



# Jasnopis - demonstracja narzędzia do badania zrozumiałości polskiego tekstu

*Bartłomiej Nitoń  
bartek.niton@gmail.com*

# Plan prezentacji

- Formaty wejściowe
- Co wskazuje Jasnopis?
- Klasy trudności
- Podpowiedzi
- Indeksy zrozumiałości
- Statystyki
- Wyszukiwanie modelu językowego tekstu
- Grafy podobieństwa leksykalnego
- Wtyczka OpenOffice

# Formaty wejściowe

- Tekst wprowadzany do pola tekstowego
- Adres strony internetowej
- Plik załadowany z komputera użytkownika w praktycznie dowolnym formacie:
  - formaty wspierane przez Open Office (m.in. .doc, .rtf, .odt, .txt)
  - pdf
  - inne

# Co wskazuje Jasnopis?

- Klasy trudności
- Trudne słowa
  - O 4 lub większej liczbie sylab z pominięciem słów uznanych za ogólnie znane
- Zbyt długie zdania
- Zbyt trudne akapity:
  - W stosunku do trudności całego tekstu
  - W stosunku do wybranej trudności
- Propozycje zmian
- Indeksy zrozumiałości i statystyki
- Grafy podobieństwa leksykalnego
- Model językowy akapitów
  - Jego zgodność ze stylem całego tekstu
  - Jego zgodność z wybranym stylem

# Interpretacja klas trudności tekstu

- 1 – klasy 1-3 szkoły podstawowej
- 2 – klasy 4-6 szkoły podstawowej
- 3 – gimnazjum
- 4 – liceum
- 5 – licencjat/inżynier (ok. 3 pierwszych lat studiów)
- 6 – magister
- 7 – doktorat i trudniejsze

# Podpowiedzi

- Zaznaczenie zdania, które należy skrócić
- Informacja, że akapit jest zbyt trudny w stosunku do całego dokumentu lub wybranej docelowo grupy odbiorców
- Propozycje zastąpienia trudnych słów (prezentowane są jedynie prostsze alternatywy):
  - Synonimami
  - Hiponimami
  - Hiperonimami

# Indeksy zrozumiałości

- Klasa trudności
- Regresja grzbietowa – autorski indeks, służący do wyprowadzenia klasy trudności
- Trzy typy indeksu FOG
  - Oparty na formach podstawowych (np. robić) lub tekstowych słów (np. robię)
  - Wygładzony
- Sześć typów indeksu Pisarka
  - Z podziałem na wersję liniową i nieliniową
  - Z podziałem na wykorzystanie form podstawowych lub tekstowych
  - Liniowa i nieliniowa wersja z wygładzeniem



# Statystyki

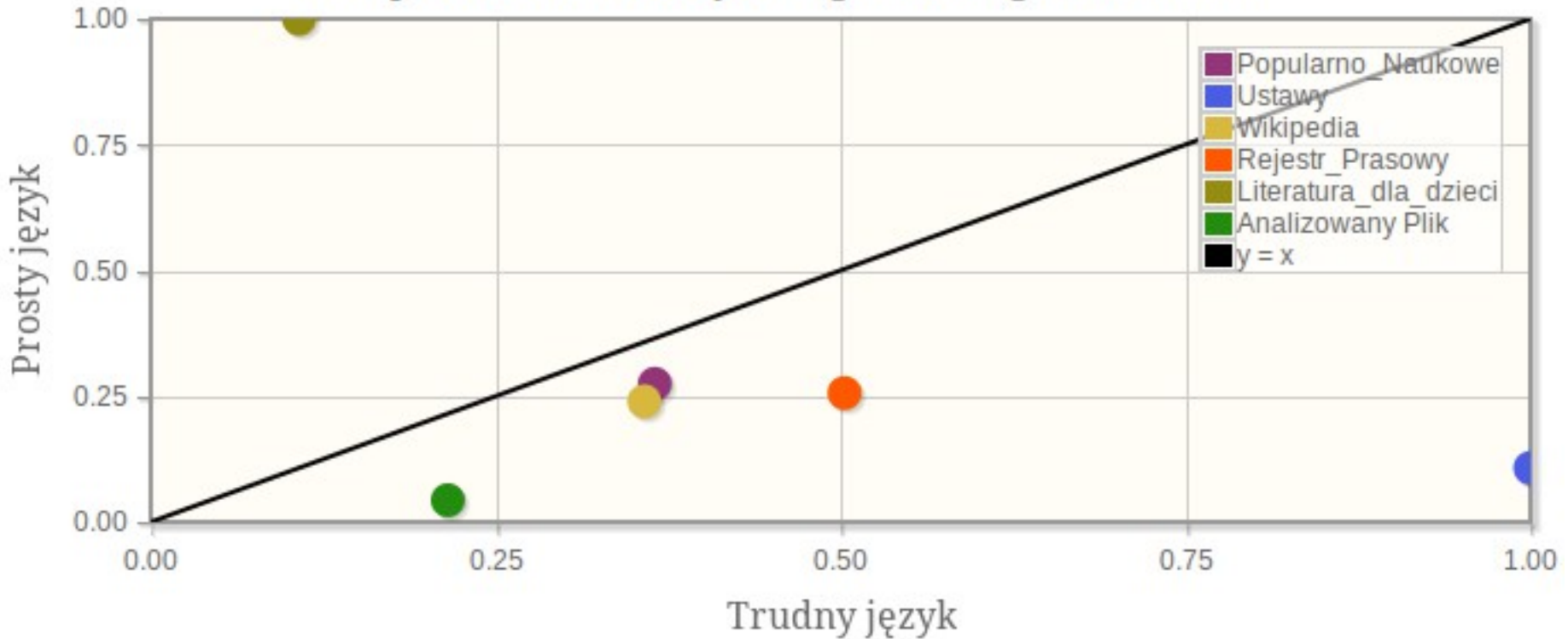
- Liczba akapitów
- Liczba zdań
- Liczba słów
- Liczba słów trudnych (z pominięciem słów uznanych za powszechnie znane)
- Średnia długość słowa (w sylabach)
- Średnia długość zdania
- Średnia długość akapitu
- Procent słów trudnych
- Procent rzeczowników
- Procent rzeczowników trudnych
- Procent czasowników
- Procent czasowników trudnych
- Procent przymiotników
- Procent przymiotników trudnych
- Stosunek rzeczowników do czasowników

# Model językowy tekstu

- Sprawdzany jest styl całego dokumentu i poszczególnych akapitów
- Korpusy, do których może być podobny tekst:
  - Literatura dla dzieci
  - Wikipedia
  - Rejestr prasowy
  - Teksty popularno-naukowe
  - Ustawy

# Grafy podobieństwa leksykalnego

Graf podobieństwa leksykalnego badanego dokumentu



# Wtyczka OpenOffice

Jeszcze tego samego dnia dzieci poznały panią Jagę. Siedziały w dużym salonie, w wielkich fotelach. Jak dorośli. Czekali, aż pani Jaga przyniesie herbatę. Dynia rozparła się wygodnie i majtała nogami. Stefanek trzymał się sztywno, jakby bał się poruszyć. Myśleli tylko o jednym. Czy ona jest czarownicą? Dynia miała nadzieję, że tak. Stefanek powtarzał sobie w duchu, że nie. Oboje rozzglądali się po mieszkaniu. Panował w nim porządek. Wszystko było czyste i ułożone w idealnym szyku. Nawet kot siedział cicho w rogu kanapy, a jego futro błyszczało. Tylko w kącie znajdowało się coś, co burzyło wygląd salonu. Śliczna, nowa miotła.

O! — wykrzyknęła Dynia, pokazując palcem na miotłę.

Pani J  
moment oczo  
kolorach i typ

krzyknęła  
wrzasnęła

Ignore  
Ignore All

Add

AutoCorrect

ABC Spelling and Grammar

Set Language for Selection

Set Language for Paragraph

a drzwi szafy i wrzuciła ją do środka. Na krótki  
i stały poukładane równo miotły, w różnych  
można było latać.

FOG-Base: 2.88
Readability Today, 20:57
FOG-Base: 8.51
Readability Today, 20:57
Zdanie jest za długie.
FOG-Base: 5.33
Readability Today, 20:57

Dziękuję za uwagę