

Ujednoznacznianie sensu słów

(ang. word-sense disambiguation — WSD)

Niejednoznaczność języka

- Segmentacja tekstu

Czy **kropka** kończy zdanie czy jest elementem skrótu, np. "ul.":

- a. Na polu stoi **ul.** Pszczoły zbierają miód z rzepaku.
- b. Na polu przy **ul.** Pszczelej stoi rozbity samochód.

- Analiza morfologiczna

Czy słowo **miał** jest rzeczownikiem czy czasownikiem?

- a. Tomek **miał** kupić masło, ale zapomniał.
- b. Tomek kupił **miał** do ogrzewania domu.

- Znaczenie słów

Czy słowo **mysz** odnosi się do zwierzęcia czy urządzenia?

- a. Na strychu grasuje **mysz**.
- b. **Mysz** umożliwia poruszanie kursorem po ekranie.

Ujednoznacznianie sensu słów (WSD)

Ujednoznacznianie sensu słów zajmuje się problemem **niejednoznaczności** słów na poziomie ich znaczeń.



Podjęcia do WSD (1/2)

WSD



```
graph TD; WSD --> Nienadzorowane; WSD --> Nadzorowane;
```

Nienadzorowane

1. **Grupowanie wystąpień słów** z wykorzystaniem miary podobieństwa treści
2. Opiera się na założeniu, że **słowa o tym samym znaczeniu występują w podobnych kontekstach**.
3. **Nie wymaga danych uczących**, tj. tekstów z przypisanymi wzorcowymi znaczeniami słów.
4. **Główne trudności:**
 - a. określenie liczby znaczeń,
 - b. wyodrębnienie rzadkich znaczeń.

Nadzorowane

1. Zakłada, że **wszystkie znaczenia są znane i opisane**.
2. Zadanie sprowadza się do problemu **klasyfikacji**, tj. przypisania do danego słowa jednego ze znanych znaczeń.
3. **Wymaga danych uczących**, tj. tekstów z przypisanymi wzorcowymi znaczeniami słów.
4. **Główne trudności:**
 - a. opis wszystkich możliwych znaczeń słów,
 - b. dane uczące.

Podejścia do WSD (2/2)

Podejście częściowo nadzorowane:

1. Wykorzystanie **wzorcowej bazy znaczeń słów**.
2. Wykorzystanie **danych oznaczonych i nieoznaczonych** (znaczeniami słów).
3. Dane oznaczone są wykorzystywane głównie do testowania modelu.
4. Brak danych uczących rekompensowany jest **innymi zasobami językowymi**, np. **wordnet** jako źródło informacji o relacjach semantycznych między słowami.

Wordnet

Wordnet to relacyjny słownik semantyczny. Słownik ten odzwierciedla system leksykalny danego języka.

Główne elementy wordnetu:

1. **Jednostki leksykalne** — wyrażenia językowe mające określone znaczenie globalne, np. *mysz* (gryzoń) i *mysz* (urządzenie) to dwie różne jednostki leksykalne,
2. **Synsety** — zbiór jednostek leksykalnych o takim samym znaczeniu lub bardzo podobnym, nieodróżnialnym znaczeniu, np. *mysz* i *mysz komputerowa*
3. **Relacje** — wskazują relacje znaczeniowe między synsetami i jednostkami leksykalnymi.

Relacje w wordnecie

Podstawową strukturę wordnetu tworzą relacje między synsetami.

Główne typy relacji to:

1. **Hiperonimia** — np. *gryzoń* jest hiperonimem dla *mysz*,
2. **Hiponimia** — np. *mysz* jest hiponimem dla *gryzoń*,
3. **Meronomia** — np. *radiator* jest meronimem dla *silnik*,
4. **Holonimia** — np. *silnik* jest holonimem dla *radiator*.

Wordnety

Princeton Wordnet — wordnet dla j. angielskiego

- <https://wordnet.princeton.edu>
- jeden z pierwszych wordnetów,
- był podstawą tworzenia wordnetów dla innych języków.

SłowoSieć (plWordNet) — wordnet dla j. polskiego

- <http://plwordnet.pwr.edu.pl>
- obecnie największy wordnet,
- był tworzony od zera przez lingwistów.



Wykorzystanie Słownosieci do WSD (1/3)

Zbiór znaczeń słów

<http://plwordnet.pwr.edu.pl/wordnet/32131a4c-28e9-11eb-8b29-8b91085c3154>



Wykorzystanie Słownosieci do WSD (2/3)

Definicje i przykłady dla znaczeń

<http://plwordnet.pwr.edu.pl/wordnet/32131a4c-28e9-11eb-8b29-8b91085c3154>

🔍 kot

Oglądasz 1 z 6 dostępnych znaczeń tego słowa

RZECZOWNIK

kot 1

każdy ssak z rodziny kotowatych

definicja

DOMENA

zwierzęta

PRZYKŁADY

Na koniec wycieczki do zoo poszliśmy do pawilonu kotów, bardzo podobał mi się tam nowy wybieg dla lwów.

przykład

ŹRÓDŁO

Słownosieć

HIPERONIMY

[Pokaż ścieżkę do najwyższego hiperonimu](#)

GRAPH

[Wyświetl wizualizację graficzną](#)

ANNOTACJE
EMOCJONALNE

NACECHOWANIE Brak nacechowania emocjonalnego



NACECHOWANIE Brak nacechowania emocjonalnego

Wykorzystanie Słownosieci do WSD (3/3)

Słowa powiązane znaczeniowo

<http://plwordnet.pwr.edu.pl/wordnet/32131a4c-28e9-11eb-8b29-8b91085c3154>

<http://plwordnet.pwr.edu.pl/wordnet/99e9ba18-28e9-11eb-b594-83324b52f696>


Słownosieć


Oglądasz 1 z 6 dostępnych znaczeń tego słowa
Następne znaczenie --

Hiperonimy

ssak drapieżny 1

ssak z rzędu lądowych i wodnych ssaków lo...

Meronymy (części)

poduszczyk 6

zgrubienie, wzniesienie anatomiczne na pys...

Holonimy (taksonomiczne)

kotowate 1

domena: systematyka, klasyfikacja

Kolokacyjność

dziki 2

o roślach lub zwierzętach: taki, który nie je...

Synonimy międzyjęzykowe

cat 1

feline mammal usually having thick soft fur ...

RZECZOWNIK

kot 1

każdy ssak z rodziny kotowatych

DOMENA

zwierzęta

PRZYKŁADY

Na koniec wycieczki do zoo poszliśmy do pawilonu kotów, bardzo podobał mi się tam nowy wybieg dla lwów.

ŹRÓDŁO

Słownosieć

HIPERONIMY

Pokaż ścieżkę do najwyższego hiperonimu

GRAPH

Wyświetl wizualizację graficzną

ANNOTACJE EMOCJONALNE

NACECHOWANIE Brak nacechowania emocjonalnego

NACECHOWANIE Brak nacechowania emocjonalnego

Hiponimy

tygrew 1

domena: zwierzęta

jaglion 1

domena: zwierzęta

liard 1

domena: zwierzęta

pantera mglista 1

domena: zwierzęta

... Pokaż pozostałe 32 połączenia

Obszar

felinologia 1

dział zoologii koncentrujący się na badaniu...

Subiekt przy niewyrażonym predykanie

felinolog 1

specjalista w dziedzinie felinologii



Subiekt (subiekt V-N)

zamruczeć 2

domena: czasowniki mówienia, śpiewania itp.

mruczeć 2

o niektórych zwierzętach (szczególnie koc...


Słownosieć


-- Poprzednie znaczenie
Oglądasz 3 z 6 dostępnych znaczeń tego słowa
Następne znaczenie --

Hiperonimy

nowy 1

człowiek, który jest nowy w jakiej grupie, w ...

Hiponimy międzyjęzykowe

freshman 1

a first-year undergraduate

sprog 1

a new military recruit

Międzyjęzykowe synonimy międzyrejestrowe

raw recruit 1

an inexperienced and untrained recruit

Międzyjęzykowe synonimy częściowe

grunt 2

an unskilled or low-ranking soldier or other ...

RZECZOWNIK

kot 4

ktos nowo przyjeżdżający do danej grupy, jeszcze przed inicjacją, świeżak

DOMENA

ludzie

PRZYKŁADY

Otrzęsiny będą za tydzień, ale do tego czasu wszystkie koty powinny na siebie uważać.

Fala w wojsku dotyczy najczęściej tzw. kotów, ale jej ofiarą padają często słabsi żołnierze już po przysiędze.

ŹRÓDŁO

Słownosieć

HIPERONIMY

Pokaż ścieżkę do najwyższego hiperonimu

GRAPH

Wyświetl wizualizację graficzną

ANNOTACJE EMOCJONALNE

NACECHOWANIE Słabe negatywne

EMOCJE wstręt, złość

WARTOŚCIOWANIE niewiedza

PRZYKŁADY Marysia jest z drugiej, czy to kot?

NACECHOWANIE Słabe negatywne

EMOCJE złość, wstręt

WARTOŚCIOWANIE nieużyteczność, niewiedza

PRZYKŁADY Patrz na tę grupę przestraszonych kotów.

Fuzzynymy synsetów

fala 6

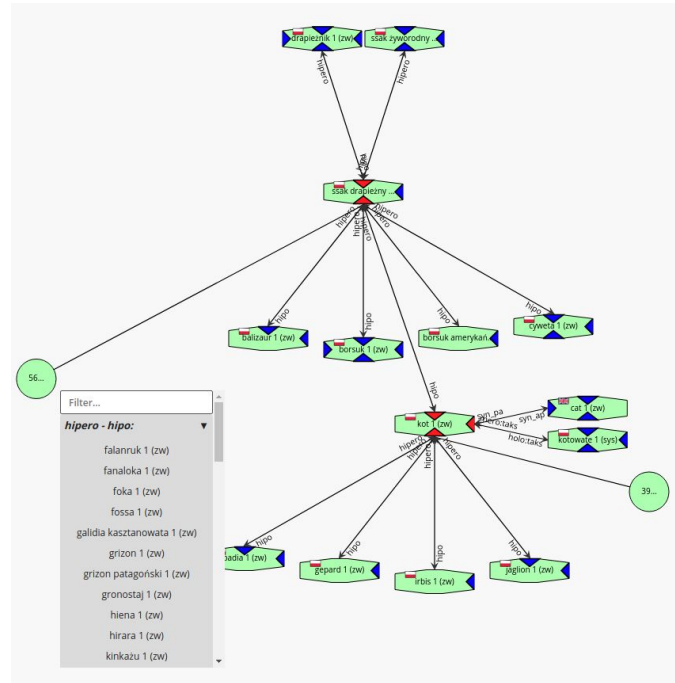
przemoc w wojsku lub w szkole w stosunku ...

otrzęsiny 1

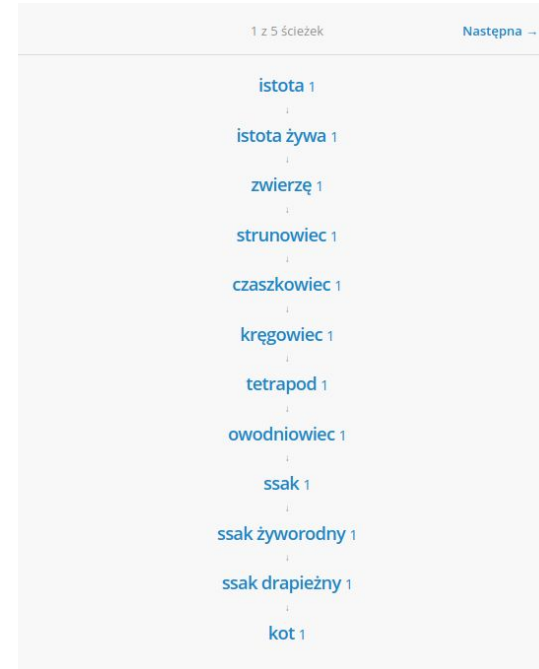
przyjście nowego pracownika, grupy praco...

Wykorzystanie Słownosieci do WSD (3/3)

Wizualizacja graficzna



Ścieżka do najwyższego hiperonimu



Dostępność Słownosieci

1. Dostęp on-line:

<http://plwordnet.pwr.edu.pl/wordnet>

2. Do pobrania

<http://plwordnet.pwr.edu.pl/wordnet/download>