



AKADEMIA GÓRNICZO-HUTNICZA  
IM. STANISŁAWA STASZICA W KRAKOWIE



# Praca magisterska

**Automatyczna klasyfikacja, grupowanie  
i znajdowanie uogólnień dla języka polskiego.**

Autor: Łukasz Wróbel

Promotor: dr Adrian Horzyk

**Wydział Elektrotechniki, Automatyki, Informatyki i Elektroniki  
Katedra Automatyki**

## Cele pracy:



- Utworzenie specjalistycznego grafu przyzwyczajzeń lingwistycznych obrazującego relacje między wyrazami w języku polskim
- Wykorzystanie korpusów tekstów oraz Internetu
- Zamodelowanie przechodniości cech w ramach hierarchicznej struktury wyrazów w oparciu o relacje grupowania i uogólniania

## Założenia projektowe:



- Relacje tworzone są tylko pomiędzy rzeczownikami, czasownikami, przymiotnikami, przysłówkami i liczebnikami.
- Podawane przez użytkownika wyrazy są poprawne i zawarte w słowniku biblioteki CLP

## Opis rozwiązania

- Pająk internetowy przeszukujący strony o wybranej tematyce
- Ekstraktor danych tworzący listę potencjalnych składników relacji
- Kreator relacji oceniający ich poprawność
- Uproszczony graf przyzwyczajień lingwistycznych (LHG) obrazujący strukturę relacji
- Mechanizm propagujący relacje w ramach hierarchii

Plik Pomoc

Generacja słownika

Online  Offline  Symuluj

Pobierz zdania z bazy danych

Pobierz relacje z bazy danych

Wyniki:

Zapisz zdania do bazy danych

Zapisz relacje do bazy danych

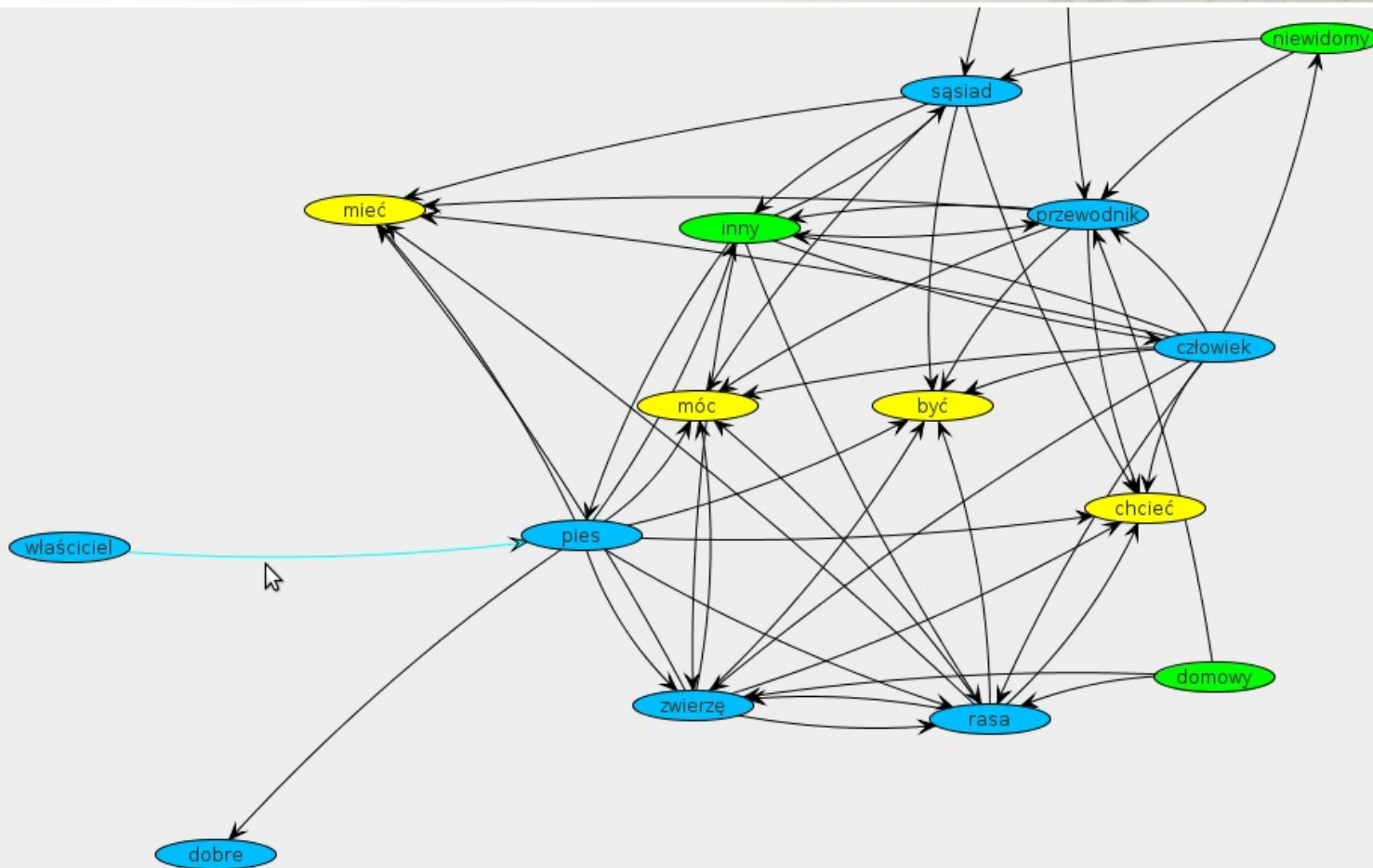
Ilość wyników z google wykorzystywanych w tworzeniu słownika:

Maksymalna ilość powiązań dla każdego wyrazu drzewa:

Maksymalna ilość relacji spropagowanych:

Słowo startowe (korzeń drzewa):

# Utworzony specjalistyczny graf LHG



Zaznaczone relacje i słowa:

- Grupowanie: właściciel to pies - 371
- Posiadanie: właściciel psa - 480000
- Posiadanie: właściciel ma psa - 8190

## Najważniejsze funkcjonalności

- Przeszukiwanie stron o tematyce związanej z podanym słowem
- Wybór potencjalnych relacji na podstawie częstości występowania w analizowanym tekście
- Badanie poprawności relacji z wykorzystaniem popularności wyrażenia w internecie
- Tworzenie potencjalnych relacji na podstawie przechodniości cech w ramach uogólnień wewnątrz powstałej struktury

## Podsumowanie

- Utworzone relacje, jak i mechanizm ich tworzenia mogą zostać wykorzystane przez inteligentnego czatbota
- Uproszczona postać grafu LHG obrazuje wycinek struktury relacji w sposób czytelny dla użytkownika
- Odwzorowanie trendów językowych panujących w internecie