKANCHANUNT

IF STATEMENT REVIEW

```
if (condition 1) {  
    Statement #1  
} elseif (condition 2) {  
    Statement #2  
} else {  
    Statement #3  
}
```

ROAD TO GOOGLE PART I

- Refer to DEMO for solution

ROAD TO GOOGLE PART I [2]

- ถ้าเราลองค้นหาข้อมูลเกี่ยวกับฟังก์ชัน `str_word_count` เราจะพบว่า `str_word_count` สามารถรับ Argument ตัวที่สองได้
- ถ้าเราให้ Argument ตัวที่สองเป็น 1 เราจะได้ Array ที่มี Key เป็นคำ และ Value เป็นจำนวนคำ ในข้อความนั้น
- ถ้าเราใช้โครงสร้างแบบมีเงื่อนไข เราสามารถทำ Road to Google Part I ได้ง่ายมาก

ARRAY SORTING

Function Name	Description
<code>asort()</code>	เรียงลำดับด้วย Value จากน้อยไปมาก
<code>arsort()</code>	เรียงลำดับด้วย Value จากมากไปน้อย
<code>ksort()</code>	เรียงลำดับด้วย Key จากน้อยไปมาก
<code>krsort()</code>	เรียงลำดับด้วย Key จากมากไปน้อย

ROAD TO GOOGLE PART II

- ในการที่เราจะบอกว่าเว็บไซต์นี้คือเว็บที่ต้องการ เราต้องรู้ก่อนว่าเว็บไซต์นี้พูดถึงเรื่องอะไร
- แต่ละเว็บไซต์ก็จะมีข้อความและคำมากมาย ซึ่งถ้าจะเอาคำทั้งหมดมาพิจารณาว่าเว็บไซต์นี้คือสิ่งที่เราต้องการนั้นทำได้ยากและเสียเวลามาก
- ดังนั้นเราควรจะสรุปเนื้อหาหรือดึงเฉพาะคำสำคัญออกมา

ROAD TO GOOGLE PART II [2]

- คำสำคัญ หรือ Keyword ควรจะเป็นคำที่ปรากฏอยู่ในเว็บไซต์นั้น ๆ
- คำถามคือคำไหนควรเป็น Keyword หรือ Keyword ต้องมีลักษณะอย่างไร

ROAD TO GOOGLE PART 2 [3]

- Keyword ควรเป็นคำที่เห็นบ่อย
- ถ้าเราจะให้คะแนนแต่ละคำว่าเป็น Keyword หรือไม่ เราสามารถให้คะแนนได้หลายวิธี วิธีที่เราจะใช้ในปฏิบัติการนี้ (และอีกหลาย ๆ ปฏิบัติการถัดไป) จะเรียกว่า TF-IDF (Term Frequency – Inverse Document Frequency)

TERM FREQUENCY

- Keyword จะปรากฏบ่อยดั่งนั้น คะแนนของคำนั้นจะสูงถ้าคำนั้นปรากฏบ่อย ถ้าเราให้ $TF(w,d)$ แทนคะแนน TF ของคำ w ของข้อความ d เราจะได้สมการของคะแนนดังนี้

$$TF(w, d) = 0.5 + 0.5 \times \frac{f(w, d)}{\max\{f(w, d): w \in d\}}$$

เมื่อ $f(w,d)$ คือ จำนวนคำ w ในข้อความ d

INVERSE DOCUMENT FREQUENCY

- แต่ถ้าคำที่ปรากฏบ่อย ๆ ในข้อความหนึ่ง ปรากฏในข้อความอื่น ๆ บ่อย เช่นกัน คำนั้นอาจจะไม่ใช่ Keyword อีกต่อไป ดังนั้นคำที่ปรากฏในหลาย ๆ ข้อความควรจะได้คะแนนน้อย

$$IDF(w) = \log \left(\frac{\text{จำนวนข้อความทั้งหมด}}{\text{จำนวนข้อความที่มีคำ } w} \right)$$

TF-IDF SCORE

- เมื่อนำทั้งสองส่วนมารวมกัน เราจะได้คะแนนของแต่ละคำเป็น

$$S(w, d) = TF(w, d) \times IDF(w)$$

- ถ้าค่าไหนของข้อความใดมีคะแนนสูง คำนั้นก็จะเป็น Keyword ของข้อความนั้น

ROAD TO GOOGLE PART 2 [4]

- ในแถบนี้เราจะเขียนฟังก์ชันคำนวณคะแนน TF และคะแนน IDF