Ocena wyszukiwania

- Ocena systemu wyszukiwania to miara tego, jak bardzo system spełnia potrzeby informacyjne użytkowników
  - Problem stanowi inna ocena jakości tego samego zbioru wynikowego przez różnych użytkowników
  - Aby rozwiązać problem wprowadzono różne miary jakości wyszukanego zbioru skorelowane z preferencjami różnych grup użytkowników

- Aby oceniać systemy wyszukiwania (indeksowania) opracowano testowe kolekcje referencyjne bazujące na założeniach wynikających z badań Cranfielda (1958-1966)

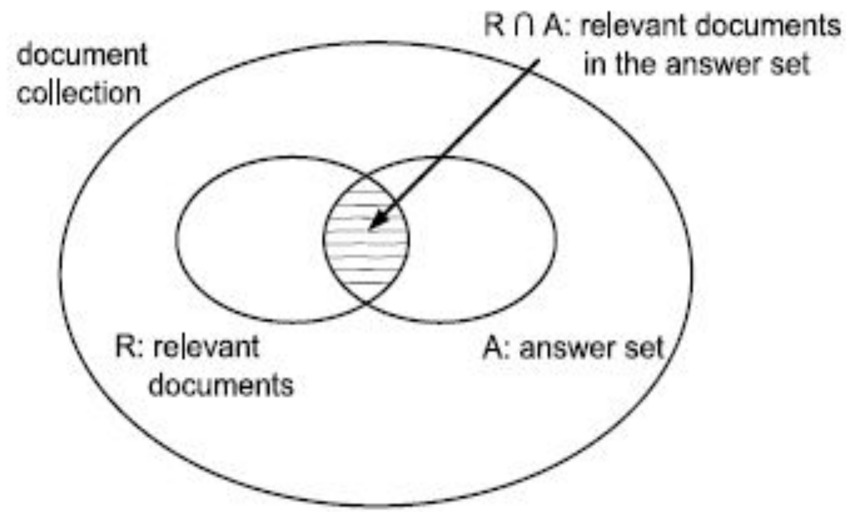
- Pozwalają one porównywać różne metody rankingu na tych samych zbiorach pytań i dokumentów

Ocena wyszukiwania

- Kolekcja referencyjna składa się z:
  - Zbioru *D* preselekcjonowanych dokumentów,
  - Zbioru *I* opisów informacji używanej do testowania,
  - Zbioru ocen istotności związanych z każdą parą $[i_m, d_j]$ gdzie $i_m \in I$, $d_j \in D$.
- Ocena istotności dokumentu ma wartość 0 gdy dokument d_j jest nieistotny dla i_m i 1 w przeciwnym przypadku
- Taka ocena dokonywana jest przez specjalistów

dokładność i kompletność

- Dane są:
  - *I*: żądanie informacji,
  - *R*: zbiór dokumentów istotnych dla *I,*
  - *A*: zbiór odpowiedzi dla *I*, wygenerowanych przez system wyszukiwania IR,
  - $R \cap A$: iloczyn zbiorów *R* i *A.*



dokładność i kompletność

- Kompletność jest częścią istotnych dokumentów (zbiór *R*), które zostały wyszukane:

$$Recall = \frac{|R \cap A|}{|R|}$$

- Dokładność jest częścią wyszukanych dokumentów (zbiór *A*), które są istotne:

$$Precision = \frac{|R \cap A|}{|A|}$$

- Definicja dokładności i kompletności zakłada, że wszystkie dokumenty w zbiorze *A* są testowane
- Użytkownik widzi uszeregowany zbiór dokumentów i sprawdza je od początku, więc dokładność i kompletność zmieniają się podczas przeglądania zbioru *A*.

dokładność i kompletność

- Wskazane jest wykreślenie zależności dokładności (precision) od kompletności (recall)
- Niech R_{q1} będzie zbiorem dokumentów istotnych dla zapytania q_1:

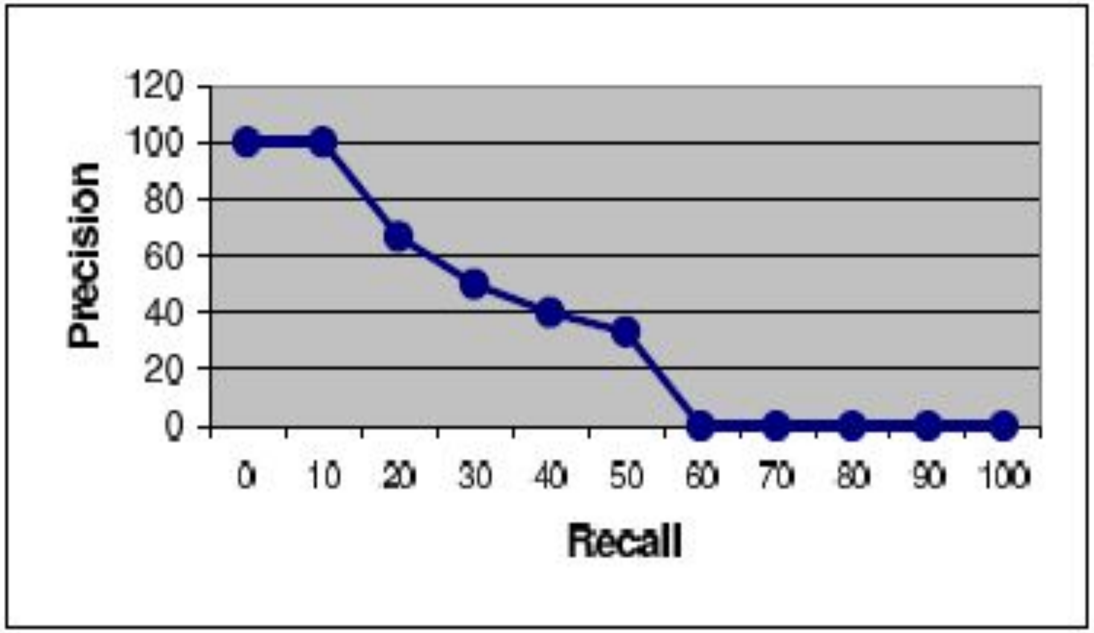
$$R_{q1} = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}\}$$

- Algorytm wyszukiwania daje zbiór odpowiedzi (dokumeny istotne oznaczone kółkiem)

01. d_{123} •	06. d_9 •	11. d_{38}
02. d_{84}	07. d_{511}	12. d_{48}
03. d_{56} •	08. d_{129}	13. d_{250}
04. d_6	09. d_{187}	14. d_{113}
05. d_8	10. d_{25} •	15. d_3 •

dokładność i kompletność

- Wykres dokładności w funkcji kompletności dla kolejnych dokumentów odpowiedzi



Recall	Precision
0	100
10	100
20	66.6
30	50
40	40
50	33.3
60	0
70	0
80	0
90	0
100	0

dokładność i kompletność

- Rozważmy zapytanie q_2 dla którego zbiór istotnych odpowiedzi jest dany jako:

$$R_{q2} = \{d_3, d_{56}, d_{129}\}$$

- Algorytm IR przetwarza zapytanie i zwraca następujący ranking:

01. d_{425}	06. d_{615}	11. d_{193}
02. d_{87}	07. d_{512}	12. d_{715}
03. d_{56} •	08. d_{129} •	13. d_{810}
04. d_{32}	09. d_{4}	14. d_{5}
05. d_{124}	10. d_{130}	15. d_{3} •

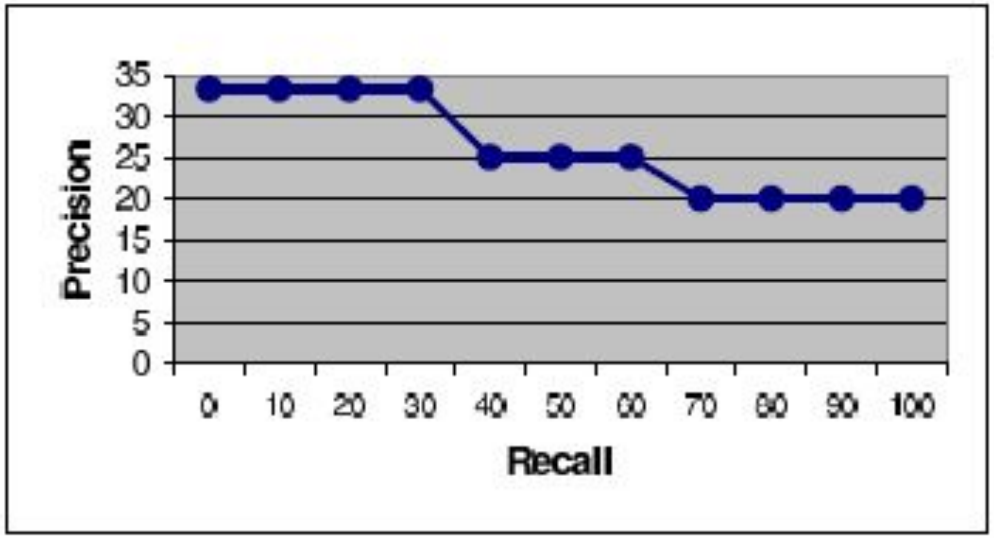
- Pierwszy istotny dokument d_{56} zapewnia dokładność i kompletność 33,3%, drugi d_{129} daje kompletność 66,6% i dokładność 25%, trzeci d_3 zapewnia 100% kompletności z dokładnością 20%

dokładność i kompletność

- Dla standardowych poziomów kompletności $r_j = \{0,1,\ldots,10\}$ dokładność jest interpolowana następująco

$$P(r_j) = \max_{\forall r \mid r \geq r_j} P(r)$$

- Ta zasada interpolacji umożliwia zbudowanie dla ostatniego przykładu następującego wykresu *Precision*(*Recall*)



Recall	Precision
0	33.3
10	33.3
20	33.3
30	33.3
40	25
50	25
60	25
70	20
80	20
90	20
100	20

dokładność i kompletność

- Zazwyczaj ocenia się średnią jakość wyszukiwania dla zbioru N_q pytań testujących

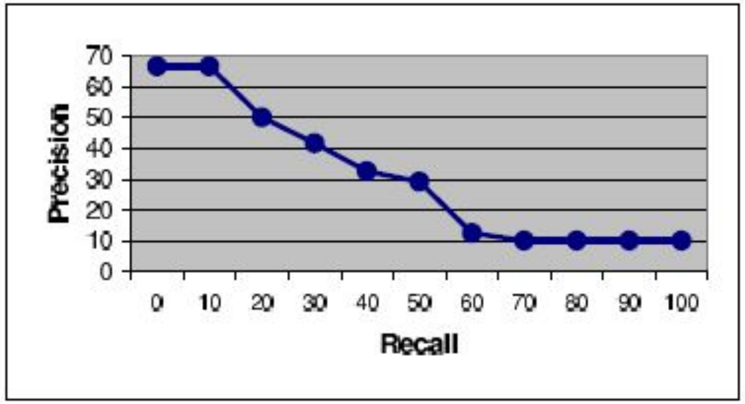
$$\overline{P(r_j)} = \sum_{i=1}^{N_q} \frac{P_i(r_j)}{N_q}$$

gdzie:

- $\overline{P(r_j)}$ - średnia dokładność na poziomie kompletności r_j,
- $P_i(r_j)$ – dokładność odpowiedzi na pytanie q_i z poziomiem kompletności r_j.

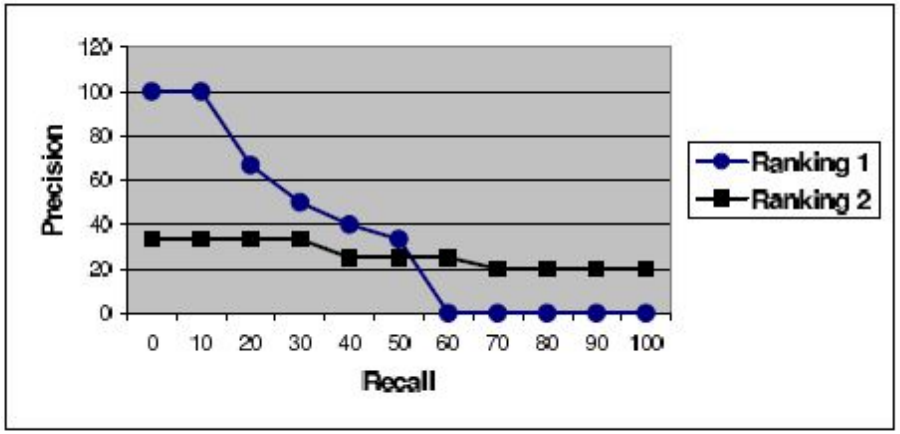
dokładność i kompletność

- Wykres *Precision*(*Recall*) uśredniony dla dwóch zapytań q_1 i q_2:



Recall	Precision
0	66.6
10	66.6
20	49.9
30	41.6
40	32.5
50	29.1
60	12.5
70	10
80	10
90	10
100	10

- Uśrednione przebiegi *Precision*(*Recall*) dla 2 różnych algorytmów wyszukiwania:



Dokładność i kompletność

- Dokładność i kompletność są często stosowane do oceny jakości wyszukiwania przez algorytmy IR
- Poprawna estymacja kompletności zapytania wymaga znajomości wszystkich dokumentów kolekcji testowej
- Miara *Precision-Recall* nie jest skalarna
- Jakość algorytmu IR jest mierzona na zbiorze zapytań w trybie wsadowym
- Dla systemów IR wymagających słabego uporządkowania dokumentów miara *Presision-Recall* może nie być najlepsza

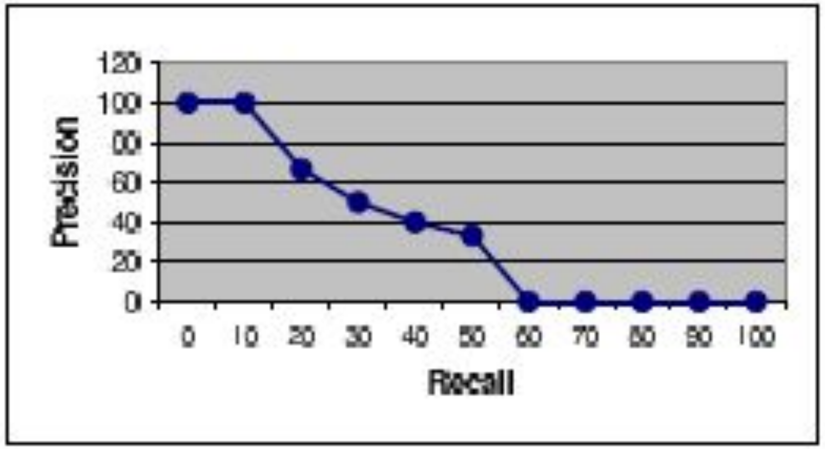
Miary *P@5* i *P@10*

- Użytkownicy wyszukiwarek cenią przede wszystkim dużo istotnych dokumentów na początku rankingu
- Miary *P@5* i *P@10* wyznaczają dokładność odpowiednio dla 5 i 10 pierwszych dokumentów; pozwalają ocenić czy użytkownicy uzyskują istotne dokumenty na początku rankingu
- Rozpatrujemy listę dokumentów dla przykładowego pytania q_1:

01. d_{123} $\bullet$	06. d_{9} $\bullet$	11. d_{38}
02. d_{84}	07. d_{511}	12. d_{48}
03. d_{56} $\bullet$	08. d_{129}	13. d_{250}
04. d_{6}	09. d_{187}	14. d_{113}
05. d_{8}	10. d_{25} $\bullet$	15. d_{3} $\bullet$

MAP: Mean average precision

- Dla tego pytania *P@5* = 40% i *P@10* = 40%
- *P@5* i *P@10* można uśredniać w zbiorze 100 zapytań
- MAP to średnia dokładność z wykresu *Precision-Recall*, przy standardowym zestawie kompletności
- Dla zapytania q_1:



Recall	Precision
0	100
10	100
20	66.6
30	50
40	40
50	33.3
60	0
70	0
80	0
90	0
100	0

$$MAP_1 = \frac{1 + 0.66 + 0.5 + 0.4 + 0.33 + 0 + 0 + 0 + 0 + 0}{10} = 0.28$$

R-precision

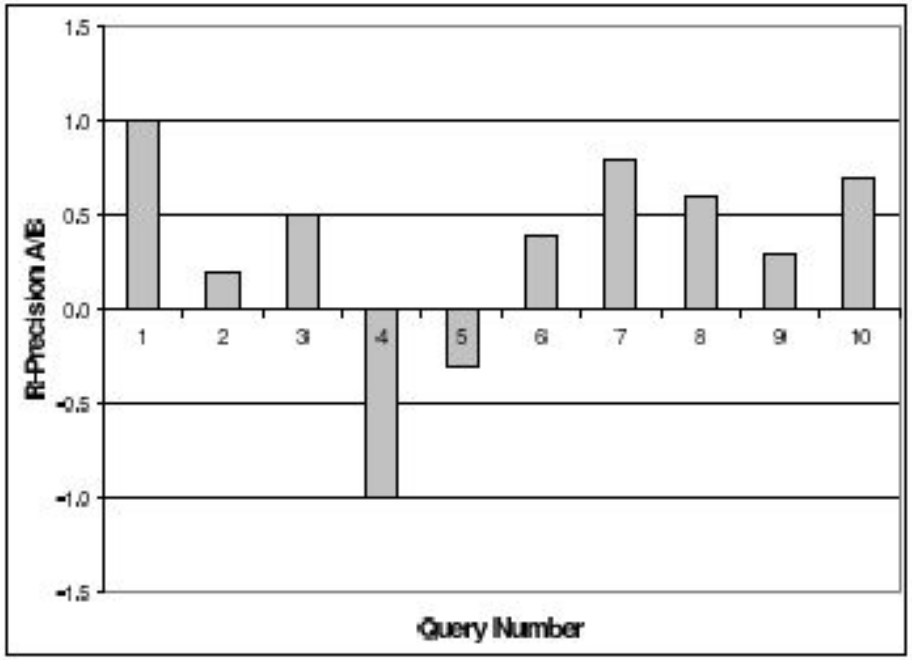
- R-dokładność to dokładność wyznaczona na r-tej pozycji w rankingu odpowiedzi
- Dla pytania q_1 są cztery dokumenty istotne wśród pierwszych 10-ciu zwróconych dokumentów więc R-dokładność wynosi 4/10=0,4
- R-dokładność może być także uśredniana w zbiorze zapytań
- Może być stosowana do porównywania 2 algorytmów wyszukiwania dla każdego zapytania

$$RP_{A/B}(i) = RP_A(i) - RP_B(i)$$

  - $RP_A(i)$ – R-dokładność algorytmu *A* dla i-tego zapytania
  - $RP_B(i)$ –R-dokładność algorytmu *B* dla i-tego zapytania

Histogram dokładności

- Przykładowy histogram porównania algorytmów $RP_{A/B}(i)$ dla 10 różnych zapytań



- Algorytm *A* jest lepszy w 8 przypadkach na 10

MRR: Mean Reciprocal Rank

- MRR – dobra miara jakości wyszukiwania gdy interesuje nas pierwsza poprawna odpowiedź – np. adresy URL, strony główne w sieci
- Jeżeli przyjąć:
  - R_i: ranking odpowiedzi dla zapytania q_i,
  - $S_\mathrm{correct}(R_i)$: pozycja pierwszej poprawnej odpowiedzi w R_i,
  - S_h: próg pozycji rankingu,
- to odwrotna ranga $RR(R_i)$ dla pytania q_i jest wyrażona jakc

$$RR(\mathcal{R}_i) = \begin{cases} \frac{1}{S_{correct}(\mathcal{R}_i)} & \text{if } S_{correct}(\mathcal{R}_i) \leq S_h \\ 0 & \text{otherwise} \end{cases}$$

MRR: Mean Reciprocal Rank

- Średnia odwrotna ranga (MRR) dla zbioru Q z N_q zapytań:

$$\boxed{MRR(Q) = \sum_i^{N_q} RR(\mathcal{R}_i)}$$

Miara E

- Łączy dokładność i kompletność w jednym wskaźniku skalarnym uwzględniającym ich proporcje:

$$E(j) = 1 - \frac{1 + b^2}{\frac{b^2}{r(j)} + \frac{1}{P(j)}}$$

- gdzie
  - *r*(*j*) – kompletność na j-tej pozycji w rankingu,
  - *P*(*j*) – dokładność na j-tej pozycji w rankingu,
  - b≥0 – ustalony parametr przewagi kompletności nad dokładnością; b=0 ⬇ *E*(*j*) =1-*P*(*j*) oraz b✹◎ ⬇ *E*(*j*) =1-*r*(*j*)
  - *E*(*j*) – E-metryka na j-tej pozycji w rankingu.

Miara F (średnia harmoniczna)

- Dla *b*=1 miara *E* staje się miarą średniej harmonicznej *F:*

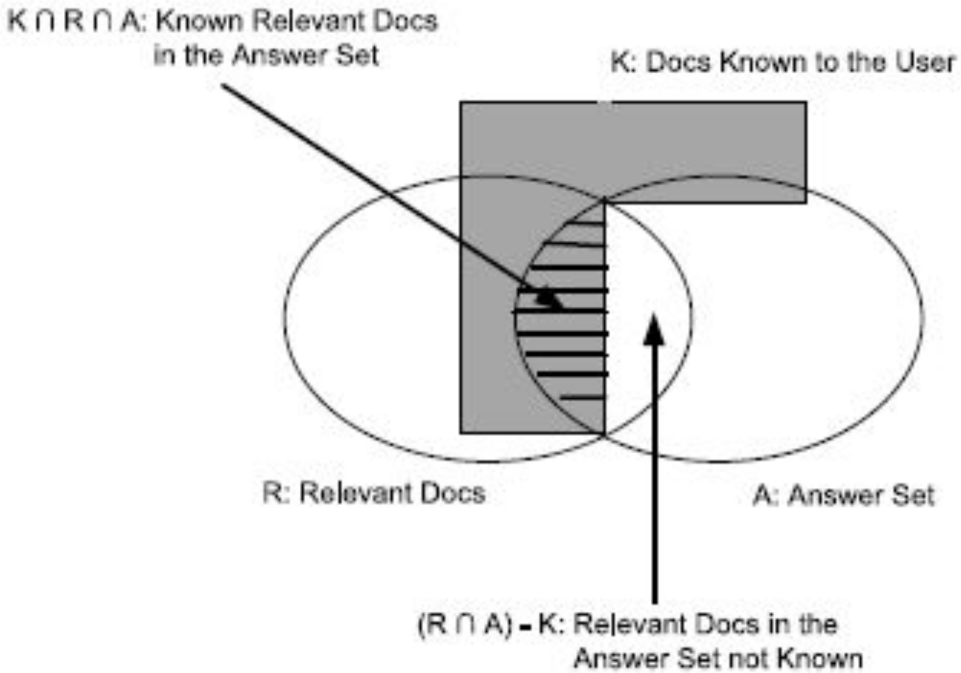
$$F(j) = \frac{2}{\frac{1}{r(j)} + \frac{1}{P(j)}}$$

- Funkcja *F* ma wartości w zakresie [0,1]; przyjmuje wartość 0 gdy żadne istotne dokumenty nie zostały znalezione lub wartość 1 gdy wszystkie dokumenty odpowiedzi są istotne
- Duże wartości *F* wiążą się z dużą dokładnością i kompletnością

Miary zorientowane na użytkownika

- Dokładność i kompletność zakładają, że zbiór istotnych dokumentów zwróconych w odpowiedzi na zapytanie nie zależy od użytkowników
- Różni użytkownicy mogą różnie rozumieć istotność dokumentów

Miary zorientowane na użytkownika



- K: zbiór dokumentów znanych użytkownikowi
- $K \cap R \cap A$: zbiór wyszukanych istotnych dokumentów znanych użytkownikowi
- $(R \cap A) - K$: zbiór wyszukanych istotnych dokumentów nieznanych użytkownikowi

Miary zorientowane na użytkownika

- Współczynnik pokrycia – frakcja dokumentów znanych i istotnych w zbiorze odpowiedzi

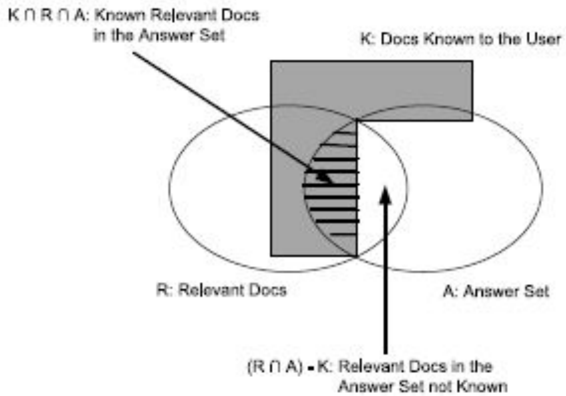
$$coverage = \frac{|K \cap R \cap A|}{|K \cap R|}$$

- Współczynnik nowości – frakcja istotnych dokumentów w zbiorze odpowiedzi nieznanych użytkownikowi

$$novelty = \frac{|(R \cap A) - K|}{|R \cap A|}$$

Wysokie pokrycie, wskazuje że system IR znalazł większość dokumentów oczekiwanych przez użytkownika

Wysoka nowość oznacza znalezienie wielu nowych, istotnych dokumentów



DCG — Discounted Cumulated Gain

- Dokładność i kompletność pozwalają jedynie na binarną ocenę istotności
- Nie ma rozróżnienia między bardzo i średnio ważnymi dokumentami
- Zredukowane skumulowane wzmocnienie DCG jest miarą uwzględniającą stopień istotności dokumentów
  - Bardzo istotne dokumenty są preferowane na początku rankingu,
  - Istotne dokumenty na końcu rankingu są mniej ważne
- Dla kilkunastu dokumentów $d[i]$ odpowiedzi na zapytanie q_j tworzy się wektor wzmocnienia $G_j[i]$ odpowiadajcy przyjętej skali istotności dokumentów np. 0-3.
- Wzmocnienie 0 mają dokumenty nieistotne.

DCG — Discounted Cumulated Gain

- Wektor skumulowany CG_j odpowiada sumowaniu wzmocnień w kolejności rankingu:

$$CG_j[i] = \begin{cases} G_j[1] & \texttt{if } i = 1; \\ G_j[i] + CG_j[i-1] & \texttt{otherwise} \end{cases}$$

- Współczynnik dyskonta stanowi logarytm pozycji rankingu np. $log_2(i)$,
- Dzieląc elementy wektora skumulowanego $CG_j[i]$ przez odpowiednie współczynniki dyskonta otrzymuje się współczynnik DCG – zdyskontowanego skumulowanego wzmocnienia.

DCG — Discounted Cumulated Gain

$$DCG_j[i] = \begin{cases} G_j[1] & \texttt{if } i = 1; \\ \frac{G_j[i]}{\log_2 i} + DCG_j[i-1] & \texttt{otherwise} \end{cases}$$

- Dla przykładowych zapytań q_1 i q_2 uzyskuje się wektory G_1, G_2 oraz DCG_1 i DCG_2.

$$\begin{aligned} G_1 &= (1, 0, 1, 0, 0, 3, 0, 0, 0, 2, 0, 0, 0, 0, 3) \\ G_2 &= (0, 0, 2, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 3) \end{aligned}$$

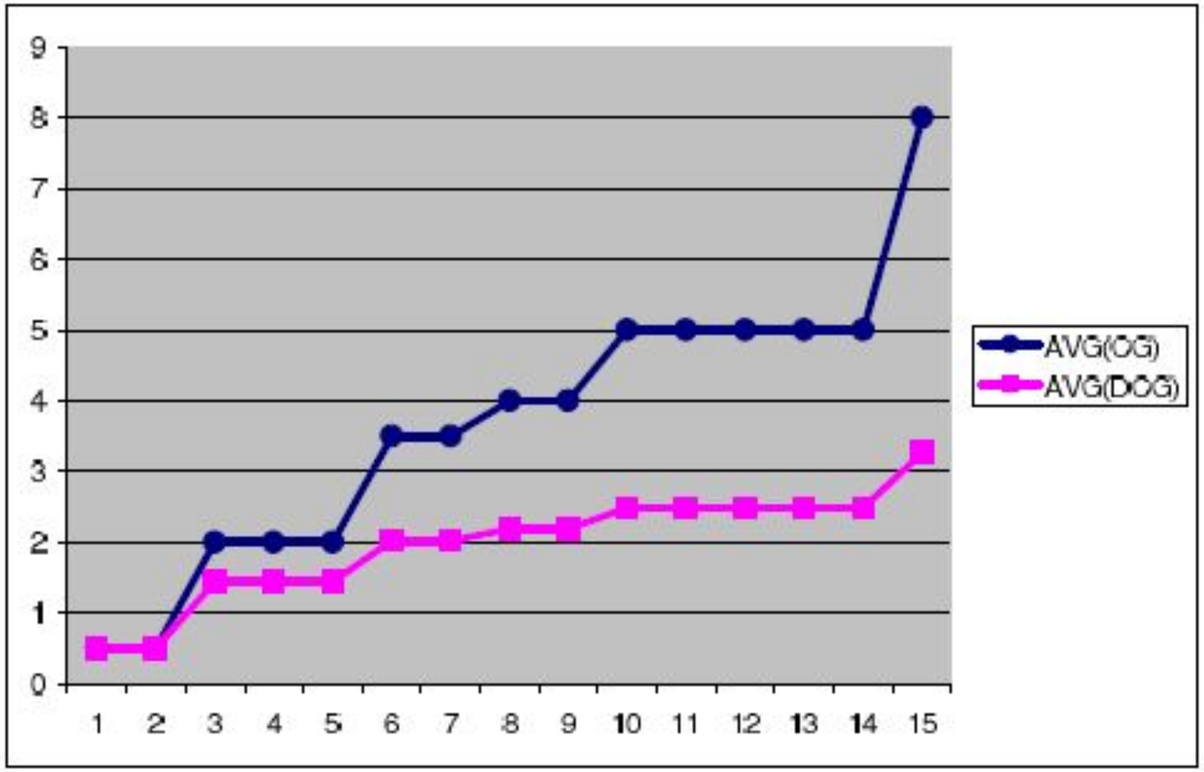
$$\begin{aligned} DCG_1 &= (1.0, 1.0, 1.6, 1.6, 1.6, 2.8, 2.8, 2.8, 2.8, 3.4, 3.4, 3.4, 3.4, 3.4, 4.2) \\ DCG_2 &= (0.0, 0.0, 1.3, 1.3, 1.3, 1.3, 1.3, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 2.4) \end{aligned}$$

- Dyskontowane skumulowane wzmocnienia są znacznie mniej wrażliwe na występowanie istotnych dokumentów pod koniec rankingu

DCG — Discounted Cumulated Gain

- Funkcje $\overline{CG}[i]$ i $\overline{DCG}[i]$ uśrednione w zbiorze N_q zapytań:

$$\overline{CG}[i] = \sum_{j=1}^{N_q} \frac{CG_j[i]}{N_q}; \qquad \overline{DCG}[i] = \sum_{j=1}^{N_q} \frac{DCG_j[i]}{N_q}$$



DCG -znormalizowane

- Dokładność i kompletność są obliczane względem zbioru istotnych dokumentów
- *CG* i *DCG* nie mają odniesienia do żadnej konkretnej bazy, co może utrudniać użycie ich do porównania metod wyszukiwania
- Niech *ICG* i *IDCG* będą idealnymi odpowiednikami *CG* i *DCG* - po wysortowaniu istotności dokumentów w porządku malejącym
- Wprowadza się znormalizowane wartości *CG* i *DCG* :

$$\boxed{NCG[i] = \frac{\overline{CG[i]}}{ICG[i]}; \qquad NDCG[i] = \frac{\overline{DCG[i]}}{IDCG[i]}}$$

DCG -znormalizowane

- Pole pod krzywymi *NCG* i *NDCG* reprezentuje jakość algorytmu wyszukiwania z rankingiem
- Większe pole oznacza lepszy algorytm
- Kumulowane wzmocnienie zapewnia skalarną miarę jakości wyszukiwania na dowolnej pozycji rankingu
- Zdyskontowane wzmocnienie pozwala na kontrolę wpływu ważnych dokumentów pod koniec rankingu na na ocenę jakości algorytmu

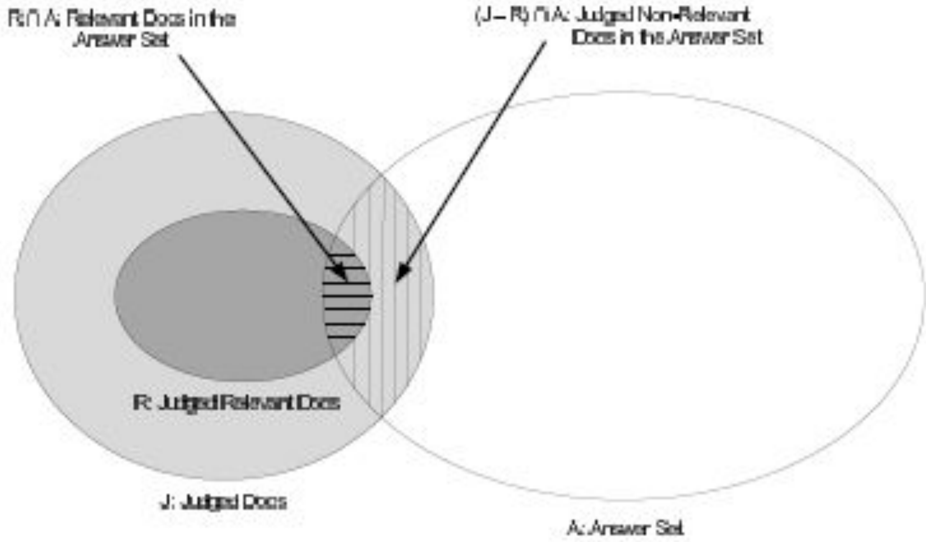
BPREF — preferencje binarne

- Paradygmat Cranfielda zakłada ocenę wszystkich dokumentów testowych ze względu na każde zapytanie
- To może być spełnione tylko dla niewielkich zbiorów
- Dla dużych zbiorów stosuje się metodę grupowania (pooling)
  - Składa ona do puli tylko najtrafniejsze wyniki uzyskane przez różne algorytmy wyszukiwania,
  - Następnie dokumenty z puli są oceniane w celu porównania metod
- Miary typu *Precision-Recall* przyjmują za nieistotne wszystkie dokumenty nie wyszukane
- Jest to nie do przyjęcia dla kolekcji złożonych z miliardów dokumentów

BPREF — preferencje binarne

- Problem rozwiązuje się poprzez ustalenie preferencji między parami wyszukanych dokumentów zamiast bezpośredniego używania ich rangi - jak to ujęto w przypadku BPREF
- <u>BPREF mierzy ilość dokumentów nieistotnych poprzedzających dokumenty istotne</u>
- Ocenia się czy dokument d_j jest preferowany w parze z d_k dla danego zapytania q.
- Ponadto każdy dokument istotny jest preferowany w parze z nieistotnym

BPREF — preferencje binarne



- Dla wymaganej informacji *I*:
  - R_A: ranking wyliczony przez system IR w odniesieniu do *I*,
  - $s_{A,j}$: pozycja dokumentu d_j w R_A,
  - $[(J - R) \wedge A]_{|R|}$: zbiór pierwszych $|R|$ dokumentów w R_A ocenionych jako nieistotne

BPREF — preferencje binarne

- Ilość nieistotnych dokumentów pojawiających się w R_A przed d_j.

$$C(\mathcal{R}_A, d_j) = \| \{d_k \mid d_k \in [(J-R) \cap A]_{|R|} \wedge s_{A,k} < s_{A,j}\} \|$$

- Miara BPREF dla rankingu R_A:

$$\boxed{Bpref(\mathcal{R}_A) = \frac{1}{|R|} \sum_{d_j \in (R \cap A)} \left(1 - \frac{C(\mathcal{R}_A, d_j)}{min(|R|, |(J-R) \cap A|)}\right)}$$

- Dla każdego istotnego dokumentu d_j w rankingu *Bpref* akumuluje wagę zmieniającą się odwrotnie do ilości nieistotnych dokumentów poprzedzających
- Jeśli liczba znanych dokumentów istotnych jest bardzo mała (1 lub 2) stosuje się skorygowaną miarę BPREF-10

BPREF — preferencje binarne

- Zapewnia ona minimum 10 istotnych dokumentów

$$C_{10}(\mathcal{R}_A, d_j) = \left\| \{d_k \mid d_k \in [(J - R) \cap A)]_{|R|+10} \wedge s_{A,k} < s_{A,j}\} \right\|$$

- Wówczas:

$$\boxed{Bpref_{10}(\mathcal{R}_A) = \frac{1}{|R|} \sum_{d_j \in (R \cap A)} \left(1 - \frac{C_{10}(\mathcal{R}_A, d_j)}{min(|R|+10,\ |(J-R) \cap A|)}\right)}$$

Miary korelacji rangowej

- Dokładność i zupełność pozwalają na porównanie istotności wyników uzyskanych przez dwie funkcje rankingowe
- W pewnych sytuacjach:
  - Nie można bezpośrednio określić istotności dokumentów,
  - Należy ustalić jak bardzo dana funkcja rankingu różni się od drugiej znanej
- Wówczas interesuje nas porównanie uporządkowania dwóch rankingów
- Można tego dokonać poprzez funkcje statystyczne zwane miarami korelacji rangowej

Miary korelacji rangowej

- Miary korelacji rankingów R_1 i R_2 można wyrazić przez współczynnik korelacji $C(R_1, R_2)$ o następujących właściwościach
  - $-1 \leq C(R_1, R_2) \leq 1$,
  - Jeżeli $C(R_1, R_2) = 1$ zgodność pomiędzy rankingami jest pełna,
  - Jeżeli $C(R_1, R_2) = -1$ niezgodność pomiędzy rankingami jest pełna, tzn. są one odwrócone względem siebie,
  - Jeżeli $C(R_1, R_2) = 0$ dwa rankingi są kompletnie niezależne,
  - Zwiększenie wartości $C(R_1, R_2)$ implikuje zwiększenie zgodności rankingów

Współczynnik Spearman'a

- Współczynnik Spearman'a jest najczęściej używaną miarą korelacji rangowej – bazuje on na różnicach pozycji tego samego dokumentu w różnych rankingach

documents	$s_{1,j}$	$s_{2,j}$	$s_{i,j} - s_{2,j}$	$(s_{1,j} - s_{2,j})^2$
d_{123}	1	2	-1	1
d_{84}	2	3	-1	1
d_{56}	3	1	+2	4
d_{6}	4	5	-1	1
d_{8}	5	4	+1	1
d_{9}	6	7	-1	1
d_{511}	7	8	-1	1
d_{129}	8	10	-2	4
d_{187}	9	6	+3	9
d_{25}	10	9	+1	1
Sum of Square Distances				24

- Dane z 10 dokumentów wyszukanych przez 2 rankingi R_1 i R_2. $s_{1,j}$ and $s_{2,j}$ są pozycjami *j*-tego dokumentu w tych rankingach

Współczynnik Spearman'a

- Przy porządkowaniu *K* dokumentów maksimum sumy kwadratów różnic pozycji rankingu wynosi

$$\frac{K \times (K^2 - 1)}{3}$$

- Przy pełnej niezgodności rankingów dla *K*=10 maksimum to wynosi $(10 \times (10^2 - 1))/3 = 330$
- Współczynnik Spearmana korelacji rang $S(R_1,R_2)$:

$$\boxed{S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^{K} (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}}$$

- zmienia się w zakresie [-1, 1], K – rozmiar porównywanych zbiorów

Współczynnik Spearman'a

- Przykład:

$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times 24}{10 \times (10^2 - 1)} = 1 - \frac{144}{990} = 0.854$$

documents	$s_{1,j}$	$s_{2,j}$	$s_{i,j} - s_{2,j}$	$(s_{1,j} - s_{2,j})^2$
d_{123}	1	2	-1	1
d_{84}	2	3	-1	1
d_{56}	3	1	+2	4
d_{6}	4	5	-1	1
d_{8}	5	4	+1	1
d_{9}	6	7	-1	1
d_{511}	7	8	-1	1
d_{129}	8	10	-2	4
d_{187}	9	6	+3	9
d_{25}	10	9	+1	1
Sum of Square Distances				24

Współczynnik tau Kendall'a

- Posiada on naturalną i intuicyjną interpretację

$$\tau(\mathcal{R}_1, \mathcal{R}_2) = P(\mathcal{R}_1 = \mathcal{R}_2) - P(\mathcal{R}_1 \neq \mathcal{R}_2)$$

- gdzie
  - $P(R_1 = R_2)$ – znormalizowana ilość zgodnych par dokumentów (concordant), dla których różnice w pozycjach rankingu par dokumentów $[d_k, d_j]$: $s_{1,k} - s_{1,j}$ oraz $s_{2,k} - s_{2,j}$ są tego samego znaku dla 2 porównywanych metod
  - $P(R_1 \neq R_2)$ – znormalizowana ilość niezgodnych par dokumentów (discordant), dla których różnice w pozycjach rankingu par dokumentów $[d_k, d_j]$: $s_{1,k} - s_{1,j}$ oraz $s_{2,k} - s_{2,j}$ są różnych znaków
- Jeżeli przyjąć
  - $\Delta(R_1, R_2)$: liczba niezgodnych par dokumentów w rankingach R_1 i R_2,
  - $K(K - 1) - \Delta(R_1, R_2)$: liczba zgodnych par dokumentów,

Współczynnik tau Kendall'a

- Wówczas:

$$P(\mathcal{R}_1 = \mathcal{R}_2) = \frac{K(K-1) - \Delta(\mathcal{R}_1, \mathcal{R}_2)}{K(K-1)}$$

$$P(\mathcal{R}_1 \neq \mathcal{R}_2) = \frac{\Delta(\mathcal{R}_1, \mathcal{R}_2)}{K(K-1)}$$

- co daje:

$$\boxed{\tau(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{2 \times \Delta(\mathcal{R}_1, \mathcal{R}_2)}{K(K-1)}}$$

- Przykład: 5 dokumentów w 2 rankingach

documents	$s_{1,j}$	$s_{2,j}$	$s_{i,j} - s_{2,j}$
d_{123}	1	2	-1
d_{84}	2	3	-1
d_{56}	3	1	+2
d_6	4	5	-1
d_8	5	4	+1

Współczynnik tau Kendall'a

- Uporządkowane pary dokumentów w rankingu R_1

$$
\begin{array}{ll}
[d_{123}, d_{84}],\ [d_{123}, d_{56}],\ [d_{123}, d_{6}],\ [d_{123}, d_{8}], & C, D, C, C, \\
[d_{84}, d_{56}],\ [d_{84}, d_{6}],\ [d_{84}, d_{8}], & D, C, C, \\
[d_{56}, d_{6}],\ [d_{56}, d_{8}], & C, C, \\
[d_{6}, d_{8}] & D
\end{array}
$$

- Uporządkowane pary dokumentów w rankingu R_2:

$$
\begin{array}{ll}
[d_{56}, d_{123}],\ [d_{56}, d_{84}],\ [d_{56}, d_{8}],\ [d_{56}, d_{6}], & D, D, C, C, \\
[d_{123}, d_{84}],\ [d_{123}, d_{8}],\ [d_{123}, d_{6}], & C, C, C, \\
[d_{84}, d_{8}],\ [d_{84}, d_{6}], & C, C, \\
[d_{8}, d_{6}] & D
\end{array}
$$

- *Dla K*=5 dokumentów *K(K-1)*=20 *oraz*

$$
\begin{aligned}
\tau(\mathcal{R}_1, \mathcal{R}_2) &= \frac{14}{20} - \frac{6}{20} \\
&= 0.4
\end{aligned}
$$