Article

Accelerating Literature Reviews with Multi-Database Information Systems for Financial Distress Research

Filipe Caetano [1,2,3,*], Rute Abreu [2,3,4,5], Pedro Brioso [6] and M. Victoria Lopez-Pérez [1]

[1] Department of Accounting and Finance, Universidad de Granada, Campus La Cartuja s/n, 18071 Granada, Spain; mvlopez@ugr.es
[2] Escola Superior de Tecnologia e Gestão, Instituto Politécnico da Guarda, Av. Dr. Francisco Sá Carneiro 50, 6300-559 Guarda, Portugal; ra@ipg.pt
[3] CICF—Center for Research in Accounting and Taxation, Instituto Politécnico do Cávado e Ave, 4750-810 Barcelos, Portugal
[4] CISeD Research Center in Digital Services, Instituto Politécnico de Viseu, 3504-510 Viseu, Portugal
[5] CiTUR Centre for Research, Development and Innovation in Tourism, Polo Instituto Politécnico da Guarda, 6270-372 Seia, Portugal
[6] Overslep, 6300-559 Guarda, Portugal; brioso@overslep.pt
[*] Correspondence: caetano@correo.ugr.es

Abstract

Literature reviews are a cornerstone of doctoral research in general, and of economic and business research, in particular. However, the exponential growth of scientific publications has made comprehensive and transparent reviews increasingly difficult. Conventional approaches, largely based on manual searches across a small number of databases, tend to be slow, error-prone, and incomplete. As a result, they constrain the scope of inquiry and, consequently, the robustness of theory development and empirical validation. This paper proposes and analyses an information system architecture driven by research questions and keyword taxonomies to automate core tasks of the literature search phase across multiple academic databases. Focusing on the domain of corporate and municipal financial distress, the authors employ a two-stage research design. First, the theoretical analysis integrates the literature on systematic reviews, automation, and financial distress prediction to derive a set of functional and non-functional requirements. Second, the experimental analysis documents a prototype front-end application designed to accelerate the literature review. The prototype is conceptualised as a socio-technical artefact that enhances IT competences and scientific resilience by enabling more efficient, reproducible, and extensible reviews. The authors conclude by discussing the scientific, technical, professional, and societal implications of the prototype, including opportunities for intellectual-property protection and avenues for future research.

Keywords: informatics; information retrieval; financial distress

1. Introduction

Literature reviews play a central role in the development of rigorous research in information systems (IS) [1,2] and economic–financial studies [3]. They provide a theoretical and empirical basis for formulating research questions, designing methodologies, and interpreting findings [4,5]. At the same time, the volume of academic publications has grown dramatically [6,7], with lots of new articles each year in domains such as corporate bankruptcy, municipal financial distress, and financial risk management [8].

In this context, the proposed system architecture directly addresses the challenge of information overload by automating and structuring the initial literature search phase [9,10]. By orchestrating parallel searches across multiple databases, consolidating results, and systematically removing duplicates, the system reduces the cognitive and operational burden associated with identifying, collecting, and organising the large volumes of potentially relevant literature. The methodological contributions have emphasised that literature reviews must be systematic [11], transparent [12], and concept-centric rather than method-centric [13]. Guidelines in information systems and software engineering stress explicit research questions, documented search strategies, clear inclusion and exclusion criteria, and reproducible procedures [14]. Frameworks such as Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) further formalise reporting standards, encouraging comprehensive documentation of how studies were identified, selected, and synthesised [15]. However, many of these guidelines implicitly assume the availability of human time and expertise to manually craft search strings, query multiple databases, download and merge results, and remove duplicates [10,14–18].

In practice, the difficulties of collecting the literature frequently limit the extensiveness of reviews [19]. Doctoral students and experienced researchers alike often default to a small number of familiar databases, stop searching when they feel they have "enough" papers, or rely on citation chasing and informal recommendations [20,21]. The resulting samples, although not necessarily invalid, may be biased towards certain journals, languages, or regions and may omit important streams of research, particularly interdisciplinary work and studies published in non-mainstream outlets [22,23].

At the same time, advances in information systems, web services, and artificial intelligence have opened opportunities to automate parts of the review process, from search strategy formulation promoted by the researcher to database querying [24,25]. Recent work in software engineering and information systems retrieval explores tools that generate search strategies, perform automated screening, or support systematic analyses of scientific corpora [26]. Yet the integration of such tools into the everyday practice of economic–financial research remains limited [27]. Many proposed solutions are generic or focused on medical [28] or computer science domains, with little explicit tailoring to the specific needs of financial distress research in companies and municipalities [29].

This paper addresses this gap by proposing a keyword- and research-question-driven information system architecture to support literature reviews in economic–financial information systems, with particular emphasis on corporate and municipal financial distress. The central idea is to treat the literature search phase as a structured information retrieval problem, implemented through a front-end application that orchestrates parallel queries across multiple high-quality databases, automatically removes duplicates and exports the resulting bibliographic dataset to a collaborative environment [30]. The architecture is not intended to replace human judgement or critical thinking in either the screening of studies or the synthesis of evidence [16,31]. But rather, it extends Human research capabilities by ensuring broad, reproducible coverage and reducing the time spent on repetitive mechanical tasks [32–34].

The research question is as follows:

What integrated, multi-database information system, driven by research questions and keyword taxonomies improves the efficiency, coverage, and transparency of literature reviews in financial distress research?

The objective of the research is to design, theoretically justify, and experimentally describe such a prototype in order to assess its potential impact on scientific rigour, IT competences, and scientific resilience [35–37].

The research question is intentionally formulated at an integrative level to guide the design and instantiation of the proposed artefact. Rather than being rhetorical, it is operationalised through the definition of functional and non-functional requirements and examined via qualitative experimental observations. Accordingly, the manuscript is positioned as a design science contribution reporting an empirical proof-of-concept, rather than as a purely conceptual paper.

This paper proposes a domain-specific information system to support literature reviews in financial distress research. Section 1 presents the introduction with context of the research. Section 2 reviews prior work on systematic reviews, automation, and financial distress as an example of the research area. Section 3 describes the materials and methods based on a design science research approach and the development of a proof-of-concept prototype. Section 4 presents qualitative results illustrating the feasibility and added value of multi-database orchestration. Section 5 discusses the methodological and practical implications. The paper concludes by summarising contributions, limitations, and directions for future research.

Furthermore, this paper addresses a timely and methodologically relevant problem and reflects broad familiarity with the literature on systematic reviews, information retrieval, and financial distress. While the reference base is strong and up to date, the contribution is positioned as a design science-oriented, proof-of-concept study with a clearly delimited research design and empirical scope.

2. Literature Review

This section reviews prior work on literature reviews, which constitute a foundational methodological pillar in information systems and management research, shaping theory development, research design, and empirical interpretation [10,38]. However, prior research has demonstrated that the epistemic and operational characteristics of literature reviews differ substantially across scientific domains, requiring domain-sensitive approaches rather than uniform methodological prescriptions [11].

In contrast to biomedical research, where systematic reviews typically rely on relatively homogeneous study designs, stable terminologies, and a limited number of dominant databases, financial distress research is inherently fragmented and interdisciplinary. It extents accounting, corporate finance, economics, public administration, law, and information systems, combining quantitative modelling, institutional analysis, regulatory interpretation, and, increasingly, artificial intelligence-based approaches [5,8].

Furthermore, relevant studies are dispersed across heterogeneous journals and databases, none of which offer comprehensive coverage of the field [21,22].

This fragmentation is further intensified by the dual focus of financial distress research on private and public entities. Corporate bankruptcy studies are traditionally embedded in finance and accounting journals, whereas municipal financial distress research is frequently published in public administration, regional science, and interdisciplinary outlets [39]. Prior methodological research has shown that reliance on a single database systematically biases the literature samples and undermines the completeness and transparency expected of high-quality reviews [19,40].

Although methodological guidelines for systematic literature reviews in information systems emphasise transparency, reproducibility, and concept-centric synthesis [14,18], they implicitly assume the feasibility of manually executing complex multi-database search strategies. Empirical evidence suggests that such assumptions are increasingly unrealistic in domains characterised by publication growth and terminological heterogeneity [6,10].

Consequently, recent research has called for the reconceptualization of the literature search phase as an information systems problem, in which search strategy formulation,

database orchestration, and result consolidation are treated as designable artefacts rather than tacit researcher skills [20,33]. This perspective is particularly relevant in financial distress research, where interdisciplinary dispersion and governance-specific terminology amplify the risks of omission and bias.

Thus, this research proposes a research-question-driven, multi-database information system specifically tailored to literature reviews in financial distress research. Positioned within a design science research framework [33,41], the system aims to enhance coverage, transparency, and reproducibility while preserving human interpretive judgement. The study forms part of an ongoing doctoral project on *Artificial Intelligence in Accounting: Development of a Collaborative Platform*, contributing directly to the systematisation of the AI-related literature and indirectly supporting research on the behavioural implications of bankruptcy and insolvency models.

2.1. The Role and Challenges of Literature Reviews

In IS and management research, literature reviews are recognised as a foundational method to consolidate existing knowledge and identify research gaps [42]. Ref. [9] argued that high-quality reviews are concept-centric, synthesising what is known about key constructs rather than merely summarising individual studies. Ref. [20] proposed a systems approach in which literature reviews are structured around inputs (sources), processes (search, analysis, synthesis), and outputs (conceptual frameworks and research agendas).

Ref. [10] highlighted difficulties of the researchers when conducting exhaustive reviews in an environment of information overload. They documented how calls for "comprehensive" reviews often clash with practical limits, leading to omitted work and apparent contradictions between guidelines and achievable practice. Refs. [18,43] synthesised guidelines across major IS journals, classifying review types (narrative, descriptive, scoping, systematic, meta-analytic) and identifying common stages such as search design, study selection, analysis, and reporting [44].

Across these contributions, two recurrent challenges emerge. First, search strategy design is complex: researchers must translate abstract constructs into keyword combinations and controlled vocabularies that vary across databases. Second, operational execution is labour-intensive: queries must be run separately in each database, results exported in heterogeneous formats, duplicates removed [45], and metadata normalised. These tasks consume time that could otherwise be devoted to critical reading and synthesis.

2.2. Systematic Review Methodologies and Reporting Standards

Systematic literature reviews (SLRs) were initially formalised in medicine and later adapted to software engineering and IS [46]. Ref. [16] provided influential guidelines for SLRs in software engineering, emphasising explicit research questions, protocol development, database selection, data extraction, and quality assessment. Subsequent work evaluated the application of these guidelines in practice, including meta-reviews of SLRs in software engineering [14].

In parallel, the PRISMA statement specified preferred reporting items for systematic reviews and meta-analyses, focusing initially on health research but increasingly adopting other domains. PRISMA 2020 updates the checklist to reflect advances in identification, selection, and synthesis methods and provides detailed flow diagrams and exemplars [17]. Although PRISMA is primarily a reporting guideline, it implicitly shapes how reviews are conducted by requiring transparent documentation of database searches and study selection [19].

In economic and financial research, SLRs and scoping reviews are increasingly used to synthesise knowledge on topics such as corporate bankruptcy prediction, financial distress

identification, and financial risk modelling [47]. Recent systematic reviews examine financial distress prediction models and artificial intelligence methods for distress identification, often relying on multi-database searches guided by PRISMA [48]. However, these studies typically describe manual search processes and offer limited detail on the technical tools used to manage search results.

2.3. Automation and Artificial Intelligence in Literature Reviews

To address the labour intensity of SLRs, researchers have developed tools and frameworks for partial or full automation. In software engineering, mapping studies and systematic reviews of tools show that automation efforts concentrate on stages such as searching, screening, and data extraction [26]. Ref. [19] evaluated the retrieval performance of different academic search systems, revealing substantial variation in coverage and precision, which reinforces the need to search multiple databases when aiming for comprehensive reviews.

Recent advances in natural language processing (NLP) and artificial intelligence (AI) have stimulated new approaches. Ref. [24] proposed an automated method for developing search strategies using keyword co-occurrence networks to suggest additional terms, thereby mitigating the risk of missing relevant terminology. Surveys of AI techniques for automating systematic reviews document applications of machine learning to document classification, active learning for screening, and topic modelling for synthesis [31].

More recently, Large Language Models (LLMs) have been integrated into Retrieval-Augmented Generation (RAG) frameworks to support the automation of literature reviews [31,49]. One stream of research proposes multi-agent systems in which specialised agents generate research questions, retrieve documents, extract data, and draft narrative text. Another stream introduces LLM-based pipelines that ingest metadata and abstracts to produce structured review sections, while seeking to mitigate hallucination risks through the use of RAG architectures and human-in-the-loop validation mechanisms [31].

Despite promising results, these approaches often require substantial computational resources and technical expertise [50]. Moreover, they typically assume access to a consolidated corpus of documents, without addressing the upstream problem of orchestrating searches across heterogeneous databases with different Application Programming Interfaces (APIs), rate limits, and metadata standards [40]. For economic–financial researchers, especially those working in public-sector financial distress where documents may be scattered across multidisciplinary outlets, this upstream orchestration is a crucial bottleneck.

2.4. Financial Distress Research and the Need for Integrated Information Systems

Financial distress prediction has a long tradition in corporate finance, beginning with discriminant analysis models such as Altman's Z-score and evolving towards logit, hazard, machine learning, and deep learning approaches [51,52]. Systematic reviews show an expanding landscape of models and predictors, with increasing attention to non-financial information such as corporate governance, textual disclosures and network measures [44]. Recent studies propose multimodal deep learning architectures combining financial indicators, filings, and inter-firm network data [53].

In the public sector, research on municipal financial distress has grown more recently. Data mining and machine learning models have been developed to predict distress in local governments, often combining accounting data with socio-economic indicators [54]. Studies highlight the importance of early-warning systems to support fiscally responsible decision-making and prevent service disruptions [55].

This research is inherently interdisciplinary, spanning accounting, economics, public administration, information systems, and computer science. Relevant studies may appear in journals indexed in different databases and classified under distinct subject categories.

Consequently, relying on a single database such as Web of Science or Scopus may systematically under-represent certain perspectives or geographies [21,22]. Integrated information systems that orchestrate searches across multiple databases are therefore essential to obtain a balanced view of the state of knowledge.

2.5. Information Systems, IT Competences, and Scientific Resilience

In this paper, scientific resilience is understood as the capability of research activity to withstand disruption and to adapt effectively by incorporating new evidence and methods over time. Resilience research commonly conceptualises this capability in terms of timely sensing, learning, and reconfiguration processes, which provide a practical basis for operationalisation in socio-technical settings. In the context of literature review workflows, scientific resilience can therefore be reflected in process-oriented indicators such as the traceability and reproducibility of search strategies, the ability to update and re-run searches consistently, and the continuity of evidence curation under changing information conditions.

The concept of scientific resilience and the capacity of research communities to adapt rapidly to new evidence, methods, and shocks depend crucially on IT infrastructure and competences [56]. Organisational resilience research suggests that systems that support timely sensing, learning, and reconfiguration can enhance adaptability in turbulent environments [57,58]. By analogy, information systems that streamline access to up-to-date literature, support collaborative curation, and enable reproducible workflows develop the resilience of research programmes.

In this sense, the value proposition of the proposed system is not limited to efficiency, but also concerns the robustness and continuity of evidence-based research workflows.

From a skills perspective, doctoral training increasingly requires not only domain knowledge but also competence in data management, scripting, and automation [59,60]. Designing and using literature review information systems strengthens these competences by exposing researchers to APIs, metadata standards, and basic programming [61]. At the same time, such systems help to institutionalise knowledge by making search strategies and datasets shareable and reusable across projects and teams [62].

Figure 1 provides an overview of the conceptual framework that underpins this study. It integrates three interrelated strands of prior research, such as contemporary challenges in conducting literature reviews, established systematic review methodologies, and advances in automation and artificial intelligence, and shows how they jointly inform the two-stage research design adopted in this paper. The framework clarifies the theoretical logic guiding both the derivation of system requirements and the subsequent development and experimental analysis of the proposed information system. Figure 1 synthesises the main theoretical strands discussed in the literature review and illustrates how they jointly underpin the research design adopted in this study. Specifically, the framework links the challenges associated with contemporary literature reviews, established systematic review methodologies, and advances in automation and artificial intelligence to the two-stage research design, thereby clarifying the conceptual logic that guides the development and experimental analysis of the proposed information system.

In summary, the literature indicates a clear need for tools that (i) operationalise systematic review guidelines, (ii) integrate multiple high-quality databases, (iii) are tailored to specific domains such as financial distress, and (iv) are accessible to doctoral researchers in economics and business studies. The next sections describe the design and experimental analysis of such an information system.
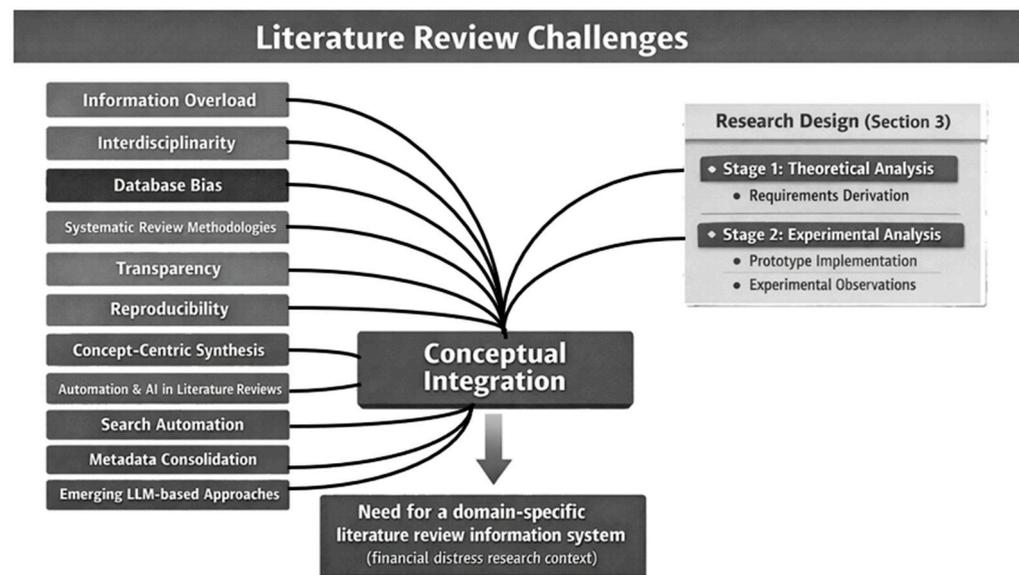
Figure 1. Conceptual framework linking literature review challenges, systematic review methodologies, and automation to the two-stage research design. Source: own elaboration.

3. Materials and Methods

The research methodology follows a two-stage research design combining theoretical and experimental analysis [36,63,64]. The construction of the keyword taxonomy followed an iterative and theory-informed process. Initial constructs were elicited directly from the exemplar research question and informed by established terminology in the financial distress and information systems literature. These constructs were then expanded through an exploratory content analysis of titles, abstracts, and author keywords from seminal and highly cited studies, and subsequently refined through iterative validation to ensure conceptual coherence and coverage.

The research design is grounded in design science principles, with a structured progression from theoretical analysis to artefact instantiation and experimental assessment. While these stages are closely connected, the revised manuscript clarifies their analytical distinction and aligns the presentation of the artefact, its implementation, and its evaluation with established design science reporting conventions.

The first stage is theoretical analysis in which, drawing on design science principles, the authors conducted a narrative yet concept-centric review of guidelines on literature reviews [20,38], automation tools [31,32], and financial distress research [51,52]. This stage aimed to derive the functional and non-functional requirements of a literature review information system. The analysis was guided by an information systems perspective on economic–financial governance, in which the system is conceived as a socio-technical artefact that mediates between researchers and dispersed bodies of scientific knowledge [41,65].

The second stage is an experimental analysis based on these requirements, the authors implemented a prototype front-end application and executed proof-of-concept searches in the domain of corporate and municipal financial distress, following a systems-development style of inquiry in which the artefact itself becomes a vehicle for empirical insight [66,67]. The experimental analysis focuses on describing and evaluating the architecture and workflow rather than on quantitative performance metrics. The implemented architecture is therefore treated as an experimental artefact, aimed at demonstrating feasibility and clarifying intended usage scenarios, rather than at providing benchmark-based performance validation [66,67].

In line with literature review guidelines recommending multiple databases [68], the prototype integrates major sources, such as Web of Science (WoS), ScienceDirect, Sco-

pus, Emerald Insight, SpringerLink, IEEE Xplore, ACM Digital Library, and MDPI. These databases were selected because, collectively, they cover top-tier journals in accounting, finance, public administration, information systems, software engineering, and computer science, including outlets where financial distress research is commonly published [50]. They also offer APIs or export functions that can be programmatically invoked via Hyper-Text Transfer Protocol (HTTP) requests, a prerequisite for automation.

The prototype is driven by the combination of a formal research question and a structured keyword taxonomy [29,68–70]. For the purpose of this paper, the exemplar research question is as follows:

Which Models Can Predict Financial Distress in Municipalities?

From this question, the authors elicit primary constructs, in line with concept-driven approaches in design science and systems thinking ("modelling", "prediction", "financial distress", "municipalities"). Each construct is associated with clusters of synonyms and related terms, for example: for public entities: financial and economic analysis, local government, and sustainability; for information systems: decision-support system, early-warning system, data mining, machine learning, and deep learning.

The taxonomy is stored in a configuration file that is accessible to the front-end application and simultaneously serves as a historical record of the conceptual evolution of the search strategy. For each database, mapping rules transform generic constructs into database-specific query strings, taking into account field codes (e.g., title, abstract, keywords), boolean operators, and phrase search syntax [71–74].

The prototype was developed using standard web-engineering practices and principles of requirements engineering [75]. Functional requirements included the following: user interface for entering research questions, selecting constructs, and specifying filters (e.g., publication years, document types as [76] has studied); parallel execution of HTTP requests to each database; collection of results in a unified internal representation (e.g., JavaScript Object Notation (JSON) with standardised fields such as title, authors, abstract, keywords, Digital Object Identifier (DOI), source, year); automated de-duplication based on identifiers (DOI or Uniform Resource Locator (URL)), titles, and author combinations; export of the consolidated dataset in formats suitable for reference managers (research information systems (RIS), bibliographic reference management tool—often used with the TeX/LaTeX typesetting system (BibTeX), Comma-Separated Values (CSV)); and upload of exported files to a collaborative repository (e.g., shared Google Drive folder).

Non-functional requirements included robustness to partial failures, such as one database temporarily unavailable, transparent logging of queries and responses, and respect for usage policies and rate limits of each database [77,78].

The experimental evaluation consisted of deploying the prototype in a controlled environment and executing test searches corresponding to the financial distress research question above [79]. The focus was on validating that the system successfully retrieved records from multiple databases, removed duplicates, and generated a consolidated file ready for downstream screening and analysis [33,41,66].

4. Results

This section presents the results of the paper, structured around three complementary components. First, it outlines the functional and non-functional requirements derived from software engineering principles and the specific characteristics of financial distress research. Second, it describes the architecture of the experimental artefact, detailing the design and operation of the proposed multi-database information system. Third, it reports qualitative experimental observations obtained from the application of the prototype, illustrating its feasibility, coverage, and implications for literature review workflows.

4.1. Requirements from Software Engineering and Financial Distress Research

From the perspective of software engineering and requirements, the proposed prototype addresses a classic information retrieval problem framed within a domain-specific context [80]. Requirements were elicited through: (i) analysis of existing SLR guidelines; (ii) examination of automation tools described in the literature; and (iii) reflection on the practical needs of researchers working on corporate and municipal financial distress.

The key functional requirements derived from this analysis included the following [24]:

- Multi-database coverage to mitigate database bias and ensure that both corporate and municipal financial distress studies are captured;
- Keyword taxonomy management to reflect domain-specific terminology, including local-government terminology and variations of "financial distress";
- Traceability and reproducibility, with logs of queries, parameters and dates, aligning with PRISMA and design science expectations;
- Extensibility, allowing new databases to be added as modules without redesigning the entire system.

Non-functional requirements included user-friendliness for researchers without advanced programming skills, security, and respect for intellectual-property and database-usage licences.

4.2. System Architecture: Description of the Experimental Artefact

The architecture implemented in the prototype is depicted conceptually in Figure 2 as a pipeline with three main components.
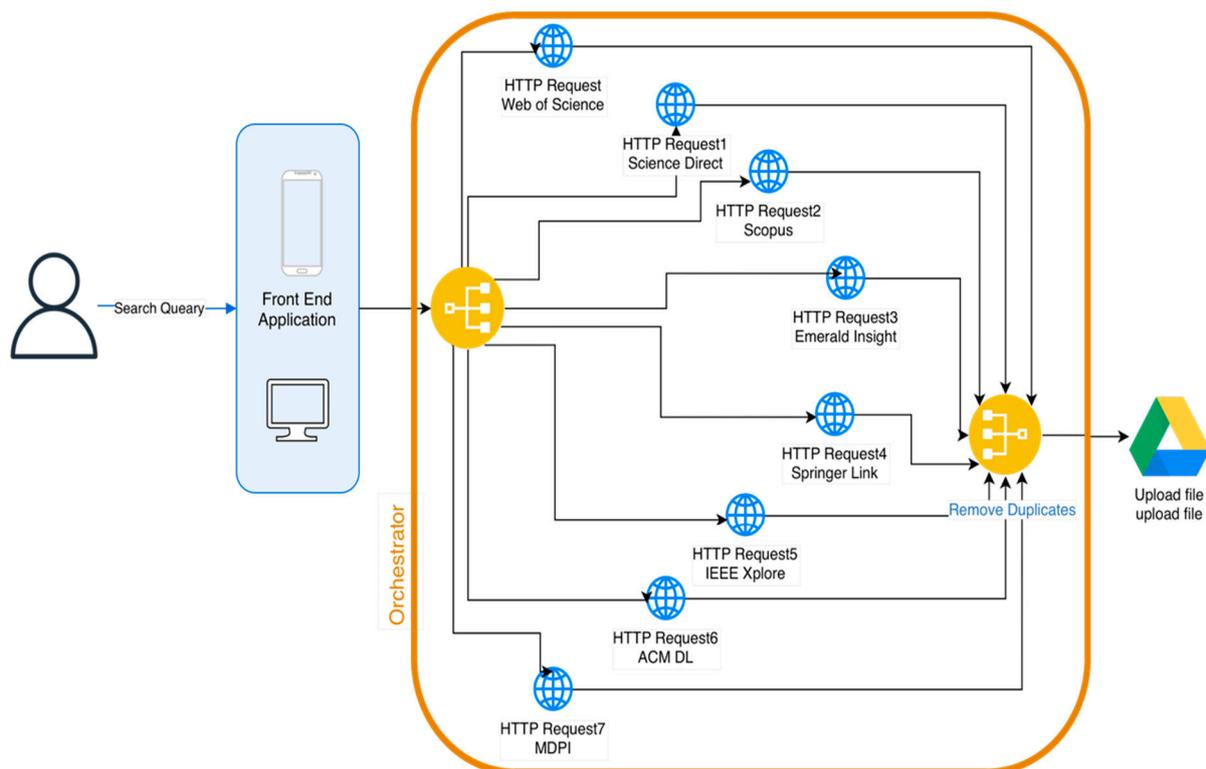
Figure 2. Prototype of accelerating literature reviews. Source: own elaboration.

The process begins when the researcher interacts with a web-based front-end application from a desktop or mobile device. The user inputs the research question, selects or edits keyword groups and specifies optional filters such as time window or document type [81].

Internally, the front-end sends these selections to the orchestrator. The orchestrator, rather than the front-end, generates and executes the API calls to each database.

From the orchestrator, the prototype sends asynchronous HTTP requests to each target database's API or search endpoint:

HTTP Request 0—Web of Science: queries the Web of Science Core Collection using appropriate field tags.

HTTP Request 1—ScienceDirect: uses the Elsevier API to search journals and books within the ScienceDirect platform.

HTTP Request 2—Scopus: issues queries via the Scopus API, potentially retrieving broader disciplinary coverage.

HTTP Request 3—Emerald Insight: focuses on the disciplinary breadth and coverage of journals.

HTTP Request 4—SpringerLink: provides coverage of journals and books spanning a broad range of disciplinary domains.

HTTP Request 5—IEEE Xplore: targets conference papers and journals in computing and engineering, capturing work on machine learning approaches to distress prediction.

HTTP Request 6—ACM Digital Library: retrieves software engineering and information systems research, particularly on tools, architectures, and algorithms.

HTTP Request 7—MDPI: enables systematic querying across its portfolio of open-access journals, providing valuable disciplinary breadth for this study.

The requests are executed in parallel, exploiting asynchronous communication to reduce overall latency. Each response is parsed into the internal metadata schema, with fields such as source_database, title, abstract, authors, year, keywords, DOI, and URL.

All metadata records are then passed to a Remove Duplicates module. This module performs the following: exact matching on DOI or URL, when available; fuzzy matching on normalised titles and author list to capture cases where DOIs are missing or formatted differently; and preferential retention of records from the database offering the most complete metadata, for example, full abstracts or keywords.

The output is a consolidated set of unique publications, annotated with the list of databases in which each record was found. This information supports later sensitivity analysis on database coverage.

Finally, the consolidated dataset is exported in user-selected formats (RIS, BibTeX, CSV) and automatically uploaded to a shared repository, for example, Google Drive folder, as indicated in Figure 2. The repository serves as the collaborative hub where authors and co-authors can access and annotate the dataset, import it into reference managers, and perform screening and coding.

4.3. Experimental Observations

This section reports the main qualitative findings derived from the experimental application of the prototype. The results indicate that the proposed architecture improves literature review coverage and transparency by enabling the systematic retrieval and consolidation of publications dispersed across heterogeneous databases. The experimental observations further suggest efficiency gains in the initial search phase by reducing manual, repetitive tasks related to query execution, result merging, and de-duplication [82]. Together, these findings provide qualitative evidence that the proposed system addresses the research question posed in the Introduction, particularly with respect to coverage, transparency, and workflow efficiency.

In proof-of-concept searches on topics, for example, financial distress, the prototype successfully retrieved and consolidated records across several (relevant) databases. Qualitative inspection of the resulting corpus indicated the following: certain streams

of research, such as deep learning models and multimodal data for distress prediction, were predominately indexed in engineering-oriented databases (IEEE Xplore, ACM DL), whereas traditional corporate finance models were concentrated in Web of Science, Scopus, and ScienceDirect. Also, municipal financial distress studies often appeared in public administration or regional science journals indexed in Emerald and SpringerLink. And the de-duplication module substantially reduced manual effort by automatically merging multiple entries of the same article retrieved from different databases.

To provide a more concrete illustration of this effect, an exploratory comparison was conducted between a single-database search (Scopus) and the proposed multi-database strategy using the same research question and keyword configuration. The single-database search predominantly retrieved accounting- and finance-oriented studies, whereas the multi-database approach additionally captured relevant contributions from information systems, public administration, and computer science outlets. Although no quantitative recall metrics are reported at this stage, this qualitative comparison supports the conclusion that the multi-database strategy yields a richer and more diverse initial literature corpus than reliance on a single database alone.

While the authors refrain from making quantitative claims about performance improvement, the prototype demonstrates the feasibility of the proposed architecture and its capacity to generate a richer and more diverse initial literature corpus than would likely be obtained using any single database alone [83].

To further support the feasibility of the proposed system with respect to efficiency, an exploratory comparison was conducted between the prototype-assisted workflow and a traditional manual literature search process. In the manual approach, equivalent searches required repeated query formulation, sequential execution across individual databases, manual export of results, and time-consuming de-duplication. By contrast, the prototype enabled parallel execution of database queries and automated consolidation of results, substantially reducing the time and manual effort required for the initial search phase. Although no formal time measurements or user studies are reported at this stage, these observations provide preliminary qualitative evidence that the system improves workflow efficiency in the early stages of literature reviews.

5. Discussion

The contribution of this paper is positioned as a design science-oriented proof of concept, in which the proposed system is described and analysed as a prototype. The results emphasise architectural feasibility and workflow integration rather than quantitative performance validation. Consequently, no metrics on efficiency, recall, precision, workload reduction, or comparative performance are reported. The development of a structured, metric-based evaluation framework constitutes a key avenue for future work.

From a scientific perspective, the proposed information system contributes to three main aspects. First, it operationalises systematic review guidelines for the search phase [14–17]. Rather than leaving database selection and search execution as tacit skills, the prototype is handled through the researcher's implicit expertise and ad hoc practices [9,10,18,20,38]; the prototype encodes these choices in configurable modules and logs. This reduces researcher-specific variability and supports reproducibility, complementing frameworks such as PRISMA [15] and design science recommendations on transparent artefact design [41,64].

Second, the prototype improves corpus coverage and diversity [19,21–24]. By design, it interrogates both mainstream and specialised databases, capturing disciplinary, methodological, and regional diversity in financial distress research [8,51,52,84,85]. This is especially important for municipal financial distress, where relevant studies may be scattered across public administration, regional science, accounting, and information systems

journals [39,48,55]. The multi-database strategy reduces the risk of missing key evidence due to database bias, a concern frequently raised in SLR methodology discussions [14,16,19,21,22].

Third, the architecture enables concept-centric synthesis, because metadata from all databases is standardised, researchers can more readily perform bibliometric analyses, cluster studies by constructs and trace the evolution of themes over time [23,71,72,86,87]. This supports the kind of concept-centric and theory-building reviews advocated in the IS literature, moving beyond narrative summaries towards structured theoretical contributions [9,10,18,20,38].

Technically, the prototype demonstrates how standard web technologies and APIs can be orchestrated to create a domain-specific research infrastructure built on established principles of distributed information systems and retrieval architectures [71,72,77,78,88,89]. The modularity of the design, where each database is encapsulated in its own request module, facilitates maintenance and extensibility, so that new databases or preprint servers can be added as they become relevant, and existing modules can be adapted when API endpoints change, following best practices in modular SLR tooling and automated retrieval pipelines [16,26,31,32,45,50,70].

Professionally, in a public-sector context, the prototype can be reinterpreted as a decision-support tool for evidence-informed governance [8,39,44,55,83]. For example, regulators or audit institutions interested in international best practices on taxation, accounting, and financial rules or municipal early-warning systems could use customised versions of the prototype to rapidly assemble the literature as a basis for guidelines or policy papers [39,44,90]. In private sector contexts, corporate risk management departments could exploit similar architectures to monitor research on emerging risk indicators or modelling techniques [8,27,39,51–53,83].

The prototype's design also enhances IT competences among researchers and practitioners as working effectively with the prototype requires familiarity with core information [30,72,73,91]. Engaging with architecture demands an understanding of APIs, data formats, and basic scripting, aligning with the increasing expectation that economic–financial professionals be proficient in data analytics and information systems [79,86,92]. For doctoral students, implementing or adapting such a prototype constitutes a valuable learning experience that bridges theoretical knowledge and practical skills [4,5,93]. Within the broader doctoral research framework, the system functions as a methodological infrastructure rather than a standalone technological solution. It directly supports the systematic consolidation of AI-related research in accounting and finance, while providing a robust evidentiary foundation for analysing how bankruptcy and insolvency models shape organisational behaviour across institutional contexts. This positioning aligns with calls to strengthen the methodological and infrastructural contributions of information systems research to high-impact domains [94].

From an intellectual-property perspective, the novelty of the prototype lies not in any single algorithm but in the domain-specific integration of multi-database retrieval, de-duplication, and collaborative export driven by research-question-based keyword taxonomies [32,36]. A patent application in the field of information technology could emphasise the following: the configurable mapping from research-question constructs to database-specific query syntaxes [71,72,95]; the orchestration of parallel database requests under constraints of licencing and rate limits, following best practices in the search [88,89]; the intelligent de-duplication strategy that leverages DOIs, fuzzy title matching, and database-priority rules [45,96]; and the automated export and upload to collaborative repositories with embedded provenance information [81,92].

Patentability would depend on demonstrating that this combination of features is not obvious to a person skilled in the art, given existing generic reference-manager and SLR tools [31,45]. The argument could stress its focus on economic–financial information systems, municipal financial distress, and integration with governance workflows [39,48]. Even if eventual patent protection were not sought, documenting the architecture in detail, as in this paper, establishes prior art and supports claims of originality in academic and professional contexts [41,50].

From a societal perspective, the prototype can be seen as an enabler of better-informed public policy and corporate governance in financial distress and early-warning mechanisms [51,52]. More efficient and comprehensive literature reviews on financial distress can improve early-warning systems for municipalities, helping detect financial distress vulnerabilities before they translate into service cuts or bailouts [39,48]. Also, this can support corporate restructuring and insolvency frameworks by ensuring that policy makers and practitioners base reforms on the most up-to-date evidence [53,96] and contribute to financial literacy and transparency initiatives, as researchers can more easily synthesise knowledge about risk factors and mitigation strategies [83].

Furthermore, by reducing the technical barrier to systematic reviews, the prototype may promote wider participation of researchers from less-resourced institutions or regions, supporting diversity in scientific discourse [21,94]. Open-source implementations could be particularly impactful in emerging economies where subscription-based review tools are unaffordable and access to commercial infrastructures is uneven [22].

On the other hand, increased automation raises ethical questions about the role of AI-based tools in shaping research agendas and evidence bases [29,45,96]. Over-reliance on automated tools may lead to superficial engagement with the literature or uncritical acceptance of algorithmically generated corpora. There is also a risk that concentration on databases with strong API support may marginalise journals or repositories without such infrastructure. These concerns underscore the need for critical digital literacy and human oversight in the use of automated retrieval systems [72,85,96].

Beyond the limitations already acknowledged, additional constraints should be considered. Because retrieval is driven by a keyword taxonomy mapped to database-specific query syntaxes, coverage may be sensitive to terminological drift in emerging research fronts where concepts and labels are not yet standardised. This risk is particularly salient in interdisciplinary domains such as corporate and municipal financial distress, where similar constructs may be expressed differently across accounting, finance, public administration, and information systems. To mitigate this limitation, the taxonomy should be treated as a living artefact, subject to continuous and transparent refinement, versioning, and expert validation, while preserving traceability to support reproducibility.

Furthermore, dependency on database APIs raises sustainability considerations that extend beyond availability and licencing conditions. Changes in database interfaces, endpoint deprecations, authentication mechanisms, rate limits, or metadata schemas may introduce maintenance overhead and affect the long-term stability and reproducibility of retrieval results. Although the system's modular architecture facilitates adaptation to such changes, sustainable operation requires ongoing monitoring, maintenance of database connectors, and, where necessary, documented fallback procedures.

6. Conclusions

This paper addressed the challenge of conducting rigorous, comprehensive literature reviews in economic–financial research, particularly, in the domains of corporate bankruptcy and municipal financial distress. Indeed, the authors propose and experimentally describe a keyword- and research-question-driven information system architecture

that orchestrates parallel searches across scientific databases, performs de-duplication, and exports results to a collaborative repository. The main contributions are conceptual contribution, because the paper frames the literature search as an information systems problem and positions multi-database orchestration as a solution to the limitations of manual, single-database searches, with errors and (economic) limitations due to limited access to scientific databases. Another contribution specifies functional and non-functional requirements for a literature review information system grounded in software engineering and SLR guidelines. The technical contribution presents an implemented prototype illustrating how a front-end application can send parallel HTTP requests to multiple databases, consolidate metadata, and support collaborative workflows. And the practical contribution highlights implications for doctoral training, professional practice in economic–financial governance, and the potential for intellectual-property protection.

Nevertheless, several limitations must be acknowledged. First, the experimental assessment is qualitative and proof-of-concept in nature, as quantitative evaluations of recall, precision, time savings, and user experience remain to be conducted. Second, the current prototype focuses on metadata retrieval and de-duplication. At its present stage, the system does not incorporate automated screening, quality assessment, or synthesis, nor does it include full-text mining or LLM-based summarisation functionalities. Third, the information system is constrained by the availability, stability, and usage terms of APIs which may affect long-term sustainability and require ongoing maintenance in response to interface changes, endpoint deprecations, or licencing conditions. In addition, because retrieval is driven by a keyword taxonomy, coverage may be sensitive to terminological drift in emerging research fronts where concepts and labels are not yet standardised, requiring continuous and transparent taxonomy refinement.

In this sense, future research proceeds along multiple directions. Technically, the architecture will extend to integrate AI-based screening and active-learning components to prioritise relevant studies; employ RAG-based LLMs to support semi-automatic coding and synthesis while maintaining human oversight; and implement "living reviews" where updated searches are scheduled and the corpus is incrementally refreshed. Practically, it will be allowed to assess the information system in comparative experiments with traditional manual reviews, measuring coverage, bias, and researcher workload. Also relevant is the possibility to increase domain-specific extensions that tailor keyword taxonomies and filtering rules to sub-topics such as debt crises, social impact, or green-financial distress. Furthermore, from a pedagogical direction, integrating the information system into doctoral curricula fosters both methodological rigour and IT competences, strengthening the capacity of new generations of researchers to navigate the expanding landscape of scientific knowledge with resilience and critical thinking, that will improve the knowledge in favour of more strong research.

Conflicts of Interest: Author Pedro Brioso is employed by the company Overslep. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ACM	Association for Computing Machinery
AI	Artificial intelligence
API	Application Programming Interface
BibTeX	Bibliographic reference management tool, often used with the TeX/LaTeX typesetting system
CSV	Comma-Separated Values
DOI	Digital Object Identifier
HTTP	HyperText Transfer Protocol
IEEE	Institute of Electrical and Electronics Engineers
IS	Information systems
IT	Information technology
JSON	JavaScript Object Notation
LLM	Large Language Model
NLP	Natural language processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RAG	Retrieval-Augmented Generation
RIS	Research information systems
SLR	Systematic literature reviews
URL	Uniform Resource Locator
WoS	Web of Science

References

1. Zeng, F.; Lee, S.H.N.; Lo, C.K.Y. The Role of Information Systems in the Sustainable Development of Enterprises: A Systematic Literature Network Analysis. *Sustainability* **2020**, *12*, 3337. [CrossRef]
2. Okoli, C.; Schabram, K. A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts Work. Pap. Inf. Syst.* **2010**, *10*. [CrossRef]
3. Baker, H.K.; Kumar, S.; Pattnaik, D. Twenty-five years of Review of Financial Economics: A bibliometric overview. *Rev. Financ. Econ.* **2019**, *38*, 3–23. [CrossRef]
4. Collins, A.; Joseph, D.; Bielaczyc, K. Design Research: Theoretical and Methodological Issues. *J. Learn. Sci.* **2004**, *13*, 15–42. [CrossRef]
5. Patten, M.L. *Proposing Empirical Research: A Guide to the Fundamentals*; Taylor & Francis: Abingdon, UK, 2014; ISBN 978-1-936-52330-6.
6. Bornmann, L.; Mutz, R. Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *J. Assoc. Inf. Sci. Technol.* **2015**, *66*, 2215–2222. [CrossRef]
7. Bornmann, L.; Haunschild, R.; Mutz, R. Growth rates of modern science: A latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanit. Soc. Sci. Commun.* **2021**, *8*, 224. [CrossRef]
8. Mallinguh, E.B.; Zéman, Z. Financial distress, prediction and strategies by firms: A systematic review of literature. *Period. Polytech. Soc. Manag. Sci.* **2020**, *28*, 162–176. [CrossRef]
9. Watson, R.T.; Webster, J. Analysing the past to prepare for the future: Writing a literature review—A roadmap for release 2.0. *J. Decis. Syst.* **2020**, *29*, 129–147. [CrossRef]
10. vom Brocke, J.; Simons, A.; Riemer, K.; Niehaves, B.; Plattfaut, R.; Cleven, A. Standing on the Shoulders of Giants: Challenges and Recommendations of literature reviews in information systems. *Commun. Assoc. Inf. Syst.* **2015**, *37*, 205–224.
11. Boell, S.K.; Cecez-Kecmanovic, D. On being "Systematic" in literature reviews in IS. *J. Inf. Technol.* **2015**, *30*, 161–173. [CrossRef]
12. Haddaway, N.R.; Woodcock, P.; Macura, B.; Collins, A. Making literature reviews more reliable through application of lessons from systematic reviews. *Conserv. Biol.* **2015**, *29*, 1596–1605. [CrossRef] [PubMed]
13. Tsironis, C. Concept-Centric or Method-Centric Research? On Paradigm War in the Interdisciplinary Study of/Research on Religion. *Annu. Rev. Sociol. Relig.* **2012**, *3*, 45–67. [CrossRef]
14. Kitchenham, B.; Brereton, O.P.; Budgen, D.; Turner, M.; Bailey, J.; Linkman, S. Systematic literature reviews in software engineering—A systematic literature review. *Inf. Softw. Technol.* **2009**, *51*, 7–15. [CrossRef]

15. Moher, D.; Liberati, A.; Tetzlaff, J.; Altman, D.G. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Med.* **2009**, *6*, e1000097. [CrossRef] [PubMed]

16. Kitchenham, B.; Charters, S. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Technical Report EBSE 2007-001. Durham Keele University. 2007. Available online: https://ebse.webspace.durham.ac.uk/ebse-bibliography/guidelines-for-performing-systematic-literature-reviews-in-software-engineering/ (accessed on 15 November 2025).

17. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.U.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [CrossRef] [PubMed]

18. Paré, G.; Trudel, M.-C.; Jaana, M.; Kitsiou, S. Synthesizing information systems knowledge: A typology of literature reviews. *Inf. Manag.* **2015**, *52*, 183–199. [CrossRef]

19. Gusenbauer, M.; Haddaway, N.R. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Res. Synth. Methods* **2019**, *11*, 181–217. [CrossRef]

20. Levy, Y.; Ellis, T.J. A systems approach to conducting an effective literature review in support of information systems research. *Informing Sci. J.* **2006**, *9*, 181–212. [CrossRef]

21. Mongeon, P.; Paul-Hus, A. The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics* **2016**, *106*, 213–228. [CrossRef]

22. Harzing, A.-W.; Alakangas, S. Google Scholar, Scopus and the Web of Science: A longitudinal and cross-disciplinary comparison. *Scientometrics* **2016**, *106*, 787–804. [CrossRef]

23. Zupic, I.; Čater, T. Bibliometric methods in management and organisation. *Organ. Res. Methods* **2015**, *18*, 429–472. [CrossRef]

24. van de Schoot, R.; de Bruin, J.; Schram, R.; Zahedi, P.; de Boer, J.; Weijdema, F.; Kramer, B.; Huijts, M.; Hoogerwerf, M.; Ferdinands, G.; et al. An open source machine learning framework for efficient and transparent systematic reviews. *Nat. Mach. Intell.* **2021**, *3*, 125–133. [CrossRef]

25. Maon, F.; Lindgreen, A.; Swaen, V. Designing and implementing corporate social responsibility: An integrative framework grounded in theory and practice. *J. Bus. Ethics* **2009**, *87*, 71–89. [CrossRef]

26. van Dinter, R.; Tekinerdogan, B.; Catal, C. Automation of systematic literature reviews: A systematic literature review. *Inf. Softw. Technol.* **2021**, *136*, 106589. [CrossRef]

27. Fromlet, H. Behavioral Finance-Theory and Practical Application: Systematic analysis of departures from the Homo OEeconomicus paradigm are essential for realistic financial research and analysis. *Bus. Econ.* **2001**, *36*, 63–69.

28. Tsafnat, G.; Glasziou, P.; Choong, M.K.; Dunn, A.; Galgani, F.; Coiera, E. Systematic review automation technologies. *Syst. Rev.* **2014**, *3*, 74. [CrossRef] [PubMed]

29. Bolaños, F.; Salatino, A.; Osborne, F.; Motta, E. Artificial intelligence for literature reviews: Opportunities and challenges. *Artif. Intell. Rev.* **2024**, *57*, 259. [CrossRef]

30. van Rijsbergen, C. *Information Retrieval*, 2nd ed.; Butterworth-Heinemann: Oxford, UK, 1979. Available online: http://www.dcs.gla.ac.uk/Keith/Preface.html (accessed on 21 November 2025).

31. Han, B.; Susnjak, T.; Mathrani, A. Automating systematic literature reviews with retrieval-augmented generation. *Appl. Sci.* **2024**, *14*, 9103. [CrossRef]

32. Marshall, C.; Brereton, P.; Kitchenham, B. Tools to support systematic literature reviews in software engineering: A feature analysis. In Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering, London, UK, 13–14 May 2014; pp. 1–10. [CrossRef]

33. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *MIS Q.* **2004**, *28*, 75–105. [CrossRef]

34. Thomas, J.; Noel-Storrb, A.; Marshallc, I.; Wallaced, B.; McDonalde, S.; Mavergamesf, C.; Glaszioug, P.; Shemilta, I.; Synnote, A.; Turnere, T.; et al. Living systematic reviews: 2. Combining human and machine effortt. *J. Clin. Epidemiol.* **2017**, *91*, 31–37. [CrossRef]

35. Tichy, W.F. Should computer scientists experiment more? *Computer* **1998**, *31*, 32–40. [CrossRef]

36. March, S.T.; Smith, G.F. Design and natural science research on information technology. *Decis. Support Syst.* **1995**, *15*, 251–266. [CrossRef]

37. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [CrossRef]

38. Webster, J.; Watson, R.T. Analysing the past to prepare for the future: Writing a literature review. *MIS Q.* **2002**, *26*, xiii–xxiii.

39. Antulov-Fantulin, N.; Lagravinese, R.; Resce, G. Predicting bankruptcy of local government: A machine learning approach. *J. Econ. Behav. Organ.* **2021**, *183*, 681–699. [CrossRef]

40. Adler, N.; Harzing, A.W. When knowledge wins: Transcending the sense and nonsense of academic rankings. *Acad. Manag. Learn. Educ.* **2009**, *8*, 72–95. [CrossRef]

41. Gregor, S.; Hevner, A.R. Positioning and presenting design science research for maximum impact. *MIS Q.* **2013**, *37*, 337–355. [CrossRef]

42. Oosterwyk, G.; Brown, I.; Geeling, S. A synthesis of literature review guidelines from information systems journals. In *Proceedings of the 4th International Conference on the Internet, Cyber Security and Information Systems (ICICIS) 2019, Johannesburg, South Africa, 31 October–1 November 2019*; Kalpa Publications in Computing: Manchester, UK, 2019; Volume 12, pp. 250–260.

43. Templier, M.; Paré, G. A Framework for Guiding and Evaluating Literature Reviews. *Commun. Assoc. Inf. Syst.* **2015**, *37*, 112–137. [CrossRef]

44. Elmagarmid, A.K.; Ilyas, I.F.; Verykios, V.S. Duplicate record detection: A survey. *IEEE Trans. Knowl. Data Eng.* **2007**, *19*, 1–16. [CrossRef]

45. Brereton, P.; Kitchenham, B.; Budgen, D.; Turner, M.; Khalil, M. Lessons from applying the systematic literature review process within the software engineering domain. *J. Syst. Softw.* **2007**, *80*, 571–583. [CrossRef]

46. Stead, E.; Clive Smallman, C. Understanding business failure: Learning and un-learning from industrial crises. *J. Contingencies Crisis Manag.* **1999**, *7*, 1–18. [CrossRef]

47. Radebe, N.E.; Nomlala, B.C.; Matenda, F.R. Systematic Review of Financial Distress Prediction Models for Municipalities: Key Evaluation Criteria and a Framework for Model Selection. *J. Risk Financ. Manag.* **2025**, *18*, 624. [CrossRef]

48. de la Torre-López, J.; Ramírez, A.; Romero, J.R. Artificial intelligence to automate the systematic review of scientific literature. *Computing* **2023**, *105*, 2171–2194. [CrossRef]

49. Wu, S.; Ma, X.; Luo, D.; Li, L.; Shi, X.; Chang, X.; Lin, X.; Luo, R.; Pei, C.; Du, C.; et al. Automated literature research and review-generation method based on large language models. *Natl. Sci. Rev.* **2025**, *12*, nwaf169. [CrossRef]

50. Bramer, W.M.; Rethlefsen, M.L.; Kleijnen, J.; Franco, O.H. Optimal database combinations for literature searches in systematic reviews: A prospective exploratory study. *Syst. Rev.* **2017**, *6*, 245. [CrossRef] [PubMed]

51. Altman, E.I. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *J. Financ.* **1968**, *23*, 589–609. [CrossRef]

52. Altman, E.I.; Iwanicz-Drozdowska, M.; Laitinen, E.K.; Suvas, A. Financial distress prediction in an international context: A review and empirical analysis. *J. Int. Financ. Manag. Account.* **2017**, *28*, 131–171. [CrossRef]

53. Che, W.; Wang, Z.; Jiang, C.; Abedin, M.Z. Predicting financial distress using multimodal data: An attentive and regularized deep learning method. *Inf. Process. Manag.* **2024**, *61*, 103703. [CrossRef]

54. Alaminos Aguilera, D.; Fernández, D.; García-Lopera, F.; Fernández, M.A. Data Mining for Municipal Financial Distress Prediction. In *18th Industrial Conference Proceedings, ICDM 2018, New York, NY, USA, July 11–12, 2018*; Perner, P., Ed.; Lecture Notes in Computer Science, LNAI 10933; Springer: Cham, Switzerland, 11–12 July 2018. [CrossRef]

55. Piermarini, D.; Sudoso, A.M.; Piccialli, V. Predicting municipalities in financial distress: A machine learning approach enhanced by domain expertise, Machine Learning. *arXiv* **2023**, arXiv:2302.05780. [CrossRef]

56. Linkov, I.; Trump, B.D. *The Science and Practice of Resilience*; Springer: Cham, Switzerland, 2019. [CrossRef]

57. Lengnick-Hall, C.A.; Beck, T.E.; Lengnick-Hall, M.L. Developing a capacity for organizational resilience through strategic human resource management. *Hum. Resour. Manag. Rev.* **2011**, *21*, 243–255. [CrossRef]

58. Duchek, S. Organizational resilience: A capability-based conceptualization. *Bus. Res.* **2020**, *13*, 215–246. [CrossRef]

59. Hansen, K.B.; Souseles, D. Expectations, Competencies and Domain Knowledge in Dataand Machine-driven Finance. *Econ. Soc.* **2023**, *52*, 421–448. [CrossRef]

60. Beiholder, J.; Naumann, F. Data fusion. *ACM Comput. Surv. (CSUR)* **2008**, *41*, 1–41. [CrossRef]

61. Gaitanou, P.; Andreou, I.; Sicilia, M.A.; Garoufallou, E. Linked data for libraries: Creating a global knowledge space, a systematic literature-review. *J. Inf. Sci.* **2024**, *50*, 204–244. [CrossRef]

62. Hetemi, E.; Pushkina, O.; Zerjav, V. Collaborative practices of knowledge work in IT projects. *Int. J. Proj. Manag.* **2022**, *40*, 906–920. [CrossRef]

63. Corbin, J.; Strauss, C. *Basics of Qualitative Research*, 3rd ed.; Sage: Thousand Oaks, CA, USA, 2008.

64. Baskerville, R.; Baiyere, A.; Gregor, S.; Hevner, A.; Rossi, M. Design science research contributions: Finding a balance between artifact and theory. *J. Assoc. Inf. Syst.* **2018**, *19*, 358376. [CrossRef]

65. Gregor, S. The nature of theory in information systems. *MIS Q.* **2006**, *30*, 611–642. [CrossRef]

66. Nunamaker, J.F.; Chen, M.; Purdin, T.D.M. Systems development in information systems research. *J. Manag. Inf. Syst.* **1991**, *71*, 89–106. [CrossRef]

67. Runeson, P.; Höst, M. Guidelines for conducting and reporting case study research in software engineering. *Empir. Softw. Eng.* **2009**, *14*, 131–164. [CrossRef]

68. Fagin, R.; Kumar, R.; Sivakumar, D. Comparing top k lists. *SIAM J. Discret. Math.* **2003**, *17*, 134–160. [CrossRef]

69. Checkland, P. Soft Systems Methodology: A Thirty Year Retrospective. *Syst. Res. Behav. Sci.* **2000**, *17*, S11–S58. [CrossRef]

70. Kwabena, A.E.; Wiafe, O.-B.; John, B.-D.; Bernard, A.; Frimpong, F.A.F. An automated method for developing search strategies for systematic review using Natural Language Processing (NLP). *MethodsX* **2023**, *10*, 101935. [CrossRef] [PubMed]

71. Manning, C.D.; Raghavan, P.; Schütze, H. *Introduction to Information Retrieval*; Cambridge University Press: Cambridge, UK, 2008. Available online: https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf (accessed on 25 November 2025).

72. Baeza-Yates, R.; Ribeiro-Neto, B. *Modern Information Retrieval*, 2nd ed.; Addison-Wesley: Essex, England, 1999; ISBN 0-201-39829-X. Available online: https://web.cs.ucla.edu/ miodrag/cs259-security/baeza-yates99modern.pdf (accessed on 22 November 2025).

73. Salton, G.; Buckley, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manag.* **1988**, *24*, 513–523. [CrossRef]

74. Robertson, S.; Zaragoza, H. The probabilistic relevance framework: BM25 and beyond. *Found. Trends Inf. Retr.* **2009**, *3*, 333–389. [CrossRef]

75. Giray, G. A software engineering perspective on engineering machine learning systems: State of the art and challenges. *J. Syst. Softw.* **2021**, *180*, 111031. [CrossRef]

76. Robertson, S. Understanding inverse document frequency: On theoretical arguments for IDF. *J. Doc.* **2004**, *60*, 503–520. [CrossRef]

77. Fielding, R.T. Architectural Styles and the Design of Network-Based Software Architectures. Doctoral Dissertation, University of California, Irvine, CA, USA, 2000.

78. Dean, J.; Ghemawat, S. MapReduce: Simplified data processing on large clusters. *Commun. ACM* **2008**, *51*, 107–113. [CrossRef]

79. Järvelin, K.; Kekäläinen, J. Cumulated gain-based evaluation of IR techniques. *ACM Trans. Inf. Syst.* **2002**, *20*, 422–446. [CrossRef]

80. Kotonya, G.; Sommerville, I. *Requirements Engineering*; Wiley: Chichester, UK, 1998.

81. Rahm, E.; Do, H.H. Data cleaning: Problems and current approaches. *VLDB J.* **2000**, *9*, 312–330.

82. Snyder, H. Literature review as a research methodology: An overview and guidelines. *J. Bus. Res.* **2019**, *104*, 333–339. [CrossRef]

83. Kuizinienė, D.; Krilavičius, T.; Damaševičius, R.; Maskeliūnas, R. Systematic review of financial distress identification using artificial intelligence. *Appl. Artif. Intell.* **2022**, *36*, 2138124. [CrossRef]

84. Samara, K.; Shinde, A. Predicting bankruptcy using machine learning and data preprocessing techniques. *Analytics* **2025**, *4*, 22. [CrossRef]

85. Saracevic, T. Relevance: A review of literature and a framework for thinking on the notion in information science. *J. Am. Soc. Inf. Sci. Technol.* **2007**, *58*, 2126–2144. [CrossRef]

86. Tversky, A. Features of similarity. *Psychol. Rev.* **1977**, *84*, 327–352. [CrossRef]

87. Shokouhi, M.; Si, L. Federated search. *Found. Trends Inf. Retr.* **2011**, *5*, 127–256. [CrossRef]

88. Voorhees, E.M.; Harman, D. (Eds.) *TREC: Experiment and Evaluation in Information Retrieval*; MIT Press: Cambridge, MA, USA, 2005. Available online: https://aclanthology.org/J06-4008.pdf (accessed on 20 November 2025).

89. Dias, A.; Rodrigues, F.; de Jesus, M. Data mining for municipal financial distress prediction. In *New Knowledge in Information Systems and Technologies*; Rocha, A., Adeli, H., Reis, L.P., Costanzo, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 253–265. [CrossRef]

90. Lim, T. Environmental, social, and governance (ESG) and artificial intelligence in finance: State-of-the-art and research takeaways. *Artif. Intell. Rev.* **2024**, *57*, 76. [CrossRef]

91. Haak, L.L.; Fenner, M.; Paglione, L.; Pentz, E.; Ratner, H. ORCID: A system to uniquely identify researchers. *Learn. Publ.* **2012**, *25*, 259–264. [CrossRef]

92. Ning, X.; Yim, D.; Khuntia, J. Online sustainability reporting and firm performance: Lessons learned from text mining. *Sustainability* **2021**, *13*, 1069. [CrossRef]

93. Agarwal, R.; Lucas, H. The information systems identity crisis: Focusing on high-visibility and high-impact research. *MIS Q.* **2005**, *29*, 381–398. [CrossRef]

94. Gusenbauer, M.; Gauster, S.P. How to search for literature in systematic reviews and meta-analyses: A comprehensive step-by-step guide. *Technol. Forecast. Soc. Change* **2025**, *212*, 1238332025. [CrossRef]

95. Fellegi, I.P.; Sunter, A.B. A theory for record linkage. *J. Am. Stat. Assoc.* **1969**, *64*, 1183–1210. [CrossRef]

96. Agbo, F.J.; Olivia, C.; Oguibe, G.; Sanusi, I.T. Computing education using generative artificial intelligence tools: A systematic literature review. *Comput. Educ. Open* **2025**, *9*, 100266. [CrossRef]