Journal of
Empirical Research In Accounting

Alzahra University, Faculty of Social Sciences and Economics
Empirical Research in Accounting, Winter 2025, V. 15, No. 58, pp.97-128

Textual Content Analysis of Financial Reports Using Named Entity Recognition: A Bibliometric Study and Visualization[1]

Malihe Farjami[2], Ebrahim Abbasi[3]

Received: 2025/02/16
Accepted: 2025/05/11

Research Paper

Abstract

Named entity recognition (NER), a subfield of natural language processing (NLP), is an emerging approach in financial and accounting research. This AI-based technique enables automatic extraction of key information, such as the names of individuals, organizations, and locations from financial texts, providing a more comprehensive analysis of corporate disclosures. The aim of this study is to conduct a Bibliometric review of scholarly output related to the application of NER in the financial and accounting domains, in order to familiarize researchers with its potential. First, the significance of the textual sections of corporate financial reports is outlined, followed by an introduction to NER and its role in textual content analysis. Using data from the Scopus database, this study identifies research trends, the most cited publications, and the most relevant finance and accounting articles. Furthermore, with the aid of VOSviewer software, co-authorship authors, co-authorship countries, and co-accurrence keywords are visualized and analyzed. The findings reveal a growing academic interest in this area in recent years, despite existing challenges, with China emerging as the leading contributor. Most studies focus on information extraction and entity relationship identification within financial texts.

Keywords: Bibliometric Study, Financial Reporting, Named Entity Recognition (NER), Textual Content Analysis.

JEL Classification: G14, M41, C88.

---

1. doi: 10.22051/jera.2025.50272.3524
2 Ph.D. Student, Department of Management, Faculty of Social and Economic Sciences, Alzahra University, Tehran, Iran. (malihe_farjami@yahoo.com).
3 Professor, Department of Management, Faculty of Social Sciences and Economics, Alzahra University, Tehran, Iran. (Corresponding Author). (abbasiebrahim2000@Alzahra.ac.ir).

Introduction

In recent decades, with the development of capital markets and the increasing awareness of users of accounting information, expectations regarding the role of accounting have expanded. As financial, economic, and market environments have become more complex, accounting is no longer expected to provide merely numerical and financial information; rather, stakeholders increasingly demand more comprehensive, understandable, and informative disclosures, particularly non-numerical, non-financial, and narrative information. Corporate managers use textual disclosures to convey information to investors and to enhance investors' understanding of simultaneously disclosed quantitative financial data. Textual disclosure, whether mandatory or voluntary, enhances information transparency and contributes to a reduction in information asymmetry. Examples of textual data include sections of the notes to financial statements, management discussion and analysis (MD&A), corporate press releases, website content, analyst reports, and analytical reports disseminated through social and traditional media by various corporate stakeholders, including internal managers, investors, customers, suppliers, interest groups, and policymakers. The rapid growth of textual data in financial markets is clearly evident. For instance, Dyer et al. (2017) document a 113 percent increase in the length of narrative disclosures in U.S. corporate reports. The expansion in the volume of unstructured textual data provides market participants with richer information, thereby facilitating investment decision-making. Moreover, academic researchers can leverage these data to conduct more innovative and informative research.

To date, a substantial body of research has examined the textual content of financial reports using content analysis approaches, focusing on various linguistic attributes such as tone, readability, and clarity. In this line of research, financial narratives are generally evaluated at an aggregate level with respect to their linguistic characteristics, and their effects on firm performance, stock returns, and, consequently, investors' judgments and decision-making are analyzed. However, only a limited number of studies have investigated specific and well-defined elements within financial texts, largely due to methodological challenges and the lack of appropriate analytical tools.

Advancements in artificial intelligence have recently provided researchers with powerful tools that facilitate more precise and efficient empirical analyses. In recent studies, artificial intelligence–based techniques have enabled researchers to pursue more granular objectives, such as identifying the names of customers, competing organizations, and foreign counterpart countries mentioned in corporate reports. Among artificial intelligence tools that focus on understanding and generating

human language are various natural language processing (NLP) methods. NLP, a subfield of computational linguistics, aims to develop systems capable of understanding, interpreting, and even generating human language.

In this study NER, as one of the core NLP techniques, is introduced and elaborated as a tool for analyzing the textual content of financial reports. Named entity recognition refers to the process of identifying and classifying concrete and observable textual units, such as the names of individuals, locations, organizations, and other specific entities, within a text. The primary objective of this process is to structure unstructured textual data so that they can be utilized in subsequent empirical analyses. The main function of NER is to identify and extract three broad categories of labeled information in texts, entities, temporal expressions, and numerical values, along with seven common subcategories, including persons, organizations, locations, currencies, time, dates, and percentages. Depending on the specific objectives of a study, researchers can leverage different outputs of this technique for empirical investigation.

The application of NER to the textual analysis of financial reports represents an interdisciplinary research domain that integrates finance and accounting with computer science, linguistics, and psychology. Named entity recognition is an artificial intelligence–based method situated at the intersection of computer science and computational linguistics. The outputs generated by this technique consist of clear and concrete information, which enhances both the speed and accuracy of information processing for users and facilitates information retention relative to other types of disclosures. This phenomenon is commonly referred to in psycholinguistics as the concreteness effect, a concept grounded in interdisciplinary research spanning linguistics and psychology (Wang & Yao, 2012).

Recent psycholinguistic research corroborates the cognitive implications of concreteness and highlights the role of proper nouns in improving the efficiency of information evaluation. Specifically, concrete and well-defined information activates not only the verbal coding system but also the non-verbal coding system of the brain through its inherent imagery. As the intensity of activation across both systems increases, information is processed more deeply, ultimately leading to more accurate evaluations by information recipients. Financial disclosures that contain a higher frequency of specific and identifiable named entities therefore convey more detailed explanations, enhance information transparency, and reduce information asymmetry (Paivio, 2013).

By employing named entity recognition techniques, key information embedded in narrative financial disclosures, such as management reports, notes to the financial statements, and management discussion and analysis

(MD& A), can be automatically extracted and subsequently utilized in qualitative analyses or in mixed-methods research combining quantitative and qualitative approaches.

Literature Review

Zhao et al. (2022) employ NER to measure the degree of linguistic specificity in corporate disclosures. They find that the presence of concrete and identifiable named entities in firms' annual reports is negatively associated with stock price synchronicity. Specifically, the disclosure of well-defined items, such as customer names, suppliers, currency units, and specific dates, provides firm-specific information to shareholders, thereby reducing the dependence of firm-level stock returns on market-wide returns and, consequently, lowering stock price synchronicity.

Similarly, He et al. (2024) apply named entity recognition techniques to examine how the disclosure of customer-specific risks affects initial public offering (IPO) approval outcomes. By identifying customer names in the risk disclosure sections of corporate reports, they show that firms that actively disclose customer-specific risk information are more likely to obtain IPO approval. This empirical evidence serves as a catalyst for encouraging firms to provide more granular customer-related disclosures and to reduce information asymmetry in the IPO market.

Given the rapid expansion of narrative sections in financial reports, the growing importance of concrete and specific information embedded in these disclosures, and the challenges associated with extracting such key information, the present study aims to introduce and elaborate on a novel methodological approach to address these issues. To the best of our knowledge, this method has not yet been applied to the analysis of corporate financial disclosures in the Iranian context. As such, it offers a valuable framework for researchers to conduct more in-depth and multidimensional studies that integrate both quantitative and qualitative perspectives. Accordingly, this study first reviews the theoretical foundations underlying the importance of narrative financial disclosures, introduces the named entity recognition method, and discusses its applications in the textual analysis of financial reports. Subsequently, using data retrieved from the Scopus database and employing VOSviewer software, the study conducts a bibliometric analysis of the application of NER in finance and accounting research, with the objective of identifying the most influential articles, leading authors, and core keywords in this research stream. The insights derived from this analysis provide a useful reference for researchers seeking to advance future studies in this domain.

In contemporary economic and financial environments, information constitutes one of the most critical economic assets. The full benefits of information can be realized only when it is provided in a timely, relevant,

material, complete, accurate, and understandable manner for investors. Information is not limited to quantitative and financial data; rather, narrative and textual disclosures enable a more comprehensive understanding of economic events and, by conveying managerial perspectives, attitudes, and values, help mitigate information asymmetry between information providers and users (Clatworthy & Jones, 2003). Indeed, the disclosure of non-financial information is essential for reducing information asymmetry between management and stakeholders (Narayanan & Nanda, 2005).

Across countries, capital market regulators require firms to prepare reports that include narrative explanations in addition to quantitative information. However, the manner in which narrative disclosures are presented in these reports largely depends on managerial discretion. Unlike the numerical information reported in financial statements, there are limited formal guidelines governing the structure, style, and presentation of narrative disclosures. Consequently, relative to quantitative financial reporting, managers possess greater flexibility in shaping disclosure style, tone, and presentation in narrative sections, enabling them to more effectively pursue their objectives (Henry & Leone, 2016).

According to Accounting Standard No. 37 (Iranian Accounting Standards, 2018), general disclosure practices encompass the format and ordering of financial statements, terminology and descriptive items, parenthetical information, notes to the financial statements, supplementary tables and figures, audit report paragraphs, and the board of directors' activity report. The most relevant and material information should, to the greatest extent possible, be presented directly within the financial statements, and its placement solely in the accompanying notes or other supplementary reports is generally not considered justifiable. When a complete and concise description of financial statement line items cannot be adequately conveyed through titles alone, additional explanations may be provided parenthetically following the relevant captions.

In the accompanying notes, non-quantitative information is primarily disclosed as an integral component of the financial statements, including explanations, disclosures, and limitations related to the reported line items. The non-quantitative sections of financial reports, whether subject to mandatory or voluntary disclosure, contain important and decision-relevant details. For instance, the narrative sections of the management discussion and analysis (MD&A) provide key information regarding the nature of the firm's business, managerial objectives and strategies, resources and expenditures, contractual relationships, and risk exposures. These sections often include concrete and traceable information, such as the firm's headquarters, competitors, customers, suppliers of raw

materials, export destinations, governing regulations, and legal claims brought by or against the firm (Namazi et al; 2023).

Similarly, the notes to the financial statements disclose influential non-numerical information on both a mandatory and voluntary basis, including details concerning debtors and creditors, corporate investments, and related and unrelated parties. Identifying such information is of substantial importance due to its significant impact on corporate and investor decision-making and its role in reducing information asymmetry. Moreover, this information is objective, concrete, and verifiable, thereby enhancing the credibility and verifiability of financial reports.

The concreteness effect reflects the notion that concrete words, those referring to tangible entities in the external world, are recognized more rapidly and with greater accuracy than abstract words, which refer to mentally or linguistically constructed entities that cannot be directly perceived through the five senses (Ghanaei Chamanabad et al; 2021). Clear and concrete information improves the speed and accuracy of information processing for users and is also more easily retained than other forms of information. This phenomenon is commonly referred to in psycholinguistics as the concreteness effect (Wang & Yao, 2012). To explain this effect, Paivio (1991) proposed the dual-coding theory, which accounts for why the processing of concrete information, such as words referring to specific entities or proper nouns (e.g; names of individuals or locations), exhibits superior performance in terms of recognition, comprehension, and recall compared to abstract expressions. Dual-coding theory posits the existence of two distinct yet interconnected cognitive processing systems. Abstract information primarily activates the verbal coding system of the left hemisphere of the brain, whereas concrete information simultaneously engages both the verbal coding system of the left hemisphere and the non-verbal coding system of the right hemisphere. Each system can be activated independently, yet they can also operate in parallel. Consequently, from a theoretical perspective, the processing speed of concrete information through dual coding is approximately twice that of abstract information processed through a single coding system (Zhao et al; 2022).

Recent studies examining the economic consequences of linguistic concreteness in corporate disclosures document a positive association between concrete, specific information and market reactions. Elliott et al. (2015) provide empirical evidence that the disclosure of concrete information can effectively reduce the psychological distance between investors and publicly listed firms. As a result, investors exhibit a greater willingness to invest in firms whose initial public offering (IPO) prospectuses exhibit higher linguistic clarity. This study represents the first

empirical investigation to document the effect of linguistic clarity on investors' judgments and decision-making.

Following Elliott et al. (2015), Hope et al (2016) examine stock returns over three-day event windows and document positive market reactions to the disclosure of proper nouns. Similarly, Pan et al. (2018), using quarterly earnings announcements as an initial empirical setting, find that greater linguistic concreteness in managerial disclosures leads to more favorable investor reactions and strengthens the reliability of the findings reported by Hope et al. (2016).

In the Iranian context, Nazari et al. (2016) investigate the effect of linguistic structure on investors' investment propensity and argue that the linguistic structure of financial reports constitutes one of the most important linguistic determinants influencing the presentation of concrete versus abstract information. They further show that when linguistic structure is aligned with informational content, it significantly affects investors' willingness to invest.

A review of the existing literature indicates that the extraction of concrete and specific named information plays a crucial role in finance and accounting research; however, a major challenge lies in accurately identifying and extracting such information. The vast majority of prior studies on the textual analysis of financial reports have relied on manual approaches, such as interviews and questionnaires, or on dictionary-based methods involving the identification and counting of specific words. These approaches are inherently time-consuming and prone to measurement error.

In recent years, artificial intelligence has substantially facilitated scientific research by enabling more efficient, accurate, and scalable empirical analyses. One of the key artificial intelligence–based approaches for textual content analysis is NLP. Human language is among the most complex and dynamic forms of communication, characterized by diverse syntactic, semantic, and pragmatic structures (Manning et al; 2008). The ability of machines to understand and process human language can enhance human–computer interaction, extract valuable insights from large-scale textual data, and enable the development of automated systems capable of performing complex linguistic tasks.

Natural language processing is an interdisciplinary field spanning computer science, artificial intelligence, and computational linguistics, with the objective of developing systems that can naturally understand, analyze, interpret, and generate human language (Jurafsky & Martin, 2023). Leveraging statistical methods, machine learning algorithms, and deep neural networks, NLP has achieved significant advances and has been applied across a wide range of domains, including machine translation,

sentiment analysis, intelligent assistants, and question-answering systems (Goldberg, 2017).

NER is a core task within NLP that focuses on identifying and classifying concrete and observable textual units, such as the names of individuals, locations, organizations, and other specific entities, within a text. The primary objective of NER is to structure unstructured textual data so that they can be effectively utilized in subsequent empirical analyses. The main function of NER is to identify and extract three broad categories of labeled information in texts, entities, temporal expressions, and numerical values, along with seven common subcategories, including persons, organizations, locations, currencies, time expressions, dates, and percentages. By systematically extracting such key information, NER can provide valuable insights to users and substantially enhance transparency and information dissemination. Within the framework of financial theories, disclosure quality and information transparency are regarded as key determinants in reducing the cost of capital, enhancing corporate credibility among investors, and improving economic decision-making. Given that a substantial portion of firms' value-relevant information, such as the names of customers, export destination countries, suppliers, and trading counterparties, is disclosed in the notes to the financial statements and other narrative reports, the application of NER enables the automated extraction of such information and its integration with quantitative financial data. This capability allows analysts and researchers to examine financial reports in a systematic and replicable manner with respect to linguistic clarity and conceptual precision.

Such an approach offers a significant advantage in comparative analyses of corporate financial reports, particularly in studies related to transparency rankings or the assessment of information risk. For example, a disclosure that merely refers to "a major customer" lacks identifiable named entities from the perspective of NER, whereas a report that explicitly discloses the customer's name or the counterparty country exhibits a higher level of concreteness. This distinction can serve as a foundation for constructing indices to evaluate disclosure quality in empirical accounting research. From this perspective, named entity recognition can be employed not only as a data extraction tool but also as a content analysis–based metric for assessing the objectivity and informational quality of narrative disclosures.

The NER process typically consists of the following stages

First, text preprocessing is conducted, which includes noise removal (e.g; eliminating unnecessary characters), tokenization, defined as segmenting the text into words or sentences, and part-of-speech tagging to identify the syntactic roles of tokens (Jurafsky & Martin, 2023).

Second, the entity recognition stage comprises two main steps. (a) Entity extraction, which involves identifying tokens that potentially refer to specific entities, and (b) entity classification, which assigns each extracted entity to a predefined category, such as persons, locations, or temporal expressions.

Third, NER relies on various modeling approaches and tools. These range from traditional statistical models, such as Hidden Markov Models (HMMs) and Conditional Random Fields (CRFs), to deep learning–based methods, including recurrent neural networks (RNNs) and attention-based architectures, most notably Bidirectional Encoder Representations from Transformers (BERT) (Lample et al; 2016). It is worth noting, however, that these advanced models are predominantly employed in computer science research. In finance and accounting applications, the primary focus is typically placed on the first two stages, entity extraction and classification, and the subsequent use of these outputs in empirical analyses. Accordingly, the extraction and categorization of named entities can be effectively implemented using more accessible artificial intelligence tools, such as Python-based NLP libraries, which are sufficient for the purposes of empirical finance and accounting research.

NER is a relatively nascent technique in accounting and finance research and, compared with more traditional empirical methods, has received limited scholarly attention due to the substantial methodological and implementation challenges it entails. Outside Iran, only a small number of studies in the fields of finance and accounting have employed NER-based approaches, while to the best of our knowledge, no prior study has applied this technique in the Iranian accounting and finance literature.

Among the notable international studies, El-Haj et al. (2019) examine the evolution of accounting and finance research by employing methods from computational linguistics (CL) to analyze financial content. Computational linguistics is an interdisciplinary field that integrates linguistics and computer science to study and process natural language using computational techniques. This field focuses on developing algorithms and models capable of analyzing, understanding, and generating human language.

The key findings of El-Haj et al. (2019) can be summarized as follows. First, accounting and finance research has generally lagged behind other disciplines in adopting computational linguistics methods. Second, the costs and challenges associated with implementation may outweigh the perceived benefits of computational linguistics approaches. Third, firm-level structural constraints can limit the practical relevance and applicability of such methods. Despite these challenges, the authors identify four computational linguistics tools that remain largely underutilized in mainstream accounting and finance research but offer

promising avenues for enhancing the analysis of financial disclosures. These tools include Named Entity Recognition, text summarization, semantic analysis, and corpus linguistics.

Zhao et al. (2022) examine the relationship between linguistic specificity and stock price synchronicity in a study titled *"Linguistic Specificity and Stock Price Synchronicity."* The authors employ NER to quantify the degree of linguistic specificity in corporate disclosures. Integrating the psycholinguistic compounding effect theory with asset pricing theory, they document a negative association between the presence of concrete and specific named entities in firms' annual reports and stock price synchronicity. Their findings suggest that a higher level of specific and concrete information enhances firm-specific information in prices, thereby reducing stock price synchronicity.

Similarly, He et al. (2024) utilize Named Entity Recognition techniques to investigate how the disclosure of customer-specific risks affects initial public offering (IPO) approval outcomes. By identifying customer names in the risk factor sections of corporate filings, they provide evidence from the Chinese market that firms actively disclosing customer-specific risk information are more likely to obtain IPO approval.

In Iran, a substantial body of research has examined the effects of linguistic characteristics, such as clarity, readability, and tone of financial reports, on investors' decision-making. However, to date, no study in the Iranian context has employed the NER approach.

For instance, Rahnamay Roudposhti et al. (2012), in a study titled *"Evaluating the Impact of Linguistic Judgmental and Cognitive Approaches in Accounting Explanatory Reports,"* analyze how different linguistic approaches affect investors' judgmental and cognitive behaviors in Iran. Their findings indicate that: (1) linguistic concreteness alone is not sufficient to ensure clarity; (2) increased linguistic clarity, regardless of linguistic structure or tone, has a significant effect on investors' judgmental and cognitive behaviors; (3) when concrete information is presented in an implicit manner, it does not exert a significant effect on investors' judgments, irrespective of tone; and (4) enhanced attention-capturing capacity in explicit disclosures is significantly associated with investment decisions.

Similarly, Mehrani & Nohnehal (2013), in their study entitled *"Testing the Effect of Linguistic Judgmental Approaches on Reducing the Audit Expectation Gap,"* examine the impact of linguistic clarity in audit reports on mitigating the audit expectation gap. Drawing on recent insights from psychology, linguistics, and auditing to conceptualize linguistic clarity, their results show that greater clarity in audit reports significantly reduces the audit expectation gap. They further suggest that adopting this approach in real-world auditing environments can meaningfully narrow this gap.

It is worth noting that, although the NER technique has not yet been applied to the analysis and processing of financial texts in Iran, extensive research has been conducted in other disciplines, particularly computer science and computer engineering, on implementing NER for the Persian language.

For example, Memtazi & Torabi (2019), in their study entitled *"Named Entity Recognition In Persian Texts Using Deep Learning,"* propose an NER system by employing advanced techniques such as dual semantic word representations based on both word-level and character-level embeddings within neural network architectures. Using deep learning approaches, they develop a robust framework for identifying named entities In Persian texts.

In addition, Naghavi et al. (2024) conduct a study titled *"Named Entity Recognition Using Deep Learning and a Reinforcement-Based Approach."* In this research, the authors design a proposed method based on a semantic language model for Persian combined with deep neural networks, which reduces dependence on large labeled datasets through a repetition-based logic.

Similarly, Masoud et al. (2023), in their study *"Improving Named Entity Recognition In Persian Using an Attention Mechanism,"* employ transfer learning for named entity detection. The main advantage of transfer learning becomes particularly evident when only a small dataset is available in the target research domain. This study is among the first to integrate a BERT-based language model with Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and an attention mechanism that focuses on neighboring words before and after the target token. The results indicate that the simultaneous use of these components, along with the core attention mechanism, can substantially improve system performance In Persian NER tasks.

Research Questions*:*

NER is a relatively novel approach in the textual content analysis of financial reports, and to date, this method has not been applied to the analysis of financial reporting texts in Iran. Accordingly, the present study aims to introduce this technique to researchers in the fields of finance and accounting, with particular emphasis on scholars in Iran. Achieving this objective requires familiarizing researchers with the existing body of work in this area across multiple dimensions.

Bibliometric analysis provides a comprehensive overview of a research domain and is particularly useful for identifying influential studies and tracing major research trends over time (Guerola-Navarro et al; 2020). Therefore, this study employs bibliometric methods to present an integrated picture of the international literature related to the application of

NER in finance and accounting research. In essence, bibliometric analysis serves as a systematic tool for addressing the following research questions, the answers to which facilitate the identification of dominant trends, key contributors, and influential concepts in the field:

1. What has been the trend in the use of Named Entity Recognition methods in finance and accounting research in recent years?
2. How has the trajectory of NER adoption in finance and accounting research evolved over time, and what factors have driven these changes?
3. Which articles are the most highly cited and most relevant within this research domain?
4. What are the most important and up-to-date keywords characterizing this field of study?
5. Who are the most influential authors, and from which countries do they originate?

In the following sections, the study seeks to enhance the reader's understanding of the importance of adopting NER techniques among finance and accounting scholars who have already been introduced, through the theoretical framework, to the concept of NER and its application in the textual analysis of financial reports. Furthermore, by identifying the most highly cited studies, the core keywords, and the most influential authors and countries in this domain, the study provides a structured foundation that enables researchers to more effectively employ this technique in future empirical research.

Research Methodology

This study adopts a descriptive–analytical research design and employs bibliometric analysis to identify, integrate, and analyze the structural characteristics of knowledge in the field. Based on the collected data, scientific knowledge maps were constructed to visualize networks defined by the structure of the literature and the relationships among its constituent elements.

To examine the bibliometric landscape of NER research in accounting and finance, data were retrieved from the Scopus database. Records were downloaded and stored in Excel format for subsequent processing and analysis. In addition to Scopus, searches were also conducted in other databases, such as Web of Science. However, given the novelty of the research topic, the number of relevant articles in these databases was substantially smaller, and most of the identified studies overlapped with those indexed in Scopus. Consequently, Scopus was selected as the sole data source for this study. After applying the required filters, the final sample comprised 112 articles.

According to the bibliometric literature, a minimum of approximately 200 sources is generally recommended for robust bibliometric analysis; however, for emerging research areas, a smaller number of publications may be sufficient (Rogers et al; 2020). Given the exploratory nature and novelty of NER applications in accounting and finance, the available sample size is deemed appropriate for the objectives of this study.

In line with the research questions, structured bibliographic information available in Scopus was used to identify research trends, highly cited articles, and the most relevant studies, which were then filtered and analyzed accordingly. The search query included the term "Named Entity Recognition" within the Title, Abstract, and Keywords fields. To further align the sample with the study's objectives, the following filters were applied: research subject areas of Business, Management and Accounting and Economics, Econometrics and Finance; document types limited to research articles and review papers published in peer-reviewed journals; and language restricted to English.

Although the primary focus of the study is on the accounting and finance disciplines, the search scope was expanded to include Business, Management and Accounting and Economics and Finance subject areas due to the limited number of specialized NER studies in accounting and finance. This broader scope ensured a sufficiently large dataset to conduct meaningful bibliometric analysis while remaining conceptually aligned with the research objectives. It should be noted that the relatively small number of available articles may create challenges in constructing bibliometric networks, such as co-citation and keyword co-occurrence maps, and may yield clusters that do not fully reflect the underlying intellectual structure of the field. Given the emerging nature of the research topic, this limitation is largely unavoidable. To identify the initial emergence of research in this area, no temporal restriction was imposed at the outset; the earliest publications were found to date back to 2007.

For clustering and visualization, the complete bibliographic records extracted from Scopus in CSV format were analyzed using VOSviewer (version 1.6.17). This software generates bibliometric maps that highlight different dimensions of the scientific literature and employs a unified approach to mapping and clustering based on normalized co-occurrence matrices of terms and similarity measures that quantify the strength of relationships among terms. Closely related terms are grouped into the same cluster, with each cluster represented by a distinct color. The spatial proximity of terms reflects their contextual similarity within the literature.

In addition, VOSviewer differentiates keywords according to their temporal occurrence using color gradients and represents their frequency through font size and bounding rectangles, where larger fonts and shapes indicate more frequently occurring terms. The software also enables the

visualization of keyword networks, country collaboration networks, and author co-authorship networks (Darvish et al; 2018).

Research Findings

Publication Dynamics and Trends

To examine the dynamics and trends in publication activity and to address the first and second research questions, an initial search was conducted in the Scopus database using the term *"Named Entity Recognition."* To align with the objectives of the study, the following filters were applied: subject areas "Business, Management and Accounting" and "Economics, Econometrics and Finance"; document types "Articles" and "Review Articles" published in academic journals; and language English. Although the primary focus of this study is on accounting and finance, the broader subject areas of business, management, and economics were selected due to the novelty of linguistic clarity and named entity recognition in accounting and finance and the limited number of field-specific studies, thereby enabling a meaningful bibliometric analysis.

After applying these filters, a total of 112 articles published between 2007 and 2024 were identified. As illustrated in Figure (1), extracted from Scopus, the publication trend exhibits a clear upward trajectory. It is worth noting that although the thematic scope extends beyond accounting and finance, the application of named entity recognition within business and economic research is largely aligned with issues of information transparency, disclosure quality, and evaluative analysis.

The Figure (1) indicates that 2007 marks the initial use of named entity recognition in the selected research domains. Publication activity remained relatively modest until 2018; however, a pronounced surge is observed beginning in 2019, followed by a steep upward trend from 2019 to 2024, reflecting growing scholarly interest in the application of named entity recognition.

The substantial increase in studies employing named entity recognition since 2019 can be largely attributed to a paradigmatic shift in deep learning–based language models, particularly the introduction of BERT (Bidirectional Encoder Representations from Transformers) by Google in late 2018. BERT represents one of the most powerful text analysis tools in artificial intelligence and natural language processing, as it is designed to capture the precise meaning of words within the full contextual structure of sentences or documents. Unlike traditional models that process text unidirectionally, BERT simultaneously considers both preceding and succeeding words, closely resembling human language comprehension.

Consequently, from 2019 onward, researchers across disciplines, including economics, accounting, and finance, have been able to apply named entity recognition with greater accuracy and efficiency to extract information from financial texts, annual reports, and corporate disclosures. Moreover, the growing demand for data-driven analyses and transparency-oriented research in the dynamic post–COVID-19 economic environment has played a significant role in accelerating the adoption of this methodology.

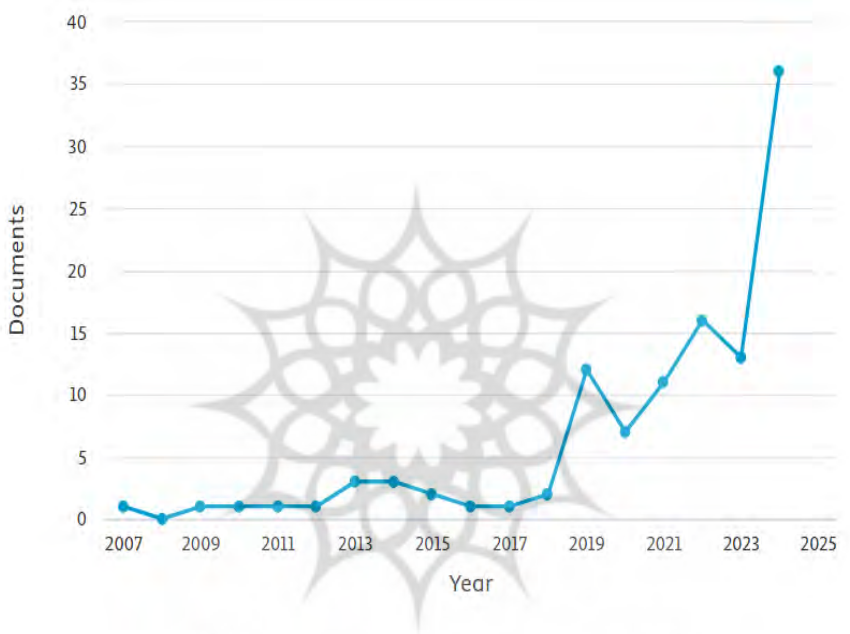

Figure 1. Publication Trend

The publication dynamics of the identified studies over the period 2007–2024 indicate a clear and substantial growth trend. During the early years, from 2007 to 2018, the annual number of publications remained very limited, ranging from one to three articles per year. In contrast, the volume of publications increased sharply in subsequent years, reaching its peak in 2024 with a total of 36 articles. Figure (2) illustrates the annual distribution of published articles.

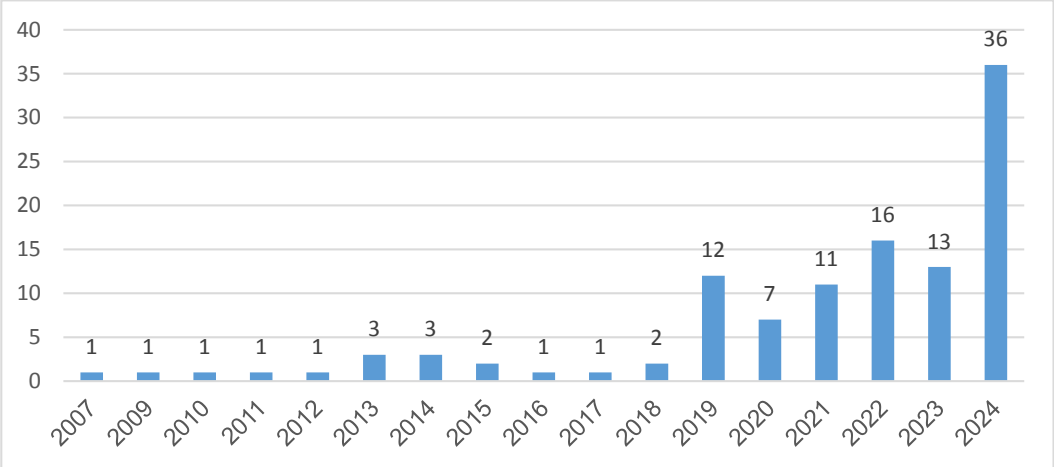

Figure 2. Number of Publications Per Year

Figure (3) presents the annual number of citations received by the identified publications. The results indicate that the highest citation counts correspond to the years 2019 and 2021, with 492 and 201 citations, respectively.

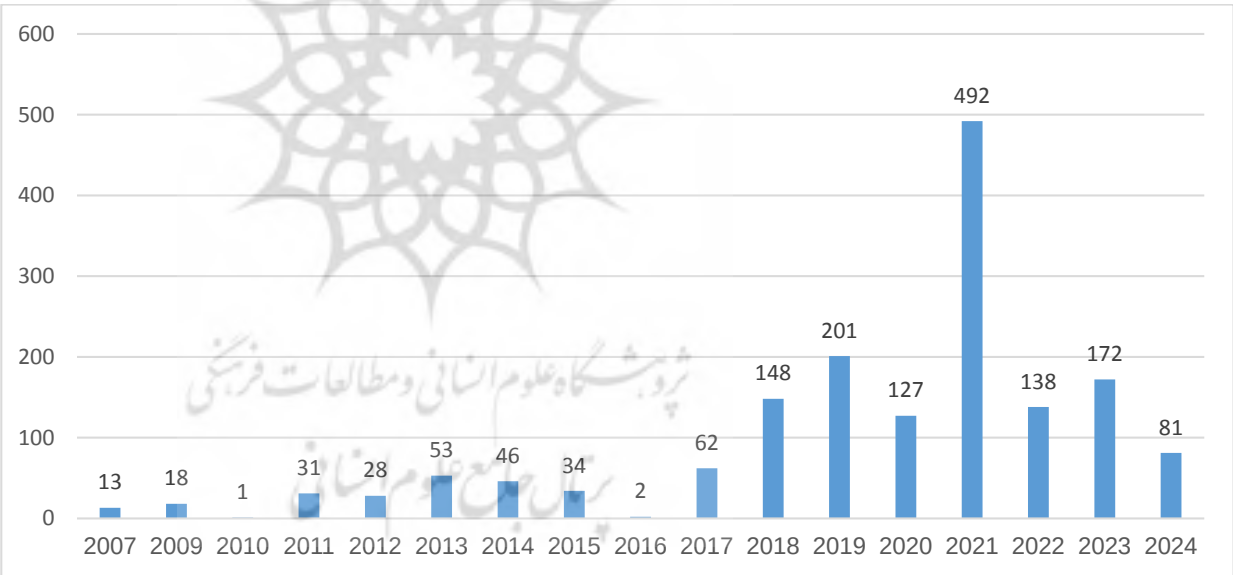

Figure 3. Number of Citation Per Year

Figures (1) and (2) are presented in response to the third research question. Table (1) displays the top ten most cited articles among the 112

publications retrieved from the Scopus database. According to this ranking, the most highly cited article is authored by Lin, Shaw, Dzinory, and Yoon (2021) and published in the journal *Knowledge-Based Systems*, with a total of 195 citations.

It is noteworthy that the third-ranked article in the list is a fully specialized study in finance and accounting, directly related to financial disclosure. This article, authored by El-Haj et al. (2019) and published in the *Journal of Business Finance & Accounting*, has received 86 citations.

The high citation counts of certain articles cannot be attributed solely to the use of the named entity recognition methodology. Rather, factors such as scientific quality, methodological innovation, and the outlet of publication, particularly the journal's ranking and impact, play a critical role in determining citation frequency. Articles published in high–impact factor journals or those indexed in Q1 quartiles (top-tier journals within a given research field) are generally more visible and more likely to be cited by subsequent studies. For instance, the article by Lin et al. (2021), which exhibits the highest citation count among the retrieved publications, was published in *Knowledge-Based Systems*, a well-established Q1 journal in the field of intelligent systems.

Table 1: Top Ten Most Cited Articles

Citations	Authors	Article title
195	Lin, J.C.-W; Shao, Y; Djenouri, Y; Yun, U.	I. ASRNN: A recurrent neural network with an attention model for sequence labeling
108	Middleton, S.E; Kordopatis-Zilos, G; Papadopoulos, S; Kompatsiaris, Y.	II. Location extraction from social media: Geoparsing, location disambiguation, and geotagging
86	El-Haj, M; Rayson, P; Walker, M; Young, S; Simaki, V.	III. In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse
65	Moon, S; Lee, G; Chi, S; Oh, H.	IV. Automated Construction Specification Review with Named Entity Recognition Using

Citations	Authors	Article title
		Natural Language Processing
65	Wang, J; Xu, W; Fu, X; Xu, G; Wu, Y.	V. ASTRAL: Adversarial Trained LSTM-CNN for Named Entity Recognition
64	Sarhan, I; Spruit, M.	VI. Open-CyKG: An Open Cyber Threat Intelligence Knowledge Graph
62	Tran, V.C; Nguyen, N.T; Fujita, H; Hoang, D.T; Hwang, D.	VII. A combination of active learning and self-learning for named entity recognition on Twitter using conditional random fields
56	Alkhatib, M; El Barachi, M; Shaalan, K.	VIII. An Arabic social media based framework for incidents and events monitoring in smart cities
52	Catelli, R; Casola, V; De Pietro, G; Fujita, H; Esposito, M.	IX. Combining contextualized word representation and sub-document level analysis through Bi-LSTM+CRF architecture for clinical de-identification
42	Kalyanathaya, K.P; Akila, D; Rajesh, P.	X. Advances in natural language processing –a survey of current research trends, development tools and industry applications

It is also worth noting that the most relevant articles applying named entity recognition in finance and accounting research are presented in Table (2). To identify these studies, the terms *"Named Entity Recognition" AND "Finance" AND "Accounting"* were searched in Scopus using the same set of filters described earlier. The resulting articles are all recent publications, spanning the period 2019–2024, and are ISI-indexed journal articles.

The starting point of this research stream coincides with the introduction of deep learning–based language models, most notably BERT, which was released by Google in late 2018. These models enabled more accurate and efficient analysis of unstructured textual data, thereby accelerating the adoption of named entity recognition methods across various disciplines, including accounting and finance.

The most recent study in this stream is by He et al. (2024), entitled *"Customer-Specific Risk Disclosure and IPO Acceptance Rates: Machine Learning–Based Evidence from Textual Analysis."* This study employs named entity recognition techniques and advanced text analysis methods to examine how the disclosure of customer-specific risks, identified through the extraction of customer names from the risk sections of corporate reports, affects IPO acceptance rates. Empirical evidence from the Chinese market indicates that firms that actively disclose specific customer risk information are more likely to obtain IPO approval.

Table 2. The Most Relevant Articles in Finance and Accounting

Publication year	journal title	Article title
2024	Applied Economics Letters	I. Specific customer risk disclosure and IPO approval rate: evidence based on machine learning and text analysis
2024	Journal of Decision Systems	II. Business insights using RAG–LLMs: a review and case study
2023	Management Communication Quarterly	III. Key Players in Corporate Social Responsibility (CSR) Institutionalization: An Analysis of Multinational Companies' Interorganizational Positioning via CSR Reports
2022	China Journal of Accounting Research	IV. Linguistic specificity and stock price synchronicity
2022	Journal of International Money and Finance	V. Media sentiment on monetary policy: Determinants and relevance for inflation expectations

Publication year	journal title	Article title
2019	Journal of Business Finance and Accounting	VI. In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse
2019	International Journal of Recent Technology and Engineering	VII. Advances in natural language processing –a survey of current research trends, development tools and industry applications

Visualization of Results

To address the fourth and fifth research questions, the study proceeds with a visualization of the bibliometric results. It should be noted that one common bibliometric visualization technique, article co-citation network analysis, could not be implemented due to the relatively limited number of available studies in this research domain. In addition, as the objective is to directly present the graphical outputs generated by the VOSviewer software, and since this software does not support the Persian language, all visualized results are displayed in English and cannot be translated into Persian.

Keyword Co-occurrence and Evolutionary Trends

The keywords used by authors in academic publications are of critical importance, as they represent the core concepts through which researchers communicate the essence of their work to the scholarly community. In the corpus analyzed in this study, authors employed a total of 403 distinct keywords, of which 40 keywords occurred at least twice.

As shown in Figure (4), which is generated using VOSviewer, the size of each node reflects the frequency of occurrence of a given keyword in the scientific literature; larger nodes indicate more frequently used keywords. The most frequently occurring keywords, in descending order, include named entity recognition, natural language processing, deep learning, information extraction, relation extraction, knowledge extraction, text mining, and machine learning.

Ultimately, eight thematic clusters were identified, as illustrated in Figure (4). The different colors represent distinct thematic clusters, while the links between keywords indicate that these terms frequently co-occur within the same publications, highlighting the intellectual structure and thematic interconnections of the research field.

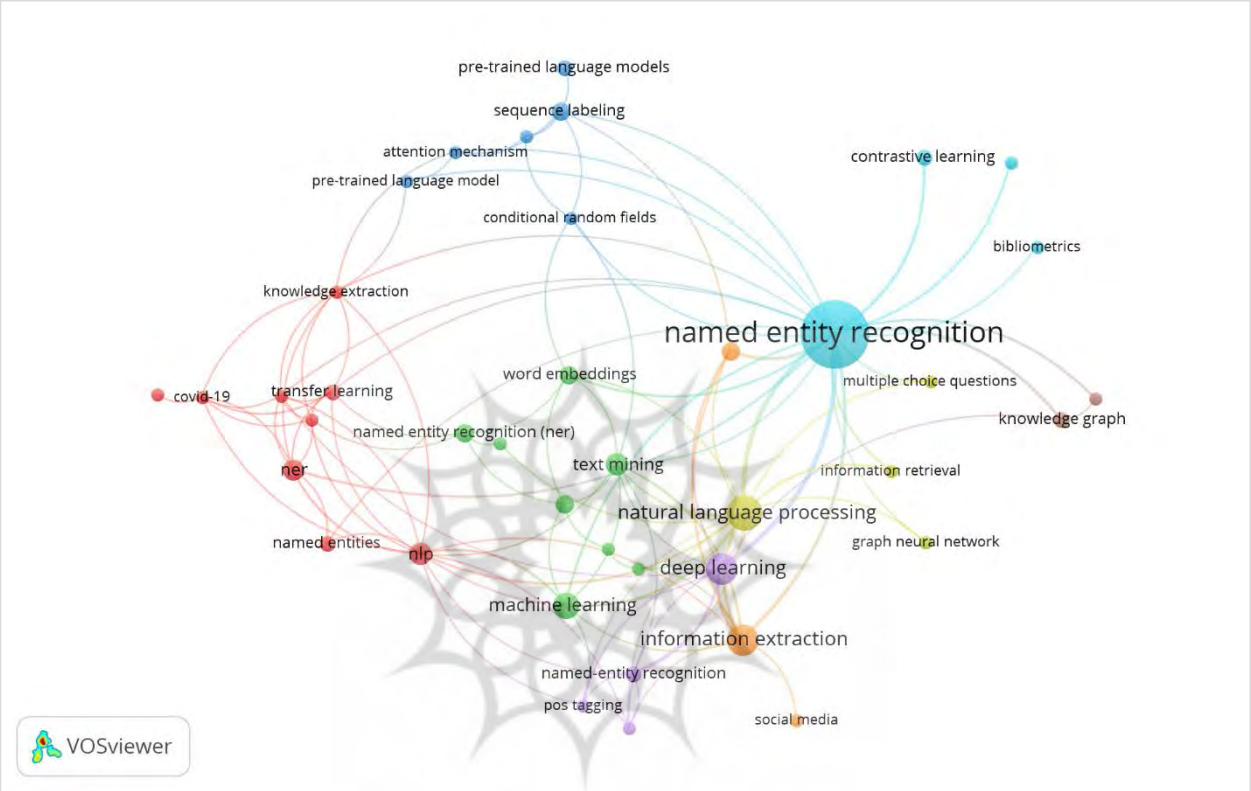

Figure 4. Keyword Co-occurrence and Evolutionary Trends

The keyword co-occurrence map presented in Figure (4) indicates that NER constitutes the central research focus in this literature stream and is closely interconnected with concepts such as machine learning, deep learning, and natural language processing. These concepts provide powerful analytical tools for extracting structured information from unstructured textual data, a capability that is particularly valuable in finance and accounting, especially for the analysis of notes to the

financial statements, management reports, and financial disclosure documents.

The linkage of these keywords with concepts such as knowledge extraction, knowledge graphs, and information retrieval reflects ongoing efforts to systematize and visualize latent relationships among financial and economic concepts. In practice, these tools facilitate the identification of hidden patterns, the recognition of key individuals or entities, and the reduction of information asymmetry in financial markets.

Moreover, the presence of advanced methodologies such as pre-trained language models, sequence labeling, and contrastive learning highlights a clear shift toward more accurate and efficient algorithms for automated information extraction. Overall, this keyword-based analysis demonstrates that artificial intelligence–driven techniques have opened a new and rapidly expanding avenue for the qualitative analysis of financial data, with significant implications for more informed and precise financial decision-making.

The historical co-occurrence analysis of keywords can also be instrumental in identifying the evolution and development of intellectual trends within the research domain. Figure (5) illustrates the historical trajectory of keywords. The color of the rectangles represents the publication year of the articles associated with each keyword.

Blue-colored keywords correspond to earlier publications (2018–2019). These include terms such as conditional random fields, word embeddings, and text mining, which reflect more traditional NER approaches. The keyword COVID-19 also appears in this period, likely reflecting research related to the pandemic that emerged in 2019.

Green-colored keywords (2020–2022) represent more advanced methods, including machine learning, deep learning, transfer learning, knowledge extraction, and information extraction.

Yellow-colored keywords (2023–2024) indicate the most recent developments in the field, such as graph neural networks and contrastive learning. Overall, as shown in Figure (5), the period from 2020 to 2024 reflects the emergence of more deep learning–based NER techniques,

facilitating the extraction of knowledge and structured information from unstructured financial texts.

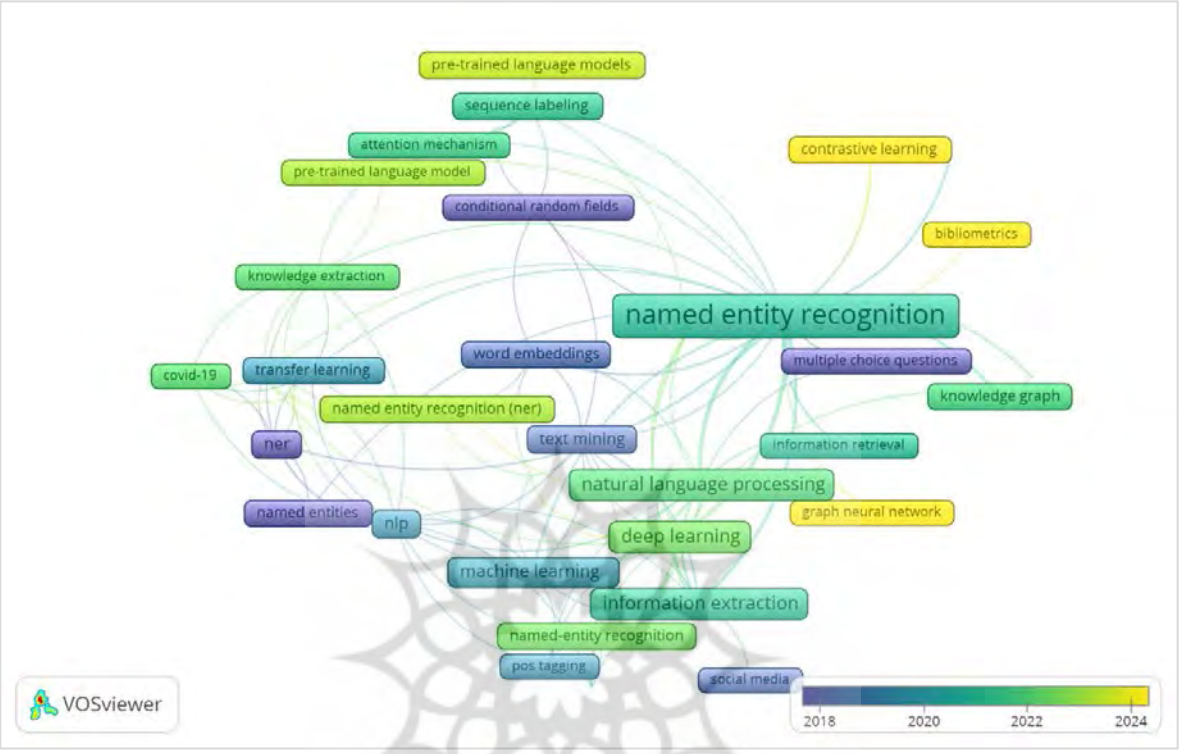

Figure 5. The Historical Trajectory of Keywords

Top authors and countries an analysis of the authors of articles in the business and economics domain employing NER reveals that, among 414 authors, only six have authored the maximum number of publications in this field, which is three articles each. Table (3) presents the names of these authors along with the number of articles they have published and the citation counts of their respective publications.

Table 3. Authors and the Number of Their Articles

Citations	Number of Paper	Author
93	3	Chiarello,filippo
93	3	Fantoni, gualtier
93	3	Giordano, vito

Citations	Number of Paper	Author
50	3	Liu, jie
50	3	Wan, qian
52	3	Saha,sujan,kumar

Subsequently, a co-authorship analysis of all authors in this field was conducted, and the resulting collaboration network is presented in Figure (6), extracted using VOSviewer. Figure (6) illustrates the strength of collaboration and connectivity among the authors, indicating a promising outlook for the future development of research in this domain.

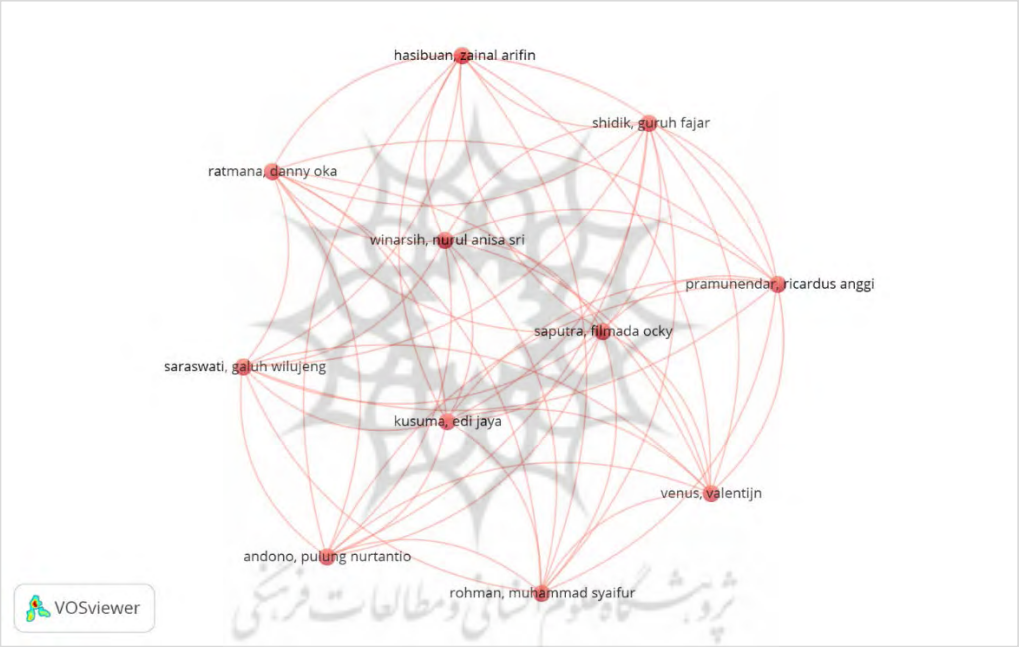

Figure 6. Collaboration and Connectivity Among the Authors

In addition to the authors, it is also important to identify the leading and influential countries in this field. According to the Scopus data, a total of 38 countries contributed to the retrieved scientific output. Among these, 12 countries had at least three publications. China ranks first with 41 articles,

followed by India with 20 articles and Italy with 9 articles, as summarized in Table (4).

Table 4. The Countries and Number of Articles

Citations	Number of Article	Country
529	41	china
162	20	india
166	9	italy
221	6	United Kingdom
141	6	spain
331	6	South Korea
71	6	Singapore
53	5	france
36	4	United States
114	3	Japan
62	3	Australia
16	3	Indonesia

Figure (7) illustrates the co-authorship network of the countries active in this field. China appears as the largest and most prominent node, reflecting the significance, volume of research, and its interactions with other countries in this domain. Following China, Italy, Spain, the United States, South Korea, and Singapore demonstrate the highest levels of international collaboration.

Figure 7. The Co-Authorship Network of the Countries

Research Gap

The aim of any research is to contribute to addressing the existing gaps within its field, and the present study is no exception. One notable research gap in accounting and finance, which has received growing attention over recent decades, is the limited focus on qualitative data and the integration of quantitative and qualitative information in financial studies. The NER technique, by identifying key information, offers a methodological avenue to analyze qualitative aspects of financial reports, thereby helping to bridge this gap.

In Iran, textual characteristics such as tone and readability are typically measured at the aggregate level of the entire text; however, no prior study has examined specific names and entities within financial disclosures. The NER approach can serve as an effective tool for identifying and quantifying such entities for targeted research purposes.

Another important research gap is the absence of a systematic method for extracting key informational features from financial reports. The output of NER provides the frequency of occurrence of specific entities, such as the names of organizations, countries, and products, which can reflect the relevance and economic importance of these entities for a particular firm or industry. For example, a high frequency of a product name identified through NER may indicate that the product is a firm's core or strategic offering. Similarly, in the banking industry, the close relationship between commercial banks and the central bank can be empirically inferred from the frequency with which the central bank is mentioned in financial disclosures, as captured by NER outputs.

Technical Challenges

These challenges include semantic ambiguity in financial texts, insufficient and imbalanced training data, and the limited adaptability of NER models to domain-specific financial language. The absence of a standardized framework for applying NER in accounting and finance has led firms and financial institutions to adopt heterogeneous information-extraction approaches, and a generally accepted methodological standard for the use of NER in this field has yet to emerge.

Another major challenge concerns the processing of multilingual financial data. Most existing NER models are trained primarily on English-language corpora, whereas in global capital markets, financial reports are disclosed in multiple languages. This mismatch limits the generalizability

and cross-country applicability of NER-based analyses in international accounting and finance research.

Organizational and Economic Challenge

In addition to the aforementioned issues, organizational and economic challenges are also highly salient. For instance, many firms are reluctant to fully disclose their financial information to the public, which constrains the accuracy and practical applicability of NER techniques. Moreover, the development and implementation of NER models require substantial financial resources and robust computational infrastructure, which may not be cost-effective or feasible for some firms.

Discussion and Conclusion

This study seeks to introduce NER to researchers in the fields of accounting and finance, given the growing emphasis in recent research on non-numerical (textual) data in financial reports and the inherent challenges associated with extracting and analyzing such information. At the same time, researchers are increasingly required to replace traditional analytical approaches with artificial intelligence–based methods in order to adopt state-of-the-art techniques and achieve more robust and reliable findings. Accordingly, a systematic review of the NER literature, combined with a bibliometric (scientometric) analysis of this technique, can substantially assist researchers in conducting more accurate and rigorous analyses of financial reporting texts.

The findings indicate that the application of NER exhibited no substantial growth between 2007 and 2018. However, a marked increase emerged in 2019, followed by a consistently upward trend from 2019 to 2024. Specifically, during the period 2007–2018, only one to three articles were published annually, whereas 12 articles were published in 2019, and the number rose to 36 articles in 2024.

A critical factor underlying the sharp increase in NER-based studies since 2019 is the introduction of deep learning–based language models, such as BERT, which were released in late 2018. The advent of these models enabled researchers to employ NER with substantially greater accuracy and efficiency for extracting information from financial texts, including annual reports, corporate disclosures, and related documents. Consequently, the results show that the most relevant and application-oriented studies of NER in accounting and finance are concentrated in the 2019–2024 period, reflecting the pivotal role of advanced language models in advancing text-based financial research.

According to the visualization results generated by the VOSviewer software, the most frequently used keywords, in descending order, include Named Entity Recognition, Natural Language Processing, Deep Learning, Information Extraction, Relation Extraction, Knowledge Extraction, Text Mining, and Machine Learning. Named Entity Recognition emerges as the central research theme in this literature and is closely connected to concepts such as machine learning, deep learning, and natural language processing. These concepts provide powerful analytical tools for extracting structured information from unstructured text, a capability that is particularly valuable in finance and accounting research, especially for the analysis of notes to the financial statements, management reports, and corporate disclosure documents. The linkage of these keywords with concepts such as knowledge extraction, knowledge graphs, and information retrieval reflects a growing effort to systematize and visualize latent relationships among financial and economic concepts. In practice, these tools facilitate the identification of hidden patterns, the recognition of key individuals or entities, and the reduction of information asymmetry in financial markets. Moreover, the emergence of advanced techniques, such as pre-trained language models, sequence labeling, and contrastive learning, indicates a clear shift in the literature toward more accurate and efficient algorithms for automated information extraction. Overall, this keyword analysis suggests that artificial intelligence–based techniques have opened a novel and rapidly expanding pathway for the qualitative analysis of financial data, thereby providing a stronger foundation for more informed and precise financial decision-making. An examination of co-authorship patterns in this research stream reveals a high degree of collaboration and interconnectedness among scholars, highlighting the development of an active and cohesive research community. In terms of geographical distribution, China accounts for the largest share of publications in this field with 41 articles, followed by India and Italy, with 20 and 9 publications, respectively.

Based on the reported findings, researchers can identify the most influential and credible studies, the core and up-to-date keywords, as well as the leading authors and countries at the forefront of this research stream, and leverage these insights to inform and strengthen their future research agendas. Moreover, this study has practical implications for capital market regulators, auditors, and financial analysts, as it facilitates automated and more efficient processing of financial information, thereby reducing analytical costs and enhancing the timeliness and consistency of financial analysis.

Limitations and Suggestions for Future Research

1. The primary limitation of this study is the very limited number of accounting and finance studies that have applied NER to the analysis of financial texts. This limitation stems from the emerging nature of NER within the accounting and finance literature. To date, no empirical study in Iran has employed NER to analyze financial reports, which underscores the novelty of this approach in the local research context.

This constraint has influenced several aspects of the research design. Specifically, due to the scarcity of finance- and accounting-specific studies, the analysis was extended to broader subject categories, namely "Business, Management and Accounting" and "Economics and Finance", rather than being restricted solely to "Finance" or "Accounting." In addition, the limited volume of relevant publications necessitated reliance on the Scopus database only, as records from other databases were too sparse and inconsistent to support a robust bibliometric analysis. Nevertheless, the final Scopus output comprised 112 articles, which is below the commonly suggested minimum threshold of approximately 200 documents for robust bibliometric analysis. This limitation prevented the construction of certain networks, most notably co-citation networks, which require a sufficiently large corpus to yield stable and interpretable structures.

Moreover, analyses such as author co-authorship, country-level collaboration, and keyword co-occurrence rely on networks with an adequate number of nodes and links to generate meaningful and cohesive results. With a limited sample size, these networks may lack density, reducing their analytical precision and internal consistency. Similarly, keyword and thematic cluster analyses require sufficient data to properly distinguish clusters; when the number of observations is small, there is a higher risk of ambiguous or misclassified clusters and the inadvertent merging of conceptually distinct research streams.

In bibliometric studies based on small samples, highly cited papers tend to exert disproportionate influence, potentially biasing the results toward a small subset of influential studies. As a consequence, a single highly cited article may unduly shape the perceived structure and evolution of the entire research field, thereby distorting the overall representation of the literature.

2. In addition, to date no accounting or finance study in Iran has employed NER to extract information from financial texts, which limits its inclusion in the domestic research background and constrains bibliometric

analysis in this context. Accordingly, in light of the challenges and limitations discussed above, the following recommendations are proposed for future research, particularly for scholars in Iran:

- Conduct more comprehensive bibliometric analyses and construct more accurate bibliometric networks as the volume of research in this area increases over time.
- Apply Named Entity Recognition to identify and extract key information, such as customer names, suppliers, countries, and institutions, to support integrated quantitative–qualitative analyses.
- Integrate NER with sentiment analysis and models for forecasting financial trends.
- Examine the combined use of NER and other NLP techniques in capital market analysis.
- Develop a standardized framework for the application of NER in accounting research and financial disclosure analysis.
- Propose standard algorithms and tools for implementing NER in the analysis of financial reports and disclosures.
- Further investigate cost-reduction strategies for implementing NER in small and medium-sized enterprises (SMEs), and evaluate efficient and cost-effective models suitable for adoption by smaller financial institutions.
- Develop multilingual NER models for international financial analysis, capable of delivering robust performance across multiple languages (e.g; Chinese, Arabic, Persian, French, and others).

References

Clatworthy, M; & Jones, M. J. (2003). Financial reporting of good news and bad news: Evidence from accounting narratives. *Accounting and Business Research*, *33*(3), 171-185. https://doi.org/10.1080/00014788.2003.9729700

Darvish, A; Tabibi, S. J; Alborzi, M; & Radfar, R. (2018). A review of scientific production trends in the field of nursing information technology. *Nursing Management Quarterly*, *7*(1), 61-72. (In Persian)

Dyer, B; Lang, M; & Stice-Lawrence, L. (2017). The evolution of 10-K textual disclosure: Evidence from latent Dirichlet allocation. *Journal of Accounting and Economics*, *64*(2-3), 221–245. https://doi.org/10.1016/j.jacceco.2017.07.002

El-Haj, M; Rayson, P; Walker, M; Young, S; & Simaki, V. (2019). In search of meaning: Lessons, resources and next steps for computational analysis of financial discourse. *Journal of Business Finance & Accounting*, *46*(3-4), 265-306. https://doi.org/10.1111/jbfa.12217

Elliott, W.B; Rennekamp, K.M; & White, B.J. (2015). Does concrete language in disclosures increase willingness to invest?. *Rev Account Stud*, 20(2), 839–865. https://doi.org/10.1007/s11142-015-9330-x

Ghanaei Chamanabad, A; Golmakani, M. R; Bigdeli, I; & Ghoshuni, M. (2022). Concreteness and Emotionality Effects on Visual Word Processing Using Event-Related Potentials in the Persian Language. *Neuropsychology*, *8*(28), 9-20. https://doi.org/10.30473/clpsy.2021.57651.1610 (In Persian)

Goldberg, Y. (2017). *Neural Network Methods for Natural Language Processing.* Morgan & Claypool Publishers.

Guerola-Navarro, V; Oltra-Badenes, R; Gil-Gomez, H; & Gil-Gomez, J.-A. (2020). Customer relationship management (CRM): A bibliometric analysis. *International Journal of Services Operations and Informatics*, *10*(3), 242-268.

Henry, E; & Leone, A. J. (2016). Measuring qualitative disclosure: The case of business risk disclosure in the annual report. *Contemporary Accounting Research*, *33*(2), 711-746. https://doi.org/10.1111/1911-3846.12156

He, X; Zhang, Y; & Lin, P. (2024). Specific customer risk disclosure and IPO approval rate: Evidence based on machine learning and text analysis. *Applied Economics Letters*.1-9. https://doi.org/10.1080/13504851.2024.2240789

Hope, O.K; Hu, D; & Lu, H. (2016). The benefits of specific risk-factor disclosures. *Rev Account Stud*, *21*(4), 1005–1045.

Jurafsky, D; & Martin, J. H. (2023). *Speech and Language Processing* (3rd Ed.). Prentice Hall.

Lample, G; Ballesteros, M; Subramanian, S; Kawakami, K; & Dyer, C. (2016). *Neural architectures for named entity recognition.* Proceedings of NAACL-HLT 2016, 260-270.

Lin, J. C.-W; Shao, Y; Djenouri, Y; & Yun, U. (2021). ASRNN: A recurrent neural network with an attention model for sequence labeling. *Knowledge-Based Systems*, *212*, 1-11. https://doi.org/10.1016/j.knosys.2020.106624

Manning, C. D; Raghavan, P; & Schütze, H. (2008). *Introduction to Information Retrieval.* Cambridge University Press.

Masoud, P, Hoor Ali, M, & Lak, B. (2023). Improving named entity recognition In Persian using attention mechanisms. *Military Information and Communication Technology*, *4*(2), 1-15. (In Persian).

Mehrani, S; & Nonahal Nahr, A. A. (2013). Evaluating the Effect of the Judgmental Approaches of Language in Reducing Audit Expectation Gap. *Accounting and Auditing Review*, *20*(3), 123-144. (In Persian)

Momtazi, S; & Torabi, F. (2019). Named entity recognition In Persian texts using deep learning. *Signal and Data Processing*, *16*(4), 93-112. (In Persian)

Nagavi, M; Hassani-Ahangar, M; & Amiri Jozeh, A. (2024). Identifying named entities using deep learning and reinforcement approach. *Signal and Data Processing*, *21*(3), 97-110. (In Persian)

Namazi, M; & Khademi, S. (2023). Designing and empirically testing the "disclosure identification" model in the management commentary report in Iran. *Accounting and Auditing Reviews*, *30*(1), 149-190. (In Persian)

Narayanan, M. P; & Neves, M. A. (2005). Disclosure and market transparency: Theory and evidence. *Review of Financial Studies*, *18*(2), 789-829. https://doi.org/10.1093/rfs/hhi029

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology / Revue canadienne de psychologie*, *45*(3), 255–287. https://doi.org/10.1037/h0084294.23

Paivio, A. (2013). *Imagery and verbal processes*. Psychology Press.

Pan, L; McNamara, G; Lee, J. J; Haleblian, J; & Devers, C. E. (2018). Give it to us straight (most of the time): Top managers' use of concrete language and its effect on investor reactions. *Strategic Management Journal*, *39*(8), 2204-2225. https://doi.org/10.1002/smj.2784

Rahnamay Roodposhti, F; Nikoomaram, H; & Nonahal Nahr, A. A. (2012). Evaluating the Effects of Language Judgmental and Cognitive Approaches in Accounting Narratives. *Accounting and Auditing Review*, *19*(2), 47-72. (In Persian)

Rogers, G; Szomszor, M; & Adams, J. (2020). Sample size in bibliometric analysis. *Scientometrics*, *125*(1), 777-794.

Wang, L; & Yao, X. (2012). Effects of concreteness on language processing: A psycholinguistic perspective. *Journal of Psycholinguistic Research*, *41*(6), 457-474. https://doi.org/10.1007/s10936-012-9210-0

Zhang, H; Wang, X; Liu, J; Zhang, L; & Ji, L. (2023). Chinese named entity recognition method for the finance domain based on enhanced features and pretrained language models. *Information Sciences*, *625*, 385-400. https://doi.org/10.1016/j.ins.2023.03.015

Zhao, X; Li, H; & Wang, J. (2022). Linguistic specificity and stock price synchronicity. *China Journal of Accounting Research*, *15*(2), 112-130. https://doi.org/10.1016/j.cjar.2022.02.001