Terminology of Computational Linguistics in Terms of Indexing and Information Retrieval in the system "iSybislaw"

Alla Taran

Cherkasy Bohdan Khmelnytskyi National University, 81 Shevchenko Boulevard, 18031, Cherkasy, Ukraine
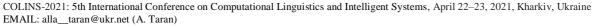
Abstract

The article explores the key words and equivalence classes of computational terms in iSybislaw. Their organization for the needs of information retrieval system has a multidimensional, multifaceted and permanent nature. It has been found that the functioning of the term as a key word in information retrieval language involves the elimination of polysemy and synonymy, which are characteristic of natural languages. However, if for descriptors as conditional normative elements unambiguity is one of the basic characteristics, then ascriptors, hierarchically subordinate to the descriptors, are capable of logical synonymy, which in turn creates conditions for improving search efficiency. The formal-semantic relations between the terms for computational linguistics as a scientific discipline in the iSybislaw information retrieval system are analyzed. Both duplicity and growing semantic differences in such terms, the need for their aspectualization are revealed. The interrelationships of the iSybislaw system with other Slavic Internet resources has been traced, which allows users to get acquainted with the state of learning a certain Slavic language in a broader temporal, genre and territorial perspective. Arranging Slavic Linguistic Terminology in Computational Linguistics.

Keywords 1

Paper template, paper information-retrieval system, national module, key words, equivalence classes, terminology of computational linguistics.

1. Introduction

New computer technologies for acquiring knowledge, including language knowledge, open up attractive prospects not only for solving practical and theoretical problems. Orientation of Ukrainian linguistics to the use of these technologies, a new way of research work marks a qualitatively new stage of its development. A new subject of communication, a new addressee of linguistic information – the Internet – needs a new apparatus for describing language, new procedures for its modeling, analysis and synthesis. As you know, the view of the object, the view of language forms the object itself. Therefore, the language system, the products of its implementation in the language of modern Ukrainian society – a variety of oral and written texts – appear in the computer environment in other guises, turn to the researcher with new facets, which makes him think about hitherto unimaginable problems, look for new, other methods of presenting linguistic information, different from traditional, tuned to human intelligence and experience, to develop means of communication with a new participant in communication – computer.

2. Prospects for the development of the Ukrainian module in iSybislaw

Today, the traditional bibliography is increasingly being turned into information retrieval systems, which allows users to obtain information in a convenient and efficient way. The annual linguistic bibliography is the most reliable source of bibliographic data, in particular on the development of linguistic thought and terminology in Ukraine. One of the most powerful information retrieval systems is the iSybislaw (Bibliograficzna baza danych językoznawstwa slawistycznego iSybislaw) [1]. Analysis of printed products in iSybislaw ensures the objectivity of modeling the state of Ukrainian and Slavic linguistic traditions and modern practice in the digital space.

Developers of national modules (including the Ukrainian module – A. A. Taran, Yu. V. Romanyuk – led by Professor Ye. A. Karpilovska) seek to provide in the iSybislaw system a bibliography on linguistic publications as detailed as possible and taking into account the current state of the national school development or the state of language research that interacts with the logic of the development of this system itself, which ideally involves the harmonious development of all its national modules, the even scale and the balance of possible modes of work with its users. Practical terminological activity is aimed at substantiating the choice of optimal, the most appropriate to their purpose terms and their meanings, which meet the need for adequate reflection of objects, processes and features in linguistics. With this approach, the user quickly will be able to extract the necessary information from iSybislaw in the amount and form that he needs. The developers have iSybislaw scientific texts on the go of modern linguistic works and do so in two ways: from new texts through their descriptions to the addition and necessary reorganization of the keyword dictionary of the information retrieval language of the system and from certain keywords to texts with them and their descriptions. Thus, the Ukrainian module in 2020 was supplemented with information about all new monographs on domestic linguistics (2016–2020), as well as the latest issues of the interdepartmental collection "Culture of the Word" and the scientific-theoretical journal "Ukrainian Language". It is also possible to search for series, in particular, many works on Ukrainian linguistics from the series "Bachelor`s rooms in Ukrainian Linguistics". Short and annotated bibliographic descriptions are added to the texts. A special key word information retrieval language, which today includes more than a thousand key words in Ukrainian, including more than 500 terms of Ukrainian linguistics, allows users to obtain a variety of information about these sources or work directly with them through the repository attached to iSybislaw.

The device for finding information are key words that are specific to a particular subject area of knowledge and represent the content of the document. The function of key words is primarily performed by linguistic terms, but not only. These can be morphemes (most often suffixes and prefixes), factual material (words like *heart, head,* etc.); names of dictionaries, atlases, conferences, essays; names of Slavic linguists, writers, compilers of dictionaries.

The multilingual iSybislaw system uses a methodology based on the establishment of equivalence classes of language units, the creation of which occurs both within the units of the language of formal description, as well as the language of subject description, which reflects the content of the document [2]. Such units seem to show the same out-of-order situation, with the same relations, but from different "sides".

The problem of synonyms (doublets) in iSybislaw solved taking into account the corresponding hierarchy. One of the terms of the conditional equivalence class is a descriptor (semantic dominant) to denote this class and to express the basic meaning of all words and phrases that belong to this class. It is highlighted in bold. And other terms of the conditional equivalence class are conditional synonyms of this descriptor and are associated with it by certain values, they are highlighted in light color. We are impressed by the definition of a descriptor in Yu. Karaulov's monograph: "A descriptor is a normalized word or phrase selected from a group of synonymous or close in meaning words, i.e. equivalence class, and intended for controlled indexing of information retrieval" [3, p. 121]; The descriptor encodes a concept expressed by a group of key words. It can be used instead of any member of the equivalence class, but not vice versa.

2.1. Key words and equivalence classes of computational linguistics terms in iSybislaw

In the article we will pay attention to the new terminology in the information retrieval system. It is represented primarily by new directions of modern Ukrainian linguistics, as well as new issues of its traditional branches. As completely new, or new in form and content, we interpret the terms of computational linguistics and other related disciplines: experimental linguistics, quantitative linguistics, computational linguistics, corpus linguistics, mathematical linguistics, applied linguistics.

Computational linguistics is already firmly established in other computational sciences. It builds on the experience gained in the pre-computer age of structural, applied and mathematical linguistics. Reading of documents with key word *językoznawstwo stosowane polski (pol) – lingwistyka stosowana polski (pol) – applied linguistics rosyjski (rus) – primijenjena lingvistika chorwacki (scr) – примењена linguistics serbski (scc) – uporabno jezikoslovje słoweński (slv)* make sure that applied linguistics is interpreted as a branch of linguistics, the subject of which is the application of methods of other scientific disciplines to study the structure of the language system and patterns of its implementation and application of linguistics in solving problems of other branches of science and social practice.

Computational linguistics is at the intersection of disciplines. For example, the term search image of a document refers to the terminology of cybernetics and computer science. However, its definition uses the concepts of computational linguistics: it is a text in information retrieval language, which is placed in unambiguous correspondence with the document and which reflects the features of the document needed to find it in the information retrieval system. In addition to the features that reveal the topic of the document, the search image of the document mostly contains some additional information (bibliographic descriptions, source data, type of document, etc.) [4, p. 290]. The interdisciplinary nature of computational linguistics is also evidenced by the key words in iSybislaw *linguistic engineering, information retrieval system, information retrieval language* (Polish descriptors *język informacyjno-wyszukiwawczy / system informacyjno-wyszukiwawczy*).

A component of the rubric in iSybislaw is the classification, according to which each record is classified according to the section of linguistics. In our opinion, there is every reason to supplement the classifier with computational linguistics, because there are many works and key words in the system of computational linguistics. According to our observations, these works are classified either from lexicography or semantics / pragmatics (Figure 1).

9.4. Ukrainian	9.4.2. Ukrainian. History of language
	9.4.3. Ukrainian. Dialectology
9.4.1. Ukrainian. Modern literary language	9.4.4. Ukrainian. Sociolinguistics
9.4.1.1. Ukrainian. Phonetics. Phonology	9.4.5. Ukrainian. Semantics. Pragmatics
9.4.1.2. Ukrainian. Morphology. Morphonology	9.4.6. Ukrainian. Lexicology
9.4.1.3. Ukrainian. Word-formation	9.4.6.1. Ukrainian. History of lexicon
9.4.1.3.1. Ukrainian. Nominal formation	9.4.6.2. Ukrainian. Dialectal lexicon
9.4.1.3.2. Ukrainian. Verbal formation	9.4.6.3. Ukrainian. Terminology
9.4.1.4. Ukrainian. Syntax	9.4.6.4. Ukrainian. Lexicography
9.4.1.4.1. Ukrainian. Word group	9.4.6.4.1. Ukrainian. Dictionaries
9.4.1.4.2. Ukrainian. Simple sentence	9.4.6.5. Ukrainian. Phraseology
9.4.1.4.3. Ukrainian. Compound sentence	9.4.7. Ukrainian. Onomastics
9.4.1.5. Ukrainian. Text linguistics	9.4.7.1. Ukrainian. Anthroponymy
9.4.1.6. Ukrainian. Stylistics	9.4.7.2. Ukrainian. Toponymy
9.4.1.7. Ukrainian. Language of an author	9.4.8. Ukrainian. Psycholinguistics
	9.4.9. Ukrainian. Ethnolinguistics

Figure 1: Classification in the information retrieval system

A description of 120 scientific papers with the key word computational linguistics (Polish descriptor *lingwistyka komputerowa*) with Russian, English, Bulgarian, Czech, Slovenian and Ukrainian descriptors is available in the information search system (Figure 2, 3).

Keyword

Descriptor:	lingwistyka komputerowa
Language:	pol (polski)

Keyword versions

Word	Language
lingwistyka komputerowa	polski (pol)
językoznawstwo komputerowe	polski (pol)
компьютерная лингвистика	rosyjski (rus)
computational linguistics	angielski (eng)
компютърна лингвистика	bułgarski (bul)
komputační lingvistika	czeski (cze)
računalniško jezikoslovje	słoweński (slv)
комп'ютерна лінгвістика	ukraiński (ukr)

Figure 2: The class of equivalence of the term computational linguistics

Documents with keyword: lingwistyka komputerowa

Description	Type	Actions
BARTELS Hauke, SPIESS Gunter: Das aktive deutsch-niedersorbische Internet-Lernerwörterbuch des verbalen Wortschatzes : elektronische Medien im Dienste des Erhalts einer bedrohten Minderheitensprache .- Copenhagen, 2002	Book article	[Show]
BARTELS Hauke: Das aktive deutsch-niedersorbische Wörterbuch (ADNW) des verbalen Wortschatzes .- Frankfurt ; Berlin ; Bern ; Bruxelles ; New York ; Oxford ; Wien, 2002	Book article	[Show]
OZBIČ Martina: Akustična spektralna FFT analiza samoglasniškega sistema slovenskega jezika .- Ljubljana, 1998	Book article	[Show]
KONČAR Nenad, ŠIPKA Danko: Algoritamska i heuristička obrada leksičkih podataka u računalnom prevodiocu "Neuro Tran" .- "Filologija" 1998	Journal article	[Show]
ROMIH Miro: "Amebis" in jezikovne tehnologije .- Ljubljana, 1998	Book article	[Show]
RABIEGA-WIŚNIEWSKA Joanna, RUDOLF Michał: AMOR - program automatycznej analizy fleksyjnej tekstu polskiego .- "Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la Société Polonaise de Linguistique" 2002	Journal article	[Show]
BIEŃ Janusz S., SZAFRAN Krzysztof: Analiza morfologiczna języka polskiego w praktyce .- "Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la Société Polonaise de Linguistique" 2001	Journal article	[Show]
IVANOVA Malina: L'aoriste en bulgare : éléments pour une caractéristique sémantique .- Warszawa, 1994	Book article	[Show]
BIEŃ Janusz S.: Aparat pojęciowy wybranych systemów przetwarzania tekstów polskich .- "Biuletyn Polskiego Towarzystwa Językoznawczego = Bulletin de la Société Polonaise de Linguistique" 2006	Journal article	[Show]
BRONK Zbigniew: Automat do odmiany i składania liczebników głównych : czyli o pewnej aplikacji internetowej .- "LingVaria : półrocznik Wydziału Polonistyki Uniwersytetu Jagiellońskiego" 2017	Journal article	[Show]

Displaying results **1-10** of **120** found.
[(1) |< First] 1 2 3 4 5 6 7 8 9 10 [Next >> (2)] [Last >| (12)]

Figure 3: Dokumenty ze słowem kluczowym: lingwistyka komputerowa

Keyword search not only represents the state of elaboration of computational linguistics problems in Slavic studies, but also plays the role of a scientometric base, which can be used to establish the citation index of a particular linguist (Figure 4). To date, there have been no such information retrieval databases with powerful Slavic language modules for the humanities.

Documents

	Role	Description	Type	Actions
☐	autor	BARTELS Hauke: Das aktive deutsch-niedersorbische Wörterbuch (ADNW) des verbalen Wortschatzes .- Frankfurt ; Berlin ; Bern ; Bruxelles ; New York ; Oxford ; Wien, 2002	Book article	[Show]
☐	autor	BARTELS Hauke: Das (diachrone) Textkorpus der niedersorbischen Schriftsprache als Grundlage für Sprachdokumentation und Sprachwandelforschung .- München ; Berlin ; Wien, 2010	Book article	[Show]
☐	autor	BARTELS Hauke: Konkurrierende Passivkonstruktionen in der niedersorbischen Schriftsprache : ein Beispiel für Sprachwandel durch Purismus .- München, 2008	Book article	[Show]
☐	autor	BARTELS Hauke: Ku koncepciji historisko-dokumentĕrujucego informaciskego systema za dolnoserbski słowosкład : plany k wótpóranju nuznego slĕźeńskego deziderata .- "Lĕtopis : časopis za rĕč, stawizny a kulturu Łužiskich Serbow = Zeitschrift für sorbische Sprache, Geschichte und Kultur " 2013	Journal article	[Show]
☐	autor	BARTELS Hauke: Lehnwörter im Niedersorbischen : Ergebnisse aus einem internationalen Forschungsprojekt .- "Lĕtopis : časopis za rĕč, stawizny a kulturu Łužiskich Serbow = Zeitschrift für sorbische Sprache, Geschichte und Kultur " 2009	Journal article	[Show]
☐	autor	BARTELS Hauke: Loanwords in Lower Sorbian, a Slavic Language of Germany .- Berlin, New York, 2009	Book article	[Show]
☐	autor	BARTELS Hauke: Massnahmen zur dokumentation des niedersorbischen .- "Slavia Occidentalis" 2012	Journal article	[Show]
☐	autor	BARTELS Hauke: Niedersorbisch : vom langen Leben einer oft totgesagten Sprache .- Bochum, 2009	Book article	[Show]
☐	autor	BARTELS Hauke: Rĕcna dokumentacija dolnoserbšćiny : pśigótowanje na cas bźez maminorĕcnych? .- "Rozhlad : serbski kulturny časopis" 2009	Journal article	[Show]
☐	autor	BARTELS Hauke: Słowniki - spĕchowanje rĕcy - nowe medije .- Budyšin, 2015	Book article	[Show]

Figure 4: Index of the linguist`s scientific works

The subject of computational linguistics research – modeling of language objects and processes with the help of a computer on the basis of formalized analysis and synthesis of language information – is becoming more and more clearly defined. We consider functionally equivalent terms *automatic word processing* and *natural language processing* natural language processing with Polish descriptors *opracowanie komputerowe tekstu* and *przetwarzanie języka naturalnego.* There is also another Polish descriptor *automatyczna analiza języka*. As you know, automatic text processing – a set of procedures for processing text using a computer, as a result of which we obtain a certain formal representation of the content and structure of the text, and vice versa - the construction of text on this representation (text synthesis) [5, p. 79]. The formal difference between the terms does not affect the semantics, but affects the change in the functional load of the relevant key words in the subsystems of the search language. If we talk about equivalence, we must remember that the compilers of the key word database do not use translation in their practice: the search for functional elements in each of the languages is carried out independently and taking into account the content of the document and syntagmatic links in a particular terminology system and search language key word subsystem. In this case, when establishing the degree of identity of key words that have identical or similar (terminological) meaning in several languages, the most important principle is semantic equivalence [6, p. 57].

Higher linguistic competence requires indexers to combine into equivalence classes of terms between which the relationship of aspect synonymy (conceptual synonymy) is established. This class of equivalence is, for example, the terms ***"ensuring the operation of the computer" / interface*** (special communication systems with the database, designed to replenish or edit it, as well as to obtain from it the information needed to perform a user task) / ***friendly interface*** (interface for interaction with the user who has no special knowledge of computer hardware, mathematics or software). Ensuring the operation of the computer today involves the creation of three components: tool, or hardware, which includes the necessary technical devices or tools to perform certain classes of tasks; algorithmic and software, which combines the procedures of automated logical processing of information, and linguistic, which includes databases or knowledge and special linguistic processors (means of linguistic analysis) for processing information in natural language. Nowadays, in order to increase the level of tasks solved with the help of computers, the importance of the linguistic component of computer support is growing. We have an illustration of a multifaceted understanding of the same process. During the indexing, the terms used in the described work should be reflected. Instead, their interpretive differences can be represented in the bibliographic description in the field of the abstract (annotation).

As O. Akhmanova rightly points out, comparing the identified doublets, triplets, etc., it is necessary to strive for a clear selection of descriptors, i.e. such words and phrases that would most accurately represent a particular concept, would most fully reveal the nature of this phenomenon, denoted by the term . … detection of descriptors in itself plays the role of normalization in the terminological series " [7, p. 509].

The term *automatic text processing* clearly demonstrates the dynamics of language norms of the Ukrainian language. We interpret such words as new in form, because changes in form are caused by elimination of deformations in the word-forming norm of modern Ukrainian literary language: for example, replacement of terms *automatic text processing (автоматична обробка тексту), generating model (породжуюча модель), compiling program (компілююча програма)* for *automatic word processing (автоматичне опрацювання тексту), generating model (породжувальна модель), compiling program (компілятивна програма).*

The term descriptor **lingvostatistika** *bułgarski*, as well as the descriptor **językoznawstwo statystyczne** *polski* and the ascriptors *lingwistyka kwantytatywna polski = statystyka językoznawcza polski = lingvistic statistics rosyjski = linguostatistics= kvantitativní lingvistika czeski.* As we can see, in the language of the subject description there are synonyms within one language: *statistical linguistics = quantitative linguistics* and doublets of *linguostatistics = statistics of Polish linguistics.* Besides, the Bulgarian descriptor *mathematical linguistics* is recorded separately. However, as we know, mathematical linguistics is a branch of linguistics, mathematics, semiotics and logic aimed at developing a formal apparatus, methods of mathematical and statistical analysis of various linguistic phenomena and the study of general laws of construction of sign system, which is natural language, through mathematical modeling. The branches of mathematical linguistics are quantitative and combinatorial linguistics, although quantitative linguistics is often identified with mathematical linguistics. One of the directions of mathematical linguistics is *linguostatistics,* the tasks of which are the study of the frequency of sounds, letters, their compounds, words, which became the basis for the creation of frequency dictionaries of different languages (Figure 5).

Keyword

| Descriptor: | językoznawstwo statystyczne |
| Language: | pol (polski) |

Keyword versions

Word	Language
językoznawstwo statystyczne	polski (pol)
lingwistyka kwantytatywna	polski (pol)
statystyka językoznawcza	polski (pol)
лингвистическая статистика	rosyjski (rus)
лингвостатистика	rosyjski (rus)
kvantitativní lingvistika	czeski (cze)

[Show documents]
[Show log]
[Check status]

Figure 5: Equivalence class of terms with a descriptor *językoznawstwo statystyczne*

Such semantic ambiguity is characteristic of a number of terminological units in computational linguistics, when there is a semantic shift / overlap, which leads to parallelism of terms and variability of relevant key words, for example: *terminologia internetowa (polski)* = www *(chorwacki)*; *lingwistyka korpusowa (polski) / korpusologia (polski).* In the class of equivalence to the Polish descriptor *corpus linguistics* there are terms-variants of the Czech language *jazykový korpus /*

lingvistický korpus, as well as ascriptors *linguistic corpus / linguistic corpus angielski / моўны корпус białoruski / jezični korpus chorwacki / jazykový korpus słowacki / linguistic corpus.* The term *electronic dictionary* has in the information search system two doublet descriptors *słownik elektroniczny* (polski) / *електронски речник* (serbski), and also the Polish descriptor has two English doublet ascriptors *digital dictionary angielski (eng) / electronic dictionary angielski.* Semantic uncertainty and non-fixation of terminological use are a source of variability of units in the information retrieval system, such as descriptors *language corpus* with the explanation that it is a set of texts, and *corpus of texts korpus językowy (zbiór tekstów zapisanych w postaci cyfrowej i opatrzonych odpowiednimi informacjami)* i *korpus tekstowy* (Figure 6, 7).

Wersje słowa kluczowego

Słowo	Język	Deskryptor
korpus tekstowy	polski (pol)	⦿
корпус текстов	rosyjski (rus)	○
text corpus	angielski (eng)	○
tekstni korpus	chorwacki (scr)	○
textový korpus	czeski (cze)	○
tekstowy korpus	dolnołużycki (dsb)	○

[Wyświetl dokumenty]

Figure 6: Equivalence class of terms with descriptors *korpus tekstowy*

Descriptor:	korpus językowy (zbiór tekstów zapisanych w postaci cyfrowej i opatrzonych odpowiednimi informacjami)
Language:	pol (polski)

Keyword versions

Word	Language	Descriptor
korpus językowy (zbiór tekstów zapisanych w postaci cyfrowej i opatrzonych odpowiednimi informacjami)	polski (pol)	⦿
лингвистический корпус	rosyjski (rus)	○
linguistic corpus	angielski (eng)	○
моўны корпус	białoruski (bel)	○
jezični korpus	chorwacki (scr)	○
jazykový korpus	czeski (cze)	○
lingvistický korpus	czeski (cze)	○
jazykový korpus	słowacki (slo)	○
лінгвістичний корпус	ukraiński (ukr)	○

[Show documents]
[Show log]
[Check status]

Figure 7: Equivalence class of terms with descriptors *korpus językowy (zbiór tekstów zapisanych w postaci cyfrowej i opatrzonych odpowiednimi informacjami*

The term *corpus of texts* acts as a generic and creative to its derivatives: *bilingual corpus* (Polish descriptor korpus dwujęzyczny), *parallel corpus* (Polish descriptors korpus równoległy, korpus równoległy), multilingual corpus (descriptor korpus wielojęzyczny), and also names of corpora: Corpus on the parliament corps (Polish descriptor Korpus Dyskursu Parlamentarnego), corpus Gigafida (Slovenian descriptor *korpus Gigafida*) Corpus of the Polish language of the scientific publishing house PWN (Polish descriptor Korpus Języka Polskiego Wydawnictwa Naukowego PWN), National corpus of the Russian language (Russian descriptor Национальный корпус

русского языка), Corpus of the language of incapacitated persons (Polish descriptor Korpus Mowy Osób Niepełnosprawnych) (Figure 8).

Keywords

Keyword starts with: korpus   Search

Status: ☐ Converted ☐ Draft ☐ For corrections ☐ Submitted ☐ Temporary ☐ Accepted

[Show languages filter]

Keyword	Language
korpus dwujęzyczny	polski (pol)
Korpus Dyskursu Parlamentarnego	polski (pol)
korpus Gigafida	słoweński (slv)
Korpus Języka Polskiego Wydawnictwa Naukowego PWN	polski (pol)
korpus językowy (zbiór tekstów zapisanych w postaci cyfrowej i opatrzonych odpowiednimi informacjami)	polski (pol)
Korpus Mowy Osób Niepełnosprawnych	polski (pol)
korpus paralelny	polski (pol)
korpus równoległy	polski (pol)
korpus tekstowy	polski (pol)
korpus wielojęzyczny	polski (pol)

Figure 8: Hyponyms for the term corpus of texts

Additional specification of the meaning of terminological phrases allows you to increase search performance when using relevant key words, narrowing the subject area pointed to by a particular key word. This applies, for example, to the key word Internet (computer network name) / Internet (global communication space) (descriptors *internet (globalna przestrzeń komunikacyjna) Internet (globalna sieć komputerowa)*. As you can see, the system focuses on the double spelling of the word: small letters are the name of the technology, and capital letters are the name of the computer network.

Between key words in iSybislaw we fix the paradigmatic relations caused by proximity of elements` functions of search language. Hyper-hyponymic relations are common, which provide a functional connection of the key words "part - whole", and this in turn allows their use in one query as options, for example: in the nest of the key term *analysis* there are the following terms-hyperonyms: *speech analysis , automated analysis, automatic analysis, automatic morphological analysis, semantic analysis, automatic parsing, corpus analysis, etc.*

Similarly, the term *database* is a hyperonym for the terms *hierarchical **database**; intelligent **database**; full-text **database**.* However, in the information search system there are also key words-hyponyms of the database names: WordNet, iSybislaw (Figure 9).

Słowo kluczowe

Deskryptor: WordNet (baza danych)
Język: pol (polski)

Wersje słowa kluczowego

Słowo	Język	Deskryptor
WordNet (baza danych)	polski (pol)	⦿
WordNet (database)	angielski (eng)	○
Уърднет (база данни)	bułgarski (bul)	○

[Wyświetl dokumenty]

Słowo kluczowe

Deskryptor: iSybislaw (baza danych)
Język: pol (polski)

Wersje słowa kluczowego

Słowo	Język	Deskryptor
iSybislaw (baza danych)	polski (pol)	⦿

[Wyświetl dokumenty]

Figure 9: Databases in iSybislaw

The key words in iSybislaw are Slavic resources that provide open access to research results, in particular: CLARIN (common language resources and technological infrastructure), DARIAH (digital research infrastructure of arts and humanities), IATE (interactive European terminology), ROAD (Open Access Scientific Resources Handbook), iSybislaw (Slavic Linguistics Information Bibliography System), iReteslaw (Online Slavic Text Repository), CLIP - an Internet portal combining related projects, tools and functions with computer processing of the Polish language, Neuro Tran - computer translator, morphological analysis system (SAM), module of morphological analysis "Computer dictionary of the Polish language" and other computer analyzers LEX POLOMOR, POLXOR, POLXOR, FIDA - project creation of a reference electronic corpus of texts in the Slovenian language, which also includes the ALEPH system, which has been used since 2013 at the Czech Institute and of linguistic languages of the Czech Academy of Sciences digitize linguistic information. Besides, there is a continuous replenishment of the Czech linguistic bibliography, in particular, due to the retrospective bibliography of Czech linguistics and bibliographic yearbooks; CoLT system * - a corpus of student translations, a hybrid of educational and parallel corpora, which they create for the training of translators. With the help of this system, students have access to original texts or their collections or to a number of student translations, which ensures a continuous communication between teacher and student without the need to send numerous e-mails. When translating texts, students have the opportunity to combine parallel and similar texts needed for translation, take advantage of quick access to external tools such as dictionaries (such as Glosbe), corpora (ruscorpora, NKJP, etc.), tools for creating a glossary. Teacher`s corrections and comments will be recorded automatically, which will improve the work and reduce the time required to check additional translations in case of repeated errors. In addition to optimizing the learning process, the system allows you to analyze students' mistakes.

The interrelationships of the iSybislaw system with other Slavic Internet resources will provide its users with an opportunity to get acquainted in a broader and deeper temporal, genre and territorial perspective with the state of learning a certain Slavic language, as well as with the state of Slavic studies in general in a certain national linguistic tradition.

The ordering of Slavic linguistic terminology in computational linguistics for the needs of the iSybislaw information retrieval system is multidimensional, multifaceted and permanent. To solve the problem of variance of terms, it is necessary to find out the degree of similarity between their meaning and the activity of use today. Taking into account these factors, as well as, apparently, the degree of closeness of national terms to Polish components of the iSybislaw shell language, influences the definition of the dominant in a number of such national-language variants of terms.

3. Acknowledgements

4. References

[1] Bibliograficzna baza danych światowego językoznawstwa slawistycznego iSybislaw, URL: http://www.isybislaw.ispan.waw.pl
[2] Z. Rudnik-Karwatowa, J. Banasiak, Z. Mikos, Językowe problemy optymalizacji wyszukiwania informacji w systemie iSybislaw, Slavia Orientalis, Vol. 62, No 4, pp. 631–646.
[3] Yu. N. Karaulov, Linguistic construction and thesaurus of literary language, Science, Moscow, 1981.

[4]   V. M. Glushkov, (Ed.), Encyclopedia of Cybernetics, Vol. 2, Kyiv, URE Main Editorial Office, Kyiv, 1973.

[5]   Ye. A. Karpilovskaya, Introduction to Computational Linguistics, South-East, Donetsk, 2006.

[6]   B. Yu. Gorodetsky, The term and its linguistic properties, in: Structural and applied linguistics, Vol. 3, Leningrad, 1987, pp. 54–62.

[7]   V. N. Yartseva, Linguistic Encyclopedic Dictionary, Soviet Encyclopedia, Moscow, 1990.

[8]   P. Kovalsky, I. Banasyak, Key words and equivalence classes in the iSybislaw information-retrieval implement and subject of research. Ukrainian language, vol. 4 (2017) 14–26.

[9]   Ye. Karpilovska, The role of national modules in the development of the iSybislaw base information-retrievallanguage, Ukrainian language, vol. 1 (2018) 17–29.

[10] B.V. Gornung, On the nature of synonymy in language and the theoretical preconditions for compiling synonymous dictionaries, Questions of Linguistics, 5 (1965) 95–99.

[11] A. A. Taran Information-retrieval system «Base of the world slavic linguistics (iSybislaw)» in language aducation, in: Proceedings 4th Colins 2020 computational linguistics and intelligent systems. Vol. 1. Lviv, Ukraine, 2020.

[12] T. L. Kandelaki, The experience of developing the principles of ordering terminology, in Modern problems of Russian terminology, Science, Moscow, 1986 pp. 12–140.

[13] Yu. N. Karaulov, Linguistic construction and thesaurus of literary language, Science, Moscow, 1981.

[14] B. Bojar, Z. Rudnik-Karwatowa, Koncepcja nowoczesnego systemu informacji slawistycznej, in: Językoznawstwo. Prace na XII Międzynarodowy Kongres Slawistów w Krakowie, Warszawa, 1998, pp. 42–49.