

Mierzenie stopnia zrozumiałości polskich tekstów użytkowych

Włodzimierz Gruszczyński, Bartosz Broda,
Łukasz Dębowski, Bartłomiej Nitoń, Maciej Ogrodniczuk

Mierzenie stopnia zrozumiałości polskich tekstów użytkowych (pozaliterackich)

- Projekt finansowany przez NCN
- DEC-2011/03/B/HS2/05799 z dn. 2012-06-20
- Czas trwania: 24 miesiące (08.2012 – 08.2014)
- Środki finansowe: 390.215,00 PLN
- Jednostka koordynująca:
Wydział Kulturoznawstwa i Filologii
Szkoly Wyższej Psychologii Społecznej.

Zespół realizujący projekt

- Włodzimierz Gruszczyński (kierownik, SWPS)
- Bartosz Broda (głównie Samsung)
- Edyta Charzyńska (SWPS, psychologia)
- Łukasz Dębowski (IPI PAN)
- Milena Hadryan (UAM)
- Bartłomiej Nitoń (IPI PAN)
- Maciej Ogrodniczuk (IPI PAN)
- Grażyna Oblas (UW, wsparcie logistyczne)

Zbyt trudne teksty w przestrzeni publicznej

- Instrukcje obsługi różnych urządzeń, m.in. instrukcje bhp.
- Ulotki dołączane do farmaceutyków i kosmetyków.
- Akty prawne, np. przepisy podatkowe.
- Pisma urzędowe.
- Niektóre teksty medialne.
- Niektóre podręczniki szkolne.

Decyzja Miejskiego Ośrodka Pomocy Rodzinie

- Miejski Ośrodek Pomocy Rodzinie podwyższa dodatek do zasiłku rodzinnego z tytułu wychowywania dziecka w rodzinie wielodzietnej na dzieci: Jadwigę B., Wandę B., i Ignacego B. o kwotę 100,- zł miesięcznie od 1 września 2007 do 31 sierpnia 2008.
- Podwyższony dodatek w całkowitej wysokości 180,- zł miesięcznie na każde z wymienionych dzieci będzie Pani otrzymywać w okresie od 01.09.2007 do 31.08.2008.
- **Uzasadnienie:** Na Pani wniosek z dnia 6.09.2007 przyznano Pani dodatek do zasiłku rodzinnego z tytułu wychowania dziecka w rodzinie wielodzietnej w wysokości 80 zł miesięcznie w okresie od 01.09.2007 do 31.08.2008 na troje dzieci: Jadwigę, Wandę i Ignacego – decyzja z dn. ...
- Dodatek ten podniesiono na podstawie Uchwały Rady Miasta CIII/1198/IV/2006 ... o kwotę 100 zł na trzecie i następne dzieci uprawnione do zasiłku rodzinnego. Podwyższony podatek finansowany będzie ze środków własnych gminy...

Jak zapobiegać powstawaniu takich tekstów?

- W Departamencie Obrony Stanów Zjednoczonych ocenia się automatycznie stopień zrozumiałości wszystkich dokumentów tworzonych na potrzeby Departamentu!
- Na całym świecie wydaje się tzw. tabloidy, które cechują się m.in. tym, że teksty w nich publikowane są zrozumiałe dla większości odbiorców.
- W wielu krajach sprawdza się (różnymi metodami) stopień zrozumiałości tekstów, które mają funkcjonować w przestrzeni publicznej (napisy informacyjne, zwłaszcza ostrzegawcze; instrukcje obsługi urządzeń, instrukcje wypełniania deklaracji podatkowej, podręczniki itp.).
- W Szwecji wdrażany jest przez instytucje państwowe i samorządowe program jasnego języka (*klarspråk*).



A w Polsce...

- Bardzo niewiele osób i instytucji zdaje sobie sprawę z problemu zrozumiałości tekstu, a zwłaszcza z tego, że zrozumiałość można obiektywnie mierzyć.
- Zazwyczaj uznaje się, że „wina” leży po stronie odbiorcy (brak wykształcenia, trudności ze skupieniem, nieznajomość podstawowej terminologii itp.).
- Bada się stopień rozumienia tekstów czytanych w ramach Międzynarodowego Badania Kompetencji Osób Dorosłych PIAAC (ang. the Programme for the International Assessment of Adult Competencies) za pomocą tekstów pełniących te same funkcje komunikacyjne w różnych społecznościach językowych, np. za pomocą instrukcji, choć są one różnie pisane w różnych językach.

A jednak coś się zmienia...

- Minister Zdrowia 26 kwietnia 2010 r. wydał rozporządzenie w sprawie badania czytelności ulotek dołączanych do lekarstw.
- Powstała inicjatywa społeczna „Język urzędowy przyjazny dla obywatela” Biura Rzecznika Praw Obywatelskich, Mazowieckiego Urzędu Wojewódzkiego i Rady Języka Polskiego (współorganizatorzy: Kancelaria Senatu RP, Departament Służby Cywilnej KPRM, Narodowe Centrum Kultury i Fundacja Języka Polskiego; Prezydent RP objął nad projektem patronat honorowy).

Inicjatywy i badania naukowe

- Już w latach 60. temat podjął prof. Walery Pisarek - językoznawca i prasoznawca: stworzył metodę obliczania stopnia zrozumiałości informacyjnych tekstów prasowych.
- W ostatnich latach powstała Pracownia Prostej Polszczyzny i Ruch Prostego Języka – powołane i prowadzone na Uniwersytecie Wrocławskim przez dra Tomasza Piekota.
- Powstają monografie, m.in. Mileny Hadryan.
- Trwają prace w ramach naszego projektu finansowanego przez NCN.

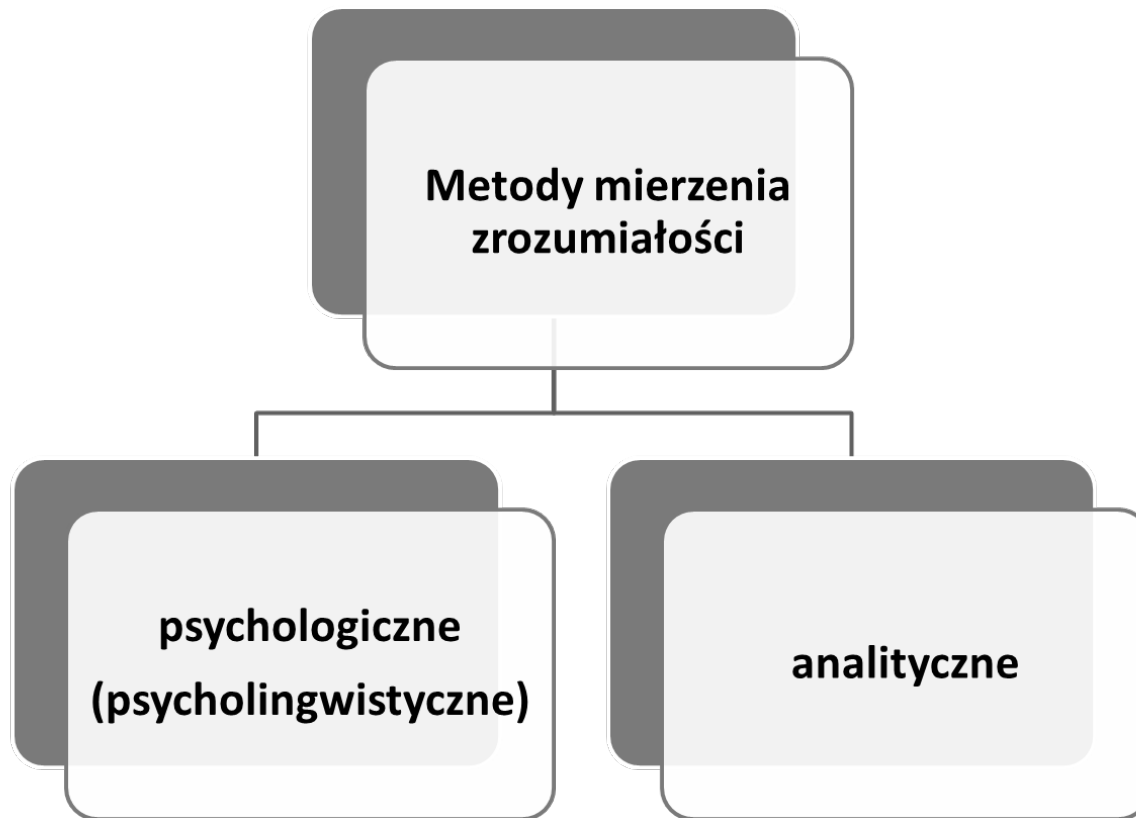
Cele projektu

- Ustalenie za pomocą analizy lingwistycznej i badań psycholingwistycznych takich cech tekstu (użytkowego), które mają wpływ na jego zrozumiałość (ang. *readability*).
- Stworzenie narzędzi informatycznych służących do mierzenia stopnia zrozumiałości tekstu oraz takich, które umożliwiają ustalenie innych wskaźników stylistyczno-statystycznych m. in.: bogactwo słownictwa, oryginalność słownictwa, subiektywność tekstu, egotyzm i ewentualnie innych (słowa kluczowe, słowa „sztandarowe”, stopień nacechowania emocjonalnego itp.).

Co to znaczy „rozumieć tekst”?

- Trudności z definicją.
- Błędne koło: „rozumieć tekst” – to ‘pojmować treść przekazywanej w nim informacji’, a „pojmować” – to ‘ogarniać rozumem, rozumieć’...
- Czy można ten sam tekst różnie rozumieć?

Metody psychologiczne i analityczne



Przykłady testów psychologicznych

- Sprawdzanie umiejętności stwierdzenia, czy tekst jest prawdziwy, czy fałszywy.
- Porównywanie czasu reakcji na polecenie związane z tekstem.
- Sprawdzanie znajomości treści tekstu za pomocą pytań kontrolnych.
- Sprawdzanie umiejętności uzupełniania brakujących fragmentów tekstu (tzw. *Cloze procedure*), czyli test Taylora.
- Prośba o intuicyjną ocenę stopnia trudności tekstu (umieszczenie na skali).

Czas potrzebny do oceny prawdziwości zdania w zależności od jego budowy

Typ zdania	Czas reakcji	Liczba błędów
Twierdzące czynne	4,5 s.	58
Twierdzące bierne	9,4 s.	55
Przeczące czynne	9,5 s.	88
Przeczące bierne	11,1 s.	117

Uzupełnij tekst...

- Wlazł kotek na płotek ____ mruga,
Ładna to piosenka ____.
Niedługa, niekrótka, jak w ____ raz,
Zaśpiewaj koteczku jeszcze ____.
- Nie będziemy omawiać bardziej ____ wszystkich tych modeli
z ____ względu, że dla psychologa ____ one wartość głównie
heurystyczną, ____ modele, z których wynikają ____
funkcjonowania umysłu ludzkiego, i ____ tego typu modeli
rozważyliśmy ____ przykładzie teorii Chomsky'ego.

Ograniczenia tego typu testów

- Czas- i pracochłonność oraz związana z tym wysoka cena badań.
- Ograniczenie do badania zrozumiałości krótkich tekstów (w wypadku niektórych metod – pojedynczych zdań).
- Trudności interpretacji wyników – wyniki testów świadczą zarówno o samych tekstach (ich trudności), jak i o poziomie inteligencji oraz wiedzy respondentów.

Wzór Pisarka (2007)

$$T = \frac{\sqrt{T_s^2 + T_w^2}}{2}$$

gdzie:

T – stopień trudności tekstu,

T_s – przeciętna długość zdania mierzona w wyrazach

T_w – procent wyrazów „trudnych” (czterosylabowych lub dłuższych)

Trudność językowa tekstów w niektórych czasopismach wg Pisarka

Tabela 1. Trudność językowa tekstów w niektórych czasopismach

W czasopismach	przeważają teksty	o wskaźniku trudności
„Przyjaciółka”, „Tina”	Bardzo łatwe	około 6,8
„Mały Gość Niedzielny”	Łatwe	około 8,3
„Przekrój”, „Gość Niedzielny”, „Wprost”, „Tyg. Powsz.”, „Polityka”	Średnie	10,6; 10,7; 10,8; 10,9; 12,4
„Zeszyty Prasoznawcze”	Trudne	około 15,2
„Biuletyn Pol. Tow. Językoznawczego”	Bardzo trudne	około 16,5

Zastosowanie metody Pisarka

Panel wysokiego szczebla – powołany w celu **doradzania** Komisji Europejskiej w zakresie **funkcjonowania** wspólnego programu Ambient Assisted Living (wspólny program w zakresie życia wspieranego przez **otoczenie**) – rozpoczął **konsultacje**, w ramach których poproszono **obywateli, przedsiębiorstwa** oraz **naukowców** o **podzielenie** się pomysłami, jak najlepiej **wykorzystać technologie informacyjne i komunikacyjne** (TIK), aby pomóc osobom starszym w bardziej **niezależnym** życiu, a w szerszym ujęciu, aby **ustanowić** nowe sposoby **wykorzystania rozwiązań** TIK na rzecz najsłabszych członków **społeczeństwa**./ Panelowi **przewodniczy** była komisarz **europejska** Meglena Kunewa./ Konsultacje publiczne stanowią pierwszy krok w kierunku **osiągnięcia** celu **wyznaczonego** przez agendę cyfrową (zob. IP/10/581), jakim jest **podwojenie** do 2015 r. liczby osób starszych **korzystających z udogodnień** umożliwiających niezależne życie./

Indeks FOG dla j. angielskiego

Gunning (1952) zaproponował wzór:

$$\text{Indeks FOG} = 0.4 \times (\text{ASL} + \text{PHW})$$

gdzie:

ASL – średnia długość zdania w słowach

PHW – procent słów dłuższych niż 2 sylaby

Automatyczny test Taylora

- Metoda Taylora – klasyczna metoda mierzenia czytelności poprzez uzupełnianie luk w tekście przez użytkowników języka.
- Wytrenowanie bigramowych modeli językowych na korpusach referencyjnych.
- Uzupełnianie luk w tekście z wykorzystaniem modeli językowych.
- Warianty:
 - Uzupełnianie co n-tego słowa poprzez model
 - Mierzenie odwrotności entropii (perplexity)

Podobieństwo

- Podobieństwo pomiędzy korpusami referencyjnymi a tekstem użytkownika
- Worek słów
- Kosinus kąta pomiędzy wektorami jako miara podobieństwa
- Dwa modele
 - tf.idf
 - model binarny
- Porównanie na poziomie leksyki

Korpusy tekstów

- **Literatura dla dzieci** – własnoręcznie zbierane bajki dla dzieci
- **Wikipedia** – teksty pochodzące z korpusu polskiej wikipedii (<http://clip.ipipan.waw.pl/PolishWikipediaCorpus>),
- **Rejestr Prasowy** – teksty pochodzące z korpusu „Rzeczpospolitej” (<http://www.cs.put.poznan.pl/dweiss/rzeczpospolita>)
- **Ustawy** – teksty polskich ustaw
- **Popularnonaukowe** – artykuły z archiwum “Wiedzy i Życia” (<http://archiwum.wiz.pl/>).

Korpusy tekstów (rozmiary)

- **Literatura dla dzieci** – mały: 33 teksty, duży: 177 tekstów;
- **Wikipedia** – mały: 41 tekstów, duży: 180 tekstów;
- **Rejestr Prasowy** – mały: 46 tekstów, duży: 180 tekstów;
- **Ustawy** – mały: 24 ustawy, duży: 175 ustaw;
- **Popularnonaukowe** – mały: 39 tekstów, duży: 183 teksty.

Badania psycholingwistyczne

W ramach projektu Edyta Charzyńska przeprowadziła obszerne studium zrozumiałości tekstów:

- 15 tekstów
- 1309 respondentów (zróżnicowanych demograficznie)
- 3 testy (cloze, multiple choice, pytania otwarte)

W następnej kolejności zbadano korelację wyników testów ze statystykami tekstów takimi jak:

- średnia długość zdania w słowach
- średnia entropia słowa
- średnia długość słowa w sylabach

Analiza badań psycholingwistycznych

Główne wyniki:

- Trudność tekstu jest silnie dodatnio skorelowana ze średnią długością zdania.
- Korelacja między trudnością tekstu a średnią długością słowa oraz średnią entropią słowa jest słaba.

Prawdopodobna przyczyna:

Średnia długość zdania jest silnie ujemnie skorelowana ze średnią długością słowa oraz ze średnią entropią słowa.
(Być może próba tekstów nie została dobrze dobrana.)

Kilka formuł regresji liniowej

Cloze = $59 - 4.6 (\text{dł. słowa}) - 1.4 (\text{dł. zdania})$

Cloze = $75 - 2.5 (\text{entropia}) - 1.4 (\text{dł. zdania})$

Otwarte = $2 - 0.2 (\text{dł. słowa}) - 0.04 (\text{dł. zdania})$

Otwarte = $2.9 - 0.13 (\text{entropia}) - 0.04 (\text{dł. zdania})$

Dla Multiple Choice współczynniki są bardzo małe albo mają nieprawidłowy znak.

Korelacja wyników badań z FOG-iem

CLOZE vs FOG-Orth: -0.710

CLOZE vs NL-Pisarek-Orth: -0.728

CLOZE_PROCENTOWY vs FOG-Orth: -0.867

CLOZE_PROCENTOWY vs NL-Pisarek-Orth: -0.875

CLOZE_SUBIEKTYWNY vs FOG-Orth: 0.796

CLOZE_SUBIEKTYWNY vs NL-Pisarek-Orth: 0.789

CLOZE_ZROZUMIENIE vs FOG-Orth: -0.867

CLOZE_ZROZUMIENIE vs NL-Pisarek-Orth: -0.860

Korelacja wyników badań z FOG-iem

MULTIPLE_CHOICE vs FOG-Orth: -0.489

MULTIPLE_CHOICE vs NL-Pisarek-Orth: -0.467

MULTIPLE_CHOICE_SUBIEKTYWNY vs FOG-Orth: -0.489

MULTIPLE_CHOICE_SUBIEKTYWNY vs NL-Pisarek-Orth: -0.507

MULTIPLE_CHOICE_TRUDNOSC vs FOG-Orth: 0.528

MULTIPLE_CHOICE_TRUDNOSC vs NL-Pisarek-Orth: 0.550

MULTIPLE_CHOICE_ZROZUMIENIE vs FOG-Orth: -0.592

MULTIPLE_CHOICE_ZROZUMIENIE vs NL-Pisarek-Orth: -0.610

Korelacja indeksów zrozumiałości

	Korelacja Spearmana
P(n,o) i P(l,o)	0,99
P(n,b) i FOG(b)	0,96
P(n,o) i FOG(o)	0,96
P(l,b) i FOG(b)	0,97
P(l,o) i FOG(o)	0,97
FOG(o) i FOG(b)	0,94

P(n,o) – nieliniowa wersja indeksu Pisarka oparta na formach ortograficznych słów;

P(l,o) – liniowa wersja indeksu Pisarka oparta na formach ortograficznych słów;

P(n,b) – nieliniowa wersja indeksu Pisarka oparta na formach bazowych słów;

P(l,b) – liniowa wersja indeksu Pisarka oparta na formach bazowych słów;

FOG(o) – indeks FOG oparty o formy ortograficzne słów;

FOG(b) – indeks FOG oparty o formy bazowe słów.

Badania przeprowadzane były z wykorzystaniem tekstów z milionowego podkorpusu NKJP (<http://clip.ipipan.waw.pl/LRT?action=AttachFile&do=view&target=NKJP-PodkorpusMilionowy-1.1.tgz>)

Korelacja indeksów zrozumiałości ze współczynnikami koreferencji

	Średnia długość łańcucha koreferencyjnego	Liczba nawiązań (bez singletonów)/ słowa
P(n,o)	-0.43	-0.43
P(l,o)	-0.46	-0.44
P(n,b)	-0.40	-0.43
P(l,b)	-0.44	-0.44
FOG(o)	-0.47	-0.44
FOG(b)	-0.45	-0.43

Badania przeprowadzane były z wykorzystaniem tekstów z polskiego korpusu koreferencyjnego
(<http://zil.ipipan.waw.pl/PolishCoreferenceCorpus>)

Badanie podobieństwa leksykalnego dokumentów

	Podobieństwo leksykalne (binary)	Podobieństwo leksykalne (tfidf)
Literatura dla dzieci	100,00%	100,00%
Wikipedia	85,37%	85,37%
Ustawy	100,00%	100,00%
Rejestr prasowy	71,74%	73,91%
Popularnonaukowe	100,00%	100,00%

Tabela prezentuje procentowy udział dokumentów dobrze zaklasyfikowanych do odpowiednich korpusów dla walidacji krzyżowej leave-one-out i małych korpusów tekstów.

Badanie podobieństwa dokumentów za pomocą automatycznych testów Taylora

	Taylor (perplexity)	Taylor (hit count)
Literatura dla dzieci	97,18%	93,79%
Wikipedia	61,11%	80,56%
Ustawy	100,00%	86,29%
Rejestr prasowy	66,11%	71,66%
Popularnonaukowe	68,31%	73,77%

Tabela prezentuje procentowy udział dokumentów dobrze zaklasyfikowanych do odpowiednich korpusów dla walidacji krzyżowej leave-one-out i dużych korpusów tekstów.

Dostępne narzędzia informatyczne

- Program stworzony w Pracowni Prostego Języka obliczający indeks mglistości tekstu (FOG Index),
- Program stworzony w ramach naszego projektu obliczający indeks mglistości i indeks wg metody Pisarka (oba w wariantach na formach bazowych i testowych) oraz wskazujący na podobieństwo gatunkowe do pewnych typów tekstów, np. ustaw lub haseł z Wikipedii.

Jasnopis (interfejs roboczy)

Readability
Szkoła Wyższa Psychologii Społecznej

APLIKACJA **INFORMACJE** **KONTAKT**

Akcje

- ☒ Analizuj tekst
- ☐ Analizuj URL
- ☐ Analizuj plik

Dane

Treść:

Panel wysokiego szczebla – powołany w celu doradzania Komisji Europejskiej w zakresie funkcjonowania wspólnego programu Ambient Assisted Living (wspólny program w zakresie życia wspieranego przez otoczenie) – rozpoczął konsultacje, w ramach których poproszono obywateli, przedsiębiorstwa oraz naukowców o podzielenie się pomysłami, jak najlepiej wykorzystać technologie informacyjne i komunikacyjne (TIK), aby pomóc osobom starszym w bardziej niezależnym życiu, a w szerszym ujęciu, aby ustanowić nowe sposoby wykorzystania rozwiązań TIK na rzecz najsłabszych członków społeczeństwa. Panelowi przewodniczy była komisarz europejska Meglena Kuneva. Konsultacje publiczne stanowią pierwszy krok w kierunku osiągnięcia celu wyznaczonego przez agendę cyfrową (zob. IP/10/581), jakim jest podwojenie do 2015 r. liczby osób starszych korzystających z udogodnień umożliwiających niezależne życie.

Typ przetwarzania:

Wykształcenie odbiorcy [lata]:

Wynik

Oczekiwanie na żądanie

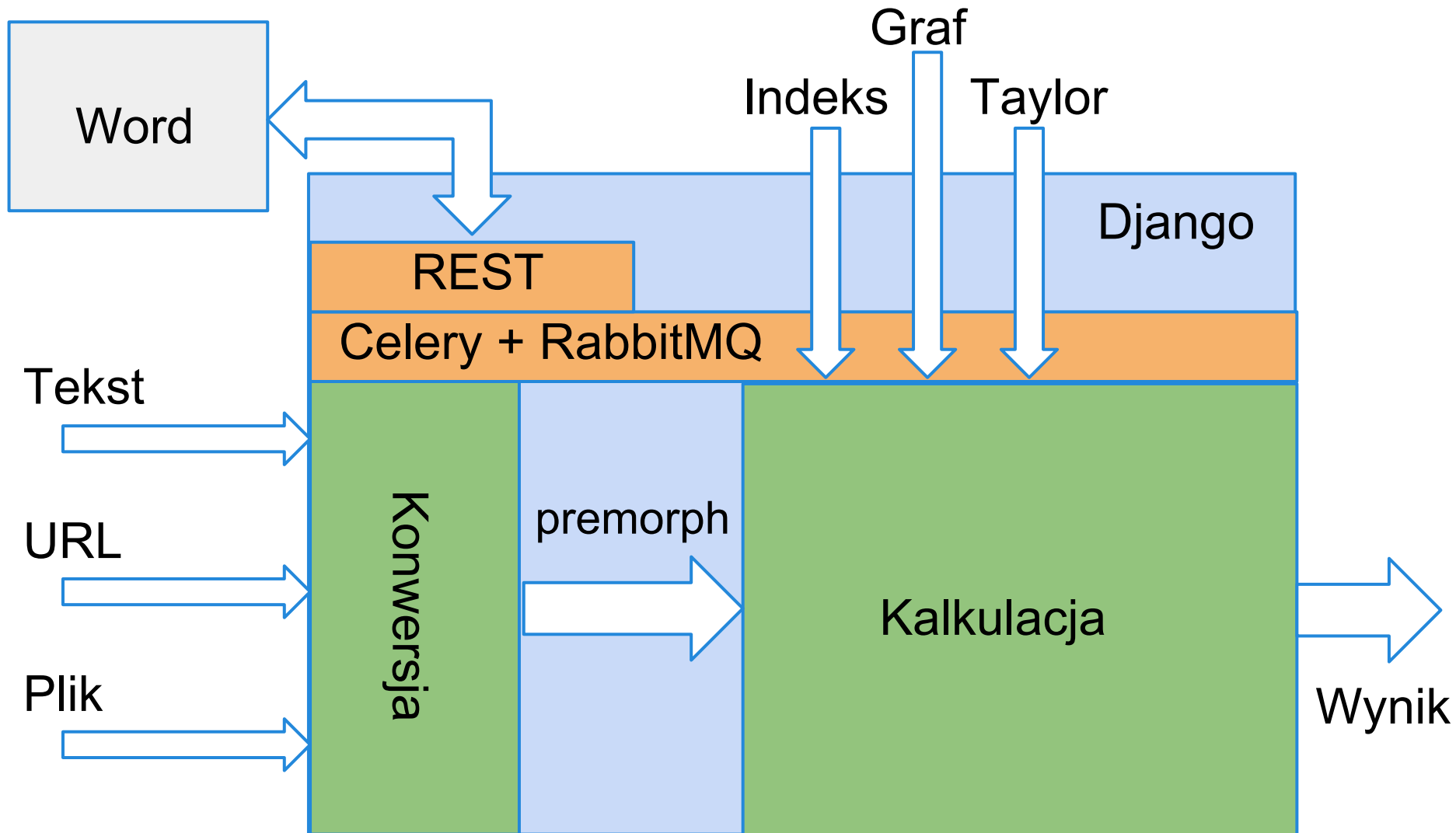
Jasnopis (formaty wejściowe)

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

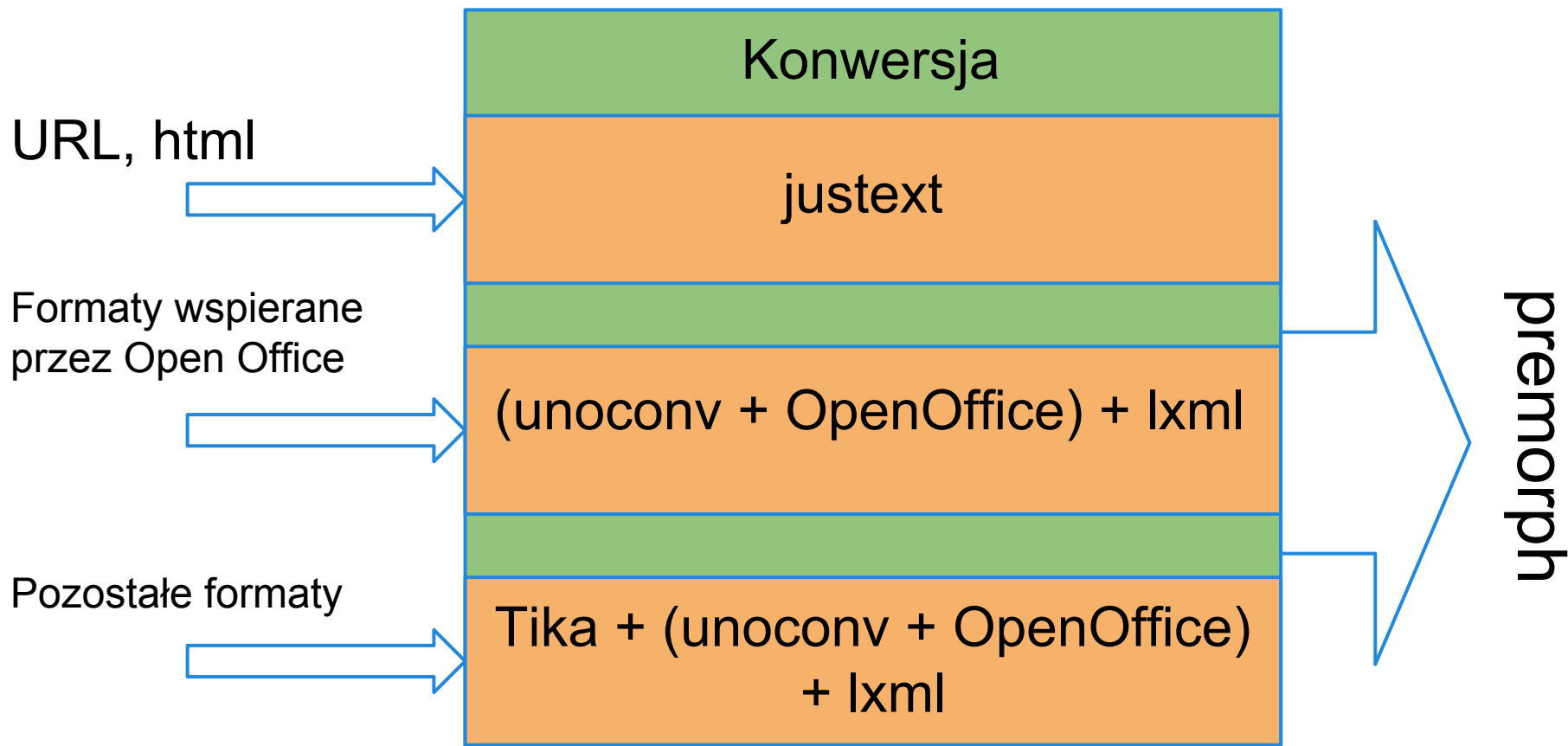
Jasnopis (typy przetwarzania)

- Dwa typy indeksu FOG.
- Cztery typy indeksu Pisarka.
- Grafy podobieństwa leksykalnego (dwa sposoby wyliczania).
- Cztery sposoby badania podobieństwa gatunkowego wprowadzonego tekstu (tzw. metody Taylora).

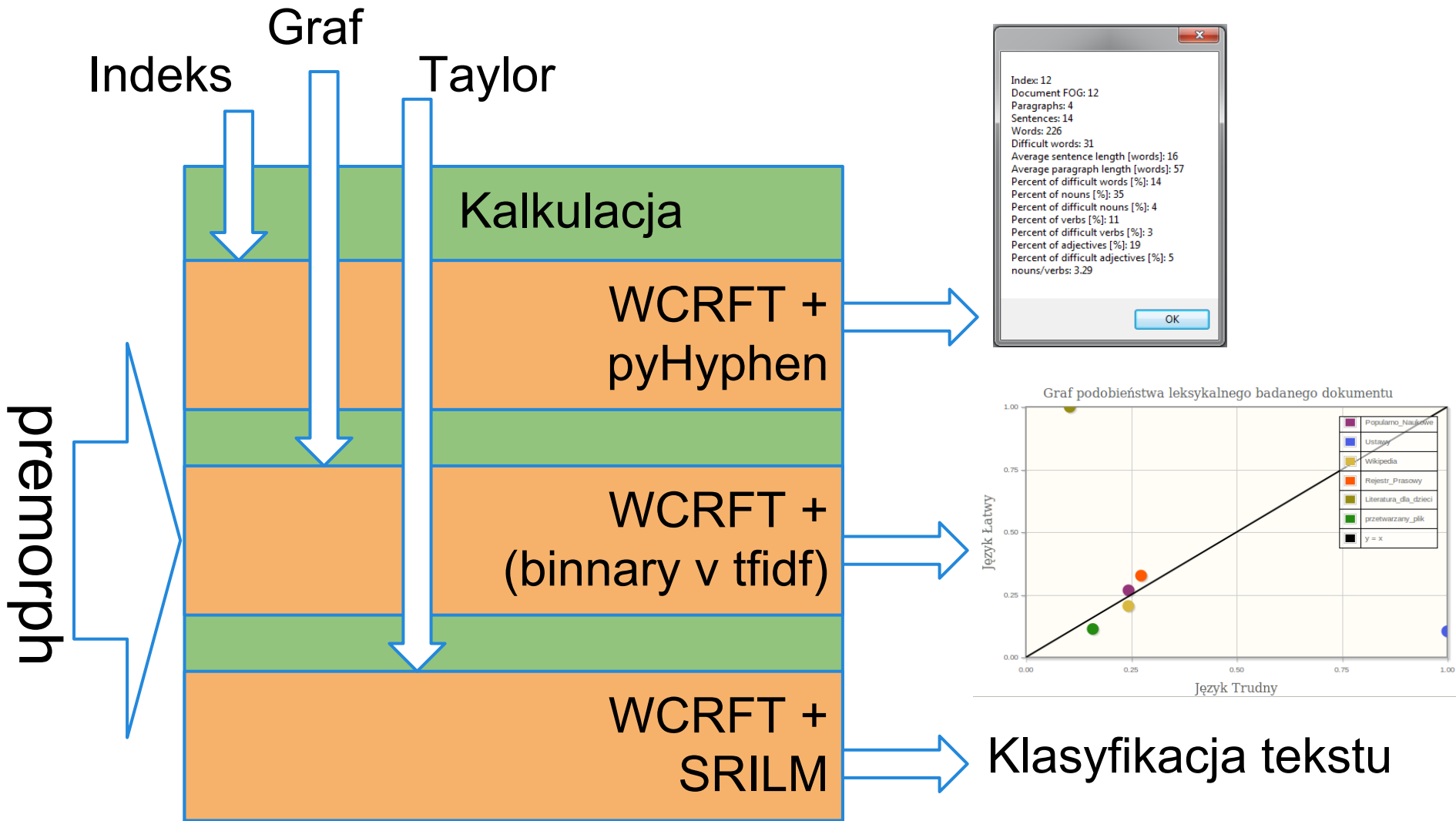
Jasnopis (architektura)



Jasnopis (moduł do konwersji formatów)



Jasnopis (moduł liczący)



Jasnopis (wykorzystane narzędzia)

- Django (<https://www.djangoproject.com/>);
- Celery (<http://www.celeryproject.org/>);
- RabbitMQ (<http://www.rabbitmq.com/>);
- Django REST (<http://django-rest-framework.org/>);
- justext (<http://code.google.com/p/justext/>);
- unoconv (<http://dag.wieers.com/home-made/unoconv/>);
- OpenOffice (<http://www.openoffice.org/pl/>);
- lxml (<http://lxml.de/>);
- Tika (<http://tika.apache.org/>);
- WCRFT (<http://nlp.pwr.wroc.pl/redmine/projects/wcrft/wiki>);
- PyHyphen (<https://pypi.python.org/pypi/PyHyphen/>);
- SRILM (<http://www.speech.sri.com/projects/srilm/>).

Jasnopis (przykład 1)

Panel wysokiego szczebla – powołany w celu doradzania Komisji Europejskiej w zakresie funkcjonowania wspólnego programu Ambient Assisted Living (wspólny program w zakresie życia wspieranego przez otoczenie) – rozpoczął konsultacje, w ramach których poproszono obywateli, przedsiębiorstwa oraz naukowców o podzielenie się pomysłami, jak najlepiej wykorzystać technologie informacyjne i komunikacyjne (TIK), aby pomóc osobom starszym w bardziej niezależnym życiu, a w szerszym ujęciu, aby ustanowić nowe sposoby wykorzystania rozwiązań TIK na rzecz najsłabszych członków społeczeństwa. Panelowi przewodniczyła komisarz europejska Meglena Kunewa. Konsultacje publiczne stanowią pierwszy krok w kierunku osiągnięcia celu wyznaczonego przez agendę cyfrową (zob. IP/10/581), jakim jest podwojenie do 2015 r. liczby osób starszych korzystających z udogodnień umożliwiających niezależne życie.

FOG: 23

Pisarek: 21

Jasnopis (przykład 2)

Wynik

Status: SUCCESSFUL

Wynik: 12,0

FOG: 12,0

Liczba akapitów: 4,0

Liczba zdań: 14,0

Liczba słów: 226,0

Liczba słów trudnych: 31,0

Średnia długość zdania: 16,0 [słowa]

Średnia długość akapitu: 57,0 [słowa]

Procent słów trudnych: 14,0 [%]

Procent rzeczowników: 35,0 [%]

Procent rzeczowników trudnych: 4,0 [%]

Procent czasowników: 11,0 [%]

Procent czasowników trudnych: 3,0 [%]

Procent przymiotników: 19,0 [%]

Procent przymiotniów trudnych: 5,0 [%]

Stosunek rzeczowników do czasowników: 3,29

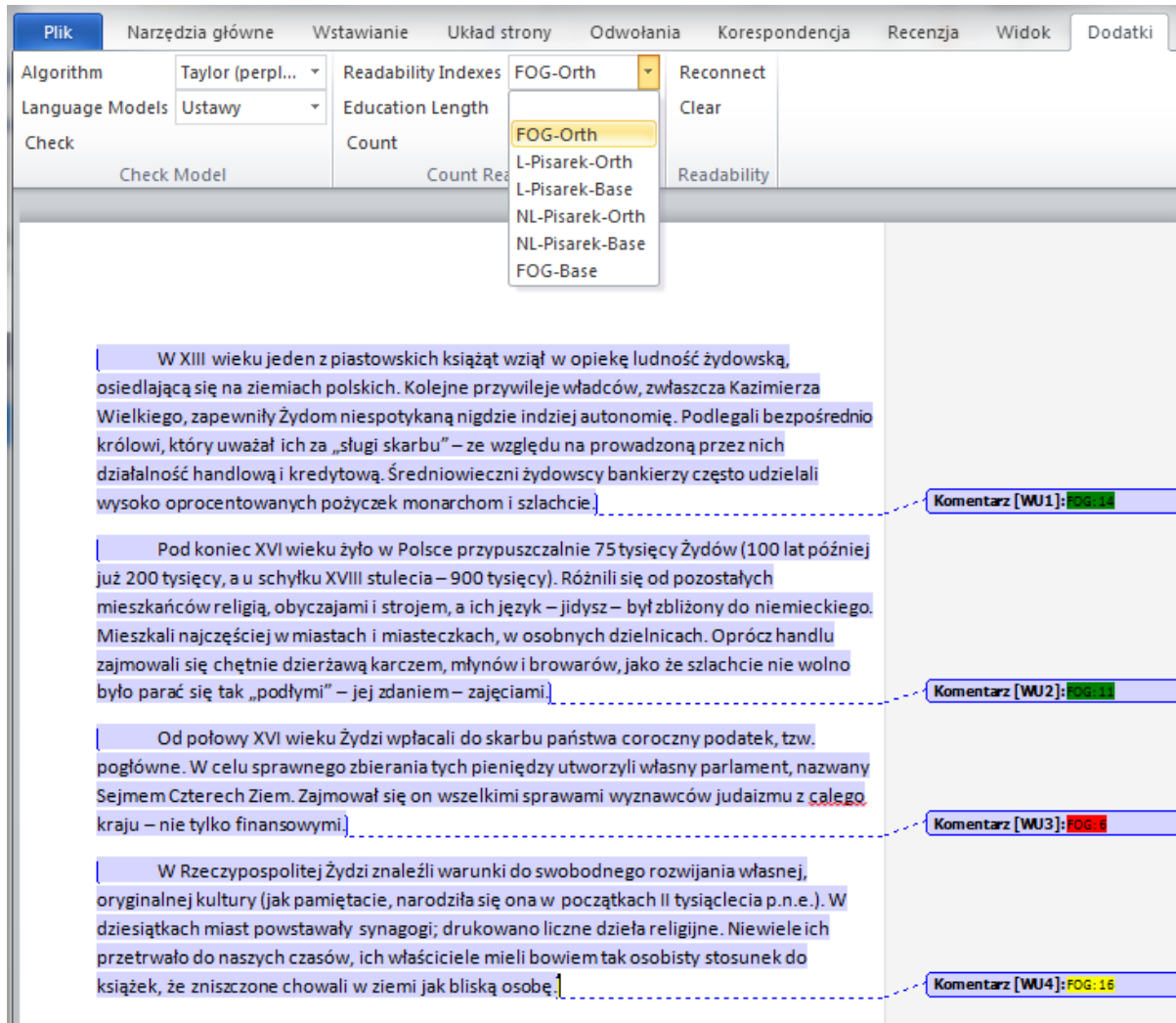
W XIII wieku jeden z piastowskich książąt wziął w opiekę ludność żydowską, osiedlającą się na ziemiach polskich. Kolejne przywileje władców, zwłaszcza Kazimierza Wielkiego, zapewniły Żydom niespotykaną nigdzie indziej autonomię. Podlegali bezpośrednio królom, który uważał ich za „sługi skarbu” – ze względu na prowadzoną przez nich działalność handlową i kredytową. Średniowieczni żydowscy bankierzy często udzielali wysoko oprocentowanych pożyczek monarchom i szlachcie.

Pod koniec XVI wieku żyło w Polsce przypuszczalnie 75 tysięcy Żydów (100 lat później już 200 tysięcy, a u schyłku XVIII stulecia – 900 tysięcy). Różnili się od pozostałych mieszkańców religią, obyczajami i strojem, a ich język – jidysz – był zbliżony do niemieckiego. Mieszkali najczęściej w miastach i miasteczkach, w osobnych dzielnicach. Oprócz handlu zajmowali się chętnie dzierżawą karczem, młynów i browarów, jako że szlachcie nie wolno było parać się tak „podłymi” – jej zdaniem – zajęciami.

Od połowy XVI wieku Żydzi wpłacali do skarbu państwa coroczny podatek, tzw. pogłówne. W celu sprawnego zbierania tych pieniędzy utworzyli własny parlament, nazwany Sejmem Czterech Ziem. Zajmował się on wszelkimi sprawami wyznawców judaizmu z całego kraju – nie tylko finansowymi.

W Rzeczypospolitej Żydzi znaleźli warunki do swobodnego rozwijania własnej, oryginalnej kultury (jak pamiętacie, narodziła się ona w początkach II tysiąclecia p.n.e.). W dziesiątkach miast powstawały synagogi; drukowano liczne dzieła religijne. Niewiele ich przetrwało do naszych czasów, ich właściciele mieli bowiem tak osobisty stosunek do książek, że zniszczone chowali w ziemi jak bliską osobę.

Wtyczka do Worda



Plik Narzędzia główne Wstawianie Układ strony Odwołania Korespondencja Recenzja Widok Dodatki

Algorithm Taylor (perpl... Readability Indexes FOG-Orth Reconnect
Language Models Ustawy Education Length Clear
Check Count Count Re Readability
Check Model Count Re

FOG-Orth
L-Pisarek-Orth
L-Pisarek-Base
NL-Pisarek-Orth
NL-Pisarek-Base
FOG-Base

W XIII wieku jeden z piastowskich książąt wziął w opiekę ludność żydowską, osiedlając się na ziemiach polskich. Kolejne przywileje władców, zwłaszcza Kazimierza Wielkiego, zapewniły Żydom niespotykaną nigdzie indziej autonomię. Podlegali bezpośrednio królowi, który uważał ich za „sługi skarbu” – ze względu na prowadzoną przez nich działalność handlową i kredytową. Średniowieczni żydowscy bankierzy często udzielali wysoko oprocentowanych pożyczek monarchom i szlachcie.]

Komentarz [WU1]: FOG: 12

Pod koniec XVI wieku żyło w Polsce przypuszczalnie 75 tysięcy Żydów (100 lat później już 200 tysięcy, a u schyłku XVIII stulecia – 900 tysięcy). Różnili się od pozostałych mieszkańców religią, obyczajami i strojem, a ich język – jidysz – był zbliżony do niemieckiego. Mieszkali najczęściej w miastach i miasteczkach, w osobnych dzielnicach. Oprócz handlu zajmowali się chętnie dzierżawą karczem, młynów i browarów, jako że szlachcie nie wolno było parać się tak „podłymi” – jej zdaniem – zajęciami.]

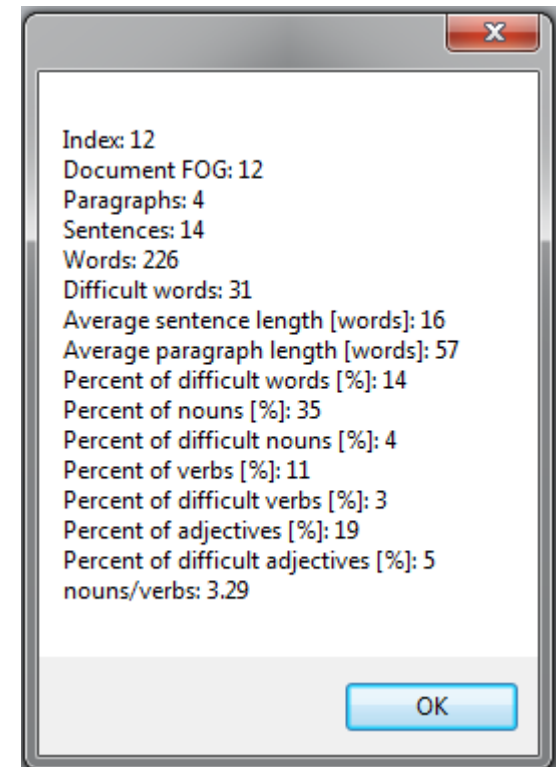
Komentarz [WU2]: FOG: 11

Od połowy XVI wieku Żydzi wpłacali do skarbu państwa coroczny podatek, tzw. pogłówne. W celu sprawnego zbierania tych pieniędzy utworzyli własny parlament, nazwany Sejmem Czterech Ziem. Zajmował się on wszelkimi sprawami wyznawców judaizmu z całego kraju – nie tylko finansowymi.]

Komentarz [WU3]: FOG: 8

W Rzeczypospolitej Żydzi znaleźli warunki do swobodnego rozwijania własnej, oryginalnej kultury (jak pamiętacie, narodziła się ona w początkach II tysiąclecia p.n.e.). W dziesiątkach miast powstawały synagogi; drukowano liczne dzieła religijne. Niewiele ich przetrwało do naszych czasów, ich właściciele mieli bowiem tak osobisty stosunek do książek, że zniszczone chowali w ziemi jak bliską osobę.]

Komentarz [WU4]: FOG: 16



Index: 12
Document FOG: 12
Paragraphs: 4
Sentences: 14
Words: 226
Difficult words: 31
Average sentence length [words]: 16
Average paragraph length [words]: 57
Percent of difficult words [%]: 14
Percent of nouns [%]: 35
Percent of difficult nouns [%]: 4
Percent of verbs [%]: 11
Percent of difficult verbs [%]: 3
Percent of adjectives [%]: 19
Percent of difficult adjectives [%]: 5
nouns/verbs: 3.29

OK

Planowane wyniki

- Narzędzie docelowe będzie obliczało stopień zrozumiałości zgodnie z naszym wzorem, który powstanie na podstawie dotychczasowych i przewidywanych w dalszym ciągu badań.
- Planujemy powtórzenie badań psycholingwistycznych na innych tekstach, być może częściowo spreparowanych.
- Obecnie pracujemy nad sprawdzeniem korelacji trudności/zrozumiałości tekstu z jego cechami subtelniejszymi niż długość zdań i procent wyrazów dłuższych niż przeciętne.

Cechy tekstu mające hipotetycznie wpływ na stopień jego zrozumiałości

Leksyka wpływająca na zwiększenie stopnia trudności:

- terminy fachowe (opatrzone kwalifikatorami w słowniku?);
- wyrazy i wyrażenia należące do stylu kancelaryjno-urzędowego;
- wyrazy archaiczne i przestarzałe;
- wyrazy nie należące do rejestru/stylu potocznego;
- wyrazy wieloznaczne.

Leksyka wpływająca na zmniejszenie stopnia trudności:

- wyrazy o wysokiej frekwencji na liście rangowej wg prawdopodobieństwa subiektywnego (listy J. Imiołczyka);
- wyrazy należące do zasobu słownictwa wspólnostylowego.

Cechy tekstu mające hipotetycznie wpływ na stopień jego zrozumiałości

Cechy morfosyntaktyczne wpływające na zwiększenie stopnia trudności:

- imiesłowy przysłówkowe uprzednie;
- imiesłowy przymiotnikowe bierne (strona bierna);
- imiesłówów przysłówkowe czynne (zwłaszcza przy zdaniach głównych bezosobowych)
- zdania zanegowane, zwłaszcza z „potrójną negacją”: *Nieprawda, że nie ma żadnych przeszkód* (również wyrazy rozpoczynające się od *nie-*);
- zdania wielokrotnie złożone podrzędnie – ogólniej skomplikowane struktury składniowe (korelacja z długością zdania?);
- struktury składniowe dające się zinterpretować na wiele sposobów (ile drzew z banku drzew odpowiada danemu zdaniu?);
- występowanie rozbudowanych członów wtrąconych;
- koreferencja w tekście (w sąsiednich zdaniach).

Dziękujemy!

