INDEKSOWANIE DANYCH

PODSTAWY TECHNIK MULTIMEDIALNYCH, A.Przelaskowski

wykorzystano m.in. materiały opracowane przez prof. W.Skarbka

- Definicje
- Miary selektywności
- Wyszukiwarki

Indeksowanie - wstęp

- W studium 'How much information UCB' (University of California Berkeley) z roku 2002 podano, że roczna produkcja nowych informacji to około 250MB średnio na każdego mężczyznę, kobietę i dziecko na ziemi (z tego 93% w postaci cyfrowej).

  $\Rightarrow$ Roczny przyrost zasobów informacyjnych $741PB$ $(1PB = 10^3TB = 10^6GB = 10^9MB)$, w tym:
  - $8 \cdot 10^9$ fotografii $-410PB$
  - $1,4 \cdot 10^9$ filmów wideo $-300PB$
  - $2 \cdot 10^9$ zdjęć rentgenowskich $-17,2PB$
  - $0,2 \cdot 10^9$ dysków twardych $-13,8PB$

Działania związane z zalewem informacji

- Wymiana/komunikacja/gromadzenie
- Kompresja/kodowanie
- Selekcja, porządkowanie
- Opis (meta-dane)
- Indeksowanie
- Wyszukiwanie
- ….

Podstawowe pojęcia

- Indeksowanie multimediów dotyczy metod budowy indeksów kolekcji (zbiorów) obiektów multimedialnych
- Obiekt multimedialny to kompozycja przestrzenna i/lub czasowa tekstu, grafiki, obrazu, dźwięku, ….
- Przykłady obiektów:
  - strona tekstu, wydzielona sekcja lub nawet akapit tekstu;
  - obraz rastrowy, obraz graficzny (grafika) lub kompozycja obrazów rastrowych i graficznych;
  - temporalna sekwencja obrazów lub temporalny segment tej sekwencji (np. między ramkami kluczowymi);
  - obraz i jego tekstowy opis;
  - sekwencja próbek dźwiękowych lub segment tej sekwencji;
  - kompozycja tekstu, dźwięku i obrazu na stronie multimedialnej;
  - film cyfrowy (tj. obraz zsynchronizowany z dźwiękiem) lub jego fragment (tzw. cięcie).

Podstawowe pojęcia

- Na indeks składają się:
  - zbiór wartości atrybutu, czyli cech opisujących określoną właściwość obiektów (atrybut to określona właściwość obiektów, wokół której budowany jest indeks)
  - listy (zestaw) identyfikatorów obiektów posiadających daną cechę
- Przykład 1: indeks książki wokół atrybutu słowa kluczowe (właściwość tekstów), gdzie na kolejnych pozycjach znajdują się zwykle w porządku alfabetycznym kolejne słowa kluczowe (cechy tekstu) z numerami stron (identyfikatorami obiektów), gdzie takie słowa występują
- Przykład 2: indeks opisujący telekonferencje (atrybut: stroje osób występujących w scenie)

Atrybut, cecha, lista obiektowa

- c=a(o), gdzie c-cecha, a-atrybut, o-obiekt multimedialny
- Lista obiektów podobnych $\mathcal{L}_c \triangleq < l_c; i_1, \ldots, i_{l_c} >$

$\Rightarrow$ Liczba elementów l_c na liście obiektowej cechy c może być kontrolowana przez trzy globalne parametry: K_{min}, K_{max} – minimalna i maksymalna liczba elementów na liście, $\rho_{min} \in [0,1]$ – minimalny próg podobieństwa cechy $a(o)$do cechy c według funkcji podobieństwa ρ wybranej dla atrybutu a.
$\Rightarrow$ Niech $l_c \leftarrow |\{o \in \mathcal{O} : \rho(a(o), c) \geq \rho_{min}\}|$.

- Jeśli $l_c > K_{max}$, to wybieranych jest na liście K_{max} identyfikatorów najbardziej podobnych obiektów. Wtedy $l_c \leftarrow K_{max}$.
- Jeśli $l_c < K_{min}$, to brakujących $K_{min} - l_c$ obiektów dobiera się spośród kolejnych najbardziej podobnych obiektów. Wtedy $l_c \leftarrow K_{min}$.

Podobieństwo

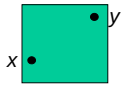
- Funkcja podobieństwa $\rho_a(a(o), a(o_1))$ zależy od atrybutu oraz doboru reprezentatywnych cech
- Przykład 1: W przypadku atrybutu a =*kolor* w modelu RGB, podobieństwo można oprzeć np. na mierze kosinusowej. Niech wektory $x = [r_1, g_1, b_1]^t$, $y = [r_2, g_2, b_2]^t$, reprezentują dwa kolory w przestrzeni RGB. Wtedy

$$\rho(x,y) \triangleq cos(x,y) \triangleq \frac{x^t y}{\|x\|\|y\|} = \frac{r_1 r_2 + g_1 g_2 + b_1 b_2}{\sqrt{r_1^2 + g_1^2 + b_1^2}\sqrt{r_2^2 + g_2^2 + b_2^2}}$$

- Przykład 2: cechami są liczby nie różniące się więcej niż o *M*

$$\rho(x,y) \triangleq 1 - \frac{|x-y|}{M}$$

- Przykład 3: cechy w kwadracie o boku 1



$$\rho(x,y) \triangleq 1 - \frac{\|x-y\|}{\sqrt{2}}$$

Podobieństwo

- edycyjne - dla cech c_1 i c_2 mamy

$$\rho(c_1, c_2) \triangleq 1 - \frac{\delta(c_1, c_2)}{\max_c |c|}$$

gdzie odległość δ pomiędzy wyrazami (słowami) określona jest jako najmniejsza liczba operacji zmiany, usuwania i dołączania pojedynczego symbolu w dowolnym miejscu, dzięki którym sekwencja c_1 przekształcana jest w sekwencję c_2 ; dzielnik normalizujący $\max_c |c|$ uwzględnia największą długość słowa w danym zbiorze cech

- przykład: obliczamy podobieństwo słów c_1=„Mama” i c_2= „Matka”; minimalna liczba operacji przekształcenia c_1 i c_2 wynosi 3, bo

USUŃ m,3; DODAJ t,3; DODAJ k,4

przy maksymalnej długości słowa 25 mamy podobieństwo:

$$\rho(\text{"Mama","Matka"}) = 1 - \frac{3}{25} = 0{,}88$$

- rangowe

$$\rho(c_1, c_2) \triangleq 1 - \frac{|\text{rank}(c_1) - \text{rank}(c_2)|}{\max_c \text{rank}(c) - \min_c \text{rank}(c)}$$

czyli

$$\rho(\text{"czerwiec", "październik"}) \triangleq 1 - \frac{|6-10|}{12-1} \approx 0,64$$

Wyszukiwanie

Wyszukiwarka oparta o cechy atrybutu a, zwracająca co najwyżej K_{max} i co najmniej K_{min} najbardziej podobnych obiektów w kolekcji $\mathcal{O}$ do cechy wejściowej c_{query} (z progiem ρ_{min}), może działać według następującego scenariusza:

1. **Zapytanie:** Na wejściu wprowadzana jest cecha $c_{query} \in \mathcal{C}_a$ (określona przez użytkownika lub przez algorytm A ekstrakcji cechy atrybutu a z pewnego obiektu o_{query});
2. **Najbardziej podobne cechy reprezentatywne:** W zbiorze cech reprezentatywnych atrybutu a znajdowanych jest co najwyżej K_{max} reprezentatywnych cech $c'_1, \ldots, c'_{K'}$ takich, że:,

$$\rho(c_{query}, c'_i) \geq \rho_{min}, i = 1, \ldots, K', \ K' \leq K_{max}.$$

   Jeśli $K' < K_{min}$, to uzupełnia się brakujące $K_{min} - K'$ cechy przez kolejne najbardziej podobne cechy oraz modyfikuje $K' \leftarrow K_{min}$;
3. **K najbardziej podobnych obiektów:** Spośród wszystkich obiektów identyfikowanych na połączonej liście $\mathcal{L} \triangleq \mathcal{L}_{c'_1} \cup \cdots \cup \mathcal{L}_{c'_{K'}}$ wybiera się K najbardziej podobnych obiektów $o_1, \ldots, o_K$, $K_{min} \leq K \leq K_{max}$, spełniających próg podobieństwa ρ_{min} lub uzupełnionych do liczności K_{min}.

Scenariusze wyszukiwań

$\Rightarrow$ Trzy parametry kontrolujące liczbę odpowiedzi na zapytanie mają w ogólności następujące zakresy (∞ oznacza brak górnego ograniczenia na dany parametr):

1. $0 \leq K_{min} \leq \infty$;
2. $0 < K_{max} \leq \infty$;
3. $0 \leq \rho_{min} \leq 1$.

$\Rightarrow$ Specjalne kategorie zapytań, często występujące w praktyce to:

1. K najbardziej podobnych obiektów (K najbliższych sąsiadów): $K_{min} = K_{max} = K > 0$, $\rho_{min} = 0$;
2. Najbliższy sąsiad: $K_{min} = K_{max} = 1$, $\rho_{min} = 0$;
3. Obiekty podobne z progiem ρ_{min} (zapytanie o zakres ρ_{min}): $K_{min} = 0$, $K_{max} = \infty$, $\rho_{min} > 0$;
4. Zapytanie zakresowe z co najmniej K wynikami: $K_{min} = K > 0$, $K_{max} = \infty$, $\rho_{min} > 0$.

Selektywność wyszukiwania

- Będziemy dalej mówić, że zwrócony przez wyszukiwarkę obiekt jest poprawny, jeśli jest on semantycznie równoważny z zapytaniem (obiektem wejściowym)
- Konieczne jest zdefiniowanie semantycznej (znaczeniowej) relacji równoważności między obiektami
- Przykłady:
  - zdjęciami twarzy zachodzi wtedy, gdy są to zdjęcia tej samej osoby;
  - nagraniami instrumentów muzycznych zachodzi wtedy, gdy nagrania dokonano na instrumentach tego samego rodzaju, np. na skrzypcach;
  - dokumentami tekstowymi zachodzi wtedy, gdy po usunięciu znaków kontrolnych i formatujących zawartość obu dokumentów jest identyczna.

Miary selektywności

- Typowe miary to:

1. *Precyzja (ang. precision):* jaka część zwróconych obiektów jest poprawna?
2. *Przywołanie (ang. recall):* jaka część poprawnych obiektów jest zwrócona?
3. *Stopa sukcesu (ang. success rate):* jak często pierwszy zwrócony obiekt jest poprawny?
4. *Średnia ranga (ang. average rank):* jaka jest średnia pozycja (ranga) poprawnych obiektów w zwróconej liście obiektów?

Miary selektywności – precyzja i przywołanie

- precyzja

$$p \triangleq \frac{1}{|Q|} \sum_{q \in Q} \frac{N(q)}{K(q)}$$

- przywołanie

$$r \triangleq \frac{1}{|Q|} \sum_{q \in Q} \frac{N(q)}{M(q)}$$

gdzie *q*-zapytanie, N(q) – liczba zwróconych obiektów poprawnych, K(q) – liczba wszystkich zwróconych obiektów, M(q) – liczba popranych obiektów w bazie

Miary selektywności – stopa sukcesu

- Stopa sukcesu

$$sr \triangleq \frac{1}{|Q|} \sum_{q \in Q} \delta(q)$$

przy czym $\delta(q) = 1$, jeśli na pierwszym miejscu zwracany jest obiekt poprawny, zaś $\delta(q) = 0$ - w przeciwnym razie

Miary selektywności – średnia ranga

- Średnia ranga:

$$\bar{s} \triangleq \frac{1}{|Q|} \sum_{q \in Q} s(q)$$

gdzie

$$s(q) \triangleq \frac{1}{M(q)} \sum_{m=1}^{M(q)} \pi(m, q)$$

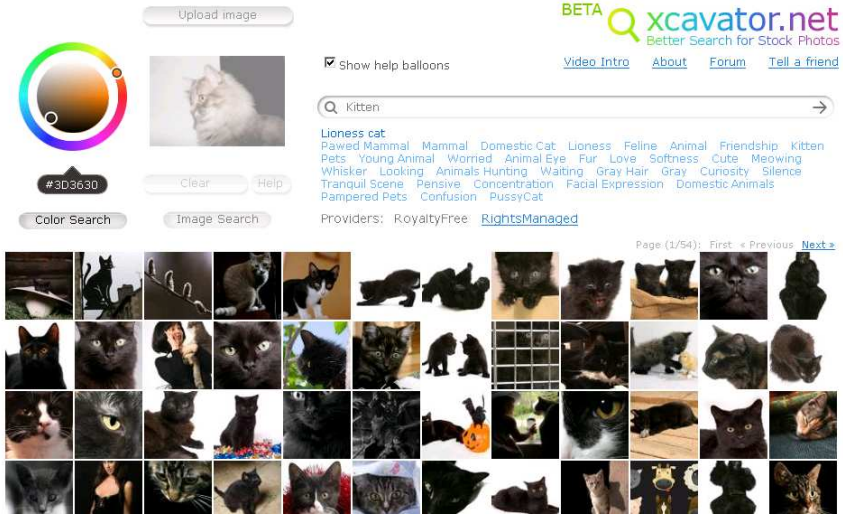
oraz

$$\pi(m,q) \triangleq \begin{cases} k, & \text{jeśli } m\text{-ty element poprawny} \\ & \text{zajmuje } k\text{-tą pozycję na liście wyników i } k \leq K(q) \\ K(q)+1, & \text{w przeciwnym razie.} \end{cases}$$

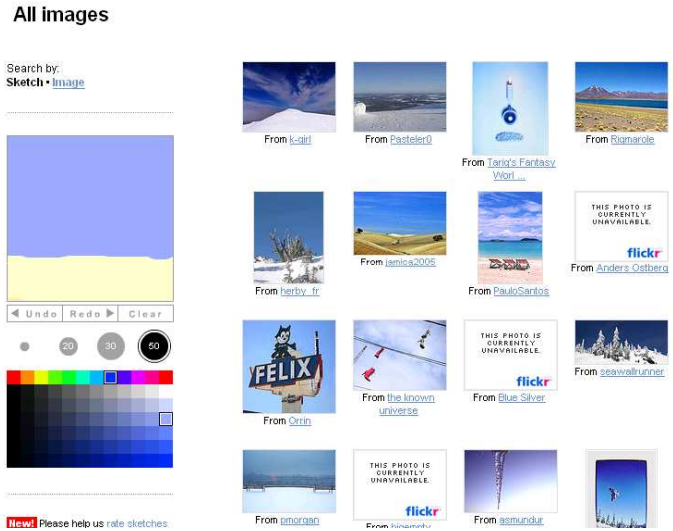
Struktury danych w indeksowaniu - słowniki

- Operacje na indeksie I(a) przy zapytaniu (c_{query}, K_{min}, K_{max}, ρ_{min}):
  - znajdź najbardziej reprezentatywne cechy
  - znajdź obiekty z listy danej cechy reprezentatywnej
- Struktury:
  - słownik cech zawierający reprezentatywne cechy danego atrybutu oraz identyfikatory (wskaźniki do) list obiektowych poszczególnych cech
  - słownik obiektów zawierający identyfikatory obiektów oraz adresy umożliwiające dostęp do zawartości obiektów
- Cechy słowników:
  - statyczny-dynamiczny
  - w pamięci operacyjnej czy dyskowej
  - cech skalarnych czy wektorowych
- Realizacje słowników: uporządkowane tablice cech, drzewa S, T, B, X, M etc

Analiza narzędzi do wyszukiwania - xcavator.net

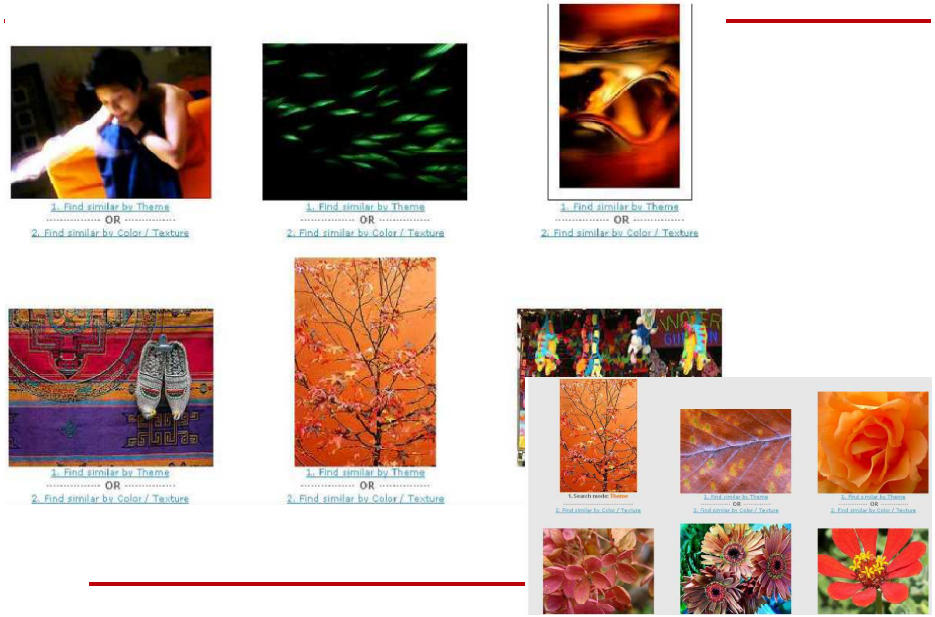


Analiza narzędzi - Retrievr

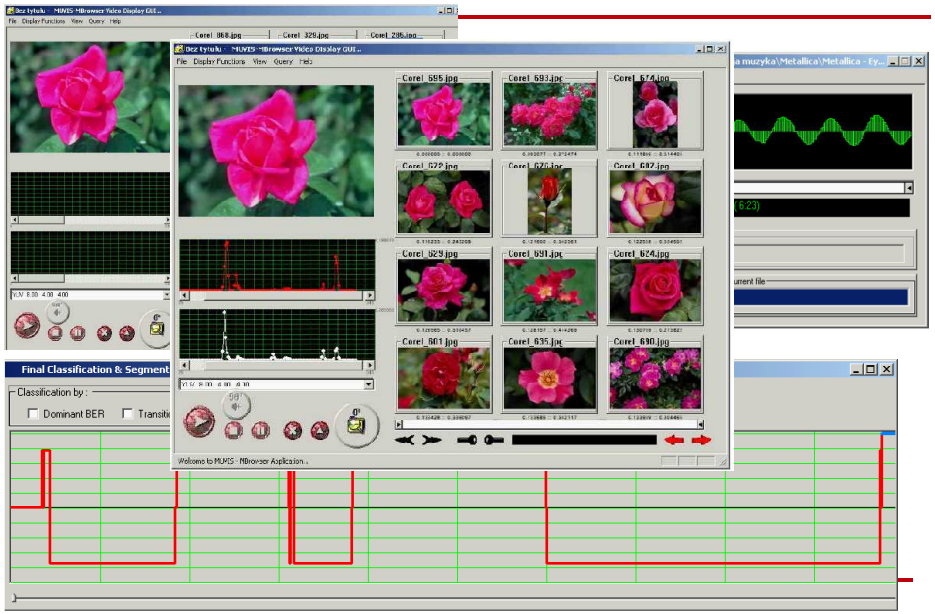


http://labs.systemone.at/retrievr/?sketchName=2007-10-09-09-22-48-807997.5

Tiltomo (www.tiltomo.com)



Analiza narzędzi – MUVIS (http://muvis.cs.tut.fi)



Analiza narzędzi – VideoGoogle (www.robots.ox.ac.uk/~vgg/research/vgoogle)

Eksploracja multimediów – przykładowe wyniki analizy narzędzi

Nazwa miary / Wyszukiwarka	Precyzja	Przywołanie	Stopa sukcesu
PIRIA	0,61	0,402	1
MFIRS	0,6	-	1
CIRES	0,68	-	1
Tiltomo	0,72	-	1
IKONA	0,85	-	1
Behold	0,7	-	0,8
Retrievr	0,86	-	1
xcavator	0,89	-	1
MUVIS	0,91	0,11	1

EKSPLORACJA DANYCH MULTIMEDIALNYCH

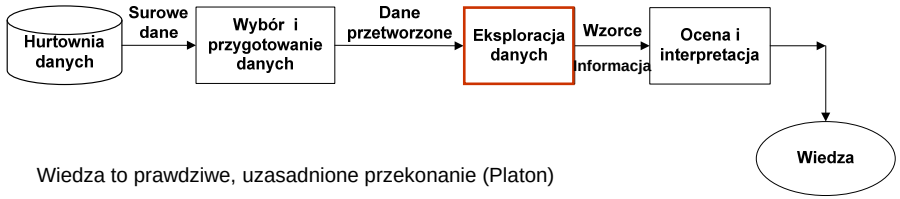
wykorzystano m.in. materiały opracowane przez prof. inż. A.Brzeską

- Podstawowe pojęcia
- Eksploracja multimediów
- Przykłady

Eksploracja danych

Eksploracja danych (*data mining)* to

- zaawansowane wydobywanie ukrytej informacji (wzorców) z rozległych zbiorów danych
- przeszukiwanie całej przestrzeni rozwiązań, znajdowanie zależności i powiązań między danymi w celu przybliżenia się do globalnego punktu – pożądanego rozwiązanie problemu
- istotny etap procesu odkrywania wiedzy z danych (*knowledge discovery in databases*)



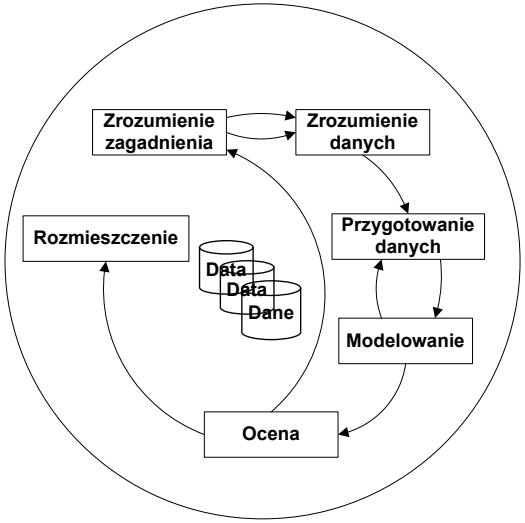
Wiedza to prawdziwe, uzasadnione przekonanie (Platon)

Wiedza to ogół wiarygodnych, powiązanych ze sobą informacji o rzeczywistości wraz z umiejętnością ich wykorzystywania

Metody – CRISP-DM

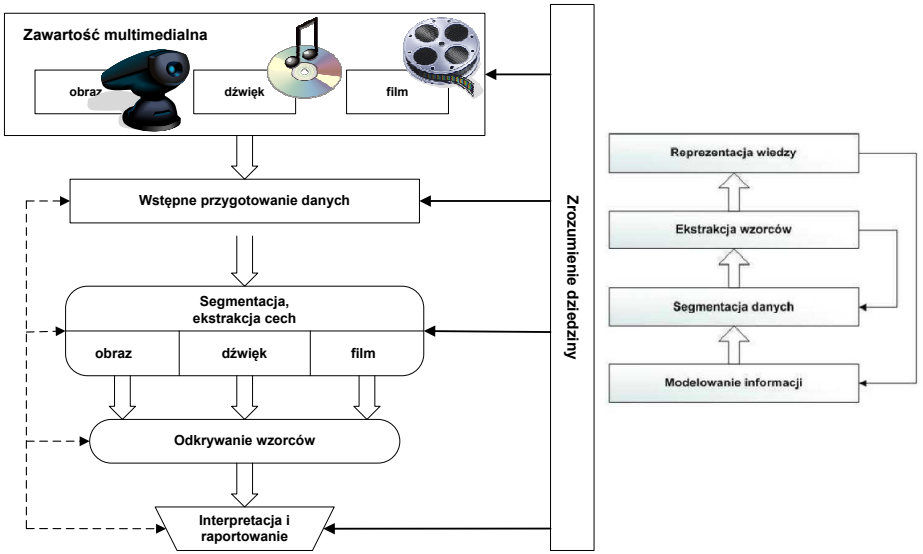
CRISP-DM (ang. *Cross-Industry Standard Process for Data Mining*)

- metodologia stworzona w 1966 roku,
- cykl życia procesu eksploracji danych



Chapman P., Clinton J., Khabaza T., Reinartz T., Wirth R., *The CRISP-DM Process Model*, marzec 1999

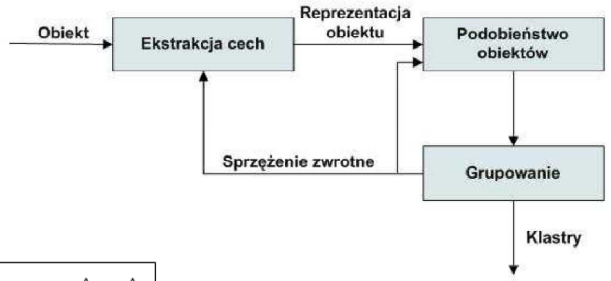
Proces eksploracji danych multimedialnych

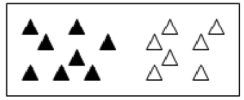


Najważniejsze metody eksploracji danych

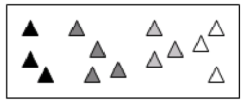
Metody eksploracji danych	Metody reprezentacji informacji – wiedzy, używane algorytmy	Zastosowanie w multimediach
Grupowanie, czyli łączenie danych w jednorodne grupy o wzajemnie różnych właściwościach	- algorytmy klasteryzacji (k-średnich, c-średnich, gmm, inne) - algorytmy hierarchiczne (na podstawie dendrogramu)	• segmentacja obrazu, podział na regiony • ekstrakcja cech
Kojarzenie danych - analiza połączeń (analiza asocjacji + analiza sekwencji)	- algorytm A-priori	• wykrywanie zależności pomiędzy wyrazami, określanie kolejności występowania wyrazów i fraz w tekstach • wykrywanie podobnych do siebie obrazów i filmów
Klasyfikacja $Y=f(x_1,\ldots,x_n, \Phi)$ f-model $(x_1,\ldots,x_n)$-dane wejściowe Φ-parametry	- drzewa decyzyjne - algorytmy probabilistyczne - algorytmy najbliższego sąsiedztwa - sieci neuronowe - maszyny wektorów nośnych	• rozpoznawanie pisma • rozpoznawanie odcisków palców • rozpoznawanie obrazów • identyfikacja osób na podstawie ich głosu • rozpoznanie patologii

Metody grupowania



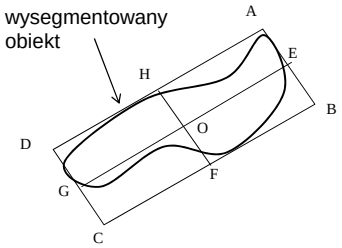


grupowanie twarde



grupowanie rozmyte

Wybrane cechy kształtu



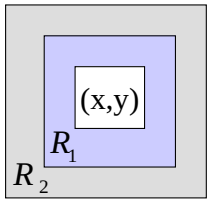
•Najdłuższa oś GE.
•Najkrótsza oś HF.
•Obwód i powierzchnia opisanego kwadratu ABCD.
•wydłużenie: GE/HF
•obwód p i powierzchnia A regionu

•kolistość $C = \frac{4\pi A}{p^2}$

•upakowanie $C_p = \frac{p^2}{A}$

SRDM (Surrounding Region - Dependence Method)

- Analiza tekstury otoczenia
- Stworzenie macierzy opisującej region



$$M(q) = [\alpha(i,j)], \qquad 0 \le i \le \overline{\overline{R_1}}, 0 \le j \le \overline{\overline{R_2}}$$

$$\alpha(i,j) = \#\left\{(x,y) \middle| C_{R_1}(x,y) = i \quad \text{oraz} \quad C_{R_2}(x,y) = j\right\}$$

$$C_{R_i}(x,y) = \#\left\{(k,l) \middle| (k,l) \in R_i \quad \text{oraz} \quad f(k,l) < f(x,y) - q\right\}$$

$M(q)-$ macierz SRDM w funkcji progu

Wybrane cechy tekstur

- pozioma suma ważona *(HWS horizontal-weighted sum)*

$$HWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} j^2 r(i,j)$$

gdzie N - suma wszystkich elementów w SRDM: $r(i,j) = \begin{cases} \frac{1}{\alpha(i,j)} & dla \ \alpha(i,j) > 0 \\ 0 & wpp \end{cases}$

- pionowa suma ważona *(VWS vertical-weighted sum)*

$$VWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} i^2 r(i,j)$$

- diagonalna suma ważona *(DWS)*

$$DWS = \frac{1}{N}\sum_{k=0}^{m+n} k^2 \left(\sum_{\substack{i,j:\ i+j=k,\\ i=0,\ldots,m \\ j=0,\ldots,n}} r(i,j) \right)$$

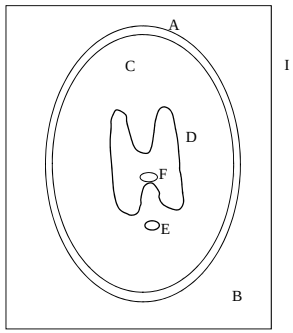
- suma ważona siatki *(GWS grid-weighted sum)*

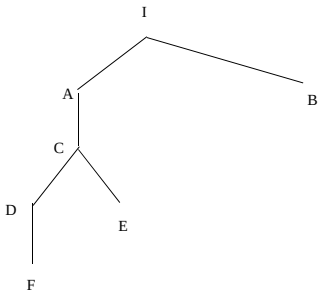
$$GWS = \frac{1}{N}\sum_{i=0}^{m}\sum_{j=0}^{n} ijr(i,j)$$

Przykładowy zestaw cech

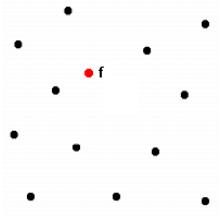
Nr	*Nazwa*	*Opis*
1	Powierzchnia	Rozmiar obiektu
2	Średnia	Średni poziom szarości, odchylenie standardowe poziomów szarości, poziom szarości tła obiektu
3	Odchylenie standardowe	
4	Tło	
5	Skontrastowanie obiektu i tła	(średnia – tło) / (średnia + tło)
6	Współczynnik zwartości	obwód² / powierzchnia
7	Moment kształtu I	
8	Moment niezmienniczy I	
9	Kontrast	Cechy wyznaczone na podstawie macierzy zdarzeń
10	Entropia	
11	Moment zwykły drugiego rzędu	
12	Odwrotny moment różnicowy	
13	Korelacja	
14	Wariancja	
15	Entropia rozkładu sumacyjnego	

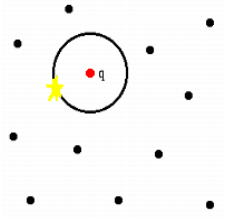
Cechy relacyjne (przykład)





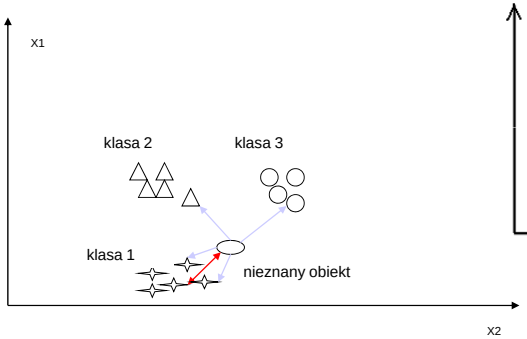
Grupowanie: metoda najbliższego sąsiada

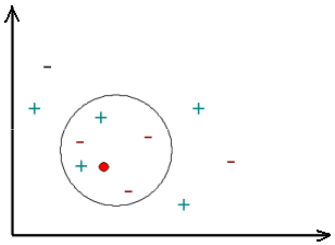




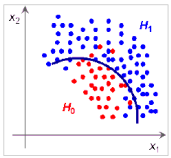
$$\mathbf{f} \in C_i \;\Leftrightarrow\; D_i(\mathbf{f}) = D_{\min}(\mathbf{f}) = \min_{j=1,\ldots,|C|}[D_j(\mathbf{f})]$$

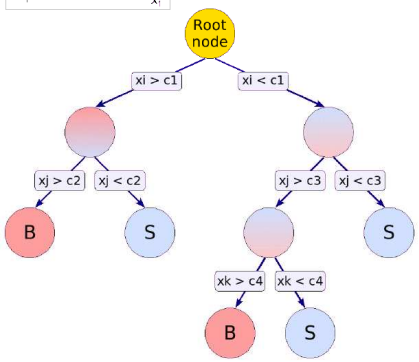
Klasyfikacja: k-najbliższych sąsiadów (kNN)



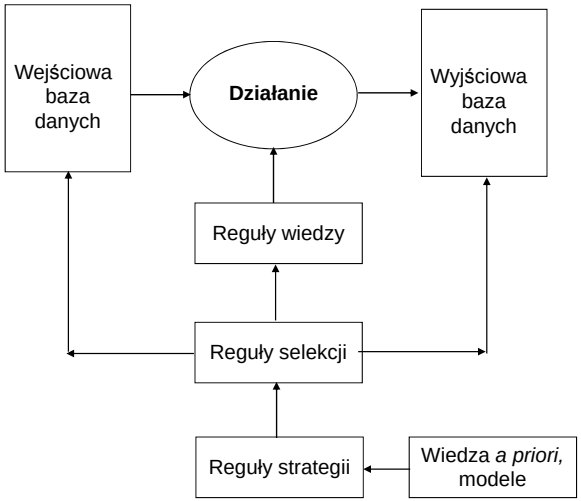


Drzewa decyzyjne





Wykorzystanie reguł semantycznych



Reguły strategiczne (region) - przykład

Strategy Rule SR1:

If
NONE REGION is ACTIVE
NONE REGION is ANALYZED
Then
ACTIVATE FOCUS in SPINAL_CORD AREA

Strategy Rule SR2:

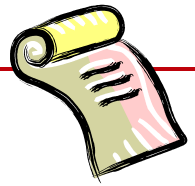
If
ANALYZED REGION is in SPINAL_CORD AREA
ALL REGIONS in SPINAL_CORD AREA are NOT ANALYZED
Then
ACTIVATE FOCUS in SPINAL_CORD AREA

Strategy Rule SR3:

If
ALL REGIONS in SPINAL_CORD AREA are ANALYZED
ALL REGION in LEFT_LUNG AREA are NOT ANALYZED

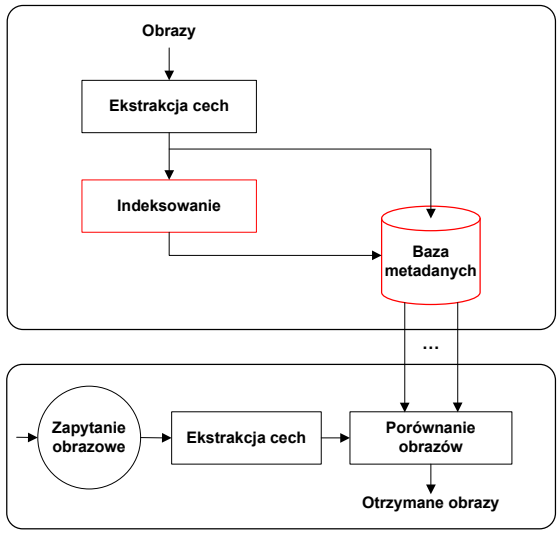
Then
ACTIVATE FOCUS in LEFT_LUNG AREA

Eksploracja tekstu



- Słowa kluczowe
- Reprezentacja wektorowa dokumentów tekstowych
  - Wektor częstości występowania słów kluczowych
- Ukryte indeksowanie semantyczne
  - Wyodrębnienie struktury semantycznej,
  - Wektor z usuniętymi zbędnymi powtórzeniami takich samych wyrażeń,
  - Większa precyzja wyszukiwania

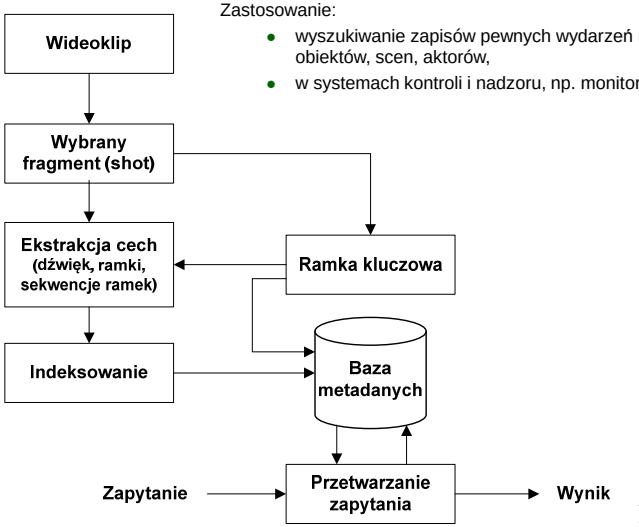
Eksploracja w oparciu o zawartość



przykład z obrazami

Źródło: Mitra S., Tinku A., *Data Mining: Multimedia, Soft Computing and Bioinformatics*, s. 331

Eksploracja filmów w oparciu o zawartość



Zastosowanie:

- wyszukiwanie zapisów pewnych wydarzeń na podstawie detekcji obiektów, scen, aktorów,
- w systemach kontroli i nadzoru, np. monitoring

Źródło: Mitra S., Tinku A., *Data Mining: Multimedia, Soft Computing and Bioinformatics*, s. 349

Eksploracja dźwięków

- Metoda słownikowa
  - przygotowanie danych, generacja pliku indeksu przechowującego słowa,
  - przeszukiwanie plików indeksu w celu znalezienia słowa lub frazy
- Fonetyczne rozpoznawanie dźwięków
  - indeksowanie, generacja pliku indeksu przechowującego fonetyczną zawartość pliku
  - przeszukiwanie plików indeksu w poszukiwaniu zadanego przez użytkownika zapytania
  - lepsze rezultaty i szybkość działania

- Zastosowanie:
  - kontrola rozmów w call center,
  - podpisy pod materiałami filmowymi

Obszary multimediów

- Reprezentacja danych (kodowanie, kompresja, transkodowanie)
- Poprawa jakości danych (filtracje, transformacje)
- Eksploracja danych (wydobywanie, indeksowanie, wyszukiwanie)
- Ochrona danych
- Sprzęt multimedialny
- Przesyłanie, wymiana danych
- Nauczanie na odległość, cyberprzestrzeń
- Integracja, standaryzacja
- …..