- ■Wykłady opracowano w oparciu o książkę Ricardo Baeza-Yates, Berthier Ribeiro-Neto, "Modern Information Retrieval, the concepts and technology behind search" 2nd edition, ACM Press Books, 2011
- ■Z tego samego źródła zaczerpnięto także różne zadania i przykłady wykorzystywane w treści wykładu.

- Rolą interfejsu użytkownika jest pomoc w sformułowaniu zapytania i odebraniu wymaganych informacji
- Interfejs powinien także umożliwić:
 - wybór źródła informacji,
 - ■zrozumienie wyników wyszukiwania,
 - śledzenie postępu w wyszukiwaniu
- ■Interakcja użytkownika z interfejsem wyszukującym zależy od:
 - ■typu postawionego zadania,
 - dotychczasowej wiedzy o użytkowniku,
 - llości czasu i wysiłku wkładanego w proces wyszukiwania.
- Rozróżnia się wyszukiwanie informacji (*information lookup*) i wyszukiwanie rozpoznawcze (*exploratory search*)

- Wyszukiwanie informacji (information lookup) :
 - ■jest podobne do wyszukiwania faktów i odpowiedzi na pytania,
 - ■może być wypełnione przez dyskretne informacje jak liczby, daty, nazwy lub strony sieciowe,
 - pracuje poprawnie w trybie standardowych interakcji z siecią.
- Wyszukiwanie rozpoznawcze (exploratory search) można podzielić na zadania śledzenia i uczenia się
- Wyszukiwanie uczące się (learning search) wymaga:
 - więcej niż jednej akcji zapytanie-odpowiedź,
 - czasu od wyszukiwacza,
 - odczytywania wielu porcji informacji,
 - złożenia różnych treści do sformowania odpowiedzi.

- Śledzenie (investigating) odnosi się do długotrwałych procesów które:
 - ■obejmują wiele iteracji w dłuższym czasie,
 - ■zwracają informacje podlegające ocenie przed uzupełnienie bazy wiedzy o użytkowniku,
 - mogą być związane z wyszukiwaniem istotnych informacji w wielkich zbiorach odpowiedzi
- Sensemaking (Karl Weick 1969) -organizowanie wiedzy: proces iteracyjny tworzenia konceptualnej reprezentacji dużych zbiorów odpowiedzi.
- Przykładowe zadania (poza wyszukiwaniem) wymagające inteligentnego wyboru
 - badanie przypadków prawnych,
 - epidemiologia (disease tracking)
 - studiowanie uwag użytkowników dla poprawy obsługi,
 - analizy danych biznesowych,

- Etapy klasycznego wyszukiwania informacji:
 - ■identyfikacja problemu,
 - sformułowanie potrzeb informacyjnych,
 - utworzenie zapytania,
 - ocena rezultatów wyszukiwania.
- Najnowsze modele uwzględniają dynamiczne aspekty wyszukiwania:
 - użytkownicy uczą się w trakcie wyszukiwania,
 - potrzeby informacyjne użytkowników zmieniają się podczas przeglądania wyników wyszukiwania.
- Model dynamicznego wyszukiwania nazywa się także "zbieraniem jagód" (berry picking) lub wyszukiwaniem zorientowanym (orienteering)

- Na podstawie analizy logów wyszukiwania Jansen stwierdził, że ok. 52% użytkowników dokonuje modyfikacji zapytań
- Modele wyszukiwania są formułowane w kategoriach dwóch zasadniczych strategii:
 - odzwierciedlającej przemyślane zachowanie specjalistów prowadzących wyszukiwanie,
 - odtwarzającej mniej planowe a bardziej odruchowe zachowania typowego poszukiwacza informacji
- Nawigacja: wyszukujący sprawdza strukturę informacji i przegląda dostępną informację (browsing)
- Taka strategia jest preferowana gdy struktura informacji jest dopasowana potrzeb użytkownika – dostępne są odpowiednie linki

- ■Jeżeli właściwe linki nie są dostępne takie przeglądanie może być denerwujące
- Przykład: poszukiwanie sterownika drukarki laserowej
- Użytkownik wybiera słowa: *printers, laser printers* i dalej sekwencję linków

HP laser printers
HP laser printers model 9750
software for HP laser printers model 9750
software drivers for HP laser printers model 9750
software drivers for HP laser printers model 9750 for the
Win98 operating system

■Taki rodzaj interakcji jest akceptowalny dokąd kolejne poprawianie dokładności wyszukiwania ma sens

- Rozliczne studia prowadzone nad procesami wyszukiwania doprowadziły do wniosku, że użytkownicy często
 - ■reformułują zapytania z niewielkimi modyfikacjami,
 - ponownie szukają informacji znalezionych poprzednio.
- ■Badacze opracowali wsparcie interfejsów wyszukiwania z odwołaniem do historii zapytań i ponawianiem zapytań (*query history and revisitation*)
- ■Badania pokazują, że wyszukujący zwracają uwagę na kilka początkowych pozycji w rankingu wyszukanych odpowiedzi.
- Użytkownicy są zazwyczaj przekonani, że pierwsze wyniki w zbiorze odpowiedzi są lepsze niż pozostałe.

Aktualne interfejsy wyszukiwania

- Sposoby rozpoczynania sesji wyszukiwania informacji online :
 - użycie wyszukiwarki internetowej,
 - wybór strony spośród stron ostatnio odwiedzonych (historia) lub zapamiętanych (zakładki),
- ■Wyszukiwarki z zakładkami online są stosowane przez mniejszą liczbę użytkowników. np. Delicious.com. Zakładki tego typu są gromadzone w sieci na odpowiednich portalach.
- Zapytania w wyszukiwarkach są zadawane w formie tekstowej
- Zapytania typowo są krótkie i zawierają jedno do trzech słów
 - jeśli wyniki nie są odpowiednie użytkownik reformułuje zapytanie,
 - ■jeśli wyniki są odpowiednie użytkownik nawiguje do stron, o których sądzi, że są istotne.

- Przed epoką sieci systemy wyszukujące typowo wspierały operatory boolowskie i składnię opartą o komendy.
- ■Badania Jansena (2007) na zbiorze 1.5M zapytań dały wyniki:
 - 2.1% zapytań zawiera operatory boolowskie,
 - ■7.6% zapytań ma inną składnię, głównie podwójne cudzysłowy.
- White (2007) badał logi interakcji ok. 600 000 użytkowników i stwierdził:
 - ■1.1% zapytań zawiera jeden lub więcej operatorów,
 - 8.7% użytkowników używało zawsze operatorów.

- Ok. 1997 roku Google wprowadził wyłącznie zapytania koniunktywne, które wkrótce stały się normą.
- Google dodał pojęcie informacji przybliżonej i wprowadził ranking stron (PageRank).
- W miarę rozwoju sieci pojawiły się poprawne odpowiedzi na dłuższe pytania stawiane w formie fraz.
- Standardowym interfejsem wejściowym zapytania jest pole edycyjne (*search box entry*).

W niektórych przeglądarkach formy wejściowe zapytań występują razem z filtrującymi np. yelp.com



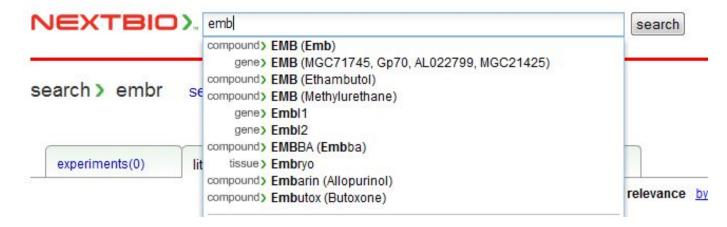
- Przewiduje się także inne specjalizowane typy wejść:
 - Np. zvents.com rozpoznaje określenia czasu typu "tomorrow" i pozwala na wprowadzanie różnych formatów daty.
 - Wyszukiwanie dla zwrotu "comedy on wed" automatycznie wybiera najbliższą środę

- Niektóre interfejsy pokazują listy sugestii zapytań i typów zapytań
- ■Nazywa się to autouzupełnianiem, autosugerowaniem, lub dynamiczną sugestią zapytań (auto-complete, auto-suggest, or dynamic query suggestions)
- Sugestie dotyczą często uzupełnienia wprowadzonych znaków traktowanych jako prefiksy, rzadziej podają uzupełnienia środkowych liter
- Podpowiadane mogą być także synonimy dotychczas używanych słów

Sugestie dynamicznych zapytań Netfix.com



- Źródła sugestii zapytań dynamicznych:
 - Historia zapytań użytkownika,
 - Zbiór metadanych, które projektant strony internetowej uważa za istotne,
 - Cały tekst zawarty na stronie internetowej.
- Sugestie dynamiczne grupowane wg typów z NextBio.com



- Przy prezentacji wyników wyszukiwania:
 - dokumenty muszą być pokazywane w całości lub
 - wyszukującemu należy przedstawić pewną reprezentację ich zawartości
- Surogat dokumentu zawiera informację streszczającą ten dokument
 - Ta informacja jest kluczowa dla sukcesu wyszukiwania,
 - Projektowanie właściwych surogatów dla dokumentów jest aktywnym polem badań,
 - Jakość surogatów może wpływać na postrzegane znaczenie wyników wyszukiwania

- Zawsze przedstawia się tytuł strony wraz z adresem URL i inne metadane
- Przy przeszukiwaniu kolekcji podawane są dane o dacie publikacji i autorze
- ■Wycięte fragmenty dokumentów zawierające wyszukiwane pojęcia lub streszczenia decydujące o ich rankingu są także istotne
- Standardowo wyniki wyświetla się jako pionowe listy *SERP* (*Search Engine Results Page*) podsumowań (streszczeń) tekstowych
- W pewnych przypadkach obok informacji tekstowej dołącza się specjalne metadane (*blended results*)

Przykład -Google

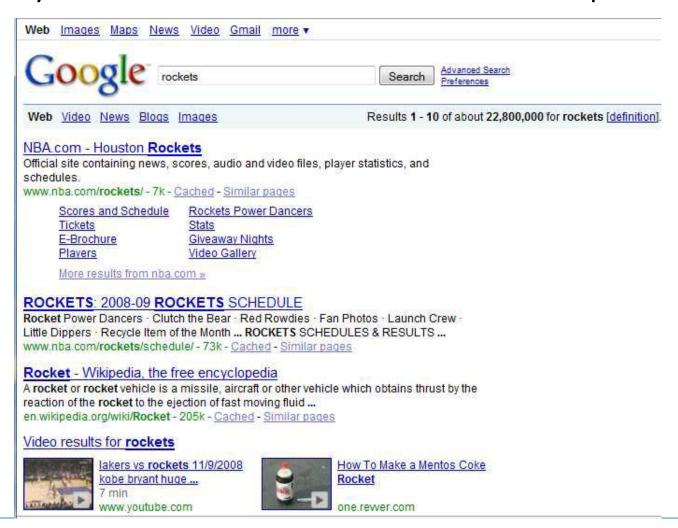
Koniunktura gospodarcza – Encyklopedia Zarządzania mfiles.pl/pl/index.php/Koniunktura_gospodarcza 8 Lis 2009 – Definicja. Koniunktura gospodarcza to wszelkie zmiany aktywności gospodarczej przejawiające się w zmianach podstawowych wskaźników ...

GUS - Główny Urząd Statystyczny - Koniunktura w przemyśle ...
www.stat.gov.pl > Strona główna > Koniunktura
23 Lut 2012 - Koniunktura w przemyśle, budownictwie, handlu i usługach ... Ogólny
klimat koniunktury w przetwórstwie przemysłowym w lutym oceniany jest ...

Np. zapytanie "rainbow" w Yahoo zwraca teksty i

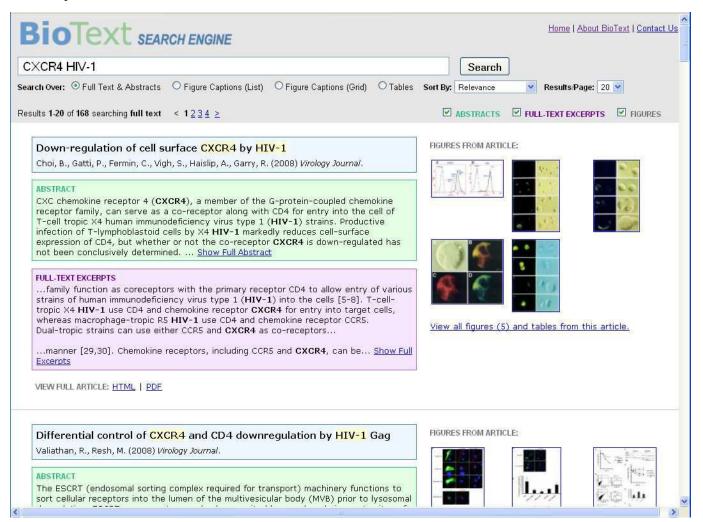


Np. zapytanie o nazwę drużyny sportowej wydobywa wyniki ostatnich meczów i linki do stron zakupu biletów



- Podkreślanie wyszukiwanych pojęć w odpowiedzi poprawia ocenę jakości wyników. Podkreślanie może być też stosowane w pełnych dokumentach.
- Wyzwanie stanowi dobór i rozmiar tekstu w streszczeniu dokumentu
- Należy wybrać pewną równowagę między pokazywaniem zdań zawierających poszukiwane termy a pokazaniem ciągłego fragmentu tekstu z tymi termami
- Najlepszy sposób wyświetlania wyników jest zawsze powiązany z intencją zapytania:
 - dla pewnych typów informacji dłuższe odpowiedzi są uważane za lepsze
 - w pytaniach nawigacyjnych o bieżące aktualności preferuje się zwięzłe odpowiedzi

Przykład równoległego wyciągania rysunków z artykułów

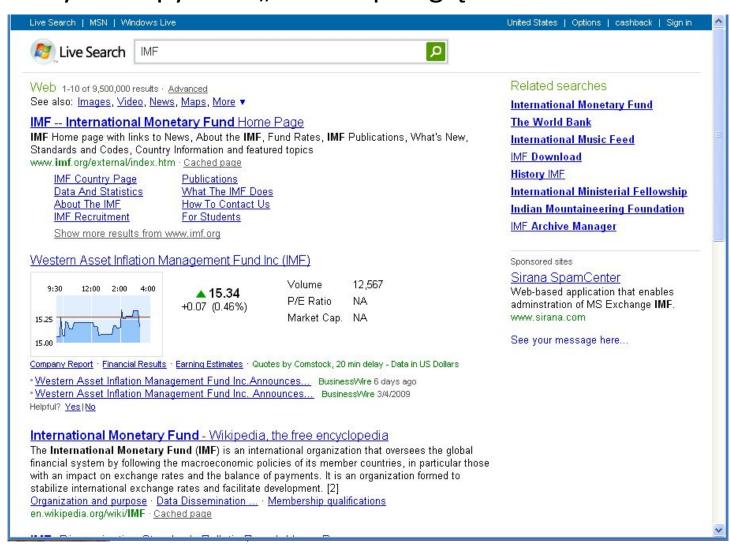


Przeformułowanie zapytań

- Pomoc w reformatowaniu zapytania
 - pokazywane są termy zapytania i związane z wyszukanym dokumentami
 - specjalny przypadek to korekty pisowni lub sugestie określeń powiązanych
- **Term expansion** interfejsy wyszukiwania w coraz większym stopniu korzystają z podpowiedzi termów powiązanych z pierwotnym zapytaniem

Przeformułowanie zapytań

Wyniki zapytania "IMF" w przeglądarce IE



Przeformułowanie zapytań

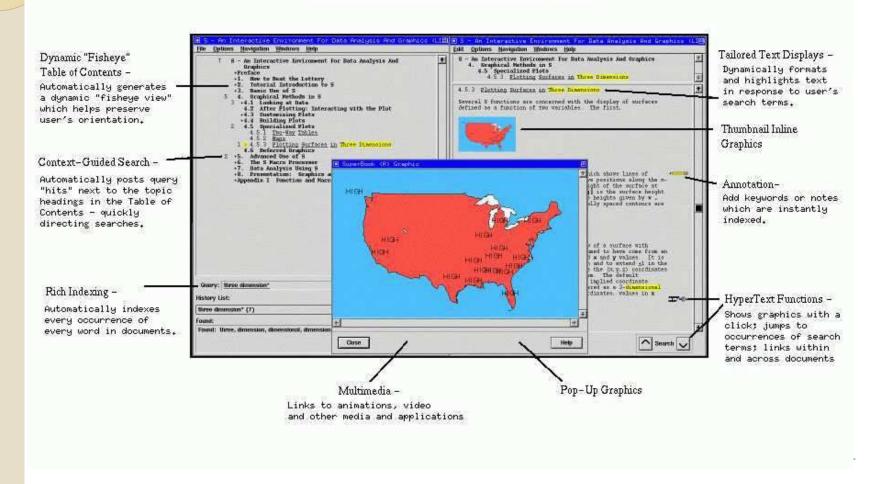
- Niektóre sugestie termów opierają się na danych całej sesji wyszukiwania danego użytkownika.
- Inne bazują na typowym zachowaniu innych użytkowników zadających podobne pytania
 - jedna strategia to wyświetlanie tych pytań
 - druga strategia: wyszukiwanie termów w dokumentach wyszukanych dla innych użytkowników
- Metoda "relevance feedback" zakłada, że użytkownik wskaże, które dokumenty wybrano trafnie lub które termy wybrane z dokumentów są istotne
- System wyszukiwania na tej podstawie formułuje nowe, poprawione zapytanie
- Metoda ta nie jest stosowana w aktualnych wyszukiwarkach ze względu na uczestnictwo użytkownika i jego ograniczoną wiedzę

- Pomaga użytkownikom ocenić wyniki i podjąć właściwe decyzje
- Wyróżnia się systemy kategorii (category systems) i klastering
- Systemy kategorii: istotne etykiety zorganizowane tak aby odwzorowywały ważne tematy w danej dziedzinie
 - powinny być możliwie spójne i kompletne,
 - ich struktura ma być przewidywalna i zgodna w całym zbiorze odpowiedzi
- Struktury kategorii: płaska, hierarchiczna lub fasetowa (faceted)
 - struktura płaska lista tematów, dokumentów
 - hierarchiczna stosowana do treści książek lub małych zbiorów

- Na zasadzie hierarchicznej działa m. in. system *Superbook*.
- W tym systemie wyniki wyszukiwania są prezentowane w kontekście spisu treści

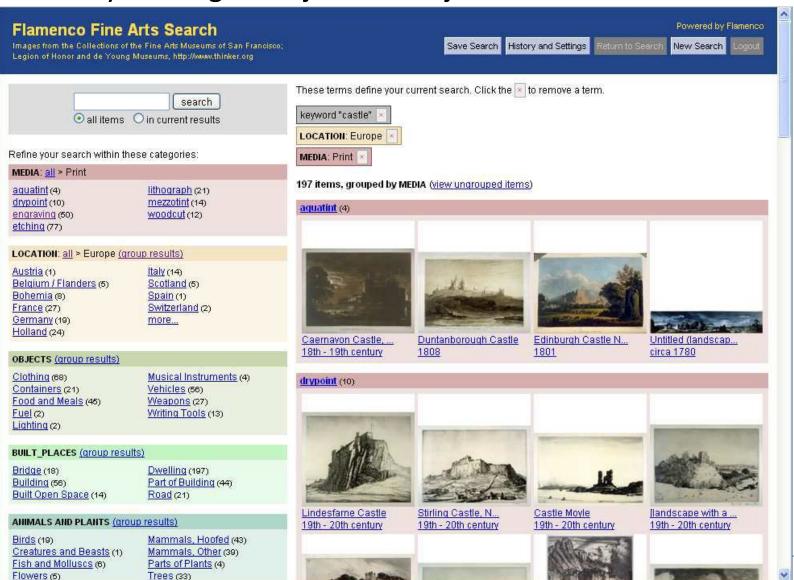
■ Interfejs Superbook

The SuperBook Document Browser Features

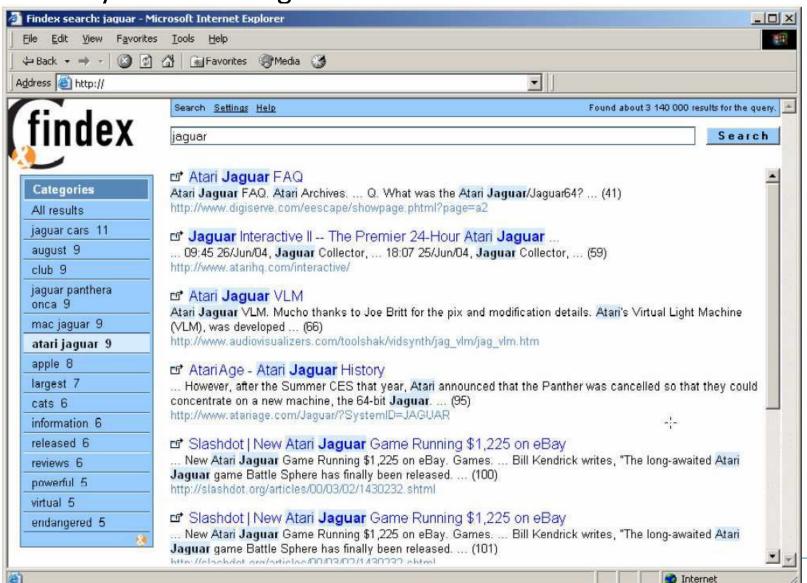


- Reprezentacja fasetowa metadanych pozwala na przypisanie wielu kategorii do pojedynczej pozycji wyników
- Każda z tych kategorii odpowiada innemu fasetowi (wymiarowi, typowi cechy) w zbiorze wyników
- Klastering to grupowanie pozycji wyników według pewnej miary podobieństwa
 - Grupuje razem dokumenty podobne do siebie i różne od pozostałych – np. dokumenty w języku japońskim w zbiorze publikacji głównie angielskich
- Klastering jest w pełni automatyczny, ale może dać wyniki grupowania niezgodne z intuicją użytkownika

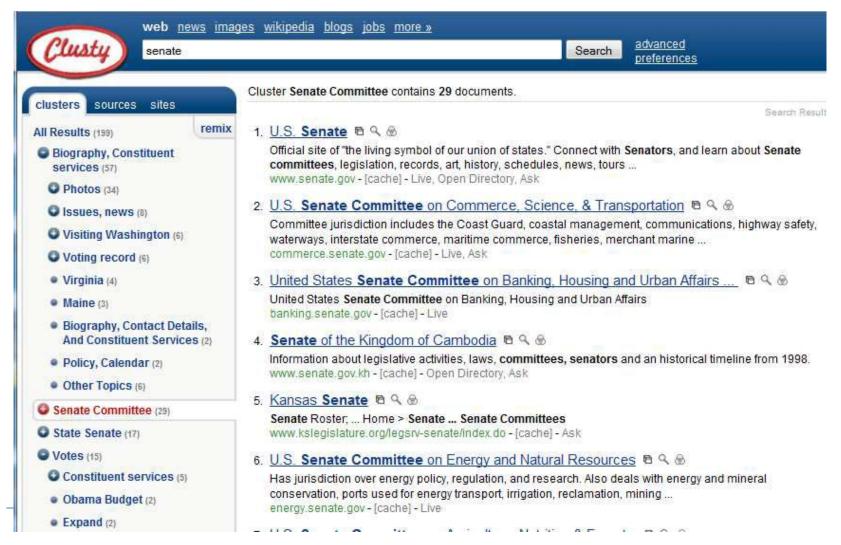
Przykład organizacji fasetowej



Wyniki klasteringu Findex



Wynik klastrowania zapytania "senate" w Clusty.com (Yippy.com)

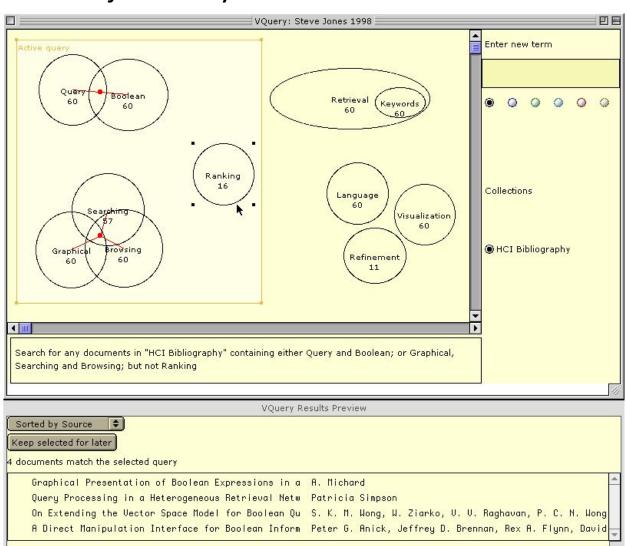


Wizualizacja w interfejsach wyszukiwania

- Wizualizacja wyników:
 - składni boolowskiej (rzadko używane),
 - termów zapytania w zbiorze wyników,
 - zależności między słowami i dokumentami
 - wizualizacja dla eksploracji tekstów
- Składnię boolowską przedstawia się przy pomocy tzw. diagramów Venna wykorzystywanych np. przez system VQuery.

Wizualizacja w interfejsach wyszukiwania

Interfejs VQuery

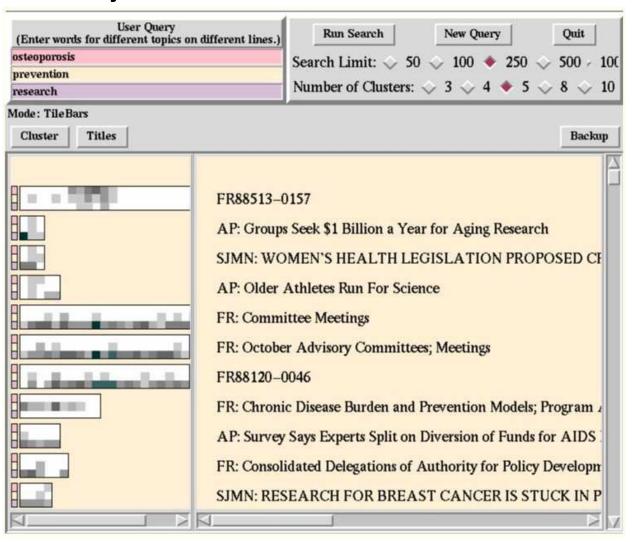


Wizualizacja termów zapytania

- ■W interfejsie TitleBar dokumenty są prezentowane jako poziome glify (*glyphs*)
- ■Położenia termów zapytania są zaznaczone wewnątrz nich
- Użytkownik jest zachęcany do rozbicia zapytania na różne fasety z jednym wariantem w linii
- ■Wtedy poziome wiersze graficznej reprezentacji dokumentów (*glyphs*) przedstawiają częstotliwość pojawiania się termów w ramach poszczególnych tematów

Wizualizacja termów zapytania

Interfejs TitleBars



Wizualizacja termów zapytania

Inne podejścia to umieszczenie termów na wykresach słupkowych (bar charts), punktowych (scatter plots) i w tablicach

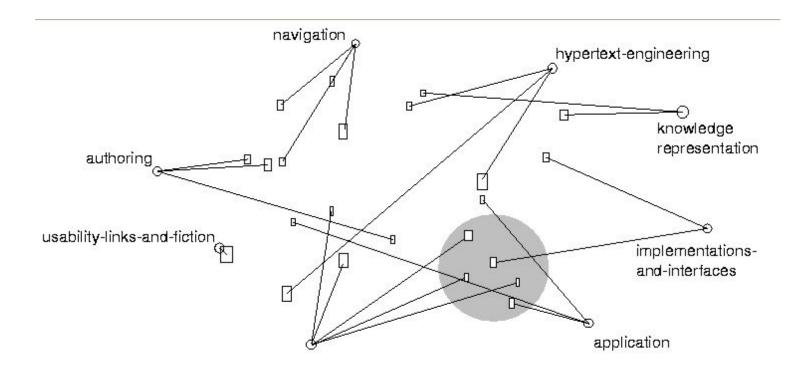
Uwypuklanie termów w miniaturach dokumentów

Powiązania termów i dokumentów

- W wielu pracach na temat interfejsów proponuje się połączenie słów (termów) i dokumentów na płótnie dwuwymiarowym
- ■Interfejs **VIBE** bliskość glifów reprezentuje powiązania semantyczne termów i dokumentów; dokumenty z kombinacjami termów są umieszczone w pół drogi pomiędzy ikonami tych termów
- Projekty Aduna Autofocus i Lyberworld rozszerzają VIBE do 3 wymiarów

Powiązania termów i dokumentów

Okno VIBE



Powiązania termów i dokumentów

Wyświetlanie typu starfield (xFIND VisIslands)

