

$N \rightarrow \text{szkoła} \{\text{Szkoła}\}$

Czytałem pańską książkę w szkole.

Tego lata koło szkoły urządzono boisko.

Szkołę jazdy żem przeszedł!

Od razu widać, z jakiej szkoły wyszedł, pismak.

Spór o historyczną szkołę krakowską.

N → klasa {Klasa}

Zademonstrował mi swoją klasę na dużej fali jeziora Ontario.  
Byli to najwyższej klasy karawaniarze specjaliści.  
Cała klasa polityczna poleciała, bo ludzie zdali...  
...że Piotrek w pierwszych klasach same piątki zbierał...  
Egzamin wstępny do klasy pierwszej nie był obowiązkowy.  
...przed komisją złożoną z nauczycieli klasy VIII i dyrektora szkoły.  
Albo: skąd wziął się w klasie zapach dymu tytoniowego?  
Jest zresztą w klasie kilku takich jak on.

$N \rightarrow \text{muł} \{\text{Muł}\}$

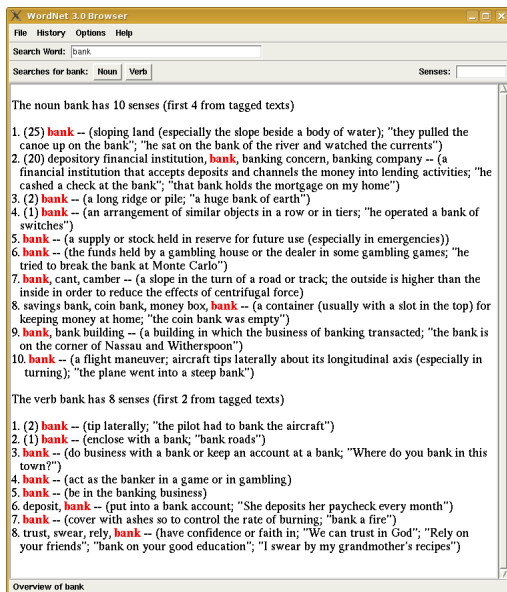
Baczyć, by muły nie parskąły!

Zażywano kąpeli siarczanych, w mule solnym, „wątrobie siarki”.

Ty mule!

# Związki między znaczeniem słów

- **synonim** (wyraz bliskoznaczny): silny – mocny.
- **antonim**: duży – mały.
- **homofon**: może – morze.
- **homogram**: on struga – ta struga – tego struga.
- **hiponim**: uczelnia → politechnika.
- **hipernim**: politechnika → uczelnia.
- **meronim**: dłoń → palec.
- **holonim**: palec → dłoń.



WordNet 3.0 Browser

File History Options Help

Search Word: bank

Searches for bank: Noun Verb Senses:

The noun bank has 10 senses (first 4 from tagged texts)

- (25) **bank** -- (sloping land (especially the slope beside a body of water); "they pulled the canoe up on the bank"; "he sat on the bank of the river and watched the currents")
- (20) depository financial institution, **bank**, banking concern, banking company -- (a financial institution that accepts deposits and channels the money into lending activities; "he cashed a check at the bank"; "that bank holds the mortgage on my home")
- (2) **bank** -- (a long ridge or pile; "a huge bank of earth")
- (1) **bank** -- (an arrangement of similar objects in a row or in tiers; "he operated a bank of switches")
- bank** -- (a supply or stock held in reserve for future use (especially in emergencies))
- bank** -- (the funds held by a gambling house or the dealer in some gambling games; "he tried to break the bank at Monte Carlo")
- bank**, cant, camber -- (a slope in the turn of a road or track; the outside is higher than the inside in order to reduce the effects of centrifugal force)
- savings bank, coin bank, money box, **bank** -- (a container (usually with a slot in the top) for keeping money at home; "the coin bank was empty")
- bank**, bank building -- (a building in which the business of banking transacted; "the bank is on the corner of Nassau and Witherspoon")
- bank** -- (a flight maneuver; aircraft tips laterally about its longitudinal axis (especially in turning); "the plane went into a steep bank")

The verb bank has 8 senses (first 2 from tagged texts)

- (2) **bank** -- (tip laterally; "the pilot had to bank the aircraft")
- (1) **bank** -- (enclose with a bank; "bank roads")
- bank** -- (do business with a bank or keep an account at a bank; "Where do you bank in this town?")
- bank** -- (act as the banker in a game or in gambling)
- bank** -- (be in the banking business)
- deposit, **bank** -- (put into a bank account; "She deposits her paycheck every month")
- bank** -- (cover with ashes so to control the rate of burning; "bank a fire")
- trust, swear, rely, **bank** -- (have confidence or faith in; "We can trust in God"; "Rely on your friends"; "bank on your good education"; "I swear by my grandmother's recipes")

Overview of bank

- *Kasia je schabowego*

$$\exists_{e,x} \text{Jedzenie}(e) \wedge \text{Jedzący}(e, \text{Kasia}) \wedge \text{Jedzony}(e, x) \wedge \text{Schabowy}(x)$$

- *Zosia myje talerz*

$$\exists_{e,x} \text{Mycie}(e) \wedge \text{Myjący}(e, \text{Zosia}) \wedge \text{Myty}(e, x) \wedge \text{Talerz}(x)$$

W obu przypadkach mamy do czynienia z wykonawcą czynności i przedmiotem czynności. Wykonawca nie musi być podmiotem:

- *Talerz jest myty przez Zosię*

$$\exists_{e,x} \text{Mycie}(e) \wedge \text{Myjący}(e, \text{Zosia}) \wedge \text{Myty}(e, x) \wedge \text{Talerz}(x)$$

- **wykonawca** – świadomie powoduje zdarzenie: *Kasia* czyta książkę
- **doznawacz** – osoba doświadczająca zdarzenia: *Kasię* boli głowa
- **zjawisko** – coś powodujące zdarzenie: *Wiatr* zerwał dach
- **przedmiot** – coś, co uczestniczyło w zdarzeniu i doświadczyło jego skutków: *Wiatr* zerwał *dach*
- **wynik** – wynik zdarzenia: *Związkowcy* ogłosili *strajk*
- **zawartość** – : *Powiedział: „wyjdz”*
- **narzędzie** – narzędzie użyte w zdarzeniu: *Pokroiła go nożem*
- **beneficjent** – ktoś odnoszący korzyść ze zdarzenia: *Zrób to dla mnie*
- **źródło** – punkt wyjściowy obiektu lub zdarzenia przemieszczania: *Wrócił z pracy*
- **Cel** – punkt docelowy obiektu lub przemieszczania: *Wrócił do domu*

Predicate *index*

*Frames file for 'index' based on survey of sentences in the WSJ corpus.*

**Roleset index.01 "to provide an index":**

**Roles:**

**Arg0:** *provider of an index*

**Arg1:** *entity indexed*

**Arg2:** *attribute of arg1, index value*

**Examples:**

person:*ns* tense:*present* aspect:*ns* voice:*passive* form:*participle*  
**passive (-)**

Based on historical rates of return , Mr. Anderson reckons [0]  
a 100 % stock portfolio , indexed [\*] to the market ,  
would have kept up with tuition and taxes in the 20th century .

**REL:** indexed

**Arg1:** [\*] a 100 % stock portfolio

**Arg2-to:** the market

*frames created by OLga*



Ustalanie znaczenia słów (ang. *word sense disambiguation* – *WSD*) polega na ustaleniu, które z możliwych znaczeń danego słowa zostało użyte w danym przypadku.

Zastosowania:

- tłumaczenie maszynowe
- automatyczne odpowiadanie na pytania
- wydobywanie informacji
- klasyfikacja tekstów

Ze względu na liczbę słów:

- z próbką słów (ang. *lexical sample*)
- dla wszystkich słów (ang. *all words*)

Ze względu na metodę uczenia:

- z nauczycielem (ang. *supervised*)
- słownikowe
- bez nauczyciela (ang. *unsupervised*)

Wektor cech może zawierać:

- **worek słów** (ang. *bag-of-words*): informacja o sąsiednich słowach słowa  $w$ , np.  $N$  słów najczęściej występujących w zdaniach z użyciem słowa  $w$ ; wektor zawiera 1 jeśli słowo z danej pozycji wektora występuje w otoczeniu słowa  $w$  w danym zdaniu.
- **wektor cech współwystępujących słów** (ang. *collocational features*): słowa z okna po  $N$  słów z każdej strony słowa  $w$  razem z wybranymi cechami tych słów, jak np. części mowy, formy podstawowe itp., np.  $[w_{i-2}, POS_{i-2}, w_{i-1}, POS_{i-1}, w_{i+1}, POS_{i+1}, w_{i+2}, POS_{i+2}]$ .

Do uczenia można wykorzystać dowolny algorytm uczenia maszynowego.

Do najpopularniejszych należą naiwny klasyfikator Bayesa i listy decyzyjne.

# Naiwny klasyfikator Bayesa

Chodzi nam o znalezienie najlepszego znaczenia  $\hat{s}$  z możliwych  $S$  mając dany wektor cech  $\vec{f}$ :

$$\hat{s} = \arg \max_{s \in S} P(s | \vec{f})$$

Stosując wzór Bayesa, pomijając stały mianownik, przy założeniu niezależności cech:

$$P(\vec{f} | s) \approx \prod_{j=1}^n P(f_j | s)$$

otrzymujemy:

$$\hat{s} = \arg \max_{s \in S} P(s) \prod_{j=1}^n P(f_j | s)$$

Wartości  $P(s_i)$  i  $P(f_j | s)$  liczymy ze wzorów:

$$P(s_i) = \frac{\text{liczba}(s_i, w_j)}{\text{liczba}(w_j)} \quad P(f_j | s) = \frac{\text{liczba}(f_j, s)}{\text{liczba}(s)}$$

```
1: function SIMPLIFIEDLESK(word,sentence)
2:   best-sense  $\leftarrow$  najczęstsze znaczenie słowa word
3:   max-overlap  $\leftarrow$  0
4:   context  $\leftarrow$  zbiór słów w zdaniu sentence
5:   for all sense in znaczenia słowa word do
6:     signature  $\leftarrow$  zbiór słów w opisie word
7:     overlap  $\leftarrow$  COMPUTEOVERLAP(signature,context)
8:     if overlap  $>$  max-overlap then
9:       max-overlap  $\leftarrow$  overlap
10:      best-sense  $\leftarrow$  sense
11:     end if
12:   end for
13:   return best-sense
14: end function
```

- **muł**<sup>1</sup> **1** Muł to zwierzę, którego matką jest klacz konia, a ojcem ogier osła. Muły są wytrzymałe i wykorzystywane przez człowieka do pracy. **2** Mówimy muł o kimś, kto jest nieinteligentny, ograniczony lub uparty. Słowo potoczne i obraźliwe.
- **muł**<sup>2</sup> Muł to grząski osad tworzący się na dnie rzeki, jeziora lub stawu. *Nogi grzęzły w mule.*

*To znaczy, nie mam nic przeciwko temu, co opisuje, ale drażni mnie za pomocą jednego słowa czegoś, co jest jak mętny osad na dnie doświadczenia, niby bezwartościowy muł, a tak naprawdę alchemiczny sekret każdej duszy godnej tego miana, cichy, niedający się wypowiedzieć wprost.*

- Jedną z możliwości jest liczenie odległości słów w słowniku na podstawie zawierania słów z opisu.
- Wersja algorytmu Leska oparta o oznaczone semantycznie zbiory tekstów (*corpus Lesk*) dodaje do sygnatury wszystkie słowa z okien lub zdań w zbiorze tekstów zawierające dane słowo w danym znaczeniu. Dodatkowo słowa mają wagi zależne od częstości występowania słów. Waga jest odwrotną częstotliwością słowa (ang. *inverse document frequency*):

$$idf_i = \log\left(\frac{Ndoc}{nd_i}\right)$$

gdzie  $Ndoc$  to całkowita liczba dokumentów (opisów haseł i przykładów), a  $nd_i$  to liczba tych dokumentów zawierających słowo  $i$ .