

Hybrydowa metoda rekomendacji dokumentów w środowisku hipertekstowym

Paweł Szoltysek

09 listopada 2009

Agenda

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja algorytmu

Definicja metod składowych

Opis

Definicja

Praktyka

O pracy

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

- ✓ 199 stron, 2 cytowania własne
- ✓ 7 rozdziałów
- ✓ Promotor: NT Nguyen

O pracy

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

✓ 199 stron, 2 cytowania własne

✓ 7 rozdziałów

✓ Promotor: NT Nguyen

O pracy

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

- ✓ 199 stron, 2 cytowania własne
- ✓ 7 rozdziałów
- ✓ Promotor: NT Nguyen

Wstęp

System rekomendujący

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

- ✓ Co to jest system rekomendujący?
System, który sugeruje użytkownikowi potencjalnie interesujące go elementy.
- ✓ Jaki problem rozwiązuje system rekomendujący?
- ✓ Do czego możemy wykorzystać systemy rekomendujące?

Wstęp

System rekomendujący

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

- ✓ Co to jest system rekomendujący?
- ✓ Jaki problem rozwiązuje system rekomendujący?
Na przykład: problem nieuporządkowania informacji w sieci Internet.
- ✓ Do czego możemy wykorzystać systemy rekomendujące?

Wstęp

System rekomendujący

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

- ✓ Co to jest system rekomendujący?
- ✓ Jaki problem rozwiązuje system rekomendujący?
- ✓ Do czego możemy wykorzystać systemy rekomendujące?
Nie tylko hipertekst - biblionetka, lastfm, pandora, wovd.

Wstęp

Wyszukiwarka a system rekomendujący

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Wstęp

Wyszukiwarka a system rekomendujący

- Agenda
- O pracy
- Wstęp**
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

podobieństwa	różnice
wspólny cel	sposób prezentacji wyników
wykorzystanie rankingu	sposób charakteryzowania wyników
feedback użytkownika	ingerencja użytkownika

Wstęp

Wyszukiwanie informacji a wyszukiwarki (1)

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

Niechęć użytkownika do zadawania pytań wynika m.in. z:

- ✓ braku umiejętności w posługiwaniu się wyszukiwarką;
- ✓ lenistwa użytkownika;
- ✓ trudności ujęcia swojej potrzeby w słowach kluczowych;
- ✓ niewiedzy dotyczącej istnienia w systemie elementów, które odpowiadają potrzebom.

Wstęp

Wyszukiwanie informacji a wyszukiwarki (1)

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

Niechęć użytkownika do zadawania pytań wynika m.in. z:

- ✓ braku umiejętności w posługiwaniu się wyszukiwarką;
- ✓ lenistwa użytkownika;
- ✓ trudności ujęcia swojej potrzeby w słowach kluczowych;
- ✓ niewiedzy dotyczącej istnienia w systemie elementów, które odpowiadają potrzebom.

Wstęp

Wyszukiwanie informacji a wyszukiwarki (1)

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

Niechęć użytkownika do zadawania pytań wynika m.in. z:

- ✓ braku umiejętności w posługiwaniu się wyszukiwarką;
- ✓ lenistwa użytkownika;
- ✓ trudności ujęcia swojej potrzeby w słowach kluczowych;
- ✓ niewiedzy dotyczącej istnienia w systemie elementów, które odpowiadają potrzebom.

Wstęp

Wyszukiwanie informacji a wyszukiwarki (1)

- Agenda
- O pracy
- Wstęp**
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

Niechęć użytkownika do zadawania pytań wynika m.in. z:

- ✓ braku umiejętności w posługiwaniu się wyszukiwarką;
- ✓ lenistwa użytkownika;
- ✓ trudności ujęcia swojej potrzeby w słowach kluczowych;
- ✓ niewiedzy dotyczącej istnienia w systemie elementów, które odpowiadają potrzebom.

Wstęp

Wyszukiwanie informacji a wyszukiwarki (1)

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

Niechęć użytkownika do zadawania pytań wynika m.in. z:

- ✓ braku umiejętności w posługiwaniu się wyszukiwarką;
- ✓ lenistwa użytkownika;
- ✓ trudności ujęcia swojej potrzeby w słowach kluczowych;
- ✓ niewiedzy dotyczącej istnienia w systemie elementów, które odpowiadają potrzebom.

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

System rekomendujący nie tylko ułatwia dotarcie do szukanej informacji, ale także pełni funkcję informacyjną.

- ✓ Rekomendując - powiadamiamy użytkownika o posiadaniu informacji
- ✓ Rekomendując - uzupełniamy wiedzę użytkownika

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

System rekomendujący nie tylko ułatwia dotarcie do szukanej informacji, ale także pełni funkcję informacyjną.

- ✓ Rekomendując - powiadamiamy użytkownika o posiadaniu informacji
- ✓ Rekomendując - uzupełniamy wiedzę użytkownika

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

System rekomendujący nie tylko ułatwia dotarcie do szukanej informacji, ale także pełni funkcję informacyjną.

- ✓ Rekomendując - powiadamiamy użytkownika o posiadaniu informacji
- ✓ Rekomendując - uzupełniamy wiedzę użytkownika

System rekomendujący nie tylko ułatwia dotarcie do szukanej informacji, ale także pełni funkcję informacyjną.

- ✓ Rekomendując - powiadamiamy użytkownika o posiadaniu informacji
- ✓ Rekomendując - uzupełniamy wiedzę użytkownika

Ze względu na podobieństwa, można wykorzystywać te same miary - dokładności i kompletności

Efektywność metod rekomendacji

Czym jest "dobra" metoda rekomendacji?

Dobra metoda rekomendacji nie charakteryzuje się tylko proponowaniem dokumentów relewantnych.

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Efektywność metod rekomendacji

Dokładność i kompletność

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

R_p - wszystkie dokumenty w systemie

D_p - wyszukane dokumenty w systemie

Dokładność (ang. precision) - stosunek liczby wyszukanych dokumentów relewantnych do liczby wszystkich wyszukanych dokumentów

$$prec(p) = \frac{card(R_p \cap D_p)}{card(R_p)} \quad (1)$$

Kompletność (ang. recall) - stosunek liczby wyszukanych dokumentów relewantnych do liczby relewantnych dokumentów znajdujących się w systemie

$$rec(p) = \frac{card(R_p \cap D_p)}{card(D_p)} \quad (2)$$

Efektywność metod rekomendacji

Określenie efektywności rekomendacji

Agenda
O pracy
Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

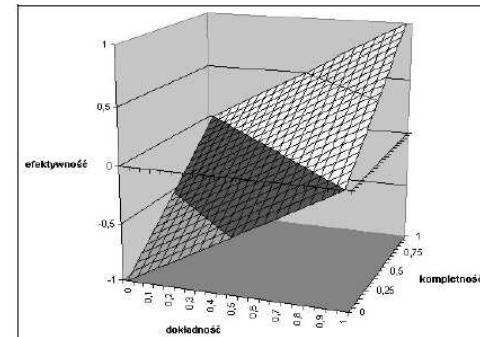
Na efektywność rekomendacji składać się będzie dokładność (*prec*) i kompletność (*rec*).

Borko: $rec + prec - 1$

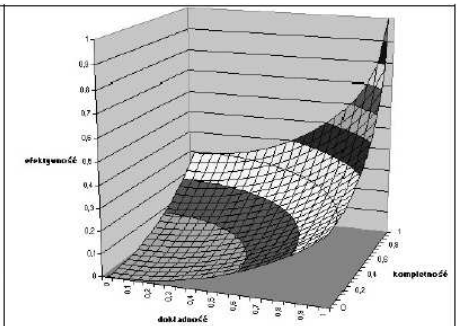
Meadowa: $1 - \frac{\sqrt{(1-prec^2)+(1-rec^2)}}{\sqrt{2}}$

Heinego: $1 - \frac{1}{\frac{1}{rec} + \frac{1}{prec} - 1}$

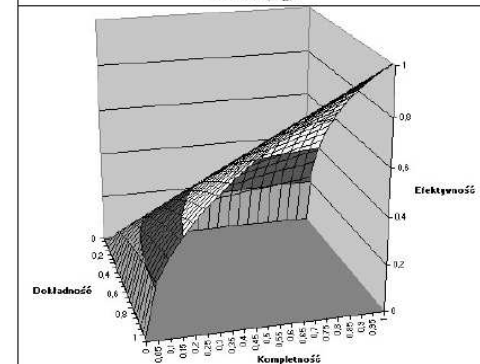
Voiskunski: $\sqrt{prec * rec}$



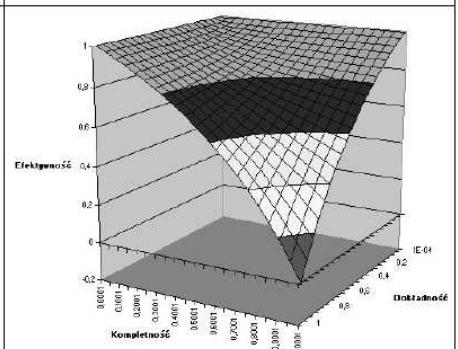
Rysunek 2.2.1a Wykres miary wprowadzonej przez Borko (E_B)



Rysunek 2.2.1b Wykres miary wprowadzonej przez Meadów (E_M)



Rysunek 2.2.1c Wykres miary, którą wprowadził Voiskunskii (E_V)



Rysunek 2.2.1d Wykres miary, którą wprowadził Heine (E_H)

Efektywność metod rekomendacji

Określenie efektywności rekomendacji

Agenda
O pracy
Wstęp

Efektywność

Klasyfikacja
Propozycja
algorytmu
Definicja metod
składowych
Opis
Definicja
Praktyka

Jak osiągnąć wysoki wskaźnik kompletności w systemach rekomendujących?

Jak zinterpretować zignorowanie dokumentu przez użytkownika?
A co, jeśli użytkownik w międzyczasie zmieni swoje zainteresowania?

Efektywność metod rekomendacji

Określenie efektywności rekomendacji

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Jak osiągnąć wysoki wskaźnik kompletności w systemach rekomendujących?

Jak zinterpretować zignorowanie dokumentu przez użytkownika?

A co, jeśli użytkownik w międzyczasie zmieni swoje zainteresowania?

Efektywność metod rekomendacji

Określenie efektywności rekomendacji

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Jak osiągnąć wysoki wskaźnik kompletności w systemach rekomendujących?

Jak zinterpretować zignorowanie dokumentu przez użytkownika?

A co, jeśli użytkownik w międzyczasie zmieni swoje zainteresowania?

Klasyfikacja systemów rekomendujących

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

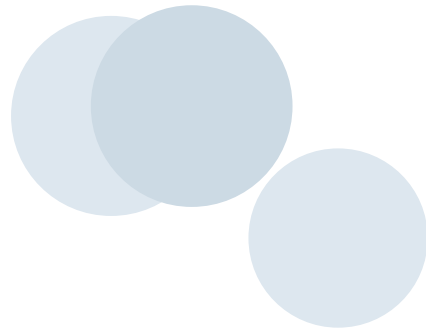
składowych

Opis

Definicja

Praktyka

- ✓ Ogólne cechy systemu rekomendacji
- ✓ Zarządzanie profilem
- ✓ Architektura



Klasyfikacja systemów rekomendujących

Ogólne cechy systemu rekomendacji

- ✓ Dziedzina
- ✓ Technika rekomendacji
- ✓ Zastosowane algorytmu uczące się
- ✓ Sposób pobierania danych
- ✓ Różnorodność rekomendacji

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja**
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

Klasyfikacja systemów rekomendujących

Zarządzanie profilem

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja**
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

- ✓ Reprezentacja profilu użytkownika
- ✓ Profil początkowy
- ✓ Adaptacja profilu
- ✓ Prywatność

Klasyfikacja systemów rekomendujących

Architektura

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

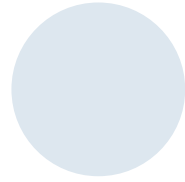
Definicja

Praktyka

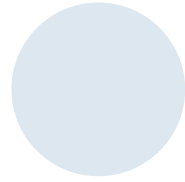
- ✓ Czasową dystrybucję obliczeń
- ✓ Miejsce instalacji
- ✓ Prezentację rekomendacji

Propozycja algorytmu

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu**
- Definicja metod składowych
- Opis
- Definicja
- Praktyka



Propozycja algorytmu

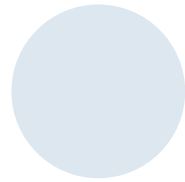


Metoda słów kluczowych



- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu**
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

Propozycja algorytmu



Metoda słów kluczowych + metoda wspólnych zachowań



- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu**
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

Propozycja algorytmu

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu**
- Definicja metod składowych
- Opis
- Definicja
- Praktyka



Metoda słów kluczowych + metoda wspólnych zachowań =
AdaptRank

Następne seminarium

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

- ✓ Opis metod składowych
- ✓ Definicja i opis hybrydowej metody rekomendacji
- ✓ Algorytm rekomendacji
- ✓ Eksperymentalna weryfikacja metody
- ✓ Podsumowanie



Hybrydowa metoda rekomendacji dokumentów w środowisku hipertekstowym

Paweł Szoltysek



15 listopada 2009

Definicja systemu rekomendującego

System rekomendujący oparty na hipertekście opisujemy trójką $SR = \langle D, S, f \rangle$, gdzie:

- ✓ D - zbiór dokumentów
- ✓ S - zbiór sesji historycznych (dwójek (s, d_s) , gdzie $s = (s_1, \dots, s_n)$)
- ✓ f - funkcja rekomendacji ($D \times S \rightarrow \mathfrak{R}$)

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Definicja metod składowych

Metoda rekomendacji oparta na podobieństwie tekstowym

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe podobne do dokumentów odwiedzonych przez użytkownika.

Dla każdego dokumentu oblicza się wektor deskryptorów $v^d = (v_1^d, \dots, v_{n_w}^d)$, gdzie v_j^d określa poziom istotności deskryptora w dokumencie.

Formuła Saltona: $v_j^d = w f_j^d \log(\frac{n}{n_j})$.

Podobieństwo tekstowe:

$$sim(d_j, d_k) = \cos(v^{d_j}, v^{d_k})$$

Funkcja rekomendacji:

$$f_t(d_j, s_b^{(i)}) = \frac{\sum_{k=1}^i sim(v^{d_j}, v^{d^{(k)}})}{i}$$

Definicja metod składowych

Metoda rekomendacji oparta na podobieństwie tekstowym

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja
algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Rekomenduje dokumenty hipertekstowe podobne do dokumentów odwiedzonych przez użytkownika.

Dla każdego dokumentu oblicza się wektor deskryptorów $v^d = (v_1^d, \dots, v_{n_w}^d)$, gdzie v_j^d określa poziom istotności deskryptora w dokumencie.

Formuła Saltona: $v_j^d = w f_j^d \log(\frac{n}{n_j})$.

Podobieństwo tekstowe:

$$sim(d_j, d_k) = \cos(v^{d_j}, v^{d_k})$$

Funkcja rekomendacji:

$$f_t(d_j, s_b^{(i)}) = \frac{\sum_{k=1}^i sim(v^{d_j}, v^{d^{(k)}})}{i}$$

Definicja metod składowych

Metoda rekomendacji oparta na podobieństwie tekstowym

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe podobne do dokumentów odwiedzonych przez użytkownika.

Dla każdego dokumentu oblicza się wektor deskryptorów $v^d = (v_1^d, \dots, v_{n_w}^d)$, gdzie v_j^d określa poziom istotności deskryptora w dokumencie.

Formuła Saltona: $v_j^d = w f_j^d \log\left(\frac{n}{n_j}\right)$.

Podobieństwo tekstowe:

$$\text{sim}(d_j, d_k) = \cos(v^{d_j}, v^{d_k})$$

Funkcja rekomendacji:

$$f_t(d_j, s_b^{(i)}) = \frac{\sum_{k=1}^i \text{sim}(v^{d_j}, v^{d^{(k)}})}{i}$$

Definicja metod składowych

Metoda rekomendacji oparta na podobieństwie tekstowym

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Rekomenduje dokumenty hipertekstowe podobne do dokumentów odwiedzonych przez użytkownika.

Dla każdego dokumentu oblicza się wektor deskryptorów $v^d = (v_1^d, \dots, v_{n_w}^d)$, gdzie v_j^d określa poziom istotności deskryptora w dokumencie.

Formuła Saltona: $v_j^d = w f_j^d \log(\frac{n}{n_j})$.

Podobieństwo tekstowe:

$$sim(d_j, d_k) = \cos(v^{d_j}, v^{d_k})$$

Funkcja rekomendacji:

$$f_t(d_j, s_b^{(i)}) = \frac{\sum_{k=1}^i sim(v^{d_j}, v^{d^{(k)}})}{i}$$

Definicja metod składowych

Metoda rekomendacji oparta na podobieństwie tekstowym

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe podobne do dokumentów odwiedzonych przez użytkownika.

Dla każdego dokumentu oblicza się wektor deskryptorów $v^d = (v_1^d, \dots, v_{n_w}^d)$, gdzie v_j^d określa poziom istotności deskryptora w dokumencie.

Formuła Saltona: $v_j^d = w f_j^d \log(\frac{n}{n_j})$.

Podobieństwo tekstowe:

$$sim(d_j, d_k) = \cos(v^{d_j}, v^{d_k})$$

Funkcja rekomendacji:

$$f_t(d_j, s_b^{(i)}) = \frac{\sum_{k=1}^i sim(v^{d_j}, v^{d^{(k)}})}{i}$$

Definicja metod składowych

Metoda rekomendacji oparta na zachowaniach użytkowników

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe w oparciu o wzorce nawigacyjne.

Wyznaczanie wzorców nawigacyjnych odbywa się przez klasteryzację, kryterium - cosinus kąta między parami wektorów historycznych sesji.

$$S = \bigcup_{j=0}^{n_c} C_j \quad C_0 \cap C_1 \cap \dots \cap C_{n_c} = \emptyset$$

Wzorzec nawigacyjny to wektor średni $c_j = (c_{j1}, \dots, c_{jn})$;

$$c_{jk} = \frac{\sum_{s \in C_j} s_k}{\text{card}(C_j)}$$

Funkcja rekomendacji:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(c_1, z^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(c_{n_c}, z^{(i)}))$$

Definicja metod składowych

Metoda rekomendacji oparta na zachowaniach użytkowników

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Rekomenduje dokumenty hipertekstowe w oparciu o wzorce nawigacyjne.

Wyznaczanie wzorców nawigacyjnych odbywa się przez klasteryzację, kryterium - cosinus kąta między parami wektorów historycznych sesji.

$$S = \bigcup_{j=0}^{n_c} C_j \quad C_0 \cap C_1 \cap \dots \cap C_{n_c} = \emptyset$$

Wzorzec nawigacyjny to wektor średni $c_j = (c_{j1}, \dots, c_{jn})$;

$$c_{jk} = \frac{\sum_{s \in C_j} s_k}{\text{card}(C_j)}$$

Funkcja rekomendacji:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(c_1, z^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(c_{n_c}, z^{(i)}))$$

Definicja metod składowych

Metoda rekomendacji oparta na zachowaniach użytkowników

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe w oparciu o wzorce nawigacyjne.

Wyznaczanie wzorców nawigacyjnych odbywa się przez klasteryzację, kryterium - cosinus kąta między parami wektorów historycznych sesji.

$$S = \bigcup_{j=0}^{n_c} C_j \quad C_0 \cap C_1 \cap \dots \cap C_{n_c} = \emptyset$$

Wzorzec nawigacyjny to wektor średni $c_j = (c_{j1}, \dots, c_{jn})$;

$$c_{jk} = \frac{\sum_{s \in C_j} s_k}{\text{card}(C_j)}$$

Funkcja rekomendacji:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(c_1, z^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(c_{n_c}, z^{(i)}))$$

Definicja metod składowych

Metoda rekomendacji oparta na zachowaniach użytkowników

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe w oparciu o wzorce nawigacyjne.

Wyznaczanie wzorców nawigacyjnych odbywa się przez klasteryzację, kryterium - cosinus kąta między parami wektorów historycznych sesji.

$$S = \bigcup_{j=0}^{n_c} C_j \quad C_0 \cap C_1 \cap \dots \cap C_{n_c} = \emptyset$$

Wzorzec nawigacyjny to wektor średni $c_j = (c_{j1}, \dots, c_{jn})$;

$$c_{jk} = \frac{\sum_{s \in C_j} s_k}{\text{card}(C_j)}$$

Funkcja rekomendacji:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(c_1, z^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(c_{n_c}, z^{(i)}))$$

Definicja metod składowych

Metoda rekomendacji oparta na zachowaniach użytkowników

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych**
- Opis
- Definicja
- Praktyka

Rekomenduje dokumenty hipertekstowe w oparciu o wzorce nawigacyjne.

Wyznaczanie wzorców nawigacyjnych odbywa się przez klasteryzację, kryterium - cosinus kąta między parami wektorów historycznych sesji.

$$S = \bigcup_{j=0}^{n_c} C_j \quad C_0 \cap C_1 \cap \dots \cap C_{n_c} = \emptyset$$

Wzorzec nawigacyjny to wektor średni $c_j = (c_{j1}, \dots, c_{jn})$;

$$c_{jk} = \frac{\sum_{s \in C_j} s_k}{\text{card}(C_j)}$$

Funkcja rekomendacji:

$$f_c(d_j, s_b^{(i)}) = \max(c_{1j} \cdot \text{sim}(c_1, z^{(i)}), \dots, c_{n_c j} \cdot \text{sim}(c_{n_c}, z^{(i)}))$$

Korzyści z korzystania z różnych źródeł danych

- ✓ Wyższa jakość rekomendacji
- ✓ Nadspecjalizacja i uogólnianie
- ✓ Niedostępność danych w poszczególnych źródłach

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod
składowych

Opis

Definicja

Praktyka

Opis hybrydowej metody rekomendacji

Cechy systemu AdaptRank

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis**
- Definicja
- Praktyka

Ogólne cechy systemu rekomendacji:

- ✓ Dziedzina: Dokumenty hipertekstowe
- ✓ Technika rekomendacji: Hybrydowa
- ✓ Zastosowane algorytmu uczące się: Grupowanie
- ✓ Sposób pobierania danych: Pośredni (obserwacja żądanych dokumentów + rekomendacja zignorowanych)
- ✓ Różnorodność rekomendacji: zmniejszenie relewancji rekomendacji

Opis hybrydowej metody rekomendacji

Cechy systemu AdaptRank

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis**
- Definicja
- Praktyka

Zarządzanie profilem:

- ✓ Reprezentacja profilu użytkownika: Model wektorowy
- ✓ Profil początkowy: Przydatność dokumentu określana za pomocą funkcji
- ✓ Adaptacja profilu: Stopniowe zapominanie
- ✓ Prywatność: Anonimowy identyfikator sesji

Opis hybrydowej metody rekomendacji

Cechy systemu AdaptRank

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis**
- Definicja
- Praktyka

Architektura:

- ✓ Czasowa dystrybucja obliczeń: Częściowa
- ✓ Miejsce instalacji: Zależne od implementacji (serwer WWW)
- ✓ Prezentacja rekomendacji: Zależna od implementacji (krótkie streszczenie + link)

Opis hybrydowej metody rekomendacji

Funkcja rekomendacji

Agenda
O pracy
Wstęp
Efektywność
Klasyfikacja
Propozycja
algorytmu
Definicja metod
składowych

Opis

Definicja

Praktyka

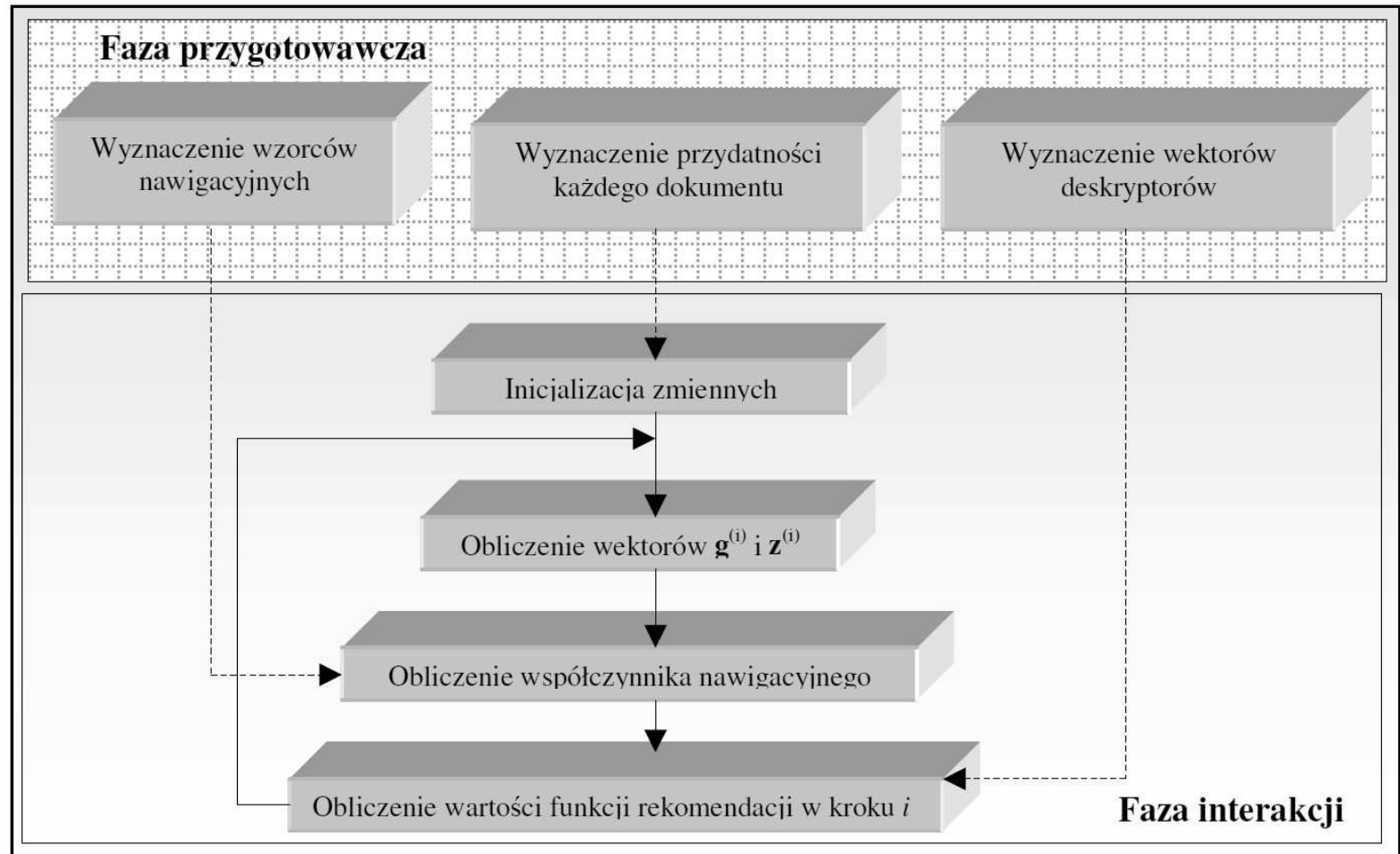
W opisywanym systemie na funkcję rekomendacji składać się będzie:

- ✓ Podobieństwo tekstowe dokumentu do dokumentów odwiedzonych w sesji
- ✓ Liczba odwiedzin dokumentu w historycznych sesjach podobnych do aktualnej
- ✓ Przydatność dokumentu (kombinacja liniowa popularności i jakości dokumentu)
- ✓ Fakt ignoracji dokumentu
- ✓ Fakt odwiedzenia dokumentu

Opis hybrydowej metody rekomendacji

Podział algorytmu na dwie fazy pracy

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis**
- Definicja
- Praktyka



Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Przydatność dokumentów jako bezkontekstowa miara jakości.
Współczynniki użytkowania - otwarcia i liczby odwiedzin
Współczynniki jakości - dostępność, nowość, liczba odsyłaczy

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Przydatność dokumentów jako bezkontekstowa miara jakości.
Współczynniki użytkowania - otwarcia i liczby odwiedzin
Współczynniki jakości - dostępność, nowość, liczba odsyłaczy

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Przydatność dokumentów jako bezkontekstowa miara jakości.
Współczynniki użytkowania - otwarcia i liczby odwiedzin
Współczynniki jakości - dostępność, nowość, liczba odsyłaczy

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

Współczynnik otwarcia

Dokumenty, które są otwierane w sesji jako pierwsze.
Stanowią "interfejs" między systemem a światem zewnętrznym.

- ✓ Identyfikator wpisany z pamięci
- ✓ Identyfikator wybrany z innego systemu hipertekstowego
- ✓ Identyfikator odnaleziony w systemie wyszukiwania

$$or(d) = \begin{cases} \frac{card(S_1^d)}{card(S^d)} & \text{dla } card(S^d) > 0 \\ 0 & \text{dla } card(S^d) = 0 \end{cases}$$

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

Współczynnik liczby odwiedzin

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Liczba odwiedzin to podstawowa wielkość wykorzystywana w statystykach aktywności.

Problem: ciągłe wzmacnianie dokumentów, które się cieszą wysoką popularnością.

$$tr(d) = \begin{cases} \frac{card(U^d)}{t_c - t_0^d} & \text{dla } t_c - t_0^d > 0 \\ 0 & \text{dla } t_c - t_0^d = 0 \end{cases}$$

Potrzebna normalizacja funkcji.

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

Współczynnik dostępności

Dokument, który nie jest zawsze dostępny, nie powinien być często rekomendowany.

$$av(d) = 1 - \frac{t_u^d}{t_c - t_0^d}$$

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

Współczynnik nowości

Nowe dokumenty nie posiadają wysokiego miejsca w rankingach wyszukiwania (mała ilość linków, mała ilość odwiedzin).

$$nr(d) = \begin{cases} \rho_t^{day(t_c) - day(t_0^d)} & \text{dla } day(t_c) - day(t_0^d) > \epsilon \\ 0 & \text{dla } day(t_c) - day(t_0^d) \leq \epsilon \end{cases}$$

gdzie $\rho_t \in (0, 1)$

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Określanie przydatności dokumentów

Współczynnik liczby działających odsyłaczy

$$gl(d) = \begin{cases} \frac{card(L_g^d)}{card(L^d)} & \text{dla } card(L^d) > 0 \\ 0 & \text{dla } card(L^d) = 0 \end{cases}$$

Agenda

O pracy

Wstęp

Efektywność

Klasyfikacja

Propozycja

algorytmu

Definicja metod

składowych

Opis

Definicja

Praktyka

Definicja hybrydowej metody rekomendacji

Definicja funkcji przydatności

Funkcja przydatności dokumentu d to znormalizowana kombinacja liniowa przedstawionych współczynników.

$$q(d) = \frac{\mu \cdot \max(nr(d), wtr(d)) + \lambda \cdot or(d) + \phi \cdot gl(d) + \chi \cdot av(d)}{\mu + \lambda + \phi + \chi}$$

gdzie $\mu + \lambda + \phi + \chi > 0$ oraz $\mu, \lambda, \phi, \chi \in [0, 1]$

Wartości parametrów są uzależnione od charakterystyki systemu.

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Wektor dokumentów odwiedzonych

Służy do formalnego zapisu aktualnych preferencji ujawnionych w bieżącej sesji.

$$z_j^{(i+1)} = \begin{cases} \rho_s \cdot z_j^{(i)} & \text{dla } d^{(i)} \neq d_j \\ 1 & \text{dla } d^{(i)} = d_j \end{cases}$$

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Wektor dokumentów zignorowanych

Określa jak dawno dokument d_j został zignorowany w bieżącej sesji.

$$g_j^{(i+1)} = \begin{cases} 1 - \alpha & \text{dla } d^{(i)} \neq d_j \wedge g_j^{(i)} = 1 \\ \gamma \cdot g_j^{(i)} & \text{dla } d^{(i)} \neq d_j \wedge g_j^{(i)} \neq 1 \\ 1 & \text{dla } d^{(i)} = d_j \end{cases}$$

gdzie $\gamma \in (\alpha, 1)$

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Wektor wzorców nawigacyjnych

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Określa zainteresowanie użytkownika dokumentami na podstawie zachowań innych użytkowników.

$$nv^{(i)} = (nv_1^{(i)}, \dots, nv_n^{(i)})$$

Składowa wektora

$$nv_j^{(i)} = \frac{1}{n_b} \sum_{k=1}^{n_c} b_k \cdot c_{kj}$$

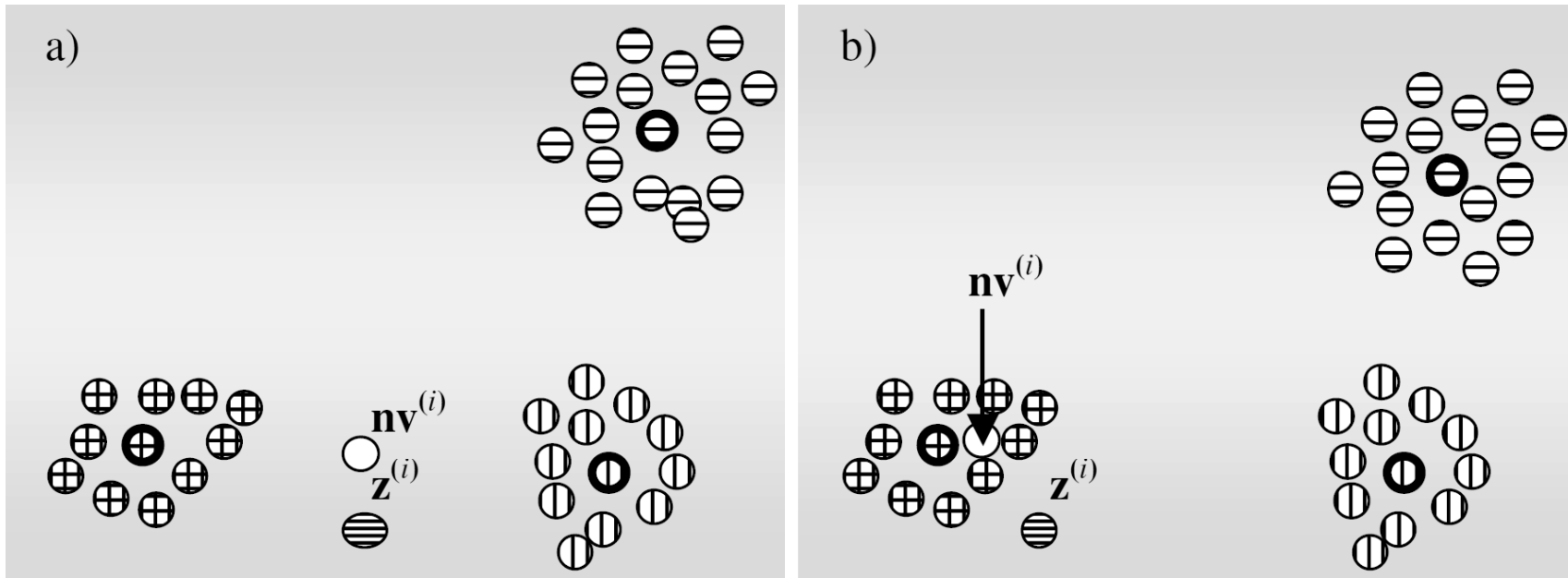
gdzie waga wzorca nawigacyjnego

$$b_k = \begin{cases} sim(z^{(i)}, c_k) & \text{gdy } sim(z^{(i)}, c_k) \geq \epsilon_b \\ 0 & \text{gdy } sim(z^{(i)}, c_k) < \epsilon_b \end{cases}$$

Definicja hybrydowej metody rekomendacji

Wektor wzorców nawigacyjnych

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka



Definicja hybrydowej metody rekomendacji

Funkcja rekomendacji (1)

$$f(d_j, s_b^{(1)}) = (0,5 \cdot (\beta + \delta) \cdot \alpha \cdot q(d_j) + \beta \cdot nv_j^{(i)} + \delta \cdot \text{sim}(v^{d^{(1)}}, v^{d_j})) \cdot (1 - z_j^{(1)}) \quad (3)$$

Gdy użytkownik wysyła żądanie dotyczące pierwszego dokumentu, wartość funkcji jest kombinacją liniową przydatności tego dokumentu, podobieństwa tekstowego i współrzędnej współczynnika wzorca.

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Funkcja rekomendacji (2)

$$f(d_j, s_b^{(i+1)}) = (\alpha \cdot f(d_j, s_b^{(i)}) + \beta \cdot nv_j^{(i+1)} + \delta \cdot \text{sim}(v^{d^{(i+1)}}, v^{d_j}) - g_j^{(i+1)}) \cdot (1 - z_j^{(i+1)}) \quad (4)$$

W kolejnych krokach dodatkowo wartość funkcji zależy od wartości w kroku poprzednim oraz współrzędnej wektora dokumentów zignorowanych.

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Definicja hybrydowej metody rekomendacji

Algorytm rekomendacji

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka

Wektor rekomendacji:

$$r^{(i)} = (r_1^{(i)}, \dots, r_n^{(i)})$$

$$r_j^{(i)} = f(d_j, s_b^{(i)})$$

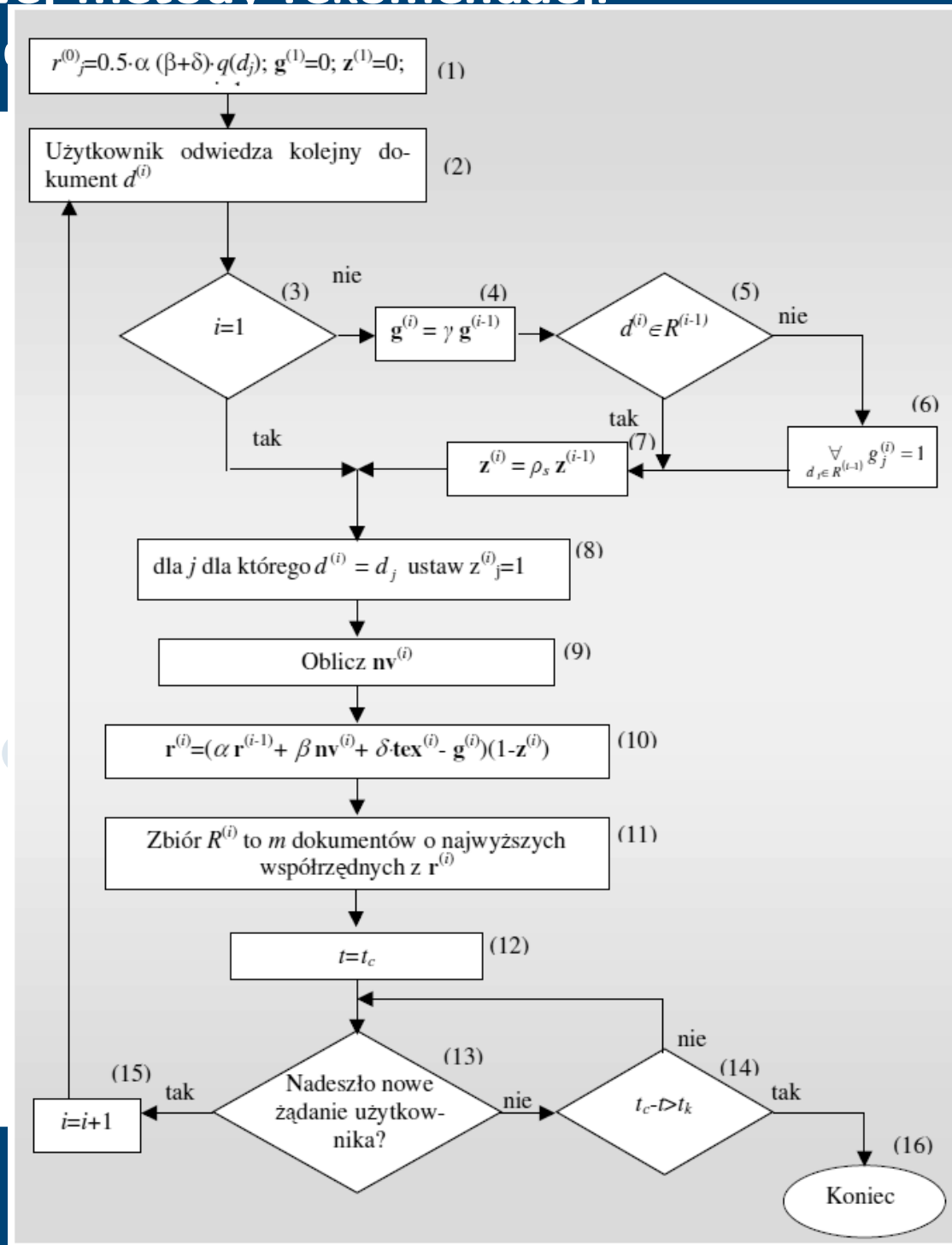
Wektor podobieństwa tekstowego:

$$tex_k^{d_j} = sim(v^{d_j}, v^{d_k})$$

Definicja hybrydowej metody rekomendacji

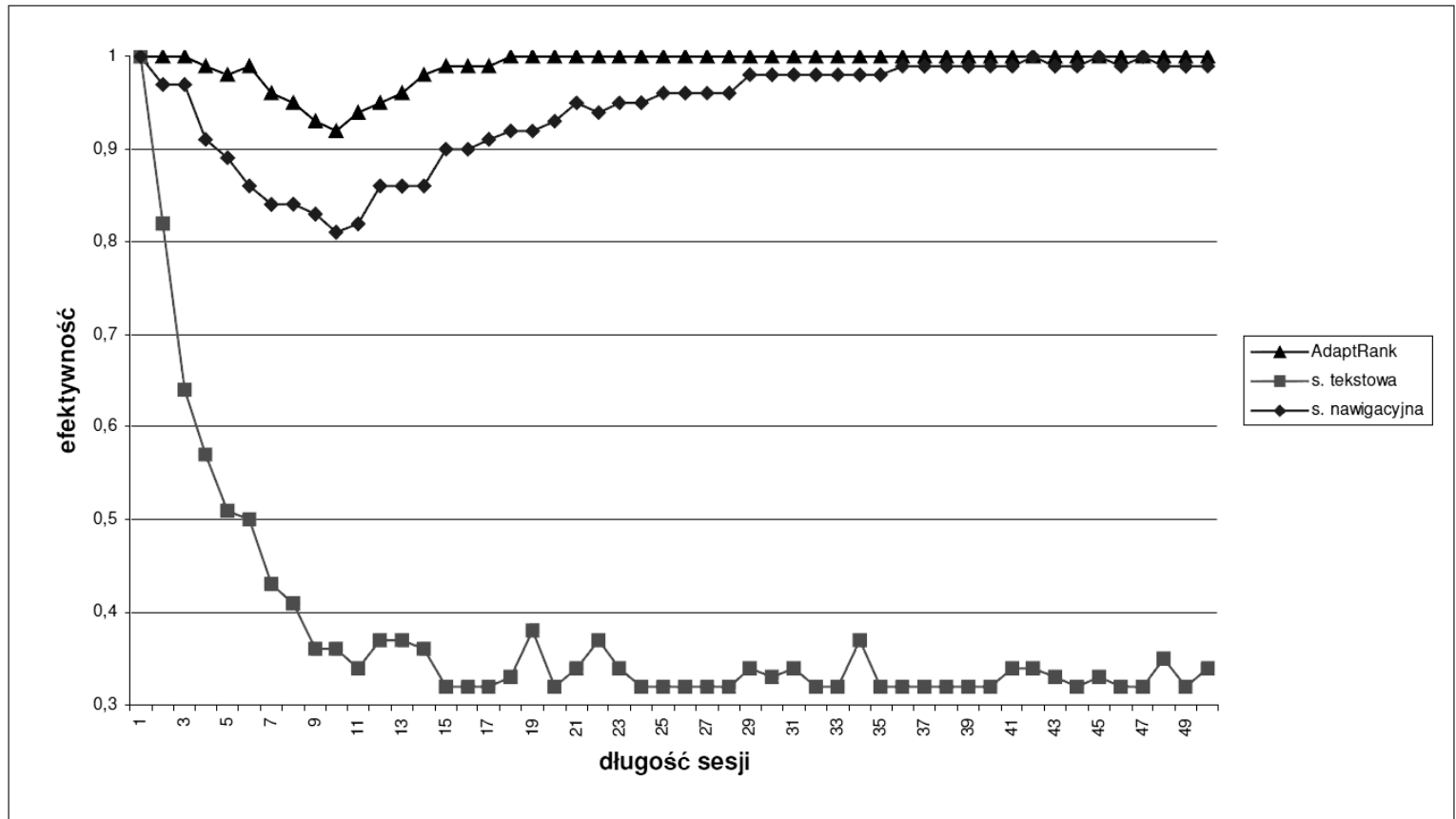
Algorytm rekomendacji

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja**
- Praktyka



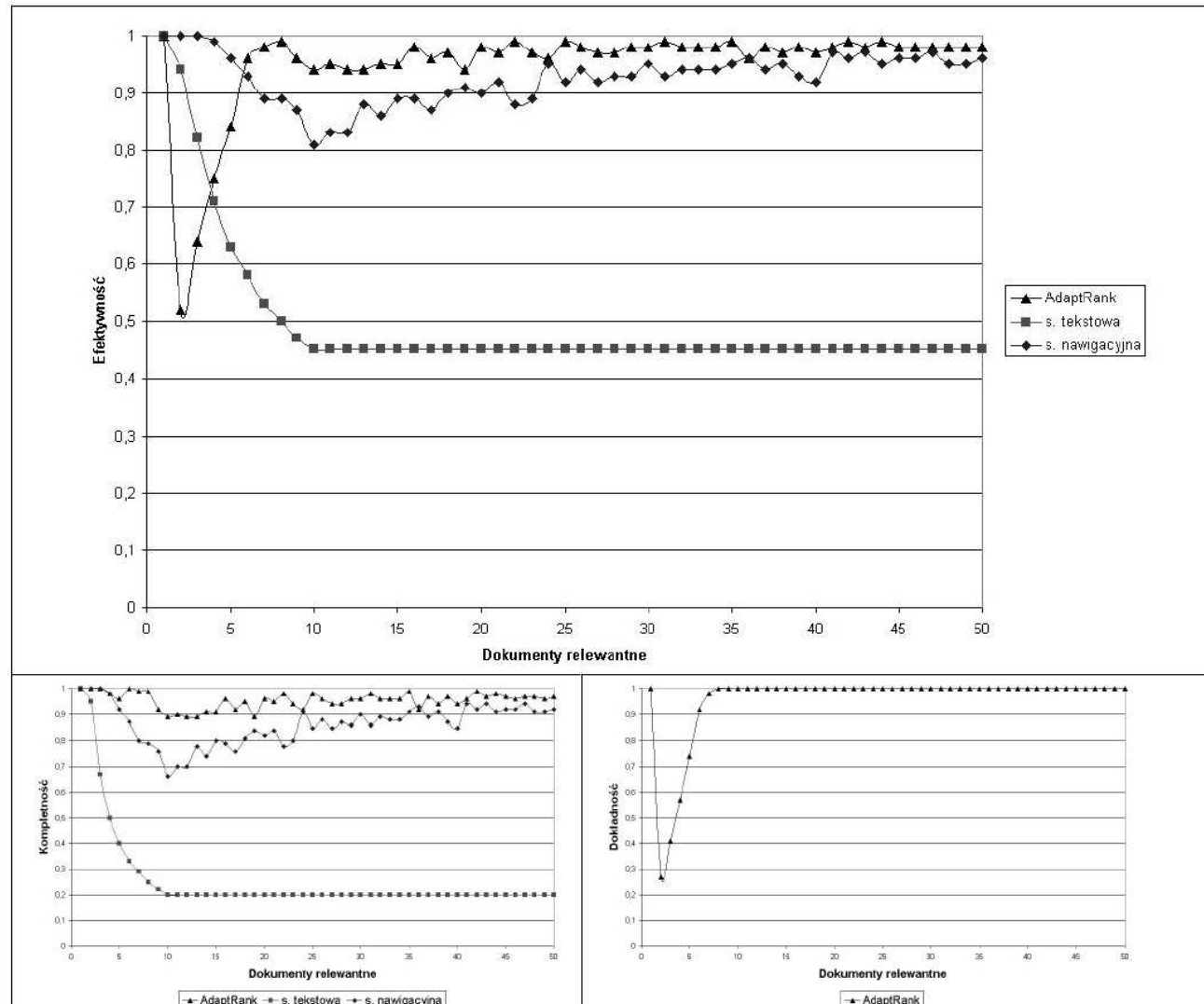
Praktyka - symulacyjne badania efektywności

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka**



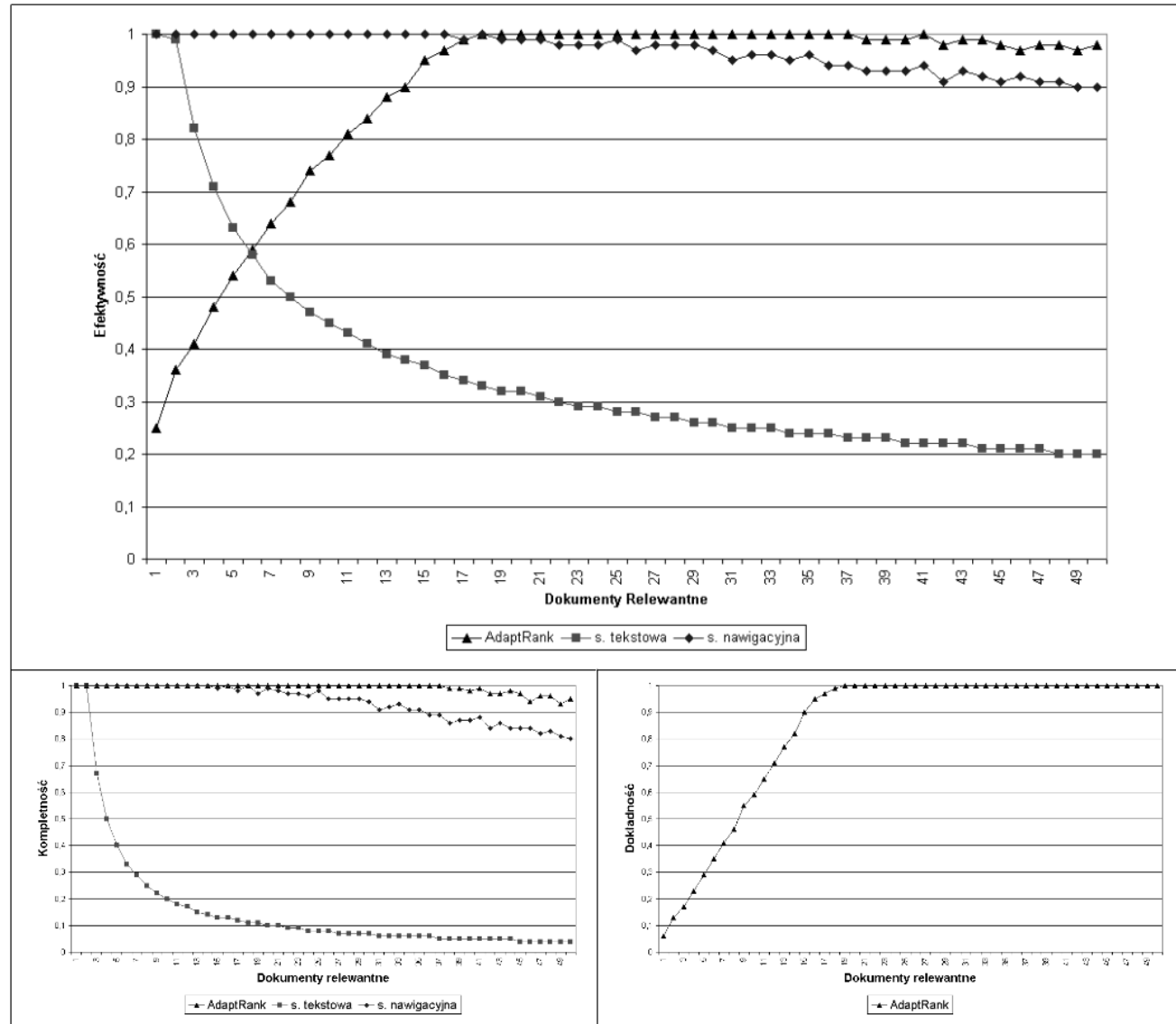
Praktyka - symulacyjne badania efektywności

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka



Praktyka - symulacyjne badania efektywności

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka



Praktyka - symulacyjne badania efektywności

- Agenda
- O pracy
- Wstęp
- Efektywność
- Klasyfikacja
- Propozycja algorytmu
- Definicja metod składowych
- Opis
- Definicja
- Praktyka

