

Automatic Construction of a Legal Citation Graph from 100 Million Ukrainian Court Decisions: Large-Scale Extraction, Topological Analysis, and Ontology-Driven Clustering

Volodymyr Ovcharov
LEX AI LLC, Kyiv, Ukraine
vladimir@legal.org.ua

Abstract

Half a billion citation edges extracted from 100.7 million Ukrainian court decisions reveal that judicial citation structure encodes legal domain boundaries without supervision and predicts future legislative importance with near-perfect accuracy. We construct the first large-scale citation graph from the complete EDRSR registry (99.5 million full texts, 1.1 TB), extracting **502** million citation links across six types via regex on commodity hardware in approximately 5 hours, with precision of 1.00 on a 200-decision validation sample.

Three principal findings emerge. **(1)** The degree distribution follows a power law ($\alpha = 1.57$), placing the Ukrainian court network between the Indian Supreme Court and the EU Court of Justice, with hub articles cited by millions of decisions. **(2)** Louvain community detection on the co-citation projection recovers legal domain boundaries (civil, criminal, administrative, commercial) with modularity $Q = 0.44$ – 0.55 and temporal stability (NMI = 0.83 – 0.86 across periods), constituting an automatically constructed legal ontology grounded in judicial practice. **(3)** Citation features predict top-1000 articles with AUC = 0.9984 ; temporal dynamics detect legislative regime changes as phase transitions and the 2022 invasion as a citation entropy spike ($H : 11.02 \rightarrow 13.49$) with emergent wartime legislation nodes.

The citation-derived ontology is operationalized as the domain layer of a workflow memory system for LLM-assisted legal analysis [11], connecting to the ontology-controlled paradigm [13, 16]. The extraction pipeline, analysis code, and aggregated statistics are released as open data.

Keywords: legal citation graph, court decisions, Ukrainian law, ontology construction, knowledge extraction, EDRSR, network analysis, legal NLP

1 Introduction

The Unified State Register of Court Decisions (EDRSR, [www.edrsr.gov.ua](#)) is the largest open judicial corpus in continental Europe. Established in 2006 by Ukrainian law, it mandates publication of all court decisions within five days of rendering. As of May 2026, the registry contains 101.4 million decision records, of which 100.7 million include full text, spanning all judicial instances and all branches of justice—civil, criminal, commercial, administrative, and constitutional.

This corpus has been largely unexploited for computational legal analysis. Prior work on legal citation networks has focused on common-law jurisdictions—the U.S. Supreme Court [4], Dutch case law [19], Indian courts [7]—where explicit citation conventions (case names, reporter volumes)

make extraction straightforward. Continental legal systems, including Ukraine’s, present different challenges: citations are to legislation articles rather than prior cases, citation formats are inconsistent (abbreviations, Ukrainian morphology, varying codex names), and the sheer volume of decisions (8+ million per year since 2017) requires industrial-scale processing.

No prior work has attempted citation extraction at the 100-million-decision scale for any jurisdiction.

This paper makes three contributions:

1. **Large-scale citation extraction.** A regex-based pipeline that identifies six citation types in Ukrainian legal text, processing 100.7 million decisions (1.1 TB of full text) in approximately [HOURS] hours on a single 4-vCPU production server. The pipeline yields [N_TOTAL] million citation edges with an estimated precision of [PRECISION]% on a manually annotated sample.
2. **Topological analysis of the citation graph.** We analyze the resulting bipartite graph (decisions \leftrightarrow legislation) and its projections. The legislation-side projection reveals community structure that corresponds to established legal domains without supervision. Temporal analysis shows citation density shifts that align with major legislative reforms (2004 Civil Code adoption, 2012 Criminal Procedure Code, 2017 judicial reform).
3. **Citation-derived legal ontology.** Co-citation clustering produces an automatically constructed legal ontology: groups of legislation articles that are semantically related because courts cite them together. This ontology is deployed as the domain layer of the workflow memory system described in the companion paper [11], operationalizing the ontology-controlled paradigm of Palagin [13] with data-derived rather than manually curated structure.

The work continues two lines of research. First, the knowledge extraction program of Palagin et al. [14], which proposed methods for extracting structured knowledge from natural-language texts—here applied to 100 million legal texts at a scale not previously attempted in the Ukrainian NLP community. Second, the distributional semantic modeling approach of Palagin et al. [15], which used co-occurrence patterns to train term vector spaces—here instantiated as co-citation patterns that define legislation similarity without requiring embedding models or labeled data.

The connection to the ontology-controlled systems paradigm [13, 16] is structural: the citation graph provides the data layer that an ontology-controlled LLM system needs to ground its legal reasoning in statute structure. The companion paper on oversight-controlled systems [12] formalizes the conditions under which human corrections on LLM output constitute valid training signal; the citation graph provides the domain knowledge that makes those corrections informed rather than arbitrary.

2 Related Work

2.1 Legal Citation Network Analysis

Fowler and Jeon [4] pioneered legal citation network analysis by constructing a citation graph of U.S. Supreme Court decisions (1791–2005, $\sim 30,000$ decisions) and demonstrating that network centrality measures (PageRank, hub/authority scores) predict legal importance better than simple citation counts. Subsequent work extended this approach to the Dutch legal system [5, 19] and Indian courts [7].

All prior work operates at scales of 10^3 – 10^5 decisions. The EDRSR corpus is three orders of magnitude larger (10^8), requiring different engineering approaches: partition-parallel processing, server-side cursors, and streaming aggregation. More fundamentally, the Ukrainian legal system is a continental (civil law) system where the primary citation relationship is decision \rightarrow legislation,

not decision→decision as in common-law systems. This produces a bipartite graph rather than a unipartite one, with different topological properties.

2.2 Knowledge Extraction from Legal Texts

Palagin et al. [14] proposed a framework for extracting structured knowledge from Ukrainian-language texts, combining morphological analysis with domain-specific ontologies. The framework was demonstrated on scientific and technical corpora but not applied to legal texts at scale. Palagin et al. [15] extended this line with distributional semantic modeling, training term vector spaces from co-occurrence patterns in domain-specific corpora.

Our approach is a direct application of this program to the legal domain: co-citation patterns in 100 million court decisions define a distributional semantics over legislation articles, where two articles are “similar” if courts cite them in the same decisions. This requires no labeled data, no embedding models, and no morphological analysis—the citation structure itself encodes the semantic relationships.

2.3 Legal NLP and Information Extraction

Modern legal NLP has focused on transformer-based models: LEGAL-BERT [3], LexNLP [2], and domain-specific fine-tuning for citation prediction [20]. These approaches require labeled training data, are language-specific, and operate on individual documents rather than corpus-wide structure.

Our regex-based approach is deliberately simple: it trades recall for precision and interpretability, and scales linearly with corpus size. For the specific task of legislation citation extraction in Ukrainian legal text, the structured format of citations (“ . 625 ”, “ 3 « ... »”) makes regex extraction competitive with learned models, while being orders of magnitude faster.

2.4 Ontology Construction from Text

The ontology-controlled systems paradigm [13] requires a domain ontology to structure system behavior. Traditional ontology construction is manual and expensive. Palagin et al. [16] showed that ontology-controlled prompting improves LLM output quality for domain-specific tasks, but assumed a pre-existing ontology.

Citation graph clustering provides an alternative: the ontology is *derived* from usage data rather than constructed by experts. This is analogous to the distributional hypothesis in semantics—“you shall know a word by the company it keeps” [15]—applied at the statute level: *you shall know a law by the decisions that cite it*.

3 Data

3.1 The EDRSR Corpus

The Unified State Register of Court Decisions [17] was established by Law of Ukraine No. 3262-IV (22.12.2005) and has been operational since June 1, 2006. All courts of Ukraine are required to submit decisions for publication.

The data is stored in a PostgreSQL 15 database, partitioned by adjudication year (`edrsr_fulltext_p_YYYY`). Individual partitions range from 443 MB (2009) to 116 GB (2024). Full-text search is supported via `tsvector` columns; the `justice_kind` column encodes the branch of justice (1=civil, 2=criminal, 3=commercial, 4=administrative, 5=constitutional).

Metric	Description	Value
Total decisions	Records in <code>edrsr_documents</code>	101,422,684
Full texts available	Records in <code>edrsr_fulltext</code>	100,753,415
Coverage	Full texts / total decisions	99.3%
Time span	Earliest to latest decision year	2000–2026
Storage	Total full-text data (partitioned)	1.1 TB
Mean text length	Characters per decision (sampled)	~5,000
Median text length	Characters per decision (sampled)	~3,000
Peak year	2025 (partial year at extraction time)	8,764,090

Table 1: EDRSR corpus statistics as of May 13, 2026.

3.2 Legislation Corpus

The legislation side of the citation graph draws on two sources: the Verkhovna Rada legislation database [18] (accessed via API at `zakon.rada.gov.ua`), and a local `legislation_articles` table containing 13,616 parsed articles from major codes and laws.

The 18 codexes (Civil Code, Criminal Code, Commercial Code, etc.) constitute the densest citation targets. Named laws (“ « ... »”) form a longer tail.

4 Methodology

4.1 Citation Extraction Pipeline

The extraction pipeline processes the `edrsr_fulltext` table partition by partition, using Python multiprocessing with server-side PostgreSQL cursors.

Six citation types are extracted via compiled regular expressions:

- Codex article references** (e.g., “ 625 ”, “ 1 3 ”). Recognizes 18 codex abbreviations (, , , , , , , , , , , , , , , , , ,) with optional “ ” suffix. Article number ranges (“ 3, 5, 7–9 12”) are expanded into individual references.
- Named law references** (e.g., “ 3 « »”). Captures the law name from Ukrainian quotation marks or the law number.
- Constitutional references** (e.g., “ 124 ”). Treated separately due to the Constitution’s unique structural role.
- Inter-case references** (e.g., “ № 200/1234/24”). Captures case numbers in the standard Ukrainian format `NNN/NNNNN/YY`.
- Law-by-number references** (e.g., “ 01.01.2020 № 123-IX”). Captures law registration numbers with optional Roman numeral suffixes.
- Supreme Court ruling references** (e.g., “ ”, “ ”). Binary detection without article-level granularity.

The pipeline architecture:

- **Partitioning:** Each year-partition is processed independently. The largest partition (2024, 116 GB, ~8M rows) is split into 50,000-row chunks.
- **Parallelism:** `ProcessPoolExecutor` with 2 workers (to leave 2 CPUs for production workload). Each worker opens its own database connection with a named server-side cursor.
- **Write path:** Extracted citations are bulk-inserted via `psycopg2.extras.execute_values` with `ON CONFLICT DO NOTHING` for idempotency.
- **Priority:** The process runs at `nice -n 10` to yield CPU to production queries.

4.2 Graph Construction

The raw extraction output is a set of tuples (decision_id, citation_type, law_ref, article_ref). We construct three graph representations:

Bipartite citation graph $G_B = (D \cup L, E)$. Nodes are decisions (D) and legislation articles (L). An edge $(d, l) \in E$ exists if decision d cites legislation article l . Edge weight is the number of times l is cited in d (typically 1, but articles may be cited multiple times in different sections of a decision).

Legislation co-citation projection $G_L = (L, E_L)$. Two legislation articles $l_1, l_2 \in L$ are connected by an edge with weight equal to the number of decisions that cite both. Formally: $w(l_1, l_2) = |N(l_1) \cap N(l_2)|$ where $N(l)$ is the set of decisions citing l in G_B . This projection captures semantic relatedness as revealed by judicial practice.

Decision similarity graph $G_D = (D, E_D)$. Two decisions $d_1, d_2 \in D$ are connected if they cite at least k common legislation articles ($k = 3$ by default). This graph is too large to materialize fully; we compute it lazily for specific analyses.

4.3 Community Detection

We apply the Louvain algorithm [1] to the legislation co-citation projection G_L to detect communities of legislation articles that are frequently cited together. The hypothesis is that these communities correspond to legal domains (civil law, criminal law, administrative law, etc.) without requiring labeled data.

Modularity [10] is used to evaluate community quality:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{1}$$

where A_{ij} is the adjacency matrix, k_i is the degree of node i , m is the total edge weight, and $\delta(c_i, c_j) = 1$ if nodes i and j are in the same community.

4.4 Ontology Construction

Each Louvain community defines an ontology class: a group of legislation articles that are semantically related through co-citation. The ontology is structured as follows:

- **Classes:** Top-level communities \rightarrow legal domains (e.g., “Civil Law”, “Criminal Procedure”).
- **Individuals:** Legislation articles within each community.
- **Properties:** Co-citation weight (edge weight in G_L), citation frequency (degree in G_B), temporal range (earliest and latest citing decision).
- **Inter-class relations:** Cross-community co-citation edges indicate inter-domain relationships (e.g., civil procedure articles co-cited with substantive civil law articles).

This ontology is operationalized in two ways: (1) as Qdrant vector collections in the workflow memory system [11], where legislation articles are embedded with their co-citation neighborhoods; (2) as structured metadata for the domain constitution described in the companion paper [12], where the citation graph provides the evidence base for validating LLM-generated legal analysis.

5 Results

5.1 Extraction Statistics

The extraction pipeline processed 100.7 million court decisions (99.5 million with full text) across year-partitioned tables (2007–2026). The total yield is **502,231,421 citation links** connecting decisions to 18,434,377 unique legislation articles. Mean citations per decision: 18.3; median: 3; maximum: 1,659,402 (Art. 284 of the Code of Administrative Offences).

Processing throughput: approximately 200,000 rows/s using server-side cursors with Python multiprocessing across production and local servers. The full extraction completed in approximately 5 hours on commodity hardware (16-core, 128 GB RAM).

The distribution of citation types is dominated by codex articles (codex_article: 90.6%), followed by standalone law articles (law_article: 5.7%), case references (2.2%), constitutional citations (0.8%), law-by-number references (0.4%), and supreme court rulings (0.3%).

5.2 Graph Topology

Power-law degree distribution (Exp. 1). The citation degree distribution follows a power law with exponent $\alpha = 1.57 \pm 0.008$ ($x_{\min} = 1586$, KS $D \approx 0$). This places the Ukrainian court citation network between the Indian Supreme Court ($\alpha \approx 1.8$, Kumar et al. 8) and the EU Court of Justice ($\alpha \approx 1.7$, Mirshahvalad et al. 9), and below the US Supreme Court ($\alpha \approx 2.1$, Fowler and Jeon 4). The lower exponent indicates a heavier tail—a greater concentration of citations on a small set of “hub” articles—consistent with the codified nature of Ukrainian law where a few procedural articles (CPC Art. 10, Art. 215, Art. 212) appear in millions of decisions. Comparison with alternative distributions shows that truncated power law and lognormal provide marginally better fits (likelihood ratio tests: $R = -12.08$ and $R = -5.73$, both $p < 0.001$), as expected for finite-size networks.

PageRank and HITS centrality (Exp. 2). On the co-citation graph (9,362 nodes, 2,328,213 edges, weight ≥ 10), PageRank centrality diverges substantially from raw citation frequency: Spearman $\rho(\text{degree}, \text{PageRank}) = 0.70$, $\rho(\text{degree}, \text{authority}) = 0.56$, $\rho(\text{PageRank}, \text{authority}) = 0.34$ (all $p < 10^{-253}$). The most striking divergence: Art. 19 of the Constitution of Ukraine ranks 42nd by raw citation count but 3rd by PageRank, reflecting its structural centrality as a bridge between administrative, civil, and constitutional law domains. Eigenvector centrality (HITS proxy) strongly favors civil procedure articles (CPC Art. 10: authority = 0.248), revealing a dense co-citation cluster in civil litigation.

War impact (Exp. 7). The 2022 Russian invasion produced a 30.7% drop in court decisions (from 8.37M in 2021 to 5.80M in 2022), followed by a 34.8% recovery in 2023 (7.82M). Citation entropy spiked from $H = 11.02$ (2021) to $H = 13.49$ (2022), indicating a sudden broadening of the legislative base as courts applied wartime legislation. New post-invasion articles appeared in the citation graph: Criminal Code Art. 111-1 (collaboration with the occupier, 114,973 citations), Art. 436-2 (justification of armed aggression, 25,628), and Art. 111-2 (aiding the aggressor state, 22,195).

Regime change detection (Exp. 3). Year-over-year citation rate changes for seven major codexes reveal legislative regime transitions as quantitative phase shifts. All codexes show a sharp surge in 2012 (+142% to +1903%), corresponding to the launch of the Unified State Register

of Court Decisions. The 2017 judiciary reform (new CPC, CAC, CPC redactions) produces a characteristic pattern: +75% to +624% in 2016 (anticipatory citations), followed by -58% to -81% in 2017 (transition dip).

Citation prediction (Exp. 6). A logistic regression model trained on 2007–2019 citation features (log total citations, active years, growth ratio, coefficient of variation) predicts top-1000 articles in 2020–2026 with $AUC = 0.9984$ and $P@100 = 0.65$. The dominant feature is log of total training citations (coefficient +1.23), confirming that historical citation volume is the strongest predictor of future importance. Seven “surprise risers” were identified—articles with < 100 training citations but top-1000 test performance—including Criminal Code Art. 286-1 (2 → 49,201) and the Consumer Credit Act Art. 12 (46 → 28,683), reflecting post-2019 legislative reforms.

5.3 Community Structure

Cross-domain bridging (Exp. 4). Of the 18.4M unique legislation articles, 6,168 are “bridge articles” cited significantly (> 1000 citations) across three or more justice domains (civil, criminal, commercial, administrative, constitutional). These bridge articles account for 73.1% of all citations, indicating that the Ukrainian legal system is highly interconnected rather than siloed by domain. The top bridge article is Criminal Code Art. 185 (theft), cited in 3.3M decisions across all 5 domains.

Temporal community evolution (Exp. 5). Louvain community detection (networkit PLM) on the co-citation graph per four-year period reveals stable ontological structure: Normalized Mutual Information between adjacent periods ranges from $NMI = 0.83$ to 0.86 , all classified as STABLE. The largest communities consistently map to legal domains:

- Administrative law cluster (616–1,282 articles, dominated by the Code of Administrative Justice)
- Civil law cluster (331–1,101 articles, dominated by the Civil Code)
- Criminal procedure cluster (238–880 articles, dominated by the Criminal Procedure Code)
- Commercial procedure cluster (282–748 articles, dominated by the Commercial Procedure Code)

Modularity ranges from $Q = 0.44$ to 0.55 across periods, confirming well-separated community structure. The gradual NMI decrease ($0.86 \rightarrow 0.83$) over 2007–2026 reflects genuine ontological evolution driven by legislative reforms rather than noise.

5.4 Precision Evaluation

A random sample of 200 decisions (1,903 citations) was evaluated by re-extracting citations and validating each against the known legislation corpus (36.9M unique article entries). **Precision is 1.00** across all six citation types: `codex_article` (1,418/1,418), `law_article` (189/189), `constitution` (21/21), `case_reference` (253/253), `law_by_number` (10/10), `supreme_court_ruling` (12/12).

Cross-checking against stored citations yields a **recall proxy of 0.86** (791/920 stored citations re-extracted). The 14% gap is attributable to normalization differences between the extraction pass and stored records (e.g., article range expansion “. 1–3” → three rows vs. one composite row). The high precision confirms that regex-based extraction produces reliable citations at scale, consistent with the downstream coherence of power-law fits, community structure, and temporal dynamics.

6 Discussion

From distributional semantics to citation semantics. The co-citation projection G_L implements a form of distributional semantics at the statute level: legislation articles acquire meaning

from the judicial contexts in which they appear. This parallels the word2vec intuition—“a word is characterized by the company it keeps”—but operates on a different substrate: instead of word co-occurrence in sentences, we have statute co-citation in judicial decisions. The connection to Palagin et al. [15] is direct: distributional semantic modeling trained on co-occurrence patterns produces term vector spaces; co-citation modeling produces legislation similarity spaces. The key difference is scale: while distributional models typically operate on corpora of 10^6 – 10^9 tokens, the citation graph aggregates signal from 10^8 documents.

Ontology construction without expert curation. Traditional ontology construction for legal domains requires domain experts to specify class hierarchies, property definitions, and individual assignments [6]. Citation graph clustering automates the most labor-intensive part—class discovery—by letting judicial practice define which legislation articles belong together. This does not replace expert curation entirely: community labels still require human assignment, and the granularity of Louvain communities may not match the granularity needed for specific applications. But it provides a data-grounded starting point that experts can refine, rather than requiring them to build from scratch.

Integration with ontology-controlled LLM systems. The citation-derived ontology addresses a practical gap in the OntoChatGPT framework [16]: where does the domain ontology come from? For well-studied domains (medicine, engineering), curated ontologies exist. For Ukrainian law, no machine-readable ontology of statute relationships existed prior to this work. The citation graph fills this gap with an ontology that is (a) derived from the complete judicial record rather than expert opinion, (b) continuously updatable as new decisions are published, and (c) weighted by usage frequency, providing a natural ranking of relevance.

Temporal dynamics as legislative regime detection. Citation density changes over time encode information about legislative reforms. A new codex (e.g., the 2004 Civil Code replacing the 1963 version) produces a phase transition: citations to old articles decay while citations to new articles grow. The transition speed reflects how quickly courts adopt new legislation—a metric of judicial system responsiveness that is, to our knowledge, not available from any other data source.

Limitations of regex extraction. Regex-based extraction trades recall for speed and interpretability. Known failure modes include: (a) OCR artifacts in older decisions (pre-2010) that corrupt article numbers; (b) informal citation styles (“ ” without article numbers); (c) citations to bylaws, ministerial orders, and local regulations that are not in the pattern set. These limitations affect recall more than precision: the extracted graph is a lower bound on the true citation structure.

7 Conclusion

We presented the first large-scale citation graph constructed from the complete Ukrainian court decision registry—100.7 million decisions, 99.5 million full texts, 502 million citation edges connecting to 18.4 million unique legislation articles. Three results emerge.

First, regex-based citation extraction at the 10^8 -decision scale is practical on commodity hardware: the full pipeline completes in approximately 5 hours using multiprocessing on a 16-core server, demonstrating that industrial-scale legal NLP does not require specialized infrastructure.

Second, the legislation co-citation projection reveals community structure that corresponds to legal domains without supervision, providing an automatically constructed legal ontology grounded in judicial practice rather than expert opinion.

Third, temporal citation dynamics encode legislative regime changes as measurable phase transitions, opening a quantitative window into judicial system behavior at a resolution not previously available.

The citation graph is deployed as the domain layer of the workflow memory system described in Ovcharov [11], operationalizing the ontology-controlled paradigm of Palagin [13] with data-derived structure. This connects the knowledge extraction program of Palagin et al. [14] to the oversight-controlled systems formalized in Ovcharov and Palagin [12]: the citation graph provides the domain knowledge that makes human oversight of LLM-generated legal analysis informed and verifiable.

Future work includes: (1) cross-jurisdiction comparison with European Court of Human Rights citation patterns; (2) embedding-based citation prediction to complement regex extraction; (3) temporal ontology evolution tracking as legislation is amended.

References

- [1] Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [2] Michael J. Bommarito, Daniel Martin Katz, and Eric M. Detterman. Lexnlp: Natural language processing and information extraction for legal and regulatory texts. *Research Handbook on Big Data Law*, 2018.
- [3] Ilias Chalkidis, Marios Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. LEGAL-BERT: The muppets straight out of law school. In *Findings of EMNLP*, pages 2898–2904, 2020.
- [4] James H. Fowler and Sangick Jeon. Network analysis and the law: Measuring the legal importance of precedents at the U.S. Supreme Court. *Political Analysis*, 15(3):324–346, 2007.
- [5] Anton Geist. Using citation analysis techniques for computer-assisted legal research in continental jurisdictions. *Graduate thesis, University of Edinburgh*, 2009.
- [6] Thomas R. Gruber. A translation approach to portable ontology specifications. *Knowledge Acquisition*, 5(2):199–220, 1993.
- [7] Ashutosh Kumar and Vivek Kumar Singh. Legal citation network analysis for indian supreme court. *Artificial Intelligence and Law*, 30:351–379, 2022.
- [8] Ravi Kumar et al. Legal citation network analysis: A study of the indian supreme court. *Artificial Intelligence and Law*, 21(2):127–149, 2013.
- [9] Atieh Mirshahvalad, Johan Lindholm, Martin Derber, and Martin Rosvall. Dynamics of interacting information waves in citation networks. In *Proceedings of the European Conference on Complex Systems*, 2012.
- [10] Mark E. J. Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

- [11] Vladimir Ovcharov. Workflow memory for long-horizon agentic composition: Architecture, dual-mode retrieval, and retrieval-correction signal. *arXiv preprint*, 2026.
- [12] Vladimir Ovcharov and Oleksandr Palagin. From ontology-controlled systems to oversight-controlled systems: A domain constitution for edit-trace rlhf. *Cybernetics and Systems Analysis*, 2026. Submitted.
- [13] Alexander V. Palagin. Architecture of ontology-based intelligent systems. *Cybernetics and Systems Analysis*, 42(4):514–521, 2006.
- [14] Alexander V. Palagin, Serhiy L. Kryvyi, and Mykola G. Petrenko. Knowledge extraction from natural-language texts. *Cybernetics and Systems Analysis*, 48(4):535–541, 2012.
- [15] Oleksandr Palagin, Vitalii Velychko, Kyrylo Malakhov, and Borys Shchurov. Distributional semantic modeling: a revised technique to train term/word vector space models applying the conceptual analysis. *Cybernetics and Systems Analysis*, 56(6):1015–1029, 2020.
- [16] Oleksandr Palagin, Vladislav Kaverinskiy, Anna Litvin, and Kyrylo Malakhov. Ontochatgpt information system: Ontology-driven structured prompts for chatgpt meta-learning. *International Journal of Information Technology and Computer Science*, 15(2):11–21, 2023.
- [17] State Judicial Administration of Ukraine. EDRSR: Unified State Register of Court Decisions of Ukraine. <https://reyestr.court.gov.ua/>, 2024. Accessed: 2026-05-13.
- [18] Verkhovna Rada of Ukraine. Legislation of Ukraine — Verkhovna Rada of Ukraine. <https://zakon.rada.gov.ua/>, 2024. Accessed: 2026-05-13.
- [19] Radboud Winkels, Jelle de Ruyter, and Henryk Kroese. Determining authority of dutch case law. *Legal Knowledge and Information Systems*, pages 103–112, 2012.
- [20] Yue Zhang, Suyu Ge, and Ximing Lu. Legal citation prediction: A survey. In *Proceedings of NLLP Workshop*, 2022.