Automatic Global Analysis Example

Example

D_1 = A,B,B,A,A,C

D_2 = D,D,C

D_3 = B,E,E

D_4 = D,E,A

Query = 2.3A+ C

$$w_{i,j} = \frac{\left(0.5+0.5\frac{f_{i,j}}{\max_j(f_{i,j})}\right) itf_j}{\sqrt{\sum_{l=1}^{N}\left(0.5+0.5\frac{f_{i,j}}{\max_l(f_{i,l})}\right)^2 itf_l^2}}$$

$$itf_j = \log\frac{t}{t_j}$$

Step1 เขียนตาราง + หา itf ทุก Doc

Automatic Global Analysis Example

① เขียนตาราง

Example

D_1 = A,B,B,A,A,C

D_2 = D,D,C

D_3 = B,E,E

D_4 = A,D,E

Query = 2.3A+ C

Term = 5

$$itf_j = \log \frac{t}{t_j}$$

$$itf_4 = \log \frac{5}{3} = 0.222$$

(5: key ทั้งหมด; 3: Doc j มี key ทั้งหมดเท่าไร)

Key/Doc	D1	D2	D3	D4
A	3	0	0	1
B	2	0	1	0
C	1	1	0	0
D	0	2	0	1
E	0	0	2	1
Max	3	2	2	1
t_j (all keyword Doc j)	3	2	2	3
itf(Doc)	0.222	0.398	0.398	0.222

Step 2

Automatic Global Analysis Example

	D1	D2	D3	D4
A	3	0	0	1
B	2	0	1	0
C	1	1	0	0
D	0	2	0	1
E	0	0	2	1
Max	3	2	2	1
tj	3	2	2	3
itf	0.222	0.398	0.398	0.222

$$w_{i,j} = \frac{\left(0.5+0.5\frac{f_{i,j}}{\max_j(f_{i,j})}\right) itf_j}{\sqrt{\sum_{l=1}^{N}\left(0.5+0.5\frac{f_{i,j}}{\max_l(f_{i,l})}\right)^2 itf_l^2}}$$

$$w_{1,3} = \frac{\left(0.5+0.5\frac{f_{1,3}}{\max(f_{d3})}\right) itf_3}{\sqrt{\left(0.5+0.5\frac{f_{1,1}}{\max(f_{d1})}\right)^2 itf_1^2 + \left(0.5+0.5\frac{f_{1,2}}{\max(f_{d2})}\right)^2 itf_2^2 + \left(0.5+0.5\frac{f_{1,3}}{\max(f_{d3})}\right)^2 itf_3^2 + \left(0.5+0.5\frac{f_{1,4}}{\max(f_{d4})}\right)^2 itf_4^2}}$$

$$w_{1,3} = \frac{(0.5+0.5*\frac{0}{2})0.398}{\sqrt{(0.5+0.5*\frac{3}{3})^2 0.222^2 + (0.5+0.5*\frac{0}{2})^2 0.398^2 + (0.5+0.5*\frac{0}{2})^2 0.398^2 + (0.5+0.5*\frac{1}{1})^2 0.222^2}}$$

$$w_{1,3} = 1.509$$

Step 3 ดับคู่ หา ค. สัมพันธ์ ทุก keyword

Automatic Global Analysis Example ตัวอย่าง

C	A	B	C	D	E
A	10.218	8.293	7.987	7.879	7.879
B	8.293	7.728	7.085	6.581	7.290
C	7.987	7.085	7.383	7.241	6.522
D	7.879	6.581	7.241	7.548	6.397
E	7.879	7.290	6.522	6.397	7.548

Term Weight

W	D_1	D_2	D_3	D_4
A	1.683	1.509	1.509	1.683
B	1.228	1.322	1.983	0.737
C	0.996	2.010	1.340	0.747
D	0.598	2.146	1.073	1.197
E	0.598	1.073	2.146	1.197

$$c_{u,v} = \vec{k_u} \bullet \vec{k_v} = \sum_{j=1}^{N} w_{u,j} \times w_{v,j}$$

$$C_{1,3} = w_{1,1} * w_{3,1} + w_{1,2} * w_{3,2} + w_{1,3} * w_{3,3} + w_{1,4} * w_{3,4} = C_{3,1}$$

$$= 1.683*0.996 + 1.509*2.010 + 1.509*1.340 + 1.683*0.747$$

$$= 7.987$$

Automatic Global Analysis Example

The relationship between two terms

C	A	B	C	D	E
A	10.218	8.293	7.987	7.879	7.879
B	8.293	7.728	7.085	6.581	7.290
C	7.987	7.085	7.383	7.241	6.522
D	7.879	6.581	7.241	7.548	6.397
E	7.879	7.290	6.522	6.397	7.548

Step 4

หา Sim ระหว่าง query กับ keyword + ปรับเปลี่ยน query

Automatic Global Analysis Example

term similarity

C	A	B	C	D	E
A	10.218	8.293	7.987	7.879	7.879
B	8.293	7.728	7.085	6.581	7.290
C	7.987	7.085	7.383	7.241	6.522
D	7.879	6.581	7.241	7.548	6.397
E	7.879	7.290	6.522	6.397	7.548
q	2.3	0	1	0	0

$Sim(q,K_j)$
31.487
26.159
25.753
25.362
24.643

(2) ปรับ query

ADD K_2 to Query

(1) หา Sim q กับทุก keyword

อยู่ใน query และ ต่ำสุดใน query

$$sim(q,k_v) = \vec{q}\cdot\vec{k}_v = \sum_{k_u \in q} w_{u,q} \times c_{u,v}$$

$$sim(q,k_3) = w_{1,q}*c_{1,3} + w_{2,q}*c_{2,3} + w_{3,q}*c_{3,3} + w_{4,q}*c_{4,3} + w_{5,q}*c_{5,3}$$

$= 2.3*7.987 + 1*7.383 = 25.753$

Steps

ถ้า query ถูกปรับ = ต้องหา Sim ใหม่

Automatic Global Analysis Example

Recompute term similarity

C	A	B	C	D	E	Sim(q,K_i)
A	10.218	8.293	7.987	7.879	7.879	39.780
B	8.293	7.728	7.085	6.581	7.290	33.887
C	7.987	7.085	7.383	7.241	6.522	32.838
D	7.879	6.581	7.241	7.548	6.397	31.942
E	7.879	7.290	6.522	6.397	7.548	31.933
q	2.3	1	1	0	0	

$$sim(q,k_3) = w_{1,q}*c_{1,3} + w_{2,q}*c_{2,3} + w_{3,q}*c_{3,3} + w_{4,q}*c_{4,3} + w_{5,q}*c_{5,3}$$

$$= 2.3*7.987 + 1*7.085 + 1*7.383 = 32.838$$

ต้องคำนวณ Sim ใหม่เพราะ k 2 มีการเปลี่ยน

Automatic Global Analysis Example

Compute new weight terms for query

Original Query

$q = 2.3K_1 + K_2 + K_3$ Sum query weight = 2.3+1+1 = 4.3

$$w_{v,q'} = \frac{sim(q,k_v)}{\sum_{k_u \in q} w_{u,q}}$$

$w_{1,q'}$ = 39.780/4.3 = 9.251
$w_{2,q'}$ = 33.887/4.3 = 7.881
$w_{3,q'}$ = 32.838/4.3 = 7.637

	A	B	C	D	E
q'	9.251	7.881	7.637	-	-

Step 7 หา Sim ระหว่าง qใหม่ กับ Doc ทุกอัน

Automatic Global Analysis Example

Arrange Relevance

$q' = 9.251A + 7.881B + 7.637C$

W	D1	D2	D3	D4
A	1.683	1.509	1.509	1.683
B	1.228	1.322	1.983	0.737
C	0.996	2.010	1.340	0.747
D	0.598	2.146	1.073	1.197
E	0.598	1.073	2.146	1.197

C	A	B	C	D	E
A	10.22	8.293	7.987	7.879	7.879
B	8.293	7.728	7.085	6.581	7.290
C	7.987	7.085	7.383	7.241	6.522
D	7.879	6.581	7.241	7.548	6.397
E	7.879	7.290	6.522	6.397	7.548

$$sim(q,d_j) \propto \sum_{k_v \in d_j} \sum_{k_u \in q} w_{i,j} \times w_{u,q} \times c_{u,v}$$

$w_{1,2} = 1.509$ $w_{1,q} = 9.251$
$w_{2,2} = 1.322$ $w_{2,q} = 7.881$
$w_{3,2} = 2.010$ $w_{3,q} = 7.637$
$w_{4,2} = 2.146$ $w_{4,q} = 0$
$w_{5,2} = 1.073$ $w_{5,q} = 0$

$$sim(q,d_2) = w_{1,2} * w_{1,q} * (c_{1,1} + c_{1,2} + c_{1,3} + c_{1,4} + c_{1,5}) +$$
$$w_{2,2} * w_{2,q} * (c_{2,1} + c_{2,2} + c_{2,3} + c_{2,4} + c_{2,5}) +$$
$$w_{3,2} * w_{3,q} * (c_{3,1} + c_{3,2} + c_{3,3} + c_{3,4} + c_{3,5}) +$$
$$w_{4,2} * w_{4,q} * (c_{4,1} + c_{4,2} + c_{4,3} + c_{4,4} + c_{4,5}) +$$
$$w_{5,2} * w_{5,q} * (c_{5,1} + c_{5,2} + c_{5,3} + c_{5,4} + c_{5,5})$$

$sim(q,d_2) = 1531.123$

Automatic Global Analysis Example

Arrange Relevance

$q' = 9.251A + 7.881B + 7.637C$

	D_1	D_2	D_3	D_4
$Sim(q,d_j)$	1,291.282	1,531.123	1,538.429	1,079.324

Answer = D_3, D_2, D_1, D_4