

SI a słowa klucze, czyli próba automatycznej ekstrakcji terminów wielowyrazowych z tekstów specjalistycznych

Miroslaw Koziarski

miroslaw.koziarski (at) gmail.com

Znaczenie leksyki

Wawrzyńczyk (2015: 7):

Leksyka to artefakty. Przyroda nie daje człowiekowi słów, on je wytwarza, wszystkie pokolenia ludzkie są ich producentami, zachowują je w swym dorobku kultury duchowej i materialnej.

Zasadne jest zainteresowanie ilością, głębsza refleksja, czy dany język ma bogate słownictwo.

Czym jest terminologia

- Według Adamo (1998: 77) terminologia służy przede wszystkim do produkcji języka. Bowker (2018: 146) natomiast twierdzi, że jej główną funkcją jest rozumieniu istniejącego już tekstu, podobnie jak w przypadku leksykologii.
- Terminologia ma też dosyć bliską relację z leksykografią – w opinii niektórych badaczy są to zupełnie odrębne dyscypliny, inni zaś twierdzą, że terminologia jest tożsama z leksykografią specjalistyczną (Bowker 2018: 138).

Funkcja terminologii

Nowicki (1986: 14):

Tak więc terminologię możemy też uważać za naukę, na wskroś praktyczną, dotyczącą metod miniaturyzowania fachowych komunikatów.

Bergenholtz (2005: 127):

Wątpliwe jest jednak, czy technicznie niepoprawna jednostka terminologiczna jest wystarczająco dobra dla laika.

Cechy (dobrego) systemu terminologicznego

Nowicki (1986: 17):

Aby system terminologiczny można było uznać za dobry, powinien on spełniać kilka warunków. Mianowicie, powinien on być:

- 1) **spójny**, tj. wewnętrznie niesprzeczny, jako że ma on być odzwierciedleniem spójnego systemu pojęciowego,*
- 2) **kompletny**, tj. obejmujący całość dziedziny oczywiście przy zadanym stopniu szczegółowości,*
- 3) **rzeczowo uporządkowany**, a więc uporządkowany stosownie do związków logicznych występujących między pojęciami,*
- 4) **harmonijny**, tj. nie faworyzujący niektórych działów dziedziny kosztem innych,*
- 5) **zrównoważony**, to jest wyważający dobrze kryteria teoretyczne (czyli względy logiczne) i kryteria praktyczne (czyli takie np. względy, jak oczekiwana częstość stosowania pojęcia) przy doborze pojęć, które zamierza się uwzględnić.*

Powody do ekstrakcji MWE

1. Głębsze poznanie danej dziedziny;
2. Uspójnienie języka specjalistycznego;
3. Wsparcie w nauce języków i tworzeniu słowników;
4. Usprawnienie tłumaczeń maszynowych i modeli NLP;
5. Lepsze wyniki wyszukiwania informacji;
6. Wspieranie postępu nauki.

Problemy w ekstrakcji MWE

1. Niejednoznaczność granic jednostek;
2. Ukryty charakter MWE;
3. Kompozycyjność a idiomatyczność;
4. Zmienność form;
5. Niska frekwencja w tekstach;
6. Niereprezentatywność korpusów.

Co jest terminem?

1. ...**end-to-end models** are significantly simpler...
1. ...in addition, **static typing** helps with large-scale project...
1. ...the **open-source software** ecosystem has seen a proliferation of...
2. ...the **open-source software ecosystem** has seen a proliferation of...
3. ...the open-source **software ecosystem** has seen a proliferation of...
1. However, adoption of **C++** in the machine learning community has been stalled...
2. However, **adoption of C++** in the machine learning community has been stalled...
3. However, adoption of C++ in the **machine learning** community has been stalled...
4. However, adoption of C++ in the **machine learning community** has been stalled...
5. However, **adoption of C++ in the machine learning community** has been stalled...
1. ...end-to-end models are significantly simpler and the **gap in accuracy** to HMM/GMM systems...

Co jest nazwą własną?

- Józef Piłsudski
- marszałek Józef Piłsudski
- Europa
- Europa Zachodnia
- Będzin
- Parafia Świętej Trójcy w Będzinie
- styl gotycki
- parafia w stylu gotyckim
- pas planetoid
- Polacy
- naród polski
- ważki równoskrzydłe (Zygoptera)

Jak stworzyć korpus?

Chlebda, Wojciech (2013):

*[K]orpus ma zawsze swego autora (zbiorowego czy indywidualnego). Korpus nie składa się sam, korpus jest zawsze złożony przez kogoś, korpus jest zawsze czyjś – a to oznacza, że korpus (...) jest zawsze, z definicji, jedynie **podmiotową interpretacją tekstowej rzeczywistości języka, a niej je lustrzanym odbiciem.***

Jak stworzyć korpus?

Chlebda, Wojciech (2013):

lingwistyczna zasada nieoznaczoneści:

[R]zeczywistość języka dana jest badaczowi wyłącznie w próbkach tekstowych, z których żadna nie może odzwierciedlić wszystkich parametrów tej rzeczywistości, każda jest rezultatem subiektywnego wyboru jej twórcy.

Pytania, które warto sobie zadać

Lower Frequency Threshold	<i>Is the term frequent enough to be relevant?</i>
Upper Frequency Threshold	<i>Is the term rare enough to be to be considered a common term?</i>
Group Use Restriction	<i>Is it used exclusively or primarily by members of a specific group?</i>
Independent Semantics	<i>Does it have a meaning separate from both constituent units and larger units?</i>
Unit as Proper Name	<i>Is it a name of an object, system, method or idea?</i>
Practical Utility	<i>Does it have a practical purpose?</i>
Exclusion from General Usage	<i>Is it, at least in its inception, not part of the general language?</i>
Functional Differentiation	<i>Has a specialized function and fit within a system of concepts, contrasting with common words, as well as opposite, broader and narrower meanings?</i>
Conceptual Clarity	<i>Does it have a well-defined meaning?</i>
Inter-Domain Unambiguity	<i>Is this term devoid of common use meanings different than domain-specific use?</i>
Semantic Stability	<i>Does it have a more or less stable meaning across time?</i>
Consistency of Scope	<i>Is it used in specific contexts or domains?</i>
Consistency of Use Across Texts	<i>Is it used consistently (in terms of semantics and formal constraints) across different texts?</i>
Corpus Documentation	<i>Is it documented in technical and scholarly publications?</i>
Accuracy & Specificity	<i>Does such segment provide maximum narrowing of scope in the context?</i>
Community Acceptance	<i>Is it accepted in the broader community?</i>
Reader Understanding	<i>Will the reader understand the passage best when focusing on such segment?</i>
Non-Evident Combinatorial Meaning	<i>Is it probable for the reader to fail to infer the meaning unambiguously from constituent elements?</i>
Contextual Relevance	<i>Does it have maximum alignment with the context of the segment?</i>
Full Adjacency Inclusion	<i>Has it exhausted the neighboring words to create a larger unit</i>
Chiefly Predicative Modifiers	<i>Modifiers create a separate entity with immutable characteristics?</i>
Periphrastic Irreplaceability	<i>The unit cannot be easily paraphrased with similar complexity?</i>
Independent Lexical Structure	<i>Does it have an independent lexical structure?</i>
Independent Syntactic Structure	<i>Is it syntactically a stand-alone unit?</i>
Noun Phrase	<i>Is the core of the unit a noun phrase?</i>

Macierz porównawcza

Criterion	the open-source software ecosystem has seen a proliferation of	the open-source software ecosystem has seen a proliferation of	the open-source software ecosystem has seen a proliferation of	end-to-end models are significantly simpler	in addition, static typing helps with large-scale projects	However, adoption of C++ in the machine learning community has been stalled	However, adoption of C++ in the machine learning community has been stalled	However, adoption of C++ in the machine learning community has been stalled	However, adoption of C++ in the machine learning community has been stalled	However, adoption of C++ in the machine learning community has been stalled	end-to-end models are significantly simpler and the gap in accuracy to HMM/GMM systems
Lower Frequency Threshold	✓	✓	✓	✓	✓	✓	✗	✓	✓	✗	✓
Upper Frequency Threshold	✗	✓	✗	✓	✓	?	✓	✗	✗	✓	?
Group Use Restriction	✗	✓	✓	✓	✓	✓	✓	✗	✗	✗	✓
Independent Semantics	✓	?	✓	✓	✓	✓	?	✓	✓	✓	✓
Unit as Proper Name	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗
Practical Utility	✓	✓	✓	✓	✓	✓	✗	✓	✗	✗	✓
Exclusion from General Usage	✗	✓	✓	✓	✓	✓	✓	✗	✗	✓	✓
Functional Differentiation	✓	?	✓	✓	✓	✓	✗	✓	✗	✗	✓
Conceptual Clarity	✓	✗	✗	✓	✓	✓	✗	✓	✗	✗	✓
Inter-Domain Unambiguity	✓	✓	✓	✓	✓	✓	✗	✓	✗	✓	✗
Semantic Stability	✓	?	?	✓	✓	✓	?	✓	✗	✗	✓
Consistency of Scope	✓	?	?	✓	✓	✓	✗	✓	✗	✗	✓
Consistency of Use Across Texts	✓	?	✓	✓	✓	✓	✗	✓	✗	✗	✓
Corpus Documentation	✓	?	✓	✓	✓	✓	✓	✓	?	?	✓
Accuracy & Specificity	✗	✓	✗	✓	✓	✗	?	✓	✗	✗	✓
Community Acceptance	✓	✓	✓	✓	✓	✓	✗	✓	?	?	✓
Reader Understanding	?	?	?	✓	✓	?	✓	✗	✓	✓	✓
Non-Evident Combinatorial Meaning	✗	✗	✗	✓	?	✗	✗	✗	✗	✗	✓
Contextual Relevance	✗	✓	✗	✓	✓	?	✓	?	?	?	✓
Full Adjacency Inclusion	✗	✓	✗	✗	✗	✓	✓	✓	✓	✗	✓
Chiefly Predicative Modifiers	✓	✓	✓	✓	✓	✗	✗	✓	?	✗	✓
Periphrastic Irreplaceability	✓	✗	✗	?	✓	✓	✗	✓	✗	✗	✗
Independent Lexical Structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Independent Syntactic Structure	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Noun Phrase	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Modele programistyczne

Metody ekstrakcji MWE (Pęzik 2018ff):

- 1. Metody pozycyjne** – współwystępowanie jednostek w sąsiedztwie.
- 2. Metody relacyjne** – elementy są powiązane syntaktycznie.
- 3. Modele językowe** – enkodery oparte o transformery.

Źródła danych

1. Wikipedia;
2. ARXIV;
3. OAGKX;
4. Portal Elektroda.pl;
5. Wybór książek specjalistycznych.

Kategorie oznaczanych elementów

Pozostałością z czasów tureckich w **<k2>Kamieńcu</k2>** jest minaret przy **<k2>katedrze Św. Piotra i Pawła</k2>**. Po odzyskaniu miasta **<k4>w 1699</k4>**, zgodnie z **<k1>traktatem pokojowym</k1>**, minaret z półksiężycem pozostawiono. Ustawiono na nim drewnianą figurę **<k3>Matki Boskiej</k3>**. **<k4>W 1756</k4>** zastąpiono ją sprowadzoną z **<k2>Gdańska</k2>**, wykonaną ze stopu miedzi i srebra, mierzącą **<k5>4,5 m</k5>** pozłacaną rzeźbą **<k3>Matki Boskiej</k3>** stojącej na kuli i księżycu. **<k2>Katedra kamieniecka</k2>** jest jedynym na świecie **<k1>kościółem katolickim</k1>**, przed którym stoi minaret. W kościele dominikanów zachowała się z kolei **<k1>kamienna konstrukcja ambony</k1>** (minbar) z czasów tureckich. **<k4>W XVIII wieku</k4>** rola miasta znacznie spadła, jednakże z woli **<k3>króla Stanisława Augusta Poniatowskiego</k3>** podjęto w tym czasie studia mające na celu **<k1>modernizację fortyfikacji</k1>** zamku i miasta, głównie wg planów polskiego **<k3>architekta Jana de Witte</k3>**.

Wysoka gęstość terminologiczna

Podczas **<k1>regulacji jasności</k1>** w górę na ekranie od góry pojawiają się ciemne **<k1>pasy poziome</k1>**. Wygląda to tak że obraz jest przerywany regularnymi pasami. Czyli ok **<k5>1cm</k5>** wizja i 1 cm wygaszone. Na małym kontraście jest ok. **<k2>TA 8658</k2>** wymieniony i **<k1>wzmacniacz wizji</k1>** też. Napięcia także właściwe. Proszę o... Wymień elektrolit w zasilaniu **<k1>wzmacniaczy wizyjnych</k1>**. To na samym początku zrobiłem. Ale dzięki, on też powoduje takie objawy. sprawdź diodę w **<k1>zasilaniu wzmacniaczy</k1>** Wszystko z tej strony jest ok. Gdy są te objawy to już **<k1>sygnał RGB</k1>** z **<k2>TA</k2>** jest przerywany - sprawdzone oscyloskopem. Pokopałbym trochę w pionie. Sprawdź **<k1>rezystor równoległy</k1>** do **<k1>cewek pionu</k1>** no i oczywiście jak zawsze elektrolity. Co ma **<k1>rezystor na cewkach</k1>** do **<k1>poszatkowanego obrazu</k1>** przy większym prądzie kineskopu (większy kontrast i jaskrawość). Przy mniejszych ustawieniach jest ok. Objaw wygląda następująco: **<k5>10</k5>** linii wyświetla **<k1>treść wizji</k1>**, następne **<k5>10</k5>** jest wyciemnione i tak przez cały ekran.

Niska gęstość terminologiczna

Tak jak w temacie, pali się tylko **<k1>zielona dioda</k1>** więc zakładam że zasilanie ok co obstawiacie??? Z góry dzięki za pomoc :lol: Nikt tutaj nie jest wruzka, poczytaj inne posty. I podaj więcej informacji.

Powywlekaj wszystko z płyty zostaw tylko procka i glosniczek z obudowy. i Restet **<k2>Cmos</k2>**'a. i tak po kolejki potem ram grafa i reszta. Płyta wyciągnięta z jednej obudowy i przełożona do drugiej i zamarła nawet procek nie był wyciągany z niej więc prawdopodobnie odpada, ja nie mam pojęcia co się stało może mi podrzucicie jakiś pomysł bo czasami człowiek poelnie banalny i błąd i nie potrafi go wyeliminować) Wyciągnij płytę z obudowy na płycie niech jest tylko procek ,grafa i pamiec .Zrób klir **<k2>Cmos</k2>** i jak masz miernik pomiesz ile ma baterja.Napisz co i jak.Pozdrawiam

Ja miałem taki numer z **<k2>A7V</k2>**: zrobiłem upgrade biosu i ten nowy , jak to w **<k2>Asusie</k2>** podobno często bywa , nie zapisał się dobrze. No i płyta is dead chwilowo. Musiałem nadpisać też boot block, bo po żadnym czasie oczeiwania stacja dysków nie odezwała się. Oczywiście płyta się włączała. Tak więc układ wymontowałem i udałem się do serwisu, gdzie za dwie dychy wrzucili mi bios. Płytę wziąłem do domku i za żadne skarby nie chciała się włączyć. Za to na drugi dzień wystartowała bez problemu.

być słowionym,
Nie mogli już słówek ob-
tak, jak ś. rzecz sama
słowa >

tytu, niby, niejako, rze-
nieki do węglarza, ten,
ukrywszy go S., wnet
Trąbeie w trąby. a to
nowie. Kochan. [A ja
plakać będę] 2. to jest,
nie, w dosłownym znacze-
dłużnika z swego dzie-
jeństwa. Ort. Iseiec, S.
cząstkę swemu pożyćcy
Przez ortyle, S. prawo.
przylubuje, S. pochle-
Edom, który S. jest Edom.
matkę S. nad niebo
roboty wymalować śmie,
Krystusem królować będą,
nie wstydzą ś. pogańską
< 3 os. lp. od * Słucé,
znaczy, SLOW >

wnie bot. P. Len.
a] p. Słowo. < Od Słowo >
lm., P. Słowianstwo: Prag-
, zbawiać tudy słowem! —
dże duchowe S. Wyp.
Por.

Słownikzek, czka, im.
Słownik >

Słownicz p. Słownikowy: Podkreślony pod-
bieństwo słownicze pomiędzy Mazurami a Kur-
piami. Krzyw. < Od Słownik >

Słownie przys. od Słowny, ustnie: Ktokolwiek g
S. i czynnie (= słowami). S. mi obiecał Tro-
S. sens tylko potrafił. Troc (= dustaument. In-
ralnie). S. (prawdźiwie) to uczynię. Troc (= p-
słowem honoru). S. to tylko mówisz. Troc. Słow
Boże jedno jest na piśmie, a drugie S. i usta
jakoby przez ręce nam podane. Karak.

Słownik, a, lm. i słowozbiór, dykcyjona, jak
kon: S. powinien mieć ile można wszystkie v
azy języka i ich znaczenie, czy są jedynak
wieloznaczne. Kopez. S. narodowy. Ko-
Towarzystwo to nowo ustanowione będzie m
najpierwsze staranie o ułożenie polskiego dyk-
narza a. słownika; wolno im będzie jak d
nazywać ten zbiór słów polskich. I. Pocz
tego słowa w Mowniku, kiedy go w słow
nie masz. Troc. S. Lindego, Wileński. S. i
klopedyczny (= encyklopedia). S. biograf
geograficzny. S. poetów polskich. S. bli
a. konkordancja = wykaz występowania wy-
zawartych w Bibliji, z ichłączeniem wersetów,
< Od Słowo lub Słowny >
Słownikarnia, l. lm. & redakcja słownika
Słownik >

Wyniki

ad Słownikarz; X SI
Słownikarskie. 23

Dobrze, ale nie beznajdziejnie

Marian Melchior Antoni Langiewicz (ur. 5 sierpnia 1827 w Krotoszynie, zm. 10 maja 1887 w Konstantynopolu) – generał i dyktator powstania styczniowego (marzec 1863), wykładowca Polskiej Szkoły Wojskowej [1].

Życiorys

Uczył się w gimnazjum w Trzemesznie. Studiował prawo we Wrocławiu i Berlinie. Oficer artylerii armii pruskiej [2].

W 1860 brał udział w wyprawie Garibaldiego na Sycylię, gdzie odznaczył się męstwem i nieustrasliwością.

Był wykładowcą Polskiej Szkoły Wojskowej w Cuneo, gdzie stał się zajadłym przeciwnikiem Ludwika Mierosławskiego.

Działal w organizacjach niepodległościowych we Włoszech. Był związany z politycznym ugrupowaniem „białych”.

Czas środkowoeuropejski (ang. Central European Time, CET) – strefa czasowa odpowiadająca czasowi słonecznemu południka 15°E, który różni się o 1 godzinę od uniwersalnego czasu koordynowanego (UTC+1:00) [1].

Czas środkowoeuropejski jest najszerszej stosowanym w krajach Unii Europejskiej, między innymi w środkowej i zachodniej Europie.

Historia

Przed wiekiem XIX w krajach Europy w powszechnym użyciu był czas słoneczny ważniejszych ośrodków miejskich.

Potrzeba standaryzacji czasu pojawiła się wraz z rozwojem kolei, która wymagała stworzenia rozkładów jazdy i obliczania czasu podróży.

W 1886 roku austriacki astronom i geodeta Robert Schram opublikował artykuł, w którym proponował wprowadzenie jednolitego systemu pomiaru czasu w kolejach austriackich, na wzór amerykański (Stany Zjednoczone wprowadziły standardowy czas kolejowy trzy lata wcześniej).

Podstawą miał być czas słoneczny południka 15°E, co oznaczało różnicę dokładnie godziny względem czasu Greenwich.

Idea ta została początkowo zignorowana, ale dwa lata później dyrekcje kolei austriackich i węgierskich przyjęły ją i przedłożyły projekt rządowi.

Propozycja została zaakceptowana, z zastrzeżeniem, że nowy standard powinien zostać przyjęty także w innych krajach:

Cesarstwie Niemieckim, Szwajcarii, Włoszech i Serbii [2].

Zapis terminograficzny

- czas letni
- czas letni (GMT+2)
- czas słoneczny
- czas urzędowy
- czas zachodnioafrykański
- czas zachodnioeuropejski letni
- czas środkowoeuropejski
- czas środkowoeuropejski letni, UTC+2
- czasu Greenwich
- czasu letniego
- czasu wschodnioeuropejskiego
- czasu środkowoeuropejskiego
- czasu środkowoeuropejskiego letniego
- czasy urzędowe państw zaborczych
- krajów okupowanych
- kryzysie naftowym
- letnim półroczu
- numerów stron
- prawem Unii Europejskiej
- przejście na czas środkowoeuropejski
- rozkłady jazdy
- rozwiniętych gospodarczo
- standardowy czas kolejowy
- standaryzacji czasu
- strefa czasowa
- strefy czasu środkowoeuropejskiego
- systemu pomiaru czasu
- uniwersalnego czasu koordynowanego (UTC+1:00)
- użytek cywilny
- wypadków drogowych

Warianty

1. Generative Adversarial Networks (GANs)
2. Generative Adversarial Networks GANs
3. Generative adversarial nets
4. Generative adversarial network (GAN)
5. Generative adversarial networks
6. Generative adversarial networks (GAN)
7. Generative adversarial networks (GANs)

Bloki terminologiczne

- dynamic pooling
- dynamic power
- dynamic pricing
- dynamic pricing strategies
- dynamic process management
- dynamic processes
- dynamic programming
- dynamic programming (DP) algorithm
- dynamic programming algorithm
- dynamic programming approach
- dynamic properties
- dynamic pushing
- dynamic random access memory
- dynamic range
- dynamic real-time holographic 3D display
- dynamic reconfiguration
- dynamic replication
- dynamic resource allocation
- dynamic resource management
- dynamic resource management strategy
- dynamic resource provisioning
- dynamic resource requirements
- dynamic response time
- dynamic reverse engineering
- dynamic route choice
- dynamic route optimization
- dynamic routing
- dynamic routing protocols
- dynamic scenarios
- dynamic scene recognition

Jak z tego korzystać

Słownikzek, czka, im. m. **Słownik** >
Słownicz p. **Słownikowy**: Podkreślony podobieństwo słownicze pomiędzy Mazurami a Kurpiami. Krzyw. <Od **Słownik**>
Słownie przys. od **Słowny**, ustnie: Kłótył g. S. i czynnie (= słowami). S. mi obiecał Troc. S. sens tylko potrafił. Troc (= dostawca, b. rabnie). S. (prawdźiwie) to uczynię. Troc (= p. słowem honoru). S. to tylko mówisz. Troc. Słow Boże jedno jest na piśmie, a drugie S. i ustnie jakoby przez ręce nam podane. Karak.
Słownik, a, lm. i **słownozbiór**, dykcyjnary, jak kon: S. powinien mieć ile można wszystkie wazy języka i ich znaczenie, czy są jedynym towarzyszem i pomocnikiem. Kopez. S. narodowy. Najpierwsze staranie o dzieło polskiego dykcyjnara a. słownika; wolno im będzie jak nazywać ten zbiór słów polskich. L. Pocz tego słowa w Mowniku, kiedy go w słowniomasz. Troc. S. Lindego, Wileński. S. kłopedyczny (= encyklopedia). S. biograficzny. S. poetów polskich. S. bli a. konkordancja = wykaz występujących w zawartych w Bibliji, z łączeniem wersetów, rycznych są załączone te wyrazy. Zdi. **Słownik** >
<Od **Słowo** lub **Słowny**>
Słownikarnia, l. lm. & redakcja słownika

być słowionym,
Nie mógł już słówek ob-
tak, jak ś. rzecz sama
słowa>
niby, niejako, rze-
nieki do węglarza, ten,
ukrywszy go S., wnet
Trąbeie w trąby. a to
nowie. Kochan. [A ja
plakać będę] 2. to jest,
nie, w dosłownym znacze-
dłużnika z swego dzie-
jeństwa. Ort. Iseiec, S.
cząstkę swemu pożyćcy
Przez orytle. S. po-
pru przylubuje. S. po-
Edom, który S. jest Edom.
matkę S. nad niebo
roboty wymalować śmie,
Krystusem królować będą,
nie wstydzą ś. pogańską
<3 os. lp. od *Słucé,
znaczy, SLOW>
wnie bot. p. Len.
a] p. **Słowo**. <Od **Słowo**>
lm., p. **Słowianstwo**: Prag-
, zbawiać tudy słowem! —
, zdaje duchowe S. Wyp.
Por.

Zastosowanie wyników

1. Budowanie słowników terminologicznych w oparciu o korpusy;
2. Traktowanie badaczy jak badaczy (a nie szperaczy);
3. Ekstrakcja kandydatów terminologicznych (niemal) w czasie rzeczywistym;
4. Konieczna interpretacja i formalny opis!

Podsumowanie

Słownik, zek, zka, im. m. Słownik >
Słownicz p. Słownikowy: Podkreślamy podobieństwo słownicze pomiędzy Mazurami a Kurpiami. Krzyw. <Od Słownik >
Słownie przys. od Słowny, ustnie: Kłótył g S. i czynnie (= słowami). S. mi obiecał Troc S. sens tylko potrafił. Troc (= dostawca, bralnie). S. (prawdźiwie) to uczynię. Troc (z p Boże jedno jest na piśmie, a drogie S. i ust jakoby przez ręce nam podane. Karu.
Słownik, a, lm. i słowozbiór, dykcyjony, jak kon: S. powinien mieć ile można wszystkie wazy języka i ich znaczenie, czy są jedynym Towarzystwo, S. Kopez. S. narodowy. Ko najpierwsze staranie o ułożenie polskiego dyk nazywać ten zbiór słów polskich. I. Pocz tego słowa w Mowniku, kiedy go w słow niemasz. Troc. S. Lindego, Wileński. S. kłopedyczny (= encyklopedia). S. biografieograficzny. S. poetów polskich. S. bl a. konkordancja = wykaz występowania w zawartych w Bibliji, z ujęciem wersetów, rzych są załączone te wyrazy. Zdi. Słowni <Od Słowo lub Słowny >
Słownikarnia, l. lm. e redakcja słownika Słownik >

być słowionym, Nie mógł już słówek ob- tak, jak s. rzecz sama słowa >
niby, niejako, rze- niekl do węglarza, ten, ukrywszy go S., wnet Trąbeie w trąby. a to nowie. Kochan. [A ja plakać będę] 2. to jest, w dosłownym znaczeniu, w dosłownym znaczeniu, z swego dziejstwa. Ort. Iseiec, S. część swemu pożyty Przez orzyle, S. przy lubuje. S. poch Ezan, który S. jest Edom. matkę S. nad niebo, a robota wymalować śmie, Krystusem królować będą, nie wstydzą s. pogańską. <3 os. lp. od * Słuc, znaczy, SLOW >
wnie bot. P. Len. a] p. Słowo. <Od Słowo > lm., P. Słowianstwo: Prag-), zbawiać tudy słowem! — dze duchowe S. Wysp. Per.

Zalety rozwiązania

1. Stosunkowo wysoka skuteczność;
2. Wystarczająco stosunkowo małe dane treningowe;
3. Równoległa detekcja wielu typów jednostek;
4. Stosunkowo szybkie przetwarzanie;
5. Wielojęzyczny charakter;
6. Przetwarzanie nieustrukturyzowanych danych.

Plany

1. Lepsze zdefiniowanie granic jednostek;
2. Lepsze rozgraniczenie typów jednostek;
3. Udoskonalenie i wyważenie korpusu treningowego;
4. Włączenie danych z różnych języków;
5. Lepsza obróbka na wyjściu;
6. Podawanie kontekstów w słownikach wyjściowych.

Bibliografia

- Adamo, Giovanni (1998): „Terminología vs. Lessicología” [w:] *Hieronymus Complutensis*, nr 8, s. 75–86. Madrid: Instituto Universitario de Lenguas Modernas y Traductores.
- Bergenholtz, Henning (2005): „Falsche und richtige lexikographische Definitionen” [w:] Gottlieb, Henrik & Mogensen, Jens Erik & Zettersten, Arne (red.): *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2–4, 2002 at the University of Copenhagen*, s. 125–132. Tübingen: Max Niemeyer Verlag.
- Bowker, Lynne (2018): „Lexicography and terminology” [w:] Fuertes-Olivera, Pedro A. (red.): *The Routledge Handbook of Lexicography*, s. 138–151. London and New York: Routledge
- Chlebda, Wojciech (2013): „Korpusologia użytkowa dla początkujących i zaawansowanych”. Na tropach korpusów: w poszukiwaniu optymalnych zbiorów tekstów. Opole: Wydawnictwo UO.
- Nowicki, Witold (1986): *Podstawy terminologii*. WROCŁAW • WARSZAWA • KRAKÓW • GDAŃSK • ŁÓDŹ: Zakład Narodowy im. Ossolińskich i Wydawnictwo Polskiej Akademii Nauk.
- Pęzik, Piotr (2018): „Facets of prefabrication. Perspectives on modelling and detecting phraseological units”. Łódź: Wydawnictwo Uniwersytetu Łódzkiego.
- Wawrzyńczyk, Jan (2015): *300 tysięcy czy milion(y)? O stanie zasobów słownictwa polskiego w dniu 31 grudnia 2000 r.* Warszawa: Mila Hoshi.