

## IR-assignment II

ให้เปรียบเทียบ search engine 2 ระบบ (นักศึกษาสามารถเลือกได้เอง) โดยใช้ queries จำนวน 2 queries ที่นักศึกษาสนใจที่จะค้นคืนข้อมูล

สำหรับ q1

นำผลลัพธ์ที่ได้ใน search engine แต่ละระบบอย่างละ 30 ลิงค์ (รวม 60 ลิงค์) มาทำเครื่องหมาย + สำหรับลิงค์ที่เกี่ยวข้อง และ - สำหรับลิงค์ที่ไม่เกี่ยวข้อง ให้รวมลิงค์ที่เกี่ยวข้องของทั้งสอง search engine ไว้ด้วยกัน สำหรับลิงค์ที่ซ้ำกันหรือลิงค์ที่ปรากฏหลายครั้งในผลลัพธ์ให้ตัดออกเหลือเพียงชุดเดียว สมมติว่าเหลือ 22 ลิงค์ ให้ถือว่า 22 เป็นจำนวนเอกสารที่เกี่ยวข้องทั้งหมด สำหรับ q1

สำหรับ q2

ทำเช่นเดียวกับ q1

สิ่งที่ต้องส่ง

1. สำหรับแต่ละ query
  - พิมพ์ค่าผลลัพธ์ 30 ลิงค์แรกของแต่ละ search engine
  - พิมพ์ค่าผลลัพธ์ของ relevant documents เมื่อนำลิงค์ที่เป็น + ของทั้งสอง search engine มา merge รวมกันแล้ว (ต่อไปจะสมมติให้เป็นเอกสารที่เกี่ยวข้องทั้งหมด)
2. สร้างตารางเพื่อคำนวณค่า precision และ recall
3. Plot graph ค่า precision และ recall ของแต่ละ search engine
4. สร้างตาราง interpolated ของค่า precision และ recall
5. วาดกราฟเพื่อเปรียบเทียบสอง search engines

ตัวอย่าง

q1 : climate change in Stockhome

หลังจากพิจารณาลิงค์ทั้งสิ้น 30 ลิงค์ จำนวนเอกสารที่เกี่ยวข้องมีทั้งหมด 22 รายการ ดังนี้

1. <http://www.bbc.co.uk/weather> .....
2. <http://.....>
- .....
22. <http://en.wikipedia.org/wiki/.....>

q2 : Biography of Micheal Jackson

หลังจากพิจารณาลิงค์ทั้งสิ้น 30 ลิงค์ จำนวนเอกสารที่เกี่ยวข้องมีทั้งหมด 9 รายการ ดังนี้

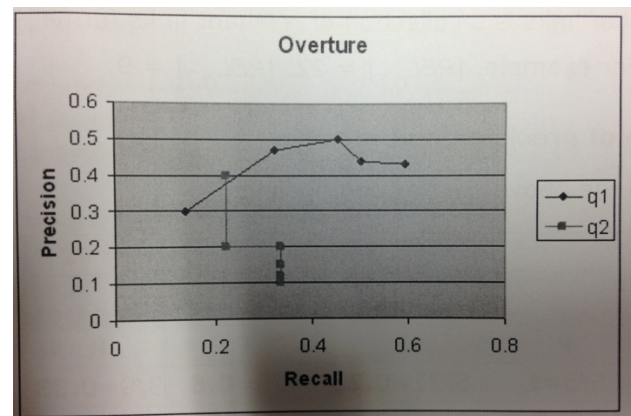
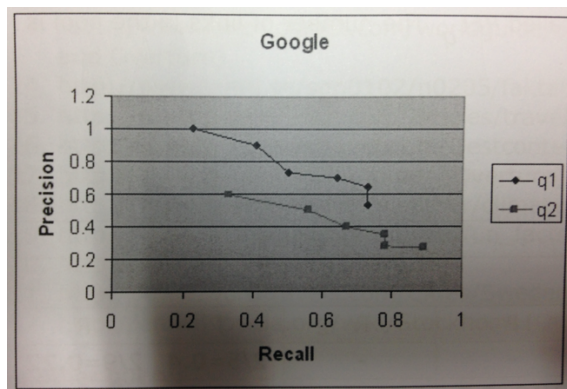
1. <http://en.wikipedia.org/wiki/.....>
2. <http://.....>
- .....
9. <http://en.wikipedia.org/wiki/.....>

ตารางแสดงผลค่า precision และ recall ของ search engine ทั้งสองสำหรับ q1 และ q2

	Google				Overture			
	q1		q2		q1		q2	
	P	R	P	R	P	R	P	R
Top 5	$5/5 = 1$	$5/22 = 0.23$	$3/5=0.6$	$3/9=0.33$			$2/5=0.4$	$2/9=0.22$
Top 10	$9/10 = 0.9$	$9/22 = 0.41$	$5/10=0.5$	$5/9=0.56$	$3/10=0.3$	$3/22=0.14$	$2/10=0.2$	$2/9=0.22$
Top 15	$11/15=0.73$	$11/22 = 0.5$	$6/15=0.4$	$6/9=0.67$	$7/15=0.47$	$7/22=0.32$	$3/15=0.2$	$3/9=0.33$
Top 20	$14/20 = 0.7$	$14/22 = 0.64$	$7/20=0.35$	$7/9=0.78$	$10/20=0.5$	$10/22=0.45$	$3/20=0.15$	$3/9=0.33$
Top 25	$16/25=0.64$	$16/22 = 0.73$	$7/25=0.28$	$7/9=0.78$	$11/25=0.44$	$11/22=0.5$	$3/25=0.12$	$3/9=0.33$
Top 30	$16/30 = 0.53$	$16/22 = 0.73$	$8/30=0.27$	$8/9=0.89$	$13/30=0.43$	$13/22=0.59$	$3/30=0.1$	$3/9=0.33$

หมายเหตุ: ค่าว่างในตารางของ Overture ไม่ได้กำหนด เนื่องจากไม่มี relevant documents ใน top5

กราฟแสดงค่า precision และ recall



Interpolated precision and recall

Standard recall	Google			Overture		
	q1	q2	average	q1	q2	average
0.1	1	0.6	0.8	0.5	0.4	0.45
0.2	1	0.6	0.8	0.5	0.4	0.45
0.3	0.9	0.6	0.73	0.5	0.2	0.35
0.4	0.9	0.5	0.7	0.5	0	0.35
0.5	0.73	0.5	0.62	0.44	0	0.22
0.6	0.7	0.4	0.55	-	-	-
0.7	0.64	0.35	0.5	-	-	-
0.8	0	0.27	0.14	-	-	-
0.9	-	-	-	-	-	-
1.0	-	-	-	-	-	-

เปรียบเทียบสอง search engines ด้วยกราฟ

