

Received 09 September 2025; revised 21 December 2025; accepted 29 December 2025; published 30 December 2025

Overview of Methods and Technologies for Obtaining Knowledge About Financial Proposals

Maksym Azarov* and Vasyl Lyashkevych

System Design Department, Faculty of Electronics and Computer Technologies, Ivan Franko National University of Lviv, Lviv, Ukraine

*Corresponding author (E-mail: maksym.azarov@lnu.edu.ua)

ABSTRACT In the contemporary world of digital finance and data-driven decision-making, the ability to effectively obtain, analyze and interpret knowledge about financial propositions (FPs) has become a critical challenge. FPs often contain critical information about investments, risk assessments, forecasts, and strategic intent. However, this data is not always complete and is often presented in semi-structured or unstructured forms, which limits its direct applicability. The motivation for this study is the exponential growth in both the volume and complexity of financial documentation, which requires more advanced and efficient approaches to context identification and knowledge obtaining (KO). By using natural language processing (NLP) techniques, machine learning (ML) algorithms, semantic analysis, and other innovative tools, public-private companies can increase the accuracy and efficiency of financial decision-making, ensure regulatory compliance, and minimize human error. The concept of a financial system is also formalized by defining its main components and describing it as a system with distinct properties and relationships. Such a systematic approach allows for a deeper understanding of the relationships between budgeting, forecasting and risk assessment, which is critical for obtaining the "right" knowledge. Based on the formalization of the FP, a conceptual scheme for KO about the FP is proposed, where the FP is considered as a multimodal information object from which structured knowledge is obtained using various methods. This knowledge is further used for analytical support of decision-making, but has significant limitations that require the involvement of experts and additional models. The article explores modern methods and technologies aimed at automating and improving the process of KO about such FPs, and also offers suggested directions for research in this area. Among them, it is planned to study the latest technologies and advanced approaches to further improve the automation and accuracy of KO about FPs, which will contribute to more informed and final decision-making in the financial sector.

KEYWORDS financial proposals, knowledge obtaining, natural language processing, machine learning, semantic analysis.

I. INTRODUCTION

The rapid digitalization of the financial sector has transformed how financial information is generated, shared and utilized. This has led to a significant increase in the volume, diversity and complexity of financial documentation, among which financial propositions (FPs) play a critical role in investment planning, risk assessment and strategic decision-making. These documents typically combine narrative descriptions, numerical estimates, forecasts and regulatory statements. They are predominantly presented in semi-structured or unstructured formats making it challenging to extract actionable insights quickly and accurately.

As the volume and complexity of financial documentation increases, manual analysis becomes increasingly inefficient and error-prone. Therefore, there is a demand for advanced methods and technologies that can automate the extraction, analysis and interpretation of knowledge about FPs from their description and context. Such opportunities exist not only to increase the efficiency of decision-making but also to ensure compliance with regulatory acts and minimize operational risks in business solutions.

Despite the increasing availability of natural language processing (NLP) and machine learning (ML) techniques, extracting actionable and reliable knowledge about FPs remains a challenging task. Existing approaches are often designed for isolated subtasks, such as sentiment analysis,

entity extraction or document classification, and lack a unified view of the FP as a holistic system of interconnected components. As a result, important dependencies between budgeting, forecasting, risk management and compliance are often ignored, which reduces the effectiveness and interpretability of automated analysis.

In addition, a lack of a generally accepted formalization of a FP concept makes it difficult to select and compare appropriate knowledge obtaining (KO) methods. Without such formalization, it is difficult to systematically evaluate existing technologies, identify their limitations or determine their applicability to real-world financial decision support systems. This study aims to investigate and analyze current methods and technologies for extracting, analyzing and interpreting knowledge about FPs in order to improve the accuracy and efficiency of decision-making processes in the financial sector. Therefore, the investigation tasks are:

- To formalize the concept of a FP by identifying its core components and system properties;
- To review the applicability of existing methods and technologies for KO about FPs;
- To analyze the challenges associated with processing semi-structured and unstructured financial data;
- Explore potential innovative hybrid approaches and emerging technologies that could enhance the automation and accuracy of KO and context extraction for financial offers.

II. ANALYSIS OF RECENT RESEARCH AND PUBLICATIONS

A. Financial proposal. The term FP has been defined in various ways across academic and professional literature, reflecting its diverse applications in different domains.

In paper [1] authors define a proposal as a comprehensive, structured document that combines a compelling scientific idea with a clear and justified financial plan. It is not merely a request for money, but a detailed representation of how funds will support the successful execution of a proposed project.

In “Methods and Mechanisms of Financing Investment Activities in Industrial Enterprises” by Nodir Giyosalievich Khidirov [2], the term FP refers to a structured set of methods and mechanisms formulated to attract and allocate financial resources for investment activities within industrial enterprises. FP outlined in the article [3] is described as a methodology for assessing financial risks, structuring investment development, and ensuring that enterprise development is sustainable in an economic, financial and environmental sense.

Also, one of the FP types is a grant proposal. In the article [4] “Competing for Research Funding: Key Elements Impacting the Evaluation of Grant Proposals” by Mihaela Mocanu, Valentina Diana Rusu, and Anca-Diana Bibiri, a grant proposal is defined as a structured request submitted by researchers aiming to secure funding for a research project.

B. FPs context and data consolidation. Data consolidation plays a critical role in financial management by transforming fragmented datasets into a unified and coherent structure. In the context of FPs, this process ensures that all relevant financial metrics, such as costs, revenue projections, investment risks and funding requirements, are consistently organized and aligned with the objectives of the project. Consolidated and consistent data enables more accurate forecasting, improves budget transparency and facilitates compliance with reporting standards.

Using intelligent tools such as automated software, web crawlers, and ML algorithms significantly enhances the accuracy and efficiency of this process. These technologies enable real-time data collection, pattern recognition, and trend analysis, making them valuable for preparing well-informed and evidence-based FPs. Several articles have focused on this topic. For example, in study [5], author uses Python crawler and k-means algorithms to grab information from the website and express it through the charts. Fan Zhang, Ye Ding, and Yuhao Liao were processing and analyzing large-scale financial information, especially trends of stock prices, with there’s visualization using matplotlib. In article [6], Liyang Wang, Yu Cheng, Xingxin Gu and Zhizhong Wu used a quite complicated approach to process the data in the risk monitoring system. Preprocessing includes:

- data cleaning, such as removing outliers and filling missing values;
- data normalization to ensure comparability across different scales;
- feature extraction to identify and construct features that are most helpful for predictive models.

Authors were using Hadoop Distributed File System (HDFS) to store market data, Apache Spark for complex

computations, Hive & Pig for complex data querying and analysis tasks, and Apache Flink for continuous data streams. In study [7] Jingru Wang, Wen Ding, Xiaotong Zhu were using NASDAQ Financial Fundamentals dataset obtained from the Kaggle platform, which comprises quarterly financial fundamental data for companies listed on the NASDAQ stock exchange.

C. Structure of FP. The structure of a FP ensures the transparency, feasibility and strategic alignment of a project with its funding objectives. A well-structured FP enables evaluators to clearly understand the goals and plans.

In study [8], the authors investigate the interrelations among components of business models, which are pertinent to understanding FPs as systems which include:

- value proposition – the benefits which a business model provides to its customers, product and service offering; the specific products and services designed to realize the value proposition;
- financial model – encompasses funding, cost, pricing, distribution, revenue, and profit models;
- customer and market segment – the target customers and market segments of the business model.

The article “Some observations on contemporary FPs” [9] conceptualizes a FP as a system composed of multiple interrelated components rather than a list of financial figures. According to the study, key components identified include: budgeting and cost estimation, financial forecasting, risk management, funding sources and contract types, and regulatory compliance. The article emphasizes that the links between these components are crucial. For example, budget decisions impact risk management strategies, which in turn influence funding approaches and compliance requirements.

D. Context extraction and KO technologies applied for FPs.

Gaining insight into FPs is no longer limited to manual reviews or basic statistical summaries. Today, researchers and practitioners leverage a diverse range of intelligent methods, tailored not only to the structure of financial products but also to the specific analytical goals in play. Depending on the context, the process may involve structured and unstructured data mining.

For instance, in [10], Shuaiqun Pan, Sergio J. Rodríguez Méndez, and Kerry Taylor introduced a creative approach to assessing grant applications by assigning “innovation and creativity” (IC) scores. Their solution employed a Random Forest (RF) classifier trained on text representations generated with a modified version of the TF-IDF algorithm - utilizing only the inverse document frequency (IDF) component. Despite this simplification, their model demonstrated impressive performance, reaching an accuracy of 84.17% across all types of grant applications. In paper [11], authors provided a well-organized and structured overview of how NLP is used in financial systems, proposing a detailed taxonomy of NLP methods used in financial applications, grouping them into categories such as probabilistic models, vector-space approaches, deep learning techniques (like LSTM and BERT), and hybrid models that combine multiple strategies. Another inventive application of NLP appears in [12], where the authors designed a system to help

investors identify significant financial events. Their approach includes segmenting text to group related content, applying co-reference resolution to track dependencies within the documents, and using Latent Dirichlet Allocation (LDA) to filter out less relevant information.

In “Extraction of Forward-looking Financial Information for Stock Price Prediction from Annual Reports Using NLP Techniques” [13] authors developed a methodology for the extraction of FLS from 10-K reports and used those for stock price growth prediction using a 4-step algorithm. In this study, Forward-Looking Statements (FLS) are first identified using a combination of rule-based methods, manual labelling, and ML models. The approach builds on previous work by incorporating the state-of-the-art NLP model DistillBERT and comparing its performance with traditional models like SVM and LSTM, resulting in three versions of FLS datasets for each report. The sentiment of the identified FLS is then assessed using FinBERT. Finally, both the sentiment and the number of FLS items are used as inputs in a Random Forest model to predict the stock price growth rate over various time horizons – one day, one week, one month, six months, and one year after the report’s publication.

In [14] authors review the evolution of automatic IE methods as applied to scientific literature, tracing key milestones in entity recognition, relation extraction, and structured data retrieval from research articles. Also in this paper scientific IE (SciIE) process is presented, and key challenges of the sphere: formats of scientific texts, the need for (and lack of) training data, and the sparsity of information of interest in literature. In work [15] provides a comprehensive systematical overview of IE of different areas and their most typical methods: entity and relation extraction (RNN, CNN, PCNNs, CAN, LSTM, K-CNN, CRN, 2ATT-BiGRU, FastText-BiGRU-Dual Attention, GREG, SKEoG, AGCN, FC-GCN, IPR-DCGAT, BioGraphSAGE, KERE, NAM, ATLOP, HDP, KSBERT, KEWE-BERT, Gated-K-BERT, GraphRel, A2DSRE, REIF), event extraction (DMCNN, JRNN, DCFEE, HNN-EE, CNN-BiGRU, DEGREE, UIE), and multimodal information extraction (IKRL, Rp-BERT, UMGF, MEGA, CAT-MNER, HVPNeT). In paper [16] author presented FinancialBERT, which has the same structure as BERT, a new pre-trained language model, which has been trained on a large corpus (TRC2-financial, Bloomberg News, Corporate Reports, Earnings Call Transcripts) and can be fine-tuned for multiple NLP tasks. Requiring minimal task-specific architectural modification, our model achieves state-of-the-art performance on the Sentiment Analysis task, significantly outperforming other compared models (BERT-base, FinBERT). In paper [17], the authors conducted a comprehensive analysis of recent approaches in the intersection of NLP and financial market prediction. They identified key methodologies for text representation, sentiment analysis, information retrieval, and predictive modelling. Specifically, they categorized text representation techniques into traditional methods such as Bag-of-Words and TF-IDF, as well as advanced approaches using contextualized word embeddings like BERT. For sentiment and emotion analysis, both lexicon-

based and transformer-based models were discussed, highlighting their importance in interpreting financial texts. The authors also emphasized the role of IE (NER, FinBERT, etc.) and graph mining techniques, including the construction of financial knowledge graphs and the use of graph embeddings.

In paper [18], the author proposed that the BERT model achieved 62.38% precision in predicting sentiment for the POSITIVE class and 55% for the NEGATIVE class of the texts of 8-K current reports published in the EDGAR system. In paper [19], the authors of the chapter explore how to extract and represent financial entities and relationships using techniques like named entity recognition (NER), entity linking, and relation extraction, and then apply knowledge graph construction, refinement, and visualization to support applications in finance and legal analysis.

In paper [20], authors proposed a semantic web-based solution to automatically extract information from banks’ annual financial reports, build a financial ontology, transform extracted data into a knowledge graph, and run smart SPARQL queries. The steps of the pipeline are: retrieving of the annual financial report from the repository, data conversion into W3C-compliant tabular data format with relevant data extraction, modelling and visualization of the knowledge base.

In paper [21], the FinTOC-2020 Shared Task on structure extraction from financial documents, its participants' results and their findings are presented. The main task was to extract the TOC (table of contents) from French and English documents. One of the team DNLP used Tesseract, an open-source OCR tool, to extract the text regions. The other team, Taxy.io, proposed a DBSCAN clustering on pages and a multilingual BERT model. The UWB team used a Maximum entropy classifier on top of a diverse set of features, including orthographic characters and character n-grams.

In paper [22], the main goal was to extract structured data from a document named KIDs (Key Information Documents). Two approaches were introduced here: rule-based and ML-based IE. For rule-based IE, around 100 rules were defined to extract data fields. In most cases, it is a simple TokensRegex. For the ML-based approach, CascadeTabNet, the OCR method Tesseract was used to extract text from segmented cells.

In paper [23], the main goal was to extract textual features from annual reports of FTSE350 companies. Several methods for resolving were introduced: multi-task classification methods, encoder: pre-trained masked language model, joint multi-task learning, weighting, sequential multi-task learning, etc. Methods that were used: RoBERTa encoder, fine-tuned on masked-language modelling with FTSE350 corpus (domain adaptation).

In paper [24], the author proposed a ViBERTgrid BiLSTM-CRF model, which demonstrates a significant improvement in performance (up to 2 percentage points) on named entity recognition from unstructured documents in the financial domain. This model combines the strengths of both transformer-based multimodal architectures (ViBERTgrid) and sequence-based models (BiLSTM-CRF).

III. RESEARCH RESULTS AND THEIR DISCUSSION.

A. Definition of FP. From a systems perspective, financial planning integrates budgeting, cost estimation, financial forecasting, risk assessment, financing mechanisms and compliance requirements. These factors are interdependent and collectively determine the feasibility, sustainability and expected outcomes.

In this work, we were focused on formalizing the concept of a FP and identifying its core characteristics. By analysing various definitions across academic literature, it was found that FPs are not only documents requesting funding, but also complex systems that encapsulate strategic intentions, budget allocations, expected outcomes and risk management approaches.

An FP is an entity that represents a complex system of interconnected financial, strategic and regulatory elements designed to justify the allocation of financial resources to a specific project, investment or activity. In this context, the FP is not only a funding request, but also a knowledge that contains explicit information, such as costs, timelines and funding sources and tacit knowledge, such as strategic intent, assumptions and risk tolerance.

Thus, an FP can be formally considered as a system, the structure, content and internal relationships of which determine both the information value and the applicability of automated methods for obtaining knowledge and supporting decisions.

B. Comparative analysis of KO methods. The concept of knowledge about FPs includes extracted entities, such as costs, funding amounts and risk types and inferred relationships between investment goals and returns. These parameters are typically hidden within unstructured text and require advanced IE or KO techniques for retrieval.

KO methods go far beyond manual review or simple rule-based systems. These methods include a range of ML and NLP approaches aimed at understanding and analysing financial text, tables and structured data. Key methods include:

- Classification algorithms (Random forests, SVMs) for evaluating proposal quality or predicting outcomes;
- Text representation models, such as Bag-of-Words, TF-IDF, contextual embeddings and others;
- Named Entity Recognition (NER) to extract financial entities such as organizations, currencies, dates or KPIs;
- Sentiment analysis to assess tone or outlook within forward-looking statements or management commentary;
- Event detection and relation extraction to identify significant financial events and relationships within proposals;
- Table and structure extraction to interpret structured data inside financial documents, such as budgets or risk tables;
- Multimodal approaches that combine text, visual layout and numerical data.

Technologies used for KO:

- Pre-trained language models (BERT, FinBERT, RoBERTa and domain-adapted models fine-tuned on financial texts);

- Optical Character Recognition (OCR). Tools like Tesseract to extract text from scanned or image-based documents;
- Deep learning architectures. LSTM, CNNs, BiLSTM-CRF and transformer-based networks for sequence labelling, classification or feature extraction;
- Topic modelling. Methods like Latent Dirichlet Allocation (LDA) for filtering or organizing content;
- Knowledge graphs and ontologies. To structure and link financial concepts and support advanced querying;
- Semantic web and SPARQL. For data integration and smart querying of structured financial knowledge;
- Clustering and rule-based systems: DBSCAN, TokensRegex or custom rule pipelines for structured extraction tasks.

While KO methods have significantly improved the accessibility and structuring of information contained in FPs, they still have fundamental limitations (Table 1) in capturing implicit intentions, validating predictions, ensuring consistency across contexts and supporting responsible decision-making. These limitations underscore the necessity of adopting a hybrid approach that combines automated context extraction with expert evaluation, domain knowledge and robust governance mechanisms.

TABLE 1. A list of unsolved problems for existing KO methods.

Unsolved problems	Types of methods	Limitation explanations
Implicit assumptions and strategic intent	All automatic methods (KG, rule-based, ML, transformers)	The methods only work with explicit content. Strategic intentions and implicit assumptions are not explicitly represented in the data
Subjective and normative assessments	ML classifiers, transformers	Models can simulate “scoring” but have no objective criterion of “quality” without external rules
Consistency between documents and contexts	Rule-based, ML, transformers	Most pipelines are document-centric and do not guarantee inter-document or temporal consistency.
Validation of predictions and causal claims	All IE/NLP/LLM methods	Obtaining is not equal to verification. Methods are unable to assess the economic plausibility of forecasts
Quantifying risks under uncertainty	Sentiment analysis, topic modelling, transformers	Qualitative descriptions of risks are not reliably transformed into numerical indicators without external models
Explainability of complex models	Transformers, multimodal KIE	High accuracy comes at the cost of low decision transparency
Resistance to manipulative or rhetorical text	Transformers, ML classifiers	Models are sensitive to stylistic changes, decorations and adversarial phrasing
Legal and regulatory interpretation	Rule-based, KG, semantic web	Formal compliance is not equal to legal adequacy. It requires legal expertise
The limitations and quality of ground truth	ML, DL, transformers	Financial marked-up corpora are limited, closed and unified

Different types of KO methods have their own advantages and complement each other. However, no single method can completely solve the problems related to implicit intent, subjective evaluation, predictive verification, legal interpretation and liability. These limitations are inherent to the decision-making environment and not specific to any particular algorithm, highlighting the necessity of adopting a human-machine hybrid approach in financial decision analysis. To put this hybrid perspective into practice and to clarify the interaction between automated methods and expert involvement, a conceptual framework for KO about FPs is proposed below (Fig. 1).

This framework integrates various KO methods, verification mechanisms, and human oversight, thereby clarifying the capabilities and limitations of automation. The proposed conceptual scheme considers a financial offer as a multimodal information object from which structured knowledge is obtained using different types of methods. This knowledge is further used for analytical support of decision-making but is subject to significant limitations that require the involvement of experts and additional models.

The proposed conceptual scheme illustrates FP analysis as a multi-layered process in which heterogeneous obtaining methods transform unstructured documents into structured knowledge. While automated techniques significantly improve accessibility and consistency of obtained knowledge, fundamental limitations related to interpretation, validation, and accountability necessitate the inclusion of expert-driven governance. Thus, effective analysis of FPs should be understood as a hybrid human-AI process rather than a fully automated pipeline.

The proposed conceptual scheme provides a system-level understanding of KO about FPs by integrating heterogeneous methods, representations, and decision-making roles into a unified analytical framework.

First, the scheme clearly defines the functional layers, distinguishing between document preprocessing, KO, knowledge representation, verification and decision

support. This division clarifies the role and limitations of each method, avoiding the erroneous assumption that any single technology can completely solve the problem of financial solution analysis.

Secondly, this paper formalizes the concept of financial planning as a complex, multimodal and interdependent system, rather than a simple financial document. This formalization provides a general analytical framework for comparing different data collection methods and understanding their applicability and limitations.

Third, a comparative analysis of KO methods was conducted based on the proposed scheme. This analysis highlights the complementary advantages of different types of methods and clearly points out the problems that automated methods cannot yet solve.

Ultimately, by introducing multi-layered verification and governance, the framework demonstrates the necessity of a hybrid human-machine approach and lays the foundation for future research on the security, interpretability, and reliability of financial decision support systems.

VII. CONCLUSIONS

This paper formally defines financial planning as a complex, multimodal, and interdependent system that integrates budgeting, forecasting, risk management, financing mechanisms and regulatory compliance requirements. This formalization provides a common analytical framework for assessing the applicability and limitations of automated KO methods and distinguishes financial planning from general financial documentation.

The results of the comparative analysis show that while different methods have their own advantages, none of them can independently address the key issues related to implicit strategic intent, subjective assessment, predictive validation, legal interpretation and accountability. Aforementioned unresolved problems are inherent to the financial decision-making context rather than to specific algorithms. To overcome these limitations, this paper proposes a conceptual framework for KO from FPs. This framework integrates heterogeneous KO methods, structured knowledge representation, verification mechanisms and human-machine collaborative management. The framework clarifies the functional roles and boundaries of automation and supports the design of hybrid human-machine decision support systems.

Standardization of financial decisions can clearly identify and categorize the knowledge contained in financial documents. By identifying different types of knowledge and their interrelationships, the scope and limitations of automated KO can be clearly defined, which will help to make financial decisions more consistent and easier to interpret.

Future research should focus on:

- enhancing the security of KO processes, protecting them from manipulative or adversarial content;
- improving the interpretability of complex models used for financial text analysis;
- increasing the reliability of KO through validation, consistency checks and integration with economic and simulation models;

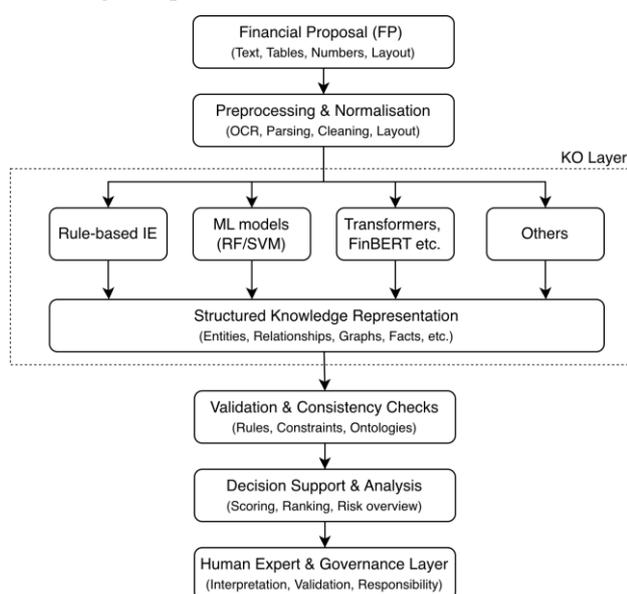


FIG. 1. Conceptual scheme of KO about FPs.

- investigation of LLaMA as language models, financial ontologies like FIBO and layout-aware models like LayoutLM and LLMs.

These directions are crucial for advancing trustworthy and responsible intelligent financial decision support systems.

AUTHOR CONTRIBUTIONS

M.A., V.L. – conceptualization, methodology, investigation, writing (original draft preparation), writing (review and editing).

COMPETING INTERESTS

The authors declare no conflict of interest.

REFERENCES

- [1] L. N. Münch, K. F. Schüttler, J. Ackermann, A. Deichsel, L. Eggeling, D. Günther et al., “Writing a research funding proposal,” [Online]. Available: <https://link.springer.com/article/10.1007/s00142-024-00690-x>. [Accessed: August 11, 2025].
- [2] N. Khidirov, “Methods and Mechanisms of Financing Investment Activities in Industrial Enterprises,” [Online]. Available: https://www.researchgate.net/publication/351675916_Methods_and_Mechanisms_of_Financing_Investment_Activities_in_Industrial_Enterprises. [Accessed: August 11, 2025].
- [3] A. Tobisova, A. Senova, G. Izarikova, and I. Krutakova, “Proposal of a Methodology for Assessing Financial Risks and Investment Development for Sustainability of Enterprises in Slovakia,” [Online]. Available: <https://www.mdpi.com/2071-1050/14/9/5068>. [Accessed: August 11, 2025].
- [4] M. Mocanu, V. D. Rusu, and A. Bibiri, “Competing for research funding: Key elements impacting the evaluation of grant proposal,” [Online]. Available: <https://www.cell.com/heliyon/fulltext/S2405-8440%2824%2912046-4>. [Accessed: August 11, 2025].
- [5] F. Zhang, Y. Ding, and Y. Liao, “Financial data collection based on big data intelligent processing,” [Online]. Available: https://www.researchgate.net/publication/369716116_Financial_Data_Collection_Based_on_Big_Data_Intelligent_Processing. [Accessed: August 11, 2025].
- [6] L. Wang, Y. Cheng, X. Gu, and Z. Wu, “Design and Optimization of Big Data and Machine Learning Based Risk Monitoring System in Financial Markets,” [Online]. Available: <https://arxiv.org/pdf/2407.19352>. [Accessed: August 11, 2025].
- [7] J. Wang, W. Ding, and X. Zhu, “Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG,” [Online]. Available: <https://arxiv.org/pdf/2504.06279>. [Accessed: August 11, 2025].
- [8] N. Schaffer, C. Drieschener, and H. Kromar, “An Analysis of Business Model Component Interrelations,” [Online]. Available: https://www.researchgate.net/publication/342304352_An_Analysis_of_Business_Model_Component_Interrelations. [Accessed: August 11, 2025].
- [9] S. Ahmed Shaikh, “Some observations on contemporary financial proposals,” [Online]. Available: https://www.researchgate.net/publication/367598580_Some_observations_on_contemporary_financial_proposals. [Accessed: August 11, 2025].
- [10] S. Pan, S. J. Rodríguez Méndez, and K. Taylor, “A pipeline for analysing grant applications,” [Online]. Available: <https://arxiv.org/pdf/2210.16843>. [Accessed: August 11, 2025].
- [11] D. Millo, B. Vika, and N. Baci, “Integrating natural language processing techniques of text mining into financial system: applications and limitations,” [Online]. Available: <https://arxiv.org/pdf/2412.20438>. [Accessed: August 11, 2025].
- [12] S. Garcia-Mendez, F. de Arriba-Perez, A. Barros-Vila, F. J. Gonzalez-Castano, and E. Costa-Montenegro, “Automatic detection of relevant information, predictions and forecasts in financial news through topic modelling with Latent Dirichlet Allocation,” [Online]. Available: <https://link.springer.com/article/10.1007/s10489-023-04452-4>. [Accessed: August 11, 2025].
- [13] A. Glodd and D. Hristova, “Extraction of Forward-looking Financial Information for Stock Price Prediction from Annual Reports Using NLP Techniques,” [Online]. Available: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/757da8dd-90c7-489f-995a-1cb29b4c9557/content>. [Accessed: August 11, 2025].
- [14] Z. Hong, L. Ward, K. Chard, B. Blaiszik, and I. Foster, “Challenges and Advances in Information Extraction From Scientific Literature: A Review,” [Online]. Available: <https://www.osti.gov/servlets/purl/1869908>. [Accessed: August 11, 2025].
- [15] Y. Yang, Z. Wu, Y. Yang, S. Lian, F. Guo, and Z. Wang, “A Survey of Information Extraction Based on Deep Learning,” [Online]. Available: <https://www.mdpi.com/2076-3417/12/19/9691>. [Accessed: August 11, 2025].
- [16] A. R. Hazourli, “FinancialBERT - A Pretrained Language Model for Financial Text Mining,” [Online]. Available: https://www.researchgate.net/publication/358284785_FinancialBERT_-_A_Pretrained_Language_Model_for_Financial_Text_Mining. [Accessed: August 11, 2025].
- [17] S. A. Farimani, M. V. Jahan, and A. M. Fard, “From Text Representation to Financial Market Prediction: A Literature Review,” [Online]. Available: <https://www.mdpi.com/2078-2489/13/10/466>. [Accessed: August 11, 2025].
- [18] M. Wujec, “Analysis of the Financial Information Contained in the Texts of Current Reports: A Deep Learning Approach,” [Online]. Available: <https://www.mdpi.com/1911-8074/14/12/582>. [Accessed: August 11, 2025].
- [19] T. Repke and R. Krestel, “Extraction and Representation of Financial Entities from Text,” [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-030-66891-4_11. [Accessed: August 11, 2025].
- [20] S. F. Mohsin, S. I. Jami, S. Wasi, and M. S. Siddiqui, “An automated information extraction system from the knowledge graph based annual financial reports,” [Online]. Available: <https://pmc.ncbi.nlm.nih.gov/articles/PMC11157543>. [Accessed: August 11, 2025].
- [21] N. Bentabet, R. Juge, I. El Maarouf, V. Moulleron, D. Valsamou-Stanislawski, and M. El-Haj, “The Financial Document Structure Extraction Shared task: FinToc 2020,” [Online]. Available: <https://aclanthology.org/2020.fnp-1.2.pdf>. [Accessed: August 11, 2025].

August 11, 2025].

- [22] F. Scafoglieri, A. Monaco, G. Neccia, D. Lembo, A. Limosani, and F. Medda, "Automatic Information Extraction from Investment Product Documents," [Online]. Available: <https://ceur-ws.org/Vol-3194/paper9.pdf>. [Accessed: August 11, 2025].
- [23] S. Montariol, M. Martinc, A. Pelicon, S. Pollak, B. Koloski, I. Loncarski et al., "Multi-Task Learning for Features Extraction in Financial Annual Reports," [Online]. Available: https://www.researchgate.net/publication/367534259_Multi-task_Learning_for_Features_Extraction_in_Financial_Annual_Reports. [Accessed: August 11, 2025].
- [24] F. Pala, M. Yasin Akpınar, O. Deniz, and G. Eryigit, "ViBERTgrid BiLSTM-CRF: Multimodal Key Information Extraction from Unstructured Financial Documents," [Online]. Available: <https://arxiv.org/pdf/2409.15004>. [Accessed: August 11, 2025].



Maksym Azarov

Bachelor's in Computer Science with an advanced foundation in both theoretical and practical aspects of software development, computer systems, and modern technologies. Has gained extensive professional experience as a Full-stack Engineer over several years, building on web-app and web-services from scratch.

ORCID ID: [0009-0003-5261-0468](https://orcid.org/0009-0003-5261-0468)



Vasyl Lyashkevych

PhD in Computer Science with excellent analytical and problem-solving skills in multi-subject domains. Constantly learning new technology, has been working for 20+ years in R&D and 15+ years in Enterprise, can build efficient, innovative business and technological solutions with poorly formalised initial requirements and business needs.

ORCID ID: [0000-0003-2810-6061](https://orcid.org/0000-0003-2810-6061)

Огляд методів і технологій видобування знань про фінансові пропозиції

Максим Азаров*, Василь Ляшкевич

Кафедра системного проектування, факультет електроніки та комп'ютерних технологій, Львівський національний університет імені Івана Франка, Львів, Україна

*Автор-кореспондент (Електронна адреса: maksym.azarov@lnu.edu.ua)

АНОТАЦІЯ У сучасному світі цифрових фінансів та прийняття рішень на основі даних здатність ефективно вилучати, аналізувати та інтерпретувати знання про фінансові пропозиції (ФП) стала критично важливим завданням. ФП часто містять важливу інформацію про інвестиції, оцінку ризиків, прогнози та стратегічні наміри. Однак ці дані не завжди є повними й нерідко представлені в напівструктурованому або неструктурованому вигляді, що обмежує можливості їх безпосереднього використання. Мотивацією цього дослідження є експоненційне зростання як обсягу, так і складності фінансової документації, що потребує застосування більш просунутих та ефективних підходів до ідентифікації контексту й видобування знань. Використання методів обробки природної мови, алгоритмів машинного навчання, семантичного аналізу та інших інноваційних інструментів дає змогу приватним компаніям і державним установам підвищити точність та ефективність прийняття фінансових рішень, забезпечити відповідність нормативним вимогам і мінімізувати людські помилки. Концепцію ФП формалізовано шляхом визначення її основних компонентів та опису як системи з окремими властивостями і взаємозв'язками. Такий системний підхід дає змогу глибше зрозуміти залежності між бюджетуванням, прогнозуванням та оцінюванням ризиків, що є критично важливим для видобування релевантних знань. На основі формалізації ФП запропоновано концептуальну схему видобування знань про ФП, у межах якої фінансова пропозиція розглядається як мультимодальний інформаційний об'єкт, з якого за допомогою різних методів отримують структуровані знання. Надалі ці знання використовуються для аналітичної підтримки прийняття рішень, однак мають певні обмеження, що зумовлює потребу в залученні експертів і додаткових моделей. У статті досліджуються сучасні методи й технології, спрямовані на автоматизацію та вдосконалення процесу видобування знань про ФП, а також пропонуються перспективні напрями подальших досліджень у цій сфері. Зокрема, планується вивчення новітніх технологій і передових підходів для підвищення рівня автоматизації та точності видобування знань, що сприятиме більш обґрунтованому й своєчасному прийняттю рішень у фінансовому секторі.

КЛЮЧОВІ СЛОВА фінансові пропозиції, видобування знань, обробка природної мови, машинне навчання, семантичний аналіз.



This article is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.