

Lösung 7

Aufgabe 27:

Precision: Anteil der relevanten Dokumente in Antwortmenge = $\frac{|Gef \cap Rel|}{|Gef|}$

Recall: Anteil der relevanten Dokumente, die gefunden werden = $\frac{|Gef \cap Rel|}{|Rel|}$

(Gef: Menge der gefundenen Dokumente für die Anfrage

Rel: Menge der relevanten Dokumente für die Anfrage)

Precision = 0.44 = 8/x $\Rightarrow$ x = 18

Insgesamt 18 Dokumente werden gefundenen.

Recall = 0.4 = 8/y $\Rightarrow$ y = 20

20 Dokumente befinden sich in der Kollektion.

Aufgabe 28:

Google: www.google.de

Rang	Abgekürzte URL	Relevant oder Irrelevant	Precision
1	www.jaguar.com/de/de/	relevant	1
2	www.jaguar.com/de/de/approved_used/	relevant	1
3	de.wikipedia.org/wiki/Jaguar_Cars	relevant	1
4	de.wikipedia.org/wiki/Jaguar	irrelevant	
5	www.jaguar-solingen.com/	irrelevant	
6	www.mobile.de › Auto	relevant	4/6
7	www.jaguar-forum.de/	relevant	5/7
8	www.jaguar-online.com/	irrelevant	
9	www.autobild.de › Alle Autos	relevant	6/9
10	www.jaguar-association.de/	relevant	7/10

Precision = 7/10 = 0.7

Average Precision = $\frac{1}{7} * (1+1+1+4/6+5/7+6/9+7/10) = 0.821$

Recall = 7/11 = 0.64

Bing: www.bing.com/?cc=de

Rang	Abgekürzte URL	Relevant oder Irrelevant	Precision
1	www.ebay.de	irrelevant	
2	www.meinauto.de	relevant	1/2
3	www.uni-kat.de	irrelevant	
4	www.jaguar.com/de/de	relevant	2/4

5	de.wikipedia.org/wiki/Jaguar	irrelevant	
6	de.wikipedia.org/wiki/Jaguar_Cars	relevant	3/6
7	www.jaguar.com/at	relevant	4/7
8	www.mobile.de/modellverzeichnisse/jaguar	relevant	5/8
9	https://media.jaguar.de/?nid=429	relevant	6/9
10	jaguar.de/jaguar.htm	irrelevant	

Precision = $6/10 = 0.6$

Average Precision = $1/6 * (1/2 + 2/4 + 3/6 + 4/7 + 5/8 + 6/9) = 0.56$

Recall = $6/11 = 0.55$

Aufgabe 29

System 1 : R N R N N N N N R R

System 2: N R N N R R R N N N

a)

R-Precision (system 1) = $r/|Rel| = 2/4 = 1/2 = 0.5$

R-Precision (system 2) = $r/|Rel| = 1/4 = 0.25$

P@10 (system 1) = $4/10 = 0.4$

P@10 (system 2) = $4/10 = 0.4$

b)

MAP (system 1) = $1/4 * (1 + 2/3 + 3/9 + 4/10) = 0.6$

MAP (system 2) = $1/4 * (1/2 + 2/5 + 3/6 + 4/7) = 0.49$

Aufgabe 30

DocID	Judge1	Judge2
1	0	0
2	0	0
3	1	1
4	1	1
5	1	0
6	1	0
7	1	0
8	1	0
9	0	1
10	0	1
11	0	1
12	0	1

1 \ 2	yes	no	total
yes	2	4	6
no	4	2	6
total	6	6	12

$$\text{Kappa} = \frac{P(A) - P(E)}{1 - P(E)}$$

$P(A)$: gemessener Übereinstimmungswert

$P(E)$: zufällig erwartete Übereinstimmung

$$P(A) = (2+2)/12 = 4/12 = 1/3 = 0.33$$

$$P(\text{nonrelevant}) = (6+6)/(12+12) = 1/2 = 0.5$$

$$P(\text{relevant}) = (6+6)/(12+12) = 1/2 = 0.5$$

$$P(E) = 1/4 + 1/4 = 1/2 = 0.5$$

$$\text{Kappa} = \frac{0.33 - 0.5}{1 - 0.5} = -0.34$$

b)

IR System hat Dokumente 4, 5, 6, 7,8 zurückgegeben.

$$| \text{Gef} | = \{4,5,6,7,8\}$$

$$| \text{Rel} | = \{3,4\}$$

$$\text{Precision} = 1/5 = 0.2$$

$$\text{Recall} = 1/2 = 0.5$$

$$F = \frac{2 * \text{Recall} * \text{Precision}}{(\text{Recall} + \text{Precision})} = \frac{2 * 0.2 * 0.5}{(0.2 + 0.5)} = \frac{2}{7} = 0.29$$