

Metody oceny zrozumiałości tekstu w j. polskim

Stanisław Zielonka

Wydział Elektryczny
Politechnika Warszawska

Seminarium dyplomowe inżynierskie, 2017

Definicja „zrozumienia”

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

RAND Reading Study Group [1] definiuje zrozumienie jako proces jednoczesnego wydobywania i budowania znaczenia poprzez interakcje i zaangażowanie w czytany tekst. Stopień w jakim czytany tekst będzie rozumiały zależy od cech samego tekstu, czytelnika oraz czynników socjologicznych i kulturalnych.

Cechy tekstu mające wpływ na jego zrozumiałość

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- terminy fachowe,

Cechy tekstu mające wpływ na jego zrozumiałość

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- terminy fachowe,
- wyrazy i wyrażenia należące do stylu kancelaryjno-urzędowego,

Cechy tekstu mające wpływ na jego zrozumiałość

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- terminy fachowe,
- wyrazy i wyrażenia należące do stylu kancelaryjno-urzędowego,
- wyrazy archaiczne i przestarzałe,

Cechy tekstu mające wpływ na jego zrozumiałość

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- terminy fachowe,
- wyrazy i wyrażenia należące do stylu kancelaryjno-urzędowego,
- wyrazy archaiczne i przestarzałe,
- wyrazy nie należące do rejestru/stylu potocznego, wyrazy wieloznaczne,

Cechy tekstu mające wpływ na jego zrozumiałość

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

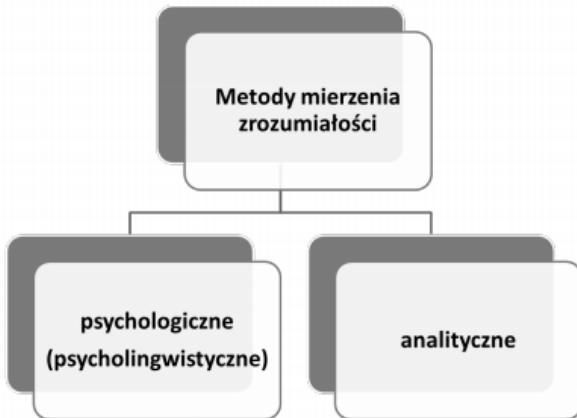
- terminy fachowe,
- wyrazy i wyrażenia należące do stylu kancelaryjno-urzędowego,
- wyrazy archaiczne i przestarzałe,
- wyrazy nie należące do rejestru/stylu potocznego, wyrazy wieloznaczne,
- cechy morfosyntaktyczne (np. struktury składniowe)

[2]

Metody mierzenia zrozumiałości

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka



Przykłady metod psychologicznych

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- Sprawdzanie umiejętności stwierdzenia, czy tekst jest prawdziwy, czy fałszywy,

Przykłady metod psychologicznych

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- Sprawdzanie umiejętności stwierdzenia, czy tekst jest prawdziwy, czy fałszywy,
- Porównywanie czasu reakcji na polecenie związane z tekstem,

Przykłady metod psychologicznych

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- Sprawdzanie umiejętności stwierdzenia, czy tekst jest prawdziwy, czy fałszywy,
- Porównywanie czasu reakcji na polecenie związane z tekstem,
- Sprawdzanie umiejętności uzupełniania brakujących fragmentów tekstu (tzw. Cloze procedure), czyli test Taylora,

Przykłady metod psychologicznych

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

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

Przykłady metod analitycznych

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

- wzory bazujące na statystyce tekstu,
 - dla języka angielskiego,
 - dla języka polskiego,
- pomiar podobieństwa pomiędzy korpusami referencyjnymi,
- automatyczna metoda Taylora - uzupełnianie luk w tekście przez modele językowe wytrenowane na tekstach o różnej zrozumiałości

Wzór Pisarka (2007) dla j. polskiego

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

$$T = \frac{\sqrt{T_w^2 + T_s^2}}{2} \quad (1)$$

T - stopień trudność tekstu,

T_w - procent wyrazów czterosylabowych lub dłuższych,

T_s - przeciętna długość zdania mierzona w wyrazach.

Trudność językowa tekstów w niektórych czasopismach wg Pisarka

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

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

Tzw. indeks mglistości (*fog index*) utworzony przez R. Gunninga dla j. angielskiego

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

$$T = 0,4 \times (T_w + T_s) \quad (2)$$

T_w - średnia liczba wyrazów w zdaniu,

T_s - procent wyrazów trójsylabowych lub dłuższych.

Przetwarzanie języka naturalnego

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

Niewątpliwie problem stworzenia oprogramowania do oceny czytelności tekstu dotyczy dziedziny jaką jest przetwarzanie języka naturalnego (*ang. natural language processing*).

NLP to interdyscyplinarna dziedzina łącząca zagadnienia sztucznej inteligencji (w tym uczenia maszynowego) oraz lingwistyki.

Aby wykorzystać potencjał komputerów w pracy z językami naturalnymi, potrzeba wzorca oraz specjalnych metod przetwarzania tekstu tak by był on „zrozumiały” dla komputera.

Języki naturalne i języki formalne

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

Mimo cech i pojęć wspólnych takich jak gramatyka, alfabet czy składnia, to co sprawia że komputer nie "rozumie" naszego języka to:

kontekstowość

Rozumienie polega na kompleksowej analizie zdań i związków między wyrazami, a nie znaczeniu poszczególnych słów

reguły gramatyczne

w j. naturalnych istnieją liczne wyjątki od reguł gramatycznych

posiadanie świadomości i umiejętności myślenia abstrakcyjnego jest konieczna do "rozumienia"

Dostępne narzędzia informatyczne do pracy z językami

Metody oceny
zrozumiałości
tekstu w j.
polskim



Stanisław
Zielonka

- Korpusy językowe np. <http://nkjp.pl/>
- Biblioteka **Natural Language Toolkit** dla języka **Python**
- Analizator morfologiczny **Morfeusz**
<http://sgjp.pl/morfeusz/>
- Wyszukiwarka **PELCRA** dla **Narodowego Korpusu Języka Polskiego** www.nkjp.uni.lodz.pl/

Bibliografia I

Metody oceny
zrozumiałości
tekstu w j.
polskim

Stanisław
Zielonka

-  RAND Reading Study Group *Reading for understanding: Toward a research and development program in reading comprehension* . Santa Monica, CA: Office of Education Research and Improvement, 2002, s. 11.
-  Gruszczyński W., Dębowski Ł., Broda B., Nitoń B., Ogrodniczuk M. *Mierzenie stopnia zrozumiałości polskich tekstów użytkowych* Seminarium "Przetwarzanie języka naturalnego", Instytut Podstaw Informatyki PAN, 4 listopada 2013, s.46-47