BioGecko Vol 12 Issue 03 2023

ISSN NO: 2230-5807

Multimedia Thesaurus: Designing and Prospects of Using in Mining

Sobirova Maftuna Jamolovna

Teacher Navoi State University of Mining and Technology Uzbekistan, Navoi

Abstract:
This article briefly analyzes the concept of a thesaurus, discusses some of the problems encountered in the creation of multimedia multilingual thesauri and provides solutions to them, shows how thesauruses can be used effectively with multimedia, and hierarchical ranking of related concepts and terms in the field of mining in the Uzbek language is revealed through examples.

Keywords:thesauri, multimedia, multilingual, broader term, narrower term, synonymic sets,mining, specialized, controlled terminology, concept.

1. INTRODUCTION

Currently, the development of modern fields of human activity that require knowledge cannot be imagined without the increasing role of computer technologies. Today, due to the significant increase in the flow of information in higher educational institutions, there is a need to search for new ways of storing, presenting, formalizing and organizing them, as well as processing which in turn leads to creation of multilingual educational environment.

In the world today, majority of higher educational institutions use English as a medium of instruction(EMI), in Uzbekistan with no exception, especially in technological universities which inevitably requires the creation of a multilingual educational environment. Sofar a lot of effort was made in terms of creating theoretical foundation and concluding the results of practice in teaching foreign languages. However, we should still admit that subject teachers do not have enough knowledge in English and teachers of foreign languages struggle with categorical-conceptual apparatus and the subject content of specialized disciplines. According to professional subject and language teachers “Content and language integrated learning (CLIL) allows for interconnection obtaining knowledge and forming creative thinking in the process of live communication, actualizing interaction in a foreign language and solving professionally oriented tasks”.

Meanwhile, in Uzbekistan, we still use XX century methods and technologies in teaching foreign languages, allowing to carry out professionally oriented integrative training, in the course of which knowledge, language skills and abilities in the form of firstly reading original authentic sources in a foreign language, comprehending, understanding and extracting information on the specialty, expressing their thoughts ideas in writing and secondly taking part in oral speech communication in a foreign language to be able to understand others and formulate thoughts in their own statements. At the same time, the advantages of the domestic concept have been convincingly proved.

According to the study of E.G, Krylov, the components of the system of foreign language professional communicative competence, means and forms of implementation of integrative bilingual education, as well as the professionally oriented content are considered in teaching a foreign language and professional disciplines. The result of such training indicate that the developed and implemented system of bilingual education not only fully achieves, its goal, but also contributes to the development of interdisciplinary thinking of the future miners, which is an indispensable attribute of his competence in any industry.

However, the most challenging part of teaching and mastering a foreign language is the lack of specialized dictionaries in chosen industry, namely, in mining field majority of subject teachers and students find it difficult to write scientific articles and case studies they do not have access to reliable sources. Therefore, there is a issue of creating a multimedia mining thesaurus, comprehensive knowledge base for miners,which can be used for various practical purposes for both workers and students.

In particular, there is a great need for systems based on neural networks that can be extracted any information from text without the human factor. In the middle of the 20th century, along with the World Wide Web, semantic webs appeared, in which additional tags carrying information about the semantics of elements on hypertext pages were provided. The most important part of the semantic web is the thesaurus, which is a lexical database consisting of a network of semantically related words.

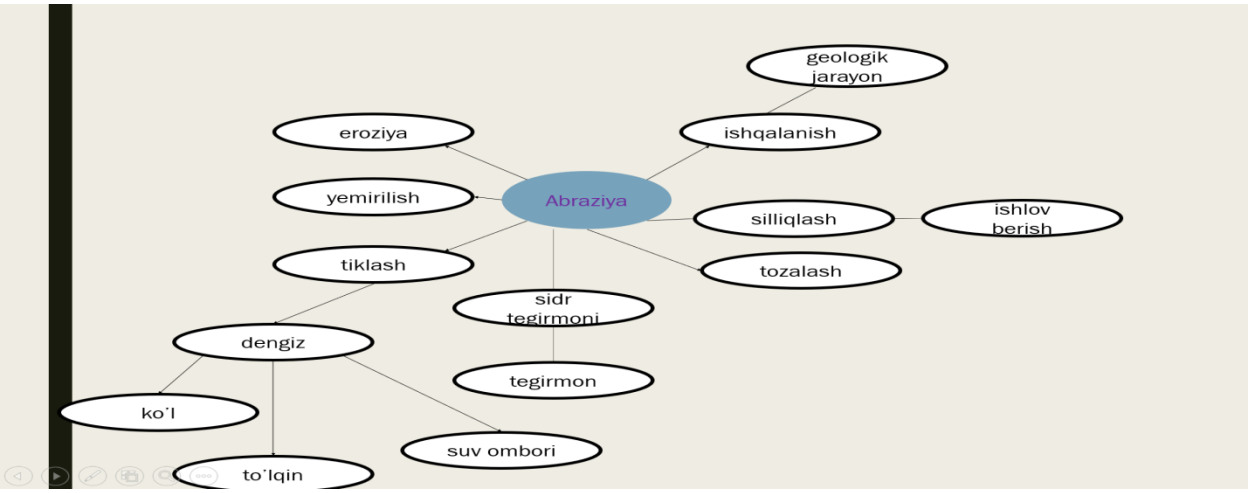
In the field of world applied linguistics and computer linguistics, thesaurus is created in order to increase the possibility of information search, improve automatic translation systems, achieve component analysis of texts, and reflect the vocabulary of the language. resulting in the creation of many other language resources based on the WordnNet English lexical database. Therefore, special attention was paid to natural language processing (NLP), linguistic modeling of language, tagging of word groups, formation of a set of semantic relations (synonymy, meronymy, hypernymy, antonymy, homonymy).

Need for developing industry-specific thesaurus

Currently formalizing the language for artificial intelligence in Uzbek linguistics and computer linguistics, creating linguistic models, achieving practical results in cyberlexicography, especially ensuring the active integration of the state language into modern information technologies and communications, popularizing the Uzbek language in the Internet global information network , to ensure that it occupies a worthy place, to create grammatical and semantic analysis systems, for this, researches and practical projects are being carried out on natural language processing in the Uzbek language. As a result, in applied linguistics and computational linguistics text processing, speech synthesizer, speech recognition, machine learning, natural language processing (NLP), computer translation, the basis for the development of the field of corpus studies, computer lexicography and language didactics is being created. Now, in order to raise the international status of the Uzbek language, especially in the field of mining, in order to widely disseminate the researches carried out in the field of mining, based on world experience, it is based on semantic relations, has a network of words, and contains the entire Uzbek language. The creation of a multimedia thesaurus of mining terms with lexical richness is considered urgent, and the deepening of scientific research in this regard is waiting for its solution (M.Abjalova, 2021)

THESAURUS (Greek thesauros - treasure, wealth) - 1) a dictionary that covers all the words of a certain language and fully reflects their usage in the text. According to the principle of word selection, dictionaries compiled according to the language of writers or some works belonging to their works are considered thesaurus.

A thesaurus is also a special terminology within certain fields of knowledge. In general, Lukashevich notes that a thesaurus is specialized terminology. Authors of thesauruses created so far state that the thesaurus is studied as a semantic system that provides information and data. we often come across the concept of ontology together with thesaurus. Bioscientists, in addition to linguists, recognize the wealth of these two languages as mutually similar. In fact, linguistic ontology covers all areas of language possibilities, while thesauruses are limited to concepts specific to a specific set or field of direction (Picture1).



Picture1. Semantic relations of the word abrasion.

And in modern linguistics, we can consider thesauruses as a special type of dictionaries that show semantic relations between vocabulary units (synonym, antonym, paronym, hyponym, hypernym, homonym, etc.). Also, a thesaurus is a dictionary in which closely related words and phrases are grouped into units such as concepts or descriptors and their semantic relations are shown in a hierarchical form between these concepts.

Although, Masterman(1961) is considered to be the first to bring the thesaurus into existence, the beginning of formation thesauri dates back to the Dictionary of Synonyms compiled by Philo of Biblsk. Also, the book "Amar-Kosha" in poetic form written in Sanskrit of the 1st-2nd centuries is recorded in the sources as the first thesaurus. However, the first modern English thesaurus was created by Peter Mark Roger.

Unlike annotated dictionaries, thesaurus does not only describe the meaning of words, but also allows to determine the meaning of words by interrelation with other concepts and their groups, thanks to which owners of certain fields, for example, in mining, from it artificial intelligence can be used effectively in filling the knowledge bases of the systems. Currently, the word meaning is well described in the dictionaries that are created in a modern form with the development of information technologies, Internet and mobile applications, various software and text processing systems with the help of computers, and the results of this word search are presented in a very convenient and perfect way. giving the user the opportunity to present. Such dictionaries can be divided into monolingual and multilingual thesauri according to the number of languages in the database.

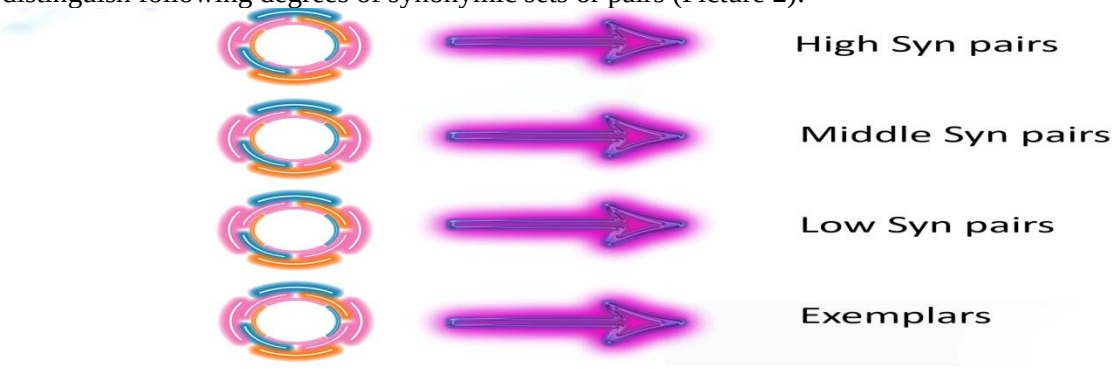
What makes thesauruses different from other linguistic systems and key graphic sources, and therefore, what increases their value, is their semantic relations. N.V. Lukashevich, the scientist who created the thesaurus of the Russian language, and his team have for years researched the technologies of word processing and their application to solve various problems of information retrieval, and made the following conclusions about the creation of thesauruses:

- the structure of thesauruses should be specially adapted to text processing tasks;
- thesaurus technologies should not be contrasted with modern word processing technologies, but the latest advances in this field should be taken into account organically;
-taking into account such requirements, the use of thesaurus significantly improves the quality of problem solving compared to the most optimal verbatim working methods.

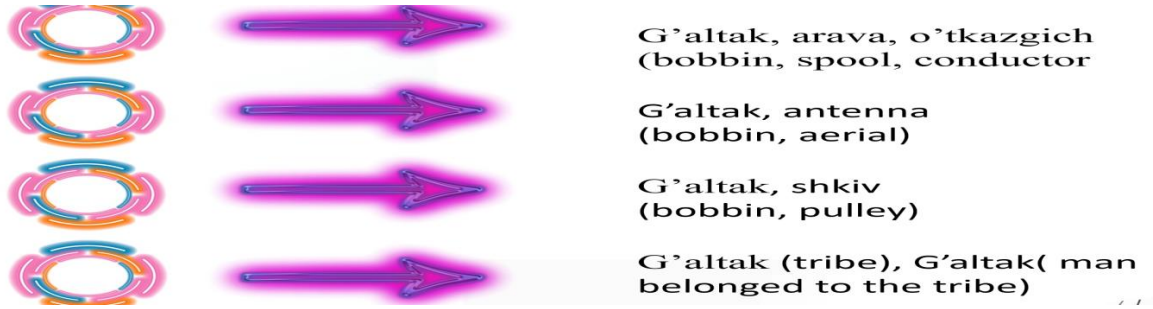
In developing thesaurus for mining, we often come across to words, concepts and terms that should be understood explicitly. Each term in the thesaurus should be used to represent only one concept, if concepts are units of thought, the terms are source of expressing them. Clarke and Zeng note the importance of this distinction between term and concept when considering thesauri in an electronic environment as humans are able to see that a thesaurus contains both concepts and terms.

We can use concepts and terms interchangeably and still draw correct inferences. A computer, however, is unable to do so and this can be easily misunderstood by the miner users as they are not

aware enough the concepts and terms of the chosen language. Once all concepts and terms are gathered, the task before the author is to clarify and distinguish them into synonyms, quasi-synonyms, hyponyms, hypernyms, homonyms, meronyms and others. Synonymic sets in thesauri in mining are of paramount importance as they are at the center of any controlled dictionaries. In this case, we distinguish following degrees of synonymic sets or pairs (Picture 2):



In the following, we can see this division in the example of the word g'altak(bobbin)which is widely used among miners in the Uzbek language (Picture 3).



The separation of concepts is not always easy task. ISO 25964-1 suggests 5 options to consider when splitting concepts and terms in degrees of use:

1. To accept the concept as it is;
2. To divide the concept into various terms of usage; mining engineering, electrical engineering, metallurgy and others;
3. To test the concept and admit or diminish it after time;
4. To reject some concepts, but keep it as a non-preferred term linking to a broader or narrower term.
5. To get rid of the concept completely if it is irrelevant or out of use.

Another common unit of thesauri is hierarchical arrangements that represent using the broader term or narrower term relationships. These terms denote relationships between the concepts in a thesaurus and indicate whether a concept contains or is contained by another concept. The broader and narrower term relationships (meronym-holonomy; hyponym-hypernym). In other words, these terms denote relationship between the conceptsin a thesaurus and indicate whether a concept contains or is contained by another concept. BT relationship are used to direct users to a concept that is higher up in a thesaurus hierarchy and therefore a broader or more general concept than the one they were seeking. Likewise, NT relationships direct users further down to the hierarchy to more specific terms (Hedden, 2010). To put it more clearly, meronym and holonomy make the whole thing, for example: book is holonomy, the parts of the book are meronym: cover, title, page, table of contents; and the text ismeronym. Hypernym and hyponym are concepts related to top-down or bottom-upor from a narrow

subject to a broad subject or from a broad subject to a narrow subjecta hypernym is a concept with a higher or broader meaning, a hyponym is a concept with a lower or narrower meaning; bird - hypernym, chicken, sparrow, parrot, etc. – hyponym. To some extent, it can be considered that that holonomy and meronyms are mainly used in relation to inanimate objects, it can be considered that hypernyms and hyponyms are used in relation to living objects.
C. Ryan states three main categories of hierarchical relationships in developing specific thesaurus:
1. The generic relationship in which genus-species or thing-king relationship is described.
2. The partitive relationship that illustrates whole-part relationship.
3. The instance relationship that uses name particular instances of a class of things (Ryan, 2014).

Multimedia multilingual thesauri: challenges and solutions
A thesaurus is a long-term project, which requires a significant amount of planning and a number of aspects of thesaurus need to be considered before the construction begins (Hedden, 2010). If your target is developing multilingual thesaurus for specific field the task gets even harder. ISO 25964-1 recommends multimedia thesaurus to plan the objectives and features of the thesaurus as well as the resources available for the thesaurus, to deciding and choose the software for the thesaurus to make it easy and accessible for users. In accordancewith a subject-logical description of the design and properties of the object are usedonly the names of types of logical-semantic relations at the level of paradigmatic and syntagmatic. For example: genus - species;whole - part; identity; the object and its components (paradigmatic relations); object characteristics, propertiesobject (syntagmatic relations), etc. In this case, it becomes that subject-matter connections come first,and the identification of the corresponding logical-semantic relations of the desired termsplays a supporting role, i.e. we are talkingabout the subsequent fixation of the matched matches (se Formation of foreign language competence of the future mining engineer requires simulation of situationsintercultural professional interaction. In this regard, individuala multimedia multilingual thesaurus is:
- subject-logical representationobject in accordance with its structure,properties, etc.;
- fixing the correspondence of the subject-logical representation of the object to the logicalsemantic relations (whole and part,generic, etc.) between terms that describe an object;
- an electronic document illustrating, with the help of hyperlinks to multimedia sources, the actual useforeign language terms in mining engineering communication.

So far, a number of multilingual thesauri are increasing day by day and Multilingual Thesaurus of Environmental Terms, EuroVocMultilingual Multidisciplinary Thesaurus, AGROVOC, The FAO Multilingual Thesaurus, Multilingual Thesaurus On Land Tenure, Multilingual Thesaurus of Geosciences by Elsevier Science, Multilingual Thesaurus of Education, Thesaurus on Tourism & Leisure Activities (Trilingual), The World of Multilingual Environmental Thesauri, Thesaurus for IR (IEKO), Multilingual Archival Terminology, Semantic-based Multilingual Islamic Finance Thesaurus, The Multilingual Thesaurus of LAURIN, INIS Multilingual Thesaurus are among them.

The most challenging part of designing a specified multilingual thesaurus is planning to give it multimedia. Thus, we are suggesting following criteria before thesaurus construction:
- To state the main purpose of specific terminology;
- To identify the users;
- To identify the scope and content of the thesaurus;
- To outline people responsible for each aspect of thesaurus construction and management;
- To source sufficient and relevant vocabulary resources and to select terms;
- To decided what type of software to choose for publication of the documentation;
- To support and test all expected thesaural relationships, features and notes;
- To integrate thesaurus with other applications and systems;
- To choose whether thesaurus will be available in paper format or online;
- To decide in which style of the display of the thesaurus will be;
- To include node labels and guide terms in the software;

- To include various types of relationships and other elements and others.

2.CONCLUSION

Modern society makes new demands on graduates of engineering universities (Especially in mining). They must be not only highly qualified specialists, but also speak foreign languages. Today, future specialists are faced with acute tasks of using a foreign language as a way of communication in the field of professional activity. From this it follows that foreign language teachers are faced with new tasks related to the study and implementation of new teaching methods (M.Sobirova, 2020). Thus, developing controlled multilingual thesaurus for mining is one of the most important task for linguists. Uncovering and using the communicative potential of the thesaurus throughsingling out speech samples from the contextual part and describing situations of professional communication is the area of responsibility of a foreign language teacher.In the process of creating a thesaurus and assimilation
based on it, students' speech samples are updated and expandedexisting knowledge about the design of the car due to the establishment of interdisciplinary connections in the course of independent workwith a variety of English-language resources. The speech patterns selected during processing of the contextual part of the thesaurus arenot a set of phrases for mechanical memorization, but a real "building material",assigned and used by trainees inlearning situations of foreign language communication in models of future situations of intercultural engineering communication.
In conclusion, we outline the immediate prospects that open up as a result of testing the technology for creating a multimedia thesaurus in line with integrative teaching of a foreign language and professional disciplines. Current stagecan be described as a stage of developmentzones of closest integration, when it is also necessary to model a number of subjectareas from previous general engineering disciplines. Development of the zone of furthersubject and competence integrationincludes an analysis of the content of textbooks, educational and industrial practice,laboratory classes for subsequentspecial disciplines: designvehicles, power units, operation and service, testing of vehicles, etc. Obviously, the specifics of mining disciplines affect the structure of the thesaurus used as a tool in engineering education. At the same time, the requirements forprofessional competence and scopethe activities of future graduates determine the criteria for the selection of speech samples typical of a foreign language mining communication in the course of performing tasks ondesigning, improving the design and characteristics of the teems.Thus, the technology of creating a multimedia thesaurus makes it possible to optimize the processes of forming foreign language competence and interdisciplinarythinking of the future engineer.

3. REFERENCES

1. Abjalova M.A. O’zbektiliontologiyasiyaratishtexnologiyasivakonsepsiyasi: Toshkent: Nodirabegim, 2021. PP. 20-25.
2. Leychik B.M. termini I terminosistemi-pogranichnaya oblast mejduestestvennimiiskusstvennim v yazike// Voprositerminologii I lingvisticheskoystatistiki. Voronej, 1981. PP. 68-115.
3. Lukashevich N.B. Tezaurusi v zadachaxinformatsionnogopoiska. -Moscow, 2010-Pp. 84.
4. Hedden, H. 2010a The accidental taxonomist. Medford, New Jersey. Information Today, Inc.
5. International Organization for Standardization 2011 ISO 25964-1:2011, information and documentation. Thesauri and interoperability with other vocabularies. Part 1: thesauri for information retrieval. Geneva. International Organization for Standardization.
6. International Organization for Standardization 2013 ISO 25964-2:2013, information and documentation. Thesauri and interoperability with other vocabularies. Part 2: interoperability with other vocabularies. Geneva. International Organization for Standardization.
7. Ryan C. Thesaurus construction guidelines: an introduction to thesauri and guidelines on their construction. Royal Irish Academy and National Library of Ireland. 2014. Pp. 40-58.

8. Krylov, E.G. (2015). Bilingvalnoyeobucheniyeinzhenernymdistsiplinamiinostrannomuyazyku v vuze [Bilingual Teaching of Engineering Disciplines and Foreign Languages in Higher Education Institution]. Moscow, Izhevsk: Computer Sciences Institute Publ., 148 p. (In Russ.).
9. Masterman, M. (1961). Semantic message detection formachine translation, using an interlingua. In Proceedings of the 1961 international conference on machine translation of languages and applied language analysis (Vol. 13, pp. 438–475). London: Her Majesty's Stationary Office.
10. Roget, Peter Mark (1824). "Explanation of an optical deception in the appearance of the spokes of a wheel when seen through vertical apertures". Philosophical Transactions of the Royal Society of London. 115: 131–140.
11. Clarke, S. and Zeng, M. 2012 From ISO 2788 to ISO 25964: the evolution of thesaurus standards towards interoperability and data modeling. Information Standards Quarterly 24(1), 20–26.
12. Hedden, H. 2010b Combining controlled vocabularies, available at: http://www.heddeninformation.com/Combining ControlledVocabularies.pdf (22 November 2013)
13. Sobirova, M. J. (2020). Methodology for increasing the efficiency of teaching engineers in English. ISJ Theoretical & Applied Science, 02 (82), 639-642
14. Marina M. Bazhutina (2021). Multimedia Thesaurus: A Case of Designing and Prospects for Using in Engineering Education. Higher Education in Russia. 2021, vol. 30, no. 1.Pp. 73-85