TASHKENT STATE UNIVERSITY OF ECONOMICS

International journal of topics

INNOVATION SCIENCE AND TECHNOLOGY

Scopus || Electronic journal specializing in Scopus

ISSUE 1

Acceptance of papers **January, 2026**

Acceptance of papers

Published monthly

Topics

economics, technology, social sciences

ISSN 3060-5229

Electronic publication, Issue 1. 109 pages.
Approved for publication on January, 2026.

THE SCIENTIFIC-POPULAR ELECTRONIC JOURNAL **"INNOVATION SCIENCE AND TECHNOLOGY"** HAS BEEN REGISTERED UNDER THE NUMBER **C-5669633** BY THE AGENCY FOR INFORMATION AND MASS COMMUNICATIONS (AOKA) OF THE REPUBLIC OF UZBEKISTAN, EFFECTIVE FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: https://ist-journal.uz

Email: innovationist2025@gmail.com

The scientific electronic journal “Innovation Science and Technology” has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.

Editorial board:

CONTENTS

LEXICAL-SEMANTIC ARCHITECTURE OF MODERN WORDNET SYSTEMS

Aynura Axmedova
Master's Student (1st Year), Department of Computational Linguistics
Tashkent State University of Uzbek Language and Literature (ToshDO'TAU)
E-mail: aynuraaxmedova345@gmail.com
ORCID: 0009-0000-0940-857X

Abstract: This article presents a scholarly analysis of the WordNet model and the lexical-semantic networks developed on an international scale. The study first examines the conceptual framework of Princeton WordNet, which is based on psycholinguistic principles and structured through synsets, glosses, and semantic relations. It further explores the architecture, functional characteristics, and role of WordNet versions 2.1 and 3.1 in natural language processing (NLP) tasks. National systems such as RuWordNet, Japanese WordNet, Korean WordNet, and Turkish WordNet are analyzed using comparative statistical indicators, and the mechanisms of their integration into the general WordNet paradigm are identified. The article also highlights the global significance of WordNet in the development of multilingual semantic networks and its functional potential as a linguistic and computational resource.

Key words: WordNet, synset, lexical-semantic network, Princeton WordNet, RuWordNet, Japanese WordNet, Korean WordNet, Turkish WordNet, Global WordNet Association, semantic relations.

Annotatsiya: Mazkur maqolada WordNet modeli va uning xalqaro miqyosda shakllangan leksik-semantik tarmoqlari ilmiy jihatdan tahlil qilinadi. Dastlab Princeton WordNetning psixolingvistik tamoyillarga asoslangan sinset tuzilishi, glosslar hamda semantik munosabatlar orqali shakllangan konseptual modeli yoritiladi. Shuningdek, WordNet 2.1 va 3.1 versiyalarining arxitekturasi, funksional xususiyatlari va tabiiy tilni qayta ishlash (NLP) jarayonlaridagi o'rni ko'rib chiqiladi. RuWordNet, Japanese WordNet, Korean WordNet va Turkish WordNet kabi milliy tizimlar qiyosiy statistik ko'rsatkichlar asosida tahlil qilinib, ularning umumiy WordNet paradigmasiga integratsiya mexanizmlari ochib beriladi. Maqolada WordNetning ko'p tilli semantik tarmoqlarni yaratishdagi global ahamiyati hamda lingvistik-hisoblash resursi sifatidagi funksional imkoniyatlari asoslab beriladi.

Kalit so'zlar: WordNet, sinset, leksik-semantik tarmoq, Princeton WordNet, RuWordNet, Japanese WordNet, Korean WordNet, Turkish WordNet, Global WordNet Association, semantik munosabatlar.

Аннотация: В статье проводится научный анализ модели WordNet и сформированных на её основе лексико-семантических сетей международного уровня. Вначале рассматривается концептуальная модель Princeton WordNet, основанная на психолингвистических принципах и формируемая посредством синсетной структуры, глосс и системы семантических отношений. Далее анализируются архитектура, функциональные особенности и роль версий WordNet 2.1 и 3.1 в процессах обработки естественного языка (NLP). Национальные системы RuWordNet, Japanese WordNet, Korean WordNet и Turkish WordNet рассматриваются на основе сравнительных статистических показателей, а также раскрываются механизмы их интеграции в общую парадигму WordNet. В статье также обоснована глобальная значимость WordNet в создании многоязычных семантических сетей и его функциональный потенциал как лингвистического и вычислительного ресурса.

Ключевые слова: WordNet, синсет, лексико-семантическая сеть, Princeton WordNet, RuWordNet, Japanese WordNet, Korean WordNet, Turkish WordNet, Global WordNet Association, семантические отношения.

INTRODUCTION

Lexical-semantic networks are among the most important resources in modern computational linguistics. Among these networks, the WordNet model occupies a distinctive position, as it aims to transfer the psychological model of language into the field of computational linguistics. The core idea of this model is to represent words not as isolated units, but as elements of a complex system interconnected through semantic relations.

The fundamental unit of WordNet is the synset, which is defined as a set of synonymous lexical items that express the same concept. Each synset is associated with a defining description (gloss), usage examples, and semantic relations linking it to other synsets [1]. These relations include hypernymy–hyponymy (general–specific relations), meronymy–holonymy (part–whole relations), antonymy (opposition), and troponymy (manner-of-action relations), among others. Together, these relations form the structural foundation of the lexical-semantic network.

One of the most significant features of WordNet is its suitability for computational processing and its ability to model the lexical-semantic structure of language in a formal and standardized manner. This model is widely used not only in linguistic research, but also in various natural language processing (NLP) applications, including machine translation, word sense disambiguation (WSD), semantic search, text analysis, chatbots, and question–answering systems. The modular, extensible, and multilingual nature of WordNet has made it one of the most successful and widely adopted lexical-semantic systems at the global level.

LITERATURE REVIEW

The theoretical foundations of WordNet originate from psycholinguistic and cognitive studies of lexical organization. George A. Miller and Philip Johnson-Laird, in their work Language and Perception published in 1976, established the idea that human lexical knowledge is structured conceptually rather than as isolated word lists. Building on this cognitive perspective, George A. Miller and his collaborators later introduced WordNet as an online lexical database, first presented in the International Journal of Lexicography in 1990. In this framework, lexical meaning is represented through synsets and systematic semantic relations, providing a structured model that reflects human mental lexicons. The formalization of WordNet as a lexical database for English was further consolidated in Miller's publication in Communications of the ACM in 1995, which positioned WordNet as a computationally usable semantic resource.

Subsequent research focused on the semantic architecture and practical applications of WordNet in natural language processing. Christiane Fellbaum, a key contributor to the development of WordNet, provided an extensive conceptual and ontological analysis of the system in Theory and Applications of Ontology published in 2010. Earlier, her work presented at the ACL Workshop in 2000 emphasized the growing impact of WordNet on NLP tasks, highlighting its role in semantic interpretation and knowledge representation. In parallel, Philip Resnik demonstrated the usefulness of WordNet's hierarchical structure for measuring semantic similarity in his IJCAI paper published in 1995, where information content was employed to quantify conceptual distance between synsets.

Research on word sense disambiguation and multilingual expansion further strengthened the significance of WordNet. Michael Lesk introduced an early dictionary-based approach to sense disambiguation in the SIGDOC Conference Proceedings of 1986, which later inspired WordNet-based extensions. Banerjee and Pedersen, in their CICLing paper published in 2003, proposed extended gloss overlap methods that explicitly relied on WordNet's semantic relations. Pedersen's technical report in 2004 systematically reviewed similarity measures derived from WordNet. At the multilingual level, Pianta, Bentivogli, and Girardi presented MultiWordNet at the Global WordNet Conference in 2002, demonstrating how aligned WordNet architectures can support cross-lingual semantic interoperability. Collectively, these studies establish WordNet as a robust lexical-semantic architecture that underpins both monolingual and multilingual NLP systems.

RESEARCH METHODOLOGY

The research methodology of this study is based on a systematic combination of qualitative and comparative analytical approaches. The primary data were obtained from established WordNet resources, including Princeton WordNet and several national WordNet systems, through their official databases, documentation, and publicly available lexical-semantic structures. Synsets, glosses, and semantic relations were extracted in structured formats such as XML and LMF to ensure consistency and comparability of the data. In addition, secondary data were collected from peer-reviewed publications that describe the theoretical foundations and technical implementations of WordNet-based systems.

The analysis was conducted using comparative and descriptive methods. Quantitative indicators, such as the number of synsets, lemmas, senses, and relation types, were analyzed to identify structural differences and similarities among WordNet systems. Qualitative analysis focused on examining the organization of semantic relations, hierarchical structures, and integration mechanisms within multilingual frameworks. This combined approach made it possible to assess the architectural characteristics of modern WordNet systems and evaluate their effectiveness as lexical-semantic resources in natural language processing.

ANALYSIS AND RESULTS

Today, numerous national and multilingual resources have been developed based on the WordNet model, including EuroWordNet, BalkaNet, Arabic WordNet, Chinese WordNet, Hindi WordNet, MultiWordNet, Open English WordNet, and other systems. These resources adapt the theoretical principles of WordNet to the linguistic and conceptual characteristics of individual languages, thereby extending the applicability of the WordNet paradigm across diverse language families.

Princeton WordNet (PWN) represents one of the most significant milestones in the development of modern lexical-semantic networks. The project was initiated in the mid-1980s at the Cognitive Science Laboratory of Princeton University under the leadership of Professor George A. Miller. Its original objective was to classify English lexical units according to a psycholinguistic and semantic model and to formally represent their conceptual interconnections in a computational framework. As a result, WordNet initially emerged as a cognitive approach situated at the intersection of linguistics and psychology.

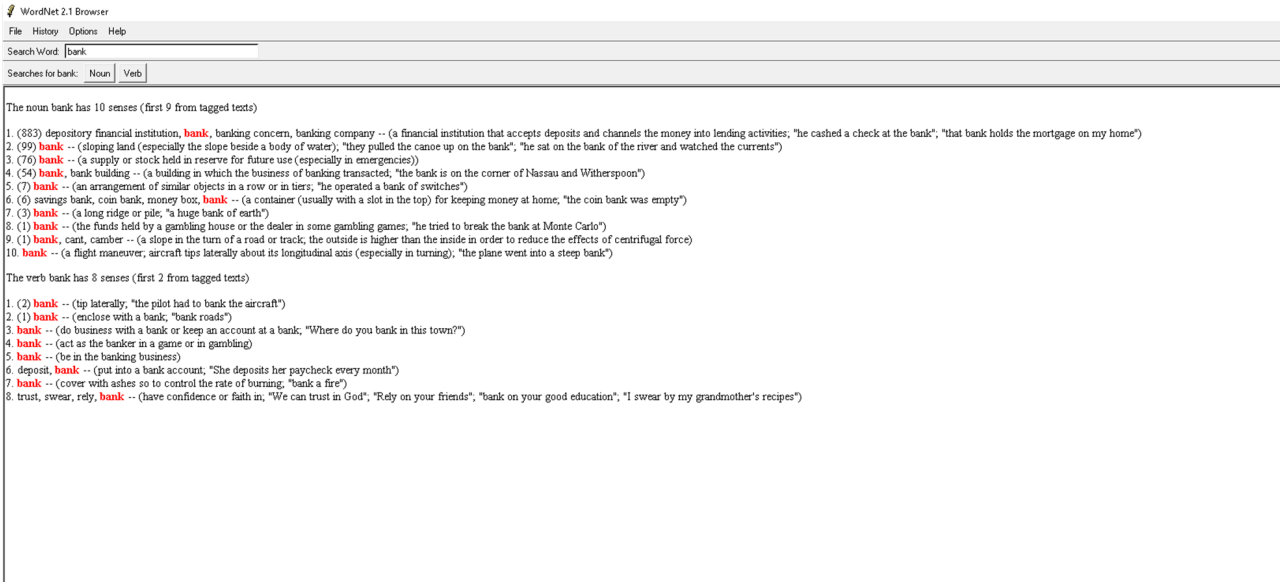

Figure 1. Search results for the word “bank” in the WordNet 2.1 Browser

The most important semantic relations in PWN are hypernymy–hyponymy relations, which provide the network with a hierarchical structure. This hierarchy significantly enhances the effectiveness of WordNet in computational linguistics, as it facilitates semantic search, automatic classification, and word sense disambiguation based on general and specific conceptual categories. In addition, PWN incorporates relations such as meronymy, antonymy, and troponymy. Each type of relation represents a specific category of semantic knowledge and contributes to the multidimensional structure of the conceptual model.

From a technical perspective, Princeton WordNet is distributed in open formats, and its structure is supported by dedicated APIs as well as XML and RDF representations. PWN has undergone multiple revisions, including versions 1.5, 2.0, 2.1, 3.0, and 3.1, all of which have been widely adopted by the scientific community. In particular, WordNet versions 3.0 and 3.1 have become de facto standards in natural language processing research, especially in areas such as Word Sense Disambiguation, Information Retrieval, Text Classification, and Machine Translation.

Throughout its development history, several major versions of WordNet have been released, among which versions 1.0, 2.0, 2.1, 3.0, 3.1, and the more recent Open English WordNet (OEWN) are of particular importance. Among these, WordNet 2.1 was distinguished by its user-oriented graphical interface, known as the WordNet Browser, while WordNet 3.1 is regarded as the version closest to an academic standard due to the richness of its semantic data, structural consistency, and suitability for formal lexical-semantic modeling. Accordingly, this article focuses on an in-depth analysis of these two versions. Released in 2005, WordNet 2.1 became one of the most widely used and extensively studied versions in the history of WordNet, playing a significant role in linguistics, computational linguistics, and educational contexts. The architecture of the WordNet 2.1 Browser includes a “Search Word” input field, part-of-speech (POS) selection buttons (Noun, Verb, Adj, Adv), a “Synset List” displaying relevant synsets, detailed glosses for each synset, usage examples, and a menu of semantic relations.

Figure 2. Global WordNet Association (GWA) interface

The Global WordNet Association (GWA) is an international scientific platform aimed at unifying WordNet systems worldwide, coordinating their theoretical and technical development, and standardizing the process of creating multilingual lexical-semantic resources. The emergence of Global WordNet is driven by the widespread adoption of WordNet, the development of WordNets for different language families, and the growing need to ensure semantic compatibility among them. The primary objectives of the association include the consistent methodological application of the WordNet conceptual model, the establishment of mapping mechanisms between WordNets of different languages, and the integration of multilingual semantic resources into thesauri, ontologies, and NLP systems. Through conferences, workshops, and technical guidelines, the organization also contributes to the continuous improvement of these resources. The biennial Global WordNet Conference provides a forum for presenting new WordNet projects, discussing updates to existing systems, and addressing standards such as LMF, RDF, and OWL, as well as issues related to multilingual synset integration, lexicographic methods, and the technical architecture of semantic networks. One of the most important initiatives of Global WordNet is the Open Multilingual WordNet (OMW) project, which interlinks WordNets of different languages by aligning their synsets with Princeton WordNet under a unified standard, thereby creating a shared database for multilingual semantic networks. In this regard, Global WordNet functions not only as a platform that brings together the scientific community, but also as a global methodological center shaping the ecosystem of lexical-semantic resources worldwide.

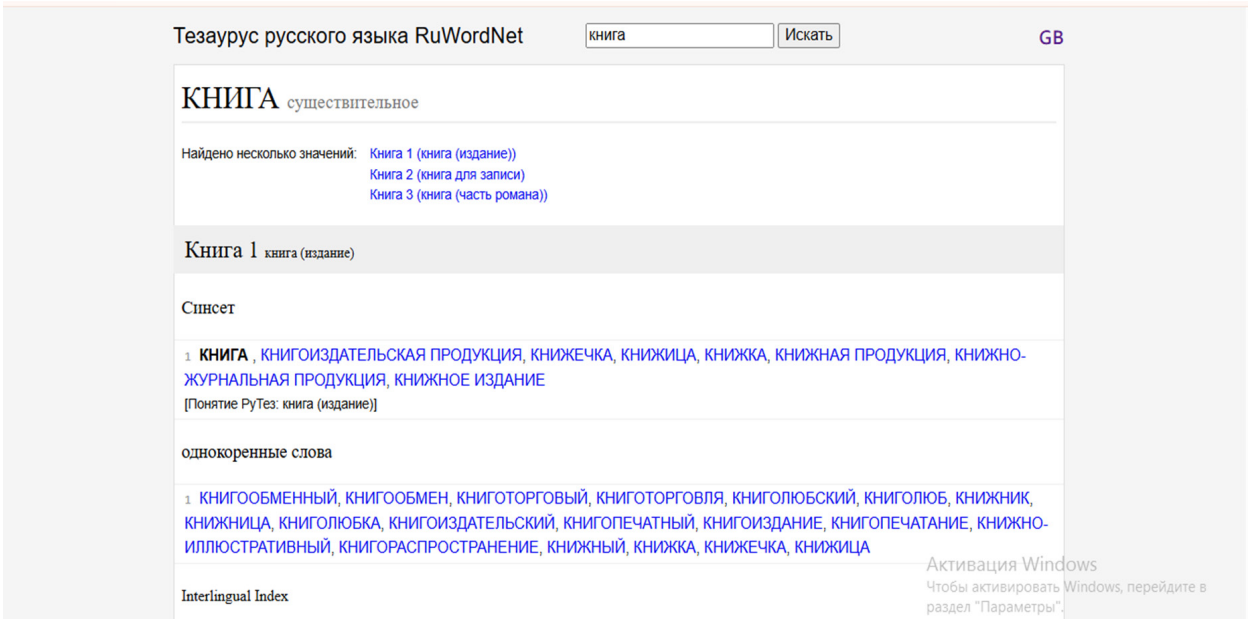

Figure 3. RuWordNet interface

RuWordNet. RuWordNet is the largest and most advanced monolingual lexical-semantic system developed for the modern Russian language. It was formed as a conceptual semantic network adapted to the WordNet format through automated transformation based on the RuThes thesaurus in Russian linguistics. Since 2011, the system has been supported by scientific grants from the Russian Foundation for Basic Research (RFBR) and the Russian Science Foundation (RNF). In accordance with international WordNet standards, RuWordNet provides synsets, glosses, semantic relations, and complex interconnections among lexical units.

The latest version, RuWordNet 2.0, contains approximately 111,500 unique words and multiword expressions, representing a total of 154,111 senses, including 29,297 noun synsets, 12,865 adjective synsets, and 7,636 verb synsets. The system integrates a wide range of paradigmatic and syntagmatic relations, such as synonymy, hyponymy–hypernymy, meronymy–holonymy, antonymy, cause–effect relations, event sequencing, class–instance relations, logical implication, and domain-specific links. A key feature of RuWordNet is that its synsets have been validated both through expert linguistic annotation and corpus-based automatic methods. High-quality conceptual descriptions are ensured through detailed glosses, usage examples, and semantic labeling. Full compliance with the WordNet LMF format facilitates seamless integration with international WordNet resources and enables applications in semantic search, word sense disambiguation (WSD), machine translation, question–answering systems, syntactic-semantic analysis, knowledge graph construction, and intelligent information retrieval.

Figure 4. Japanese WordNet interface

Japanese WordNet (JWN) is a national lexical-semantic resource developed by the National Institute of Information and Communications Technology (NICT) in Japan. Its primary objective is the formal modeling of Japanese-language concepts in the form of synsets. The project began in 2006, and by 2009 a comprehensive WordNet model for the Japanese language had been established. Although JWN was initially aligned with Princeton WordNet synsets, it is not a direct translation; rather, it is based on the reconstruction of concepts in accordance with the actual semantic structure of the Japanese language.

The system comprises more than 57,000 synsets, over 93,000 word forms, 158,000 lemma–sense pairs, and more than 135,000 semantic relations, making it one of the most extensive WordNet resources in Asia. Due to the morphological complexity of Japanese, its inflectional variability, polysemy, and the large number of word forms, JWN relies on specialized lexicographic approaches for the creation of glosses and examples. The system incorporates relations such as hyponymy, hypernymy, antonymy, attributive relations, meronymy, synset-level hierarchies, and relations based on syntactic similarity. With its rich semantic network and language-specific relation types, Japanese WordNet provides a formal representation of the conceptual map of the Japanese language and constitutes a high-quality monolingual WordNet resource.

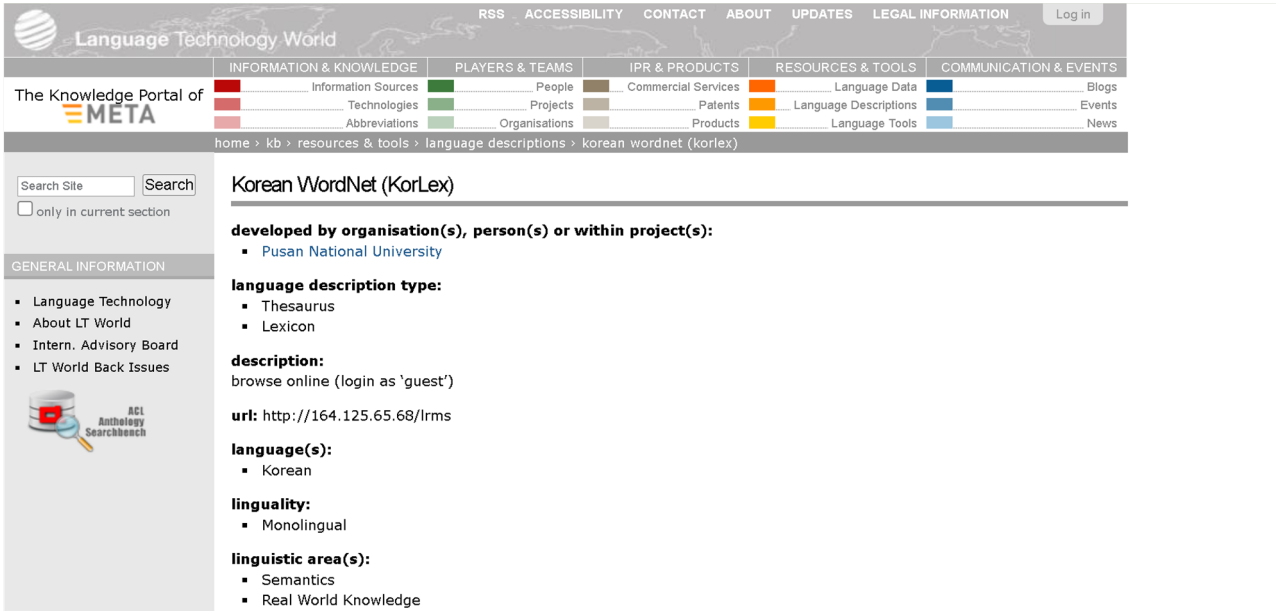

Figure 5. KorLex interface

Korean WordNet (KorLex) is a monolingual WordNet developed for the Korean language by Pusan National University. It represents the semantic system of the language in the form of a conceptual graph model based on thesauri. KorLex has been designed with careful consideration of the agglutinative morphology of Korean, its multi-form paradigms, context-dependent polysemy, and the high degree of verb valency. Each synset in the system corresponds to a real conceptual entity in Korean.

KorLex provides a wide range of semantic relations, including hyponymy, hypernymy, antonymy, meronymy, cause–effect relations, ontological “is-a” and “part-of” relations, as well as conceptual relations reflecting real-world knowledge. Technically, the system is implemented using XML-formatted synsets, lexical labels, glosses, examples, and semantic links. KorLex is regarded as one of the most authoritative resources for formalizing the national semantic landscape of the Korean language and serves as a core resource for semantic analysis, knowledge acquisition systems, information retrieval, and NLP applications.

Turkish WordNet (KeNet). Turkish WordNet (KeNet) is the largest and most comprehensive monolingual WordNet system developed for the Turkish language. It has been created over many years by Starlang Software and formally models Turkish lexical and semantic structures through synsets, glosses, and semantic relations. KeNet accounts for the agglutinative structure of Turkish, complex verb derivation processes, polysemy, derivational affixes, historical-linguistic changes, and multiple paradigms of word formation.

The synsets are enriched with relations such as synonymy, antonymy, hyponymy, hypernymy, meronymy, holonymy, troponymy, attributive relations, verb-valency relations, and various forms of semantic similarity. Each synset is systematically associated with glosses, multiple usage examples, lemma variants, morphological features, and structured semantic relations. The XML-based representation of synsets enables direct integration into machine-readable systems, while the explicit formalization of semantic and paradigmatic relations allows KeNet to be effectively used as a knowledge graph in advanced NLP applications.

WordNet System	Number of Synsets	Number of Lemmas	Number of Senses	Number of Relation Types
Princeton WordNet	117,000	155,327	206,978	21 types
RuWordNet	49,798	111,500	154,111	12 types
Japanese WordNet	57,238	93,834	158,058	20 types
Turkish WordNet (KeNet)	77,330	109,049	189,000	15 types
Korean WordNet (KorLex)	32,000	70,000	98,000	15 types

CONCLUSION AND RECOMMENDATIONS

The conducted analysis demonstrates that the WordNet model represents one of the most important and stable semantic structures in modern computational linguistics. Its synset-based conceptual framework and hierarchical system of semantic relations are widely applied across all major areas of natural language processing, including word sense disambiguation, semantic search, machine translation, and question–answering systems. National resources developed on the basis of the theoretical principles of Princeton WordNet—such as RuWordNet, Japanese WordNet, Korean WordNet, and Turkish WordNet—clearly illustrate the successful adaptation of this model to different language families.

The comparative analysis confirms that WordNet systems not only formalize the lexical richness of languages into structured representations, but also provide a robust conceptual foundation for the computational processing of semantic knowledge. Furthermore, the methodological unification principles promoted by the Global WordNet Association ensure interoperability among individual WordNets and enable the integration of multilingual semantic networks into a single global ecosystem. As a result, the WordNet model continues to be one of the most stable, widely used, and influential semantic resources in contemporary linguistic and computational research.

Recommendations. Future research should focus on expanding WordNet-based resources for underrepresented languages, enhancing the automatic construction and validation of synsets using large-scale corpora, and strengthening the integration of WordNet with ontologies and knowledge graphs. In addition, the application of WordNet in emerging NLP domains—such as large language models, semantic reasoning, and cross-lingual information retrieval—offers promising directions for further development.

REFERENCES

1. Miller G.A., Beckwith R., Fellbaum C., Gross D., and Miller K.J. Introduction to WordNet: An on-line lexical database. International journal of lexicography, 1990. –p. 219.
2. Miller G.A. WordNet: An on-line lexical database. International Journal of Lexicography, 1990;
3. Miller G.A., and Johnson-Laird P. Language and Perception. Harvard University Press, 1976;
4. Miller G.A. WordNet: A lexical database for English. Communications of the ACM, 1995;
5. Fellbaum C. WordNet. In: Theory and Applications of Ontology, 2010;
6. Fellbaum C. Impact of WordNet on natural language processing. In Proceedings of the ACL Workshop, 2000;
7. Resnik P. Using information content to evaluate semantic similarity. Proceedings of IJCAI, 1995;
8. Banerjee S., and Pedersen T. Extended gloss overlaps for sense disambiguation. Proceedings of CICLing, 2003;
9. Lesk M. Automatic sense disambiguation using dictionary overlaps. SIGDOC Conference Proceedings, 1986;
10. Pedersen T. WordNet-based similarity measures. University of Minnesota Technical Reports, 2004;
11. Pianta E., Bentivogli A., and Girardi C. MultiWordNet: Developing an aligned multilingual WordNet. Proceedings of GWC, 2002;
12. https://globalwordnet.github.io
13. https://ruwordnet.ru/ru
14. https://bond-lab.github.io/wnja/
15. http://www.lt-world.org/kb/resources-and-tools/language-descriptions/ltw_x3alanguage_x5fdescriptions_.2010-10-28.2540969419

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 1

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block, House 17.

TASHKENT STATE UNIVERSITY OF ECONOMICS

Google Scholar | OPEN ACCESS | ULRICHSWEB GLOBAL SERIALS DIRECTORY

Academic Resource Index ResearchBib | ISSN INTERNATIONAL STANDARD SERIAL NUMBER INTERNATIONAL CENTRE | CYBERLENINKA

OpenAIRE | ROAD | НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА LIBRARY.RU

BASE | INDEX COPERNICUS INTERNATIONAL | Crossref

doi Digital Object Identifier | CiteFactor Academic Scientific Journals | zenodo

Acceptance of articles

Published every monthly

Directions

Social, economic, political, technological, scientific

Scopus || Scientific electronic journal specializing in Scopus

CERTIFICATE NUMBER: №390637

ORDER NUMBER ACCORDING TO THE LICENSE REGISTER: C-5669633

CONTACT:

Contact us
+998 50 737 87 88

Telegram channel
t.me/scopus_IST2100

Journal official website
https://ist-journal.uz/index.php/IST