

Ontology-Driven Computer Systems: Elementary Senses in Domain Knowledge Processing

Mykola Petrenko^a, Ellen Cohn^b, Oleksandr Shchurov^a, Kyrylo Malakhov^a

^a Microprocessor technology lab, Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

^b Department of Communication and Rhetoric, University of Pittsburgh, Pittsburgh, PA, USA

ABSTRACT

This article delves into the evolving frontier of ontology-driven natural language information processing. Through an in-depth examination, we put forth a novel linguistic processor architecture, uniquely integrating linguistic and ontological paradigms during semantic analysis. Distancing from conventional methodologies, our approach showcases a profound merger of knowledge extraction and representation techniques. A central highlight of our research is the development of an ontology-driven information system, architected with an innate emphasis on self-enhancement and adaptability. The system's salient capability lies in its adept handling of elementary knowledge, combined with its dynamic aptitude to foster innovative concepts and relationships. A particular focus is accorded to the system's application in scientific information processing, signifying its potential in revolutionising knowledge-based applications within scientific domains. Through our endeavours, we aim to pave the way for more intuitive, precise, and expansive ontology-driven tools in the realm of knowledge extraction and representation.

Keywords Ontology engineering, Elementary sense, Knowledge representation, Commonsense knowledge, Deep artificial intelligence, Scientific model of the World

Categories • Artificial intelligence ~ Knowledge representation and reasoning, Ontology engineering

Email

Mykola Petrenko – petrng@ukr.net,

Ellen Cohn – ecohn@pitt.edu,

Oleksandr Shchurov – alexlug89@gmail.com,

Kyrylo Malakhov – malakhovks@nas.gov.ua (CORRESPONDING)

1 INTRODUCTION

The development of dynamic computerised knowledge systems and deep artificial intelligence systems are intertwined, stemming from similar foundational roots. A core aim for these systems is formulating a comprehensive scientific model of the world (SMW) and harnessing it effectively. The former systems strive to weave a global web of trans-disciplinary knowledge,

Petrenko, M., Cohn, E., Shchurov, O., and Malakhov, K. (2023). Ontology-Driven Computer Systems: Elementary Senses in Domain Knowledge Processing [Viewpoint]. *South African Computer Journal* 35(2), 127–144. <https://doi.org/10.18489/sacj.v35i2.17445>

Copyright © the author(s); published under a [Creative Commons NonCommercial 4.0 License](https://creativecommons.org/licenses/by-nc/4.0/) 
SACJ is a publication of *SAICSIT*. ISSN 1015-7999 (print) ISSN 2313-7835 (online)

targeting humanity's intricate challenges, while the latter seeks to emulate human-like common sense and a comprehensive world view.

The ambition to attain human-level artificial intelligence (AI) remains a foremost aspiration within AI research. Such an achievement could lead to revolutionary innovations with far-reaching consequences for mankind. Currently, however, AI's role is predominantly as a practical instrument with specific, albeit limited, applications.

As AI systems tailored for real-world challenges evolve, they will undeniably benefit from contemporary research breakthroughs. It is anticipated that, in the near future, AI's potential will predominantly be showcased through a surge in specialised applications. This trajectory is already evident across sectors, from industry to economics to societal facets. The expansive promise of AI is evident, as it permeates diverse fields, from scientific endeavours to facilitation of major tech innovations (Ford, 2021; Luger, 2008; OpenAI, 2023).

Reference to David Ferrucci's perspective on deep artificial intelligence is apt here. The CEO and founder of the AI startup, Elemental Cognition, postulates that mastering natural language understanding is pivotal for achieving universal intelligence. Contrasting with approaches like those by DeepMind that probe brain physiology, Ferrucci's stance is on building a system surpassing human capabilities in natural language processing and logical reasoning. He uniquely contends that the foundational elements for crafting such intelligence are already available (Ford, 2018).

To achieve this vision, Elemental Cognition is devising a hybrid system. This model synergises deep neural networks and varied machine learning strategies with software modules honed for logic and reasoning, all rooted in conventional programming paradigms. Interestingly, the current research climate favours an amalgamation, rather than a division, between symbolist and connectionist systems. Such an integrated research avenue is termed "neurosymbolic AI" (Ford, 2018, 2021).

Understanding causality is crucial for fostering creativity and devising alternative solutions. Unlike reinforcement learning algorithms in neural networks that necessitate repeated failures for successful strategising, humans inherently practice mental simulations. This cognitive act allows us to predict potential outcomes of varying decisions, rooted in our intrinsic grasp of causality. This facilitates our quest for answers to the quintessential "why?" Mastery over causality, especially the skill to frame and tackle causative questions, is integral to the evolution of universal machine intelligence (Ford, 2018).

A distinguishing trait of human intelligence is the ability to assimilate information from one source and adapt it across different realms, underpinning creativity and innovation. For universal machine intelligence to be practically impactful, it must go beyond mere textual understanding. Its true merit rests in its capability to leverage its knowledge reservoir to navigate uncharted challenges. An AI system's proficiency in applying knowledge in diverse and novel scenarios might very well be the definitive benchmark for gauging its intellectual depth (Ford, 2021).

Echoing D. Ferrucci's sentiment, the capacity to extrapolate information from one context and adeptly apply it in varied scenarios is quintessential for fostering innovation (Ford, 2018).

From our vantage point, emulating human reasoning in deep artificial intelligence requires formulating a domain-specific SMW or, minimally, a discipline-oriented SMW. Constructing such a model should be anchored in an ontological methodology, culminating in a Scientific-Ontological Model of the World (SOMW). The ensuing discourse will explore the imperative of crafting and leveraging the SOMW, underscoring the pivotal role of systematic ontological knowledge representation in replicating human reasoning via deep artificial intelligence.

Ontology-driven information processing and knowledge representation originated from the quest for a standard protocol to streamline knowledge across varied knowledge spectra. This paradigm aims to offer a unified blueprint and guiding principles for systematic knowledge depiction, categorisation, and interlinking, irrespective of the domain of expertise. The advent of ontological strategies has enabled the effective construction of knowledge-centric systems and, crucially, laid the foundation for trans-disciplinary engagement and ontological engineering within the realm of contemporary AI (Gómez-Pérez et al., 2004; Guarino, 1998; Palagin, Kaverinskiy et al., 2023; Sowa, 2000; Staab & Studer, 2009).

2 ONTOLOGICAL COMPONENTS AND KNOWLEDGE REPRESENTATION

Ontology, as a formalised structure for knowledge representation, is typically defined by the following four components:

Classes – These symbolise categories or concepts within a specific domain, offering a means to group entities sharing similar attributes.

Properties – These delineate attributes or associations of classes and individuals, serving to establish relationships with terms such as “has”, “is”, or “part of”.

Individuals – Representing tangible instances of classes, individuals can be thought of as distinct entities, concepts, or cases within a domain.

Axioms or Constraints – These constitute rules or logical assertions that dictate relationships and behaviour within an ontology, reinforcing its logical consistency.

These components underpin ontology modelling, offering a robust framework for structured knowledge representation specific to a domain. Such a structure not only aids human comprehension but also facilitates machine-based reasoning (Palagin, Petrenko et al., 2023).

Ontological methodologies grant users a holistic perspective on specific subjects or intricate projects. Utilising ontological models enables the delineation of classes, entities, functions, and formal theories. Ontological tools support the creation of analytic systems for research and organisational purposes, which span functions from multi-factorial analysis of primary data to fostering collaborative decision-making. Moreover, ontologies serve as both the manipulative medium and outcome for Semantic Web technologies.

An invaluable tool within the ontological suite is the linguistic-ontological model of the world (LOMW). Envisioned as a lexicographical system, the LOMW is an integral part of the

overarching scientific model of the world (SMW) and is pivotal for systems focused on natural language object comprehension (Palagin, 2006, 2016; Palagin, Petrenko et al., 2023).

Within this framework, the LOMW acts as a categorical scaffold, providing a semantically enriched base for domain-specific knowledge repositories. It also aids in merging diverse knowledge sources. By amalgamating linguistic and ontological components, the LOMW enhances comprehension, communication, and knowledge management, propelling both domain-specific and interdisciplinary research.

Whether it is a human's linguistic cognition or a computer system, the processing of speech or textual data hinges on a linguistic processor. Within a computer system, this processor is paramount, responsible for discerning and understanding incoming natural language data, deriving core knowledge, and presenting it in a logical format.

This processed data lays the groundwork for knowledge-based operations, aiding in problem resolution, decision-making, and various associated tasks. Essentially, the computer system linguistic processor serves as a conduit between human linguistic input and computational action, extracting and harnessing knowledge for diverse applications.

A linguistic processor, either hardware or software-based, deciphers textual data, such as a document, article, monograph, or linguistic corpus of texts, through consecutive stages of linguistic examination. This typically encompasses graphemata, morphological, syntactic, and surface-semantic evaluations, each contributing to the understanding of the structure and semantics of text (Kurgaev & Petrenko, 1995; Petrenko & Kurgaev, 2003; Petrenko & Sofiyuk, 2003).

Post-processing by the linguistic processor, the resultant information structure is primed for intensive semantic scrutiny within an extra-linguistic subsystem. Here, the primary goal is concept structuring. Essentially, it automates knowledge extraction from the natural language object, aiming to pragmatically interpret this knowledge, mirroring human understanding and response (Palagin, Petrenko et al., 2023).

The Analytical and Understanding System (AUS) architecture, with the LOMW at its core, is illustrated in Figure 1. The AUS's primary data reservoir is the corpus of text, linked to a specific knowledge domain or an array of scientific writings. The linguistic corpus, a finite text set, is represented by k , which denotes the collective number of texts in the corpus. These texts are processed sequentially, channelled initially through the graphemata analysis subsystem.

As the text progresses through the linguistic analysis algorithm, it metamorphoses across graphemata, morphological, syntactic, and semantic structures, each possessing distinct representation models and tools. An exhaustive account of this procedure is elaborated in Palagin, Petrenko et al. (2023). What differentiates this method from traditional semantic analysis is the integration of the linguistic-ontological picture of the world within the semantic review. This picture transcends dictionary-based semantic data, embedding multi-tiered patterns of both general and specific semantic structures observed in elementary sentences.

Upon completing the AUS processing, a text-sentence pattern table emerges. This repository encompasses the information structure for the entire text, serving as input for the extra-linguistic text processing subsystem. Within this subsystem, formal-logical translation of the

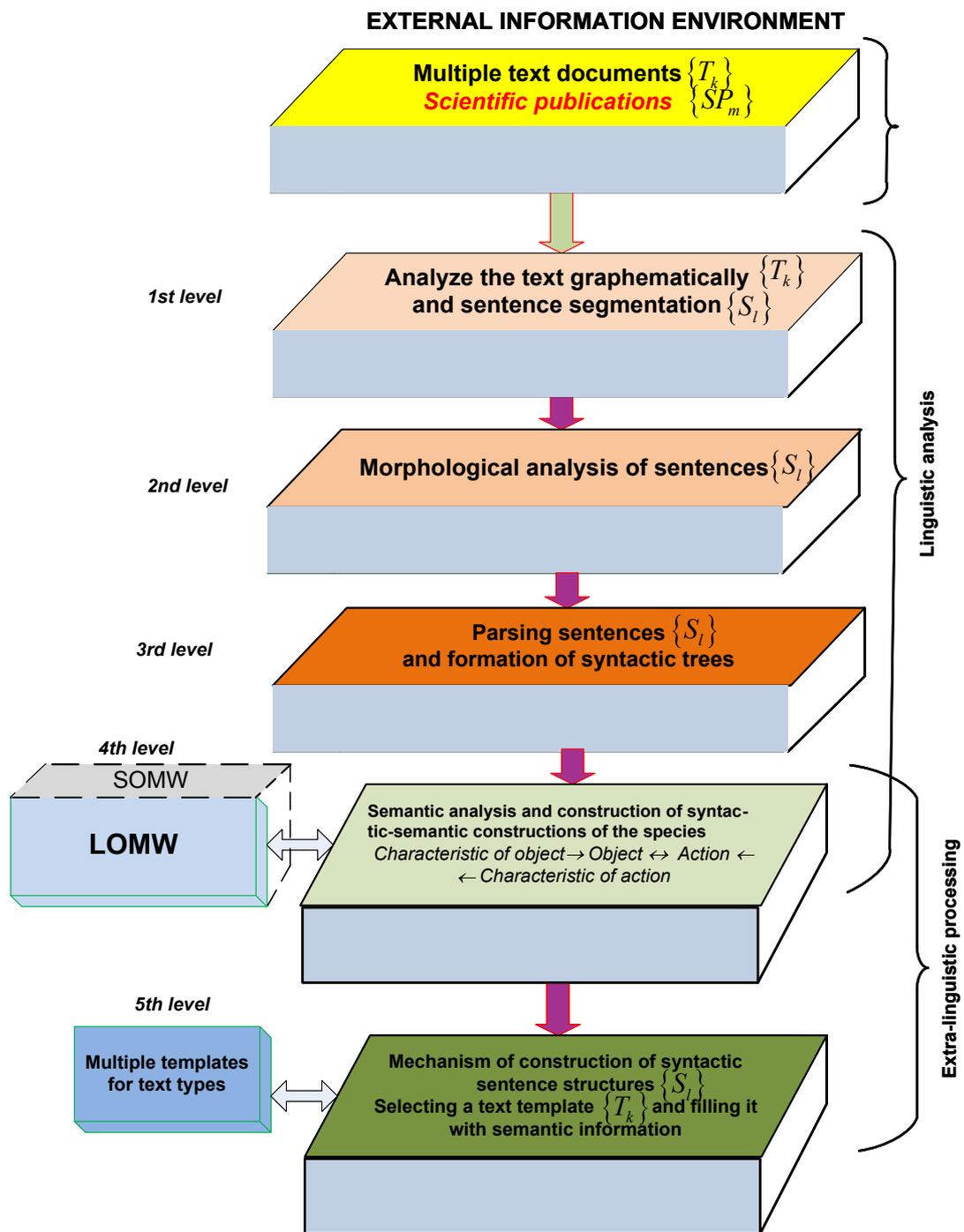


Figure 1: The Analytical and Understanding System architecture ^a

^aPalagin, Petrenko et al. (2023)

sentences and the collective text is executed, typically transforming the text into an appropriate first-order formal theory. A common technique includes an intermediary phase where data is recast into modified conceptual graphs, followed by a transition into first-order predicate logic (Palagin, Kaverinskiy et al., 2023).

The LOMW's formulation is exhaustively dissected in Palagin (2016), and Palagin (2006) furnishes intricate details about the SOMW's development. Both are indispensable references, shedding light on the methodologies integral to the creation of these vital knowledge representation and semantic analysis system components.

3 TEMPORAL AND SPATIAL DYNAMICS OF KNOWLEDGE DEVELOPMENT

Knowledge evolution within any domain, including the expansive realm of science, manifests as parameters that are temporally and spatially modulated. Historically, the pace of knowledge accumulation in nascent stages of scientific development, both broadly and within particular specialisations, paled in comparison to current exponential knowledge growth rates. While knowledge quantity and its informational representation are interlinked, they are not synonymous. Their association, at any temporal juncture, hinges largely on the equilibrium between verbal and formalised representations. The latter, which includes analytical, tabular, and graphical representations, offers brevity compared to its verbal counterpart and is more amenable to operational handling. Over time, three primary evolutionary trajectories are discerned (Palagin, 2006):

1. Augmentation in total knowledge and information
2. Expansion of the formalised knowledge segment
3. Super-session of antiquated knowledge with contemporary insights

A salient characteristic of generic knowledge evolution is the presence of dichotomous tendencies: scientific discipline differentiation followed by integration. Historically, the primacy lay with differentiation, leading to the genesis of specialised disciplines. Presently, differentiation coexists with integration, symbolised by the emergence of multidisciplinary research entities and the inception of trans-disciplinary convergence clusters. These clusters are instrumental, acting as crucibles for cross-disciplinary collaboration, thereby furthering cohesive solutions to multifaceted challenges and promoting human epistemological advancement (Palagin, Petrenko et al., 2023).

4 ONTOLOGICAL KNOWLEDGE-BASED SYSTEMS IN COMPUTER SCIENCE

The advancement of blueprints and methodologies catering to knowledge-based systems remains a focal endeavour within computer science. The genesis of ontological knowledge-based systems (OKBSs) is intrinsically woven with the maturation of theoretical foundations and design approaches.

Integral elements include:

Generalised system architecture and structure – This involves the articulation of rudimentary principles governing system architecture, ensuring a cohesive framework for knowledge representation and processing.

Formal knowledge representation models – This aspect emphasises structured knowledge representation models, promoting efficient data stewardship and retrieval.

Algorithms for knowledge processing – This pertains to the creation of algorithms facilitating structured knowledge handling, supporting a gamut of knowledge-centric operations.

Collectively, these design endeavours amplify the significance of ontological knowledge within knowledge-based systems, specifically when addressing formidable objectives such as crafting a SOMW. As underscored earlier, the SOMW is integral to deep AI and the burgeoning “neurosymbolic” AI domain. Both strive for an amalgamation of symbolic and connectionist AI paradigms, aiming for a more intricate AI embodiment.

An evolving intelligent computer system’s architectural development can be viewed from a bifocal lens, capturing both external (user-centric) and intrinsic dimensions. The harmonious orchestration between these dimensions is quintessential for maximising the efficacy of OKBSs (Palagin, 2006).

Architecting efficacious ontologically governed computer systems necessitates the assimilation of modern computer science realms, inclusive of artificial intelligence, knowledge processing, and the pragmatic model of linguistic consciousness. Conceptualised as a productive sequence, their operational synergy mirrors: “Input signal → Knowledge System → Reaction.”

An emerging OKBS functions with pre-ordained goals, spanning both long-term visions and immediate targets. This operational alignment is modulated by feedback-driven interactions with the external data milieu. Central to an OKBS’s operational tenet is its knowledge system, envisioned as an intricate subsystem synergising with a constellation of domain-specific knowledge subsystems (Palagin, Petrenko et al., 2023; Palagin et al., 2014; Palagin et al., 2018). This multifaceted interplay equips the OKBS to adeptly ingest, process, and apply knowledge spanning varied domains, resonating with its operational objectives.

Intrinsically, the OKBS capitalises on its knowledge system and specialised knowledge subsystems, dynamically fine-tuning its reactions to match overarching goals and the capricious demands of the external information ecosystem. This adaptability ensures the sustained efficacy of the system in fulfilling its delineated objectives.

The OKBS architecture, visualised in Figure 2, spotlights a pivotal component: the self-evolving mechanism of the knowledge base (KB) pertinent to a designated domain. This auto-evolutionary feature ensures the KB’s continuous adaptation and growth, resonating with the domain’s shifting contours and insights.

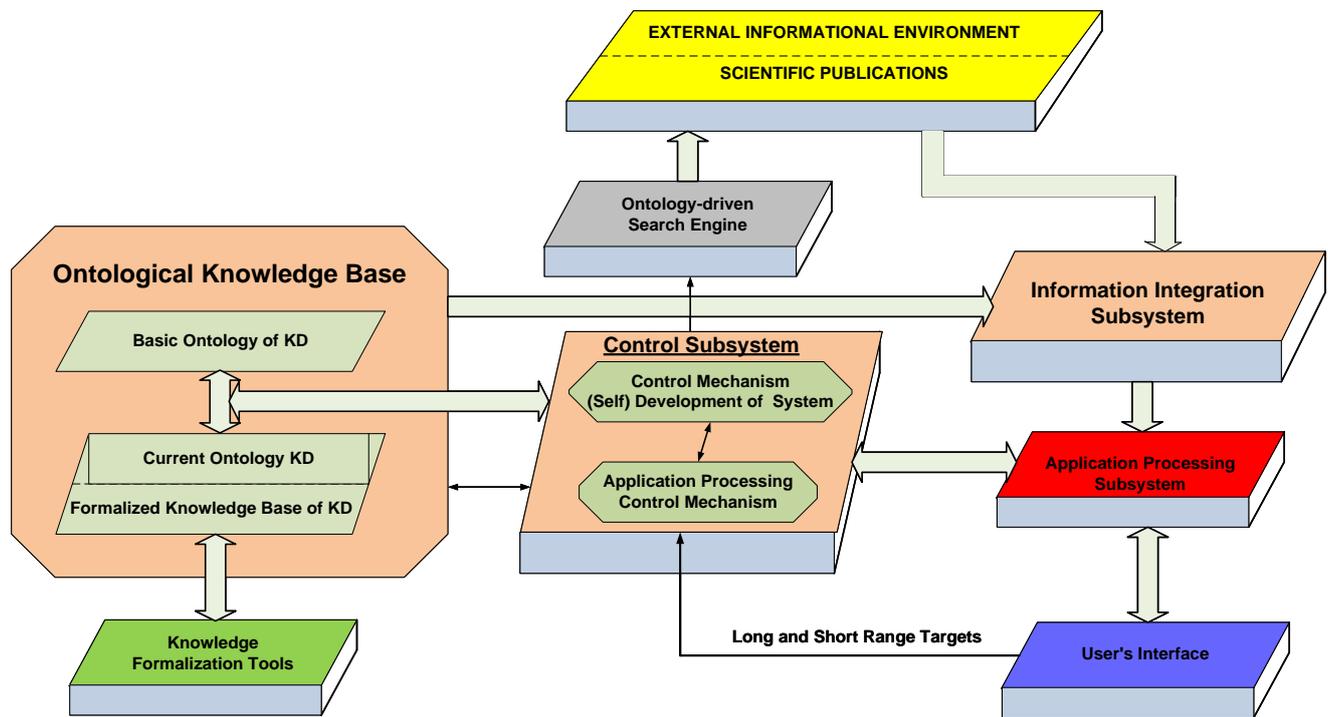


Figure 2: The Ontological Knowledge-Based System architecture

Central to this self-propagating mechanism are ontological controls governing two foundational processes: external information space exploration and formalised knowledge base formation, which materialises via two predominant channels (Palagin, 2006, 2016):

1. Data extraction from the External Information Environment (predominantly the Internet).
2. Inference-driven knowledge genesis.

Both facets of formalised KB evolution within ontological knowledge-based systems exhibit profound interplay. Fresh knowledge genesis via inferential channels is contingent upon the influx of contemporary data, predominantly sourced from the digital universe of the Internet. This perpetual data stream equips the system with foundational insights, crucial for logical derivations, inferential processes, and the proactive birth of new knowledge realms.

Furthermore, the ontological addition of novel knowledge, potentially introducing new concepts, is integral for knowledge base evolution. This entails identifying and assimilating emerging concepts and dynamics from external information spheres into the system’s ontology. This dynamic integration approach accentuates the imperativeness of a cyclic knowledge procurement methodology, ensuring the OKBS’s perpetual adaptation and absorption of evolving ontological insights (Kryvyi, 2016).

5 FORMALISING SCIENTIFIC PUBLICATIONS FOR KNOWLEDGE DEVELOPMENT

Scientific publications serve as prominent vehicles for the synthesis of new knowledge, especially given their unique scientific narrative style and structured presentation. Scientific publications inherently possess a well-defined syntactic and semantic structure. Their templated format allows for automated formalised description of content, a pivotal characteristic elucidated in (Malakhov et al., 2023).

In the developmental mode, the OKBS processes user queries directed towards the natural language processing (NLP) database, as detailed in Malakhov et al. (2023).

The ontology system for processing and enhancing scientific publications is visually summarised in Figure 3, offering a schematic of primary components and functions.

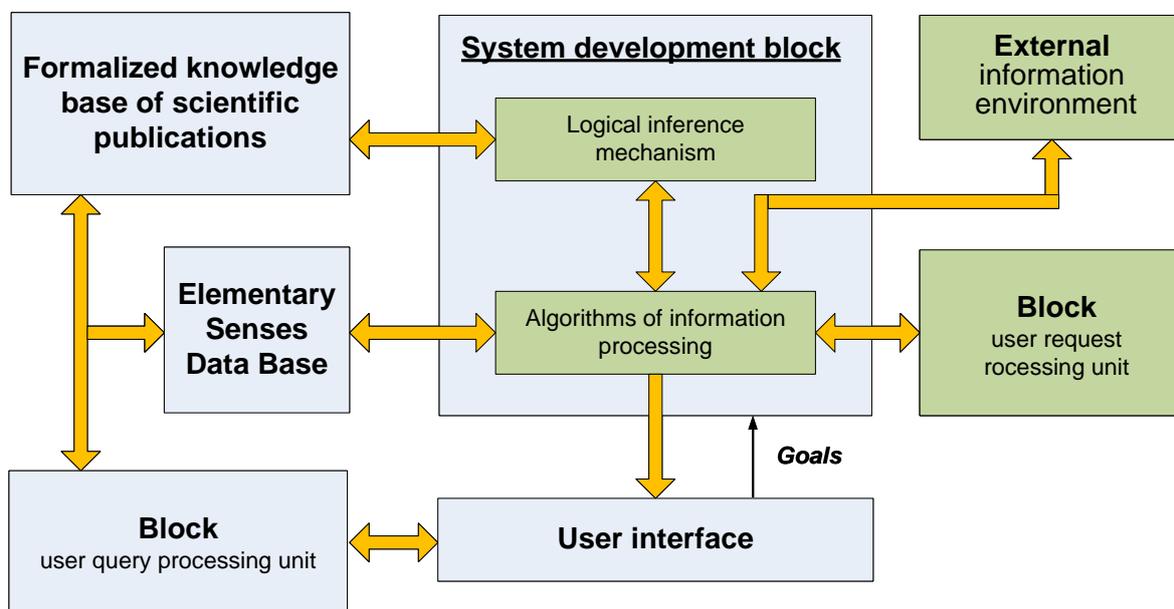


Figure 3: The architecture of the ontology system for processing and enhancing scientific publications

Ontology system development is characterised by three distinct operational modalities:

1. **Logical inference mechanism:** This mode employs reasoning tools like Pellet's reasoner, enhancing knowledge through inferential techniques. It crafts novel ontological connections and unveils nuanced relationships within the pre-existing knowledge base.
2. **Elementary sense (ES) processing algorithms:** This mode harnesses specialised algorithms tailored for elementary sense analysis. The "electronic collider" operation (Malakhov et al., 2023), exemplifies this approach. By dissecting foundational linguistic and semantic units, knowledge is extracted directly from elementary sense analysis.

3. Formula-driven NLP: Employing recognised formulas, such as the Brooks formula, underscores behaviour-centric strategies for natural language comprehension.

Selecting an approach may hinge on the system's strategic objectives, domain intricacies, and specific knowledge development aims. With the amalgamation of inferential, algorithmic, and formulaic methods, the ontology system exhibits versatility in knowledge development tasks, progressively honing its acumen.

We now investigate the correlation between *elementary sense*, *commonsense knowledge*, and *commonsense reasoning* in ontology-driven computer systems and domain knowledge processing.

Definitions

The *elementary sense* notion, introduced in Malakhov et al. (2023): a simple two-syllable statement that contains a subject, a predicate, and a direct object. These components correspond to the subject, predicate, and object of the Semantic Web RDF-triple.

The *commonsense knowledge* notion, introduced in Davis (1990): the universal understanding about the world possessed by most individuals, encompassing obvious inferences and covering a vast range of domains from natural language to high-level vision. It forms the fundamental core of human knowledge and intelligence.

The *commonsense reasoning* notion, introduced in Davis; Mueller (1990, 2015): the act of performing inference on a set of object-level information using a knowledge base and a knowledge base manager. It involves deriving new insights from existing knowledge about general worldly scenarios.

Similarities

Basis for knowledge representation – Both “Elementary Sense” and “Commonsense Knowledge” provide foundational structures for representing information. Elementary Sense offers a concise, structured format analogous to Semantic Web RDF-triples, while Commonsense Knowledge provides a comprehensive knowledge about the world.

Inference –Both Commonsense Knowledge and Commonsense Reasoning involve the process of making inferences. While the former provides the foundational knowledge, the latter is about the actual process of deriving new insights from that knowledge. Furthermore, with Elementary Sense, inference can also be performed using SPARQL queries on RDF-triples, thereby extracting precise insights from the structured data representation.

Dynamic Understanding – All three concepts emphasise the dynamic nature of knowledge. Elementary Senses offer structured representations, Commonsense Knowledge encompasses ever-growing human understanding, and Commonsense Reasoning involves the continual process of deriving new insights.

Differences

Granularity – Elementary Sense focuses on a precise representation of information as simple two-syllable statements, aligned with RDF-triples, while Commonsense Knowledge is broader, spanning a wide array of general knowledge domains.

Purpose – The primary goal of Elementary Sense is to offer structured representations for easy processing. In contrast, Commonsense Knowledge serves as a foundational base for understanding the world, and Commonsense Reasoning seeks to infer new insights based on that foundational knowledge.

Process vs. Data – Elementary Sense is about data representation and storage, aligning closely with Semantic Web structures. Commonsense Knowledge, on the other hand, is about the data itself, and Commonsense Reasoning is process-oriented, centred on the act of inference.

Extracting New Knowledge from Existing Knowledge

Role of Elementary Sense – In ontology-driven computer systems, Elementary Sense plays a crucial role in simplifying and structuring data. By organising information in a format analogous to RDF-triples, it aids in knowledge extraction, paving the way for deeper semantic and ontological analysis.

Incorporating Commonsense – Commonsense Knowledge acts as a background reservoir during knowledge processing. When systems encounter ambiguous or incomplete data, this knowledge can be leveraged to fill in gaps or infer missing components. The vast scope of Commonsense Knowledge ensures a well-rounded understanding, even in the absence of explicit information.

Reasoning and Evolution – Through Commonsense Reasoning, systems can derive new information or connections from existing knowledge. Applying inference on a knowledge base, especially when combined with the structured insights from Elementary Senses and SPARQL queries on RDF-triples, magnifies the potential for discovering new patterns, relationships, or insights. This dynamic reasoning is pivotal for the evolution and self-improvement of ontology-driven computer systems.

5.1 Connections

Elementary Sense, Commonsense Knowledge, and Commonsense Reasoning are integrally connected in the realm of ontology-driven computer systems and domain knowledge processing. Together, they contribute to the accurate representation, comprehensive understanding, and dynamic expansion of knowledge within the system.

ESs are pivotal for scientific publications analysis. To elaborate:

Scientific publications repository – This encompasses a myriad of structured scientific articles, replete with titles, bibliographic details, texts (structured as sections, paragraphs, sentences, inclusive of figures, formulas, and tables), abstracts, preludes, conclusions, and citations.

Complex sentence decomposition – Intricate sentence structures are distilled into simpler, ideally binary, constructs to streamline analysis.

Simple sentence depiction – Binary sentences are decomposed into their foundational triad: subject, predicate, and complement, analogous to the subject-predicate-object (SPO) or RDF-triples, prevalent in Semantic Web frameworks.

Segmenting intricate constructs into *Elementary Senses* encapsulated as RDF-triples ensures concise knowledge representation. Consequently, this fosters efficient and pinpoint analysis and extraction of knowledge nuances from scientific publications.

On average, a sentence in a scientific publication text, particularly in Ukrainian, may encapsulate 3 to 4 ESs. The manipulation of these elementary meanings, combined with the associated ontological knowledge domain structures, paves the way for the potential genesis of new ontological concepts or inter-relationships.

Example to illustrate the ES formation process

In the example below, the following abbreviations are employed:

S – ES subject

P – ES predicate

O – ES object

Cnt ES_m – context of the *m*th elementary sense

Y_{n-j} – where *n* signifies the level number in the context ontograph, and *j* indicates the ordinal number of the concept node at that specific level.

Original Statement – “Theory and practice of creation and use of knowledge-based systems is the most actual and intensively developing direction of Computer Science, allowing increasing the efficiency of creation and use of computer technologies, application systems and toolkits.”

Upon decomposition into ES, the sentence bifurcates into its core components: subject, predicate, and object, alongside the contextual codes of the knowledge domain (KD):

- **ES1** Computer Science is the most actual and intensively developing direction;
S – Computer Science, **P** – have, **O** – actual and intensively developing direction;
Cnt ES_m – Y7-5, Y9-5.

- **ES2** Direction of Computer Science is theory and practice of creation and use of systems;
S – Direction of Computer Science, **P** – to be, **O** – theory and practice of creation and use of systems;
Cnt *ESm* – **Y7-5, Y9-5**.
- **ES3** Systems is knowledge-based systems;
S – Systems, **P** – to be, **O** – systems of knowledge-based systems;
Cnt *ESm* – **Y7-5, Y9-3, Y9-5**.
- **ES4** Knowledge-based systems allow for an increase the efficiency of creation and use of computer technologies;
S – Knowledge-based systems, **P** – to allow for an increase, **O** – the efficiency of creation and use of computer technologies;
Cnt *ESm* – **Y7-5, Y9-5**.
- **ES5** Knowledge-based systems allow for an increase the efficiency of creation and use of application systems;
S – Knowledge-based systems, **P** – to allow for an increase, **O** – the efficiency of creation and use of application systems;
Cnt *ESm* – **Y7-5, Y10-6**.
- **ES6** Knowledge-based systems allow for an increase the efficiency of creation and use of toolkits;
S – Knowledge-based systems, **P** – to allow for an increase, **O** – the efficiency of creation and use of toolkits;
Cnt *ESm* – **Y7-5, Y10-4**.

A segment of the context ontograph is visually represented in [Figure 4](#). It should be noted that during the analysis of ES structural components, the ontograph of contexts pertaining to the knowledge domain is extensively utilised. This ensures a precise association with the respective knowledge domain. Within this framework, an ES combined with its specific context is termed *Elementary knowledge*.

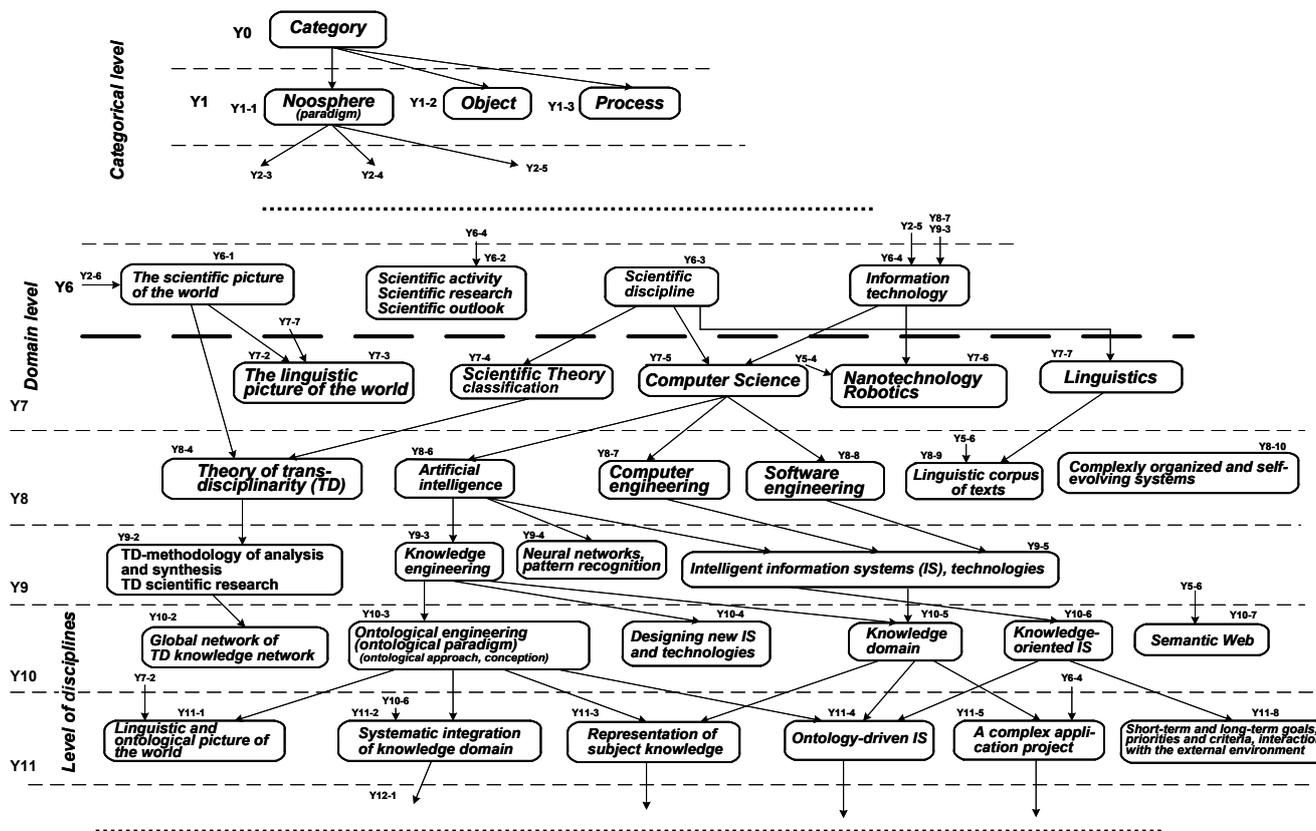


Figure 4: The fragment of the contexts ontograph

6 CONCLUSION

In this research, we have delved deep into the intersection of ontological perspectives and natural language processing to revolutionise the way knowledge is extracted and represented. By presenting a groundbreaking architecture for a linguistic processor, we have shifted away from traditional implementations, bringing together linguistic and ontological facets during the semantic analysis phase. Moreover, our venture into constructing a forward-thinking, ontology-driven information system is marked by its inherent emphasis on continuous self-enhancement. A standout feature of our approach is the enhancement of the ontological system, tailored explicitly for scientific data processing. This system’s prowess not only lies in its adept handling of elementary knowledge but also its dynamic capability to birth new concepts and weave intricate relationships. Such advancements hold immense promise in bolstering the effectiveness and relevance of our system in a myriad of scientific domains, marking a significant stride in the landscape of knowledge representation and analysis.

7 CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Mykola Petrenko: Supervision, Conceptualisation, Methodology, Writing – original draft, Validation. *Ellen Cohn*: Writing – review & editing. *Oleksandr Shchurov*: Writing – original draft. *Kyrylo Malakhov*: Validation, Resources, Term, Writing – review & editing.

8 ACKNOWLEDGEMENTS

The research team at the Microprocessor Technology Lab extends its gratitude to Oleksandr Palagin, a distinguished scholar and leader in the field. Oleksandr Palagin, who holds the esteemed titles of Academician of the National Academy of Sciences of Ukraine, Doctor of Technical Sciences, Professor, Honored Inventor of Ukraine, Deputy Director for Research at the Glushkov Institute of Cybernetics of the National Academy of Sciences of Ukraine, and Head of the Microprocessor Technology Lab, served as the guiding force and scientific supervisor for this research endeavor. We deeply appreciate his invaluable mentorship and expertise throughout this project.

The corresponding author, Kyrylo Malakhov, representing both himself and his co-author, Mykola Petrenko, along with Oleksandr Shchurov, would like to express their heartfelt gratitude to Dr. Ellen Cohn (PhD, CCC-SLP, ASHA-F) from the Department of Communication at the University of Pittsburgh, PA, USA. Dr. Cohn provided invaluable assistance in reviewing and editing the article, and her international support has played a pivotal role in preserving Ukrainian science during wartime and enhancing international scientific collaboration.

The Glushkov Institute of Cybernetics research team wishes to extend a special acknowledgment to Katherine Malan, the Editor-in-Chief of the South African Computer Journal. We deeply value her commitment to advancing Ukrainian science by facilitating the publication of research in scholarly journals, which has significantly contributed to our field's progress.

9 FUNDING

This study would not have been possible without the financial support of the National Research Foundation of Ukraine (Open Funder Registry: 10.13039/100018227). Our work was funded by Grant contract (application ID: 2021.01/0136):

Development of the cloud-based platform for patient-centered telerehabilitation of oncology patients with mathematical-related modeling (Malakhov, 2022, 2023a, 2023b; Palagin, Malakhov, Velychko, Semykopna & Shchurov, 2022; Palagin, Malakhov, Velychko & Semykopna, 2022; Stetsyuk et al., 2023).

10 DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Davis, E. (1990, January). Chapter 1 – Automating common sense. In *Representations of commonsense knowledge* (pp. 1–26). Morgan Kaufmann. <https://doi.org/10.1016/B978-1-4832-0770-4.50009-5>
- Ford, M. (2018, November). *Architects of intelligence: The truth about AI from the people building it*. Packt Publishing.
- Ford, M. (2021, September). *Rule of the robots: How artificial intelligence will transform everything* (First Edition). Basic Books.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological engineering* (1st ed.). Springer-Verlag. <https://doi.org/10.1007/b97353>
- Guarino, N. (1998, June). *Formal ontology in information systems: Proceedings of the 1st international conference June 6-8, 1998, trento, italy*. IOS Press.
- Kryvyi, S. (2016). Formal ontological models in scientific researchers. *Upravljâúšie sistemy i mašiny*, 263(3), 04–15. <https://doi.org/10.15407/usim.2016.03.004>
- Kurgaev, A. F., & Petrenko, M. G. (1995). Processor structure design. *Cybernetics and Systems Analysis*, 31(4), 618–625. <https://doi.org/10.1007/BF02366417>
- Luger, G. (2008, February). *Artificial intelligence: Structures and strategies for complex problem solving* (6th edition). Pearson.
- Malakhov, K. (2022). Letter to the editor – Update from Ukraine: Rehabilitation and research. *International Journal of Telerehabilitation*, 14(2), 1–2. <https://doi.org/10.5195/ijt.2022.6535>
- Malakhov, K. (2023a). Insight into the digital health system of Ukraine (ehealth): Trends, definitions, standards, and legislative revisions. *International Journal of Telerehabilitation*, 15(2). <https://doi.org/10.5195/ijt.2023.6599>
- Malakhov, K. (2023b). Letter to the editor – Update from Ukraine: Development of the cloud-based platform for patient-centered telerehabilitation of oncology patients with mathematical-related modeling. *International Journal of Telerehabilitation*, 15(1). <https://doi.org/10.5195/ijt.2023.6562>
- Malakhov, K., Petrenko, M., & Cohn, E. (2023). Developing an ontology-based system for semantic processing of scientific digital libraries. *South African Computer Journal*, 35(1), 19–36. <https://doi.org/10.18489/sacj.v35i1.1219>
- Mueller, E. T. (2015, January). Chapter 1 - Introduction. In *Commonsense reasoning* (2nd ed., pp. 1–16). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-801416-5.00001-2>

- OpenAI. (2023, March). *GPT-4 technical report* (tech. rep.) (arXiv:2303.08774 [cs]). OpenAI. arXiv. <https://doi.org/10.48550/arXiv.2303.08774>
- Palagin, O., Malakhov, K., Velychko, V., Semykopna, T., & Shchurov, O. (2022). Hospital information smart-system for hybrid e-rehabilitation. *CEUR Workshop Proceedings*, 3501, 140–157. <https://ceur-ws.org/Vol-3501/s50.pdf>
- Palagin, O. (2006). Architecture of ontology-controlled computer systems. *Cybernetics and Systems Analysis*, 42(2), 254–264. <https://doi.org/10.1007/s10559-006-0061-z>
- Palagin, O. (2016). An ontological conception of informatization of scientific investigations. *Cybernetics and Systems Analysis*, 52(1), 1–7. <https://doi.org/10.1007/s10559-016-9793-6>
- Palagin, O., Kaverinskiy, V., Litvin, A., & Malakhov, K. (2023). OntoChatGPT information system: Ontology-driven structured prompts for ChatGPT meta-learning. *International Journal of Computing*, 22(2), 170–183. <https://doi.org/10.47839/ijc.22.2.3086>
- Palagin, O., Malakhov, K., Velychko, V., & Semykopna, T. (2022). Hybrid e-rehabilitation services: SMART-system for remote support of rehabilitation activities and services. *International Journal of Telerehabilitation, Special Issue*(Research Status Report – Ukraine). <https://doi.org/10.5195/ijt.2022.6480>
- Palagin, O., Petrenko, M., Kryvyi, S., Boyko, M., & Malakhov, K. (2023, July). *Ontology-driven processing of transdisciplinary domain knowledge*. Iowa State University Digital Press. <https://doi.org/10.31274/isudp.2023.140>
- Palagin, O., Petrenko, M., Velychko, V., & Malakhov, K. (2014). Development of formal models, algorithms, procedures, engineering and functioning of the software system “Instrumental complex for ontological engineering purpose”. *CEUR Workshop Proceedings*, 1843, 221–232. <http://ceur-ws.org/Vol-1843/221-232.pdf>
- Palagin, O., Velychko, V., Malakhov, K., & Shchurov, O. (2018). Research and development workstation environment: The new class of current research information systems. *CEUR Workshop Proceedings*, 2139, 255–269. <http://ceur-ws.org/Vol-2139/255-269.pdf>
- Petrenko, M., & Kurgaev, A. (2003). Distinguishing features of design of a modern circuitry type processor. *Upravlyayushchie Sistemy i Mashiny*, 187(5), 16–19. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0347622333&partnerID=40&md5=7283307afdf891445ec9062c7b2ff80a>
- Petrenko, M., & Sofiyuk, A. (2003). On one approach to the transfer of an information structures interpreter to PLD-implementation. *Upravlyayushchie Sistemy i Mashiny*, 188(6), 48–57. <https://www.scopus.com/inward/record.uri?eid=2-s2.0-0442276898&partnerID=40&md5=44974b40409363e5fe4378e240149c52>
- Sowa, J. F. (2000, January). *Knowledge representation: Logical, philosophical, and computational foundations* (1st ed.). Brooks / Cole.
- Staab, S., & Studer, R. (Eds.). (2009). *Handbook on ontologies* (2nd ed.). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-92673-3>

Stetsyuk, P. I., Fischer, A., & Khomiak, O. M. (2023). Unified representation of the classical ellipsoid method. *Cybernetics and Systems Analysis*, 59(5), 784–793. <https://doi.org/10.1007/s10559-023-00614-x>