

Finding Knowledge Beyond the Algorithm: Ten Reasons and a Proposal for DDC and WebDeweySearch in 2026

Piero Cavaleri^(a)

a) LIUC University, <https://orcid.org/0000-0003-4870-1757>

Contact: Piero Cavaleri, cavaleri.piero@gmail.com

Received: 29 March 2026; **Accepted:** 08 July 2026; **First Published:** 15 September 2026

ABSTRACT

In the 2026 information landscape, information retrieval is increasingly dominated by search engines and generative artificial intelligence (GenAI). While recognizing the genuine value of these technologies, this paper argues that they exhibit documented limitations, such as hallucinated citations, visibility bias, and the progressive fragmentation of knowledge. This position paper presents a critical analysis supporting the continued relevance of the Dewey Decimal Classification (DDC) and the WebDeweySearch platform as epistemological infrastructures for organizing and exploring knowledge. Through a hierarchical and disciplinary logic, the DDC is interpreted not merely as a cataloging tool, but as a cognitive map capable of fostering structured serendipity, disciplinary contextualization, and global semantic interoperability. The analysis further highlights how the stability of numerical notation mitigates the effects of terminological shift and misinformation, ensuring transparency and accountability in knowledge organization. Finally, the paper explores the DDC's potential as a complementary framework for AI, particularly within Retrieval-Augmented Generation (RAG) architectures, where it provides the semantic grounding necessary to counteract algorithmic drift. From this perspective, the DDC emerges not as a legacy of the past, but as a critical component for preserving epistemological rigor in the contemporary digital ecosystem.

KEYWORDS

Dewey Decimal Classification (DDC); WebDeweySearch; Retrieval-Augmented Generation (RAG); Knowledge organization; Information retrieval.

Introduction

This paper is conceived as a position paper and does not aim to provide systematic empirical validation of the arguments advanced. Rather, it develops a logical-theoretical argument aimed at highlighting the strengths of classification systems—and in particular the Dewey Decimal Classification (DDC)—in facilitating comprehensive, systematic, and guided forms of documentary research based on selected and structured corpora.

From this perspective, the complementarity of other approaches to information retrieval is also acknowledged: keyword-based systems are recognized for their speed and effectiveness in searches aimed at solving practical problems, while artificial intelligence systems based on large language models (GenAI), one of the many “Artificial Intelligences” (Mitchell 2022), are recognized for their ability to produce syntactically correct and pragmatically effective texts.

The primary aim of this paper is to argue for the relevance of DDC-based systems—particularly WebDeweySearch—as tools for structured information retrieval. A secondary aim is to explore how DDC may also contribute to improving AI-based systems.

The contribution is directed in particular at information professionals, scholars of Library and Information Science, and researchers interested in the intersections between knowledge organization and artificial intelligence. It aims to argue that, in the current landscape, different systems—classification schemes, thesauri, information retrieval systems, and artificial intelligence systems—perform complementary functions in meeting information needs.

This paper adopts a conceptual and argumentative approach. While it draws on selected literature to support the discussion of limitations in contemporary information retrieval systems and generative artificial intelligence, the core of the paper develops a theoretically informed position based on analytical reasoning about the structural properties of classification systems, particularly the Dewey Decimal Classification.

The limitations that we are going to signal for keyword searches or searches using generative AI do not mean that they are not useful — to claim otherwise would be implausible given their widespread use by billions of people and many libraries have implemented or are currently implementing systems that use them (Boateng 2025; Santosa 2025) — but only that relying exclusively on these tools for any informational need risks exposing users to errors, failure to identify relevant documents, and citations of documents that do not exist.

These systems represent a genuine advance in accessibility and text synthesis, and their widespread adoption reflects real user needs. The limitations discussed here are structural rather than incidental, but do not diminish their value for many tasks, particularly in applications where accuracy of reference is not required or the truth of the statements is not essential.

The structure of the text is the following: the first section examines the limitations of contemporary information retrieval systems; the second presents the advantages of DDC-based approaches through a set of arguments; the final section outlines a proposal for integrating DDC within AI-driven architectures.

In the contemporary 2026 information landscape, searching for books using the Dewey Decimal Classification (DDC) offers a distinctive approach to information retrieval—one based on disciplinary structure rather than algorithmic ranking or statistical inference.

A keyword search on platforms such as Google or Amazon primarily identifies the occurrence of terms within documents, and while increasingly supported by semantic ranking techniques, it does not explicitly represent the underlying conceptual relationships between topics (Pasquale 2015).

Generative artificial intelligence tools can produce text and, potentially, citations based on statistical inference that conveys apparent certainty, yet lacks any systematic guarantee of authenticity, completeness, or epistemic reliability.

In their current implementations, text-generation programs based on large language models (GenAI) tend to be structurally prone to inaccuracies in citation generation. Rather than verifying the existence of a document against a real-time database, these models are designed to predict linguistically plausible text, functioning as text generators that prioritize linguistic plausibility, often at the expense of factual accuracy. This makes them prone to producing “hallucinatory” references: fabricated citations that appear authentic but do not exist, complete with fictitious authors, real journal names, and non-existent DOIs. Falsification rates can be substantial in certain contexts. For example, Dahl et al. (2024), studying legal applications of large language models, documented frequent fabrication of legal citations. Likewise, Linardon et al. (2025), in an experimental study on mental health research, showed that citation fabrication varied according to topic familiarity and prompt specificity, with unsupported references remaining a significant problem even under carefully designed prompts.

A related limitation concerns the temporal coverage and the document archives used. Most current GenAI models are trained on fixed datasets and largely limited to what is accessible online; therefore, they do not reliably incorporate recent publications, works inaccessible online because they are held in restricted archives, or even the texts of older studies that have never been digitized—all elements that could be essential for a comprehensive literature review.

One of the primary causes of these limitations lies in the way natural language models process language. Rather than retrieving verified records, they generate citations through pattern matching: they predict the statistically most likely sequence of words that resembles a reference entry (Lissack and Meagher 2024). The result is often very convincing—plausible author names, familiar journal titles, realistic formatting—which can make it difficult, particularly for non-expert users, to identify fictitious citations without manual verification (Perc 2025).

This dynamic introduces a systematic visibility bias. The GenAI-generated citation patterns perform reliably when reproducing frequently cited, high-profile articles, but they tend to underperform with niche, specialized, or low-circulation research, for example research published in languages other than English or exclusively in print publications. The result is a so-called “Matthew effect” (Merton 1968, 1988), in which already popular works are further amplified and cited at the expense of more relevant but lesser-known ones (Algaba et al. 2025).

Finally, unless a system uses Retrieval-Augmented Generation (RAG), GenAI does not, by default, query the Internet in real-time or verify the existence of a document. Even RAG-enabled models, however, are not immune to problems. Empirical analysis has identified multiple structural failure points in RAG architectures: retrieved documents may lack the relevant content entirely, relevant passages may fail to be ranked highly enough to enter the generation context, or correctly retrieved information may not be adequately integrated into the final response (Barnett et al. 2024). Moreover, hallucinations can persist even when retrieval is nominally successful, since the quality of the generated output depends not only on document selection but on ranking, chunking, and the model’s subsequent interpretation of retrieved passages (Niu et al. 2024; Afane

et al. 2026). These limitations have driven a significant body of recent research toward what has been described as ‘Retrieval and Structuring Augmented Generation’ — architectures in which retrieved information is organized through explicit semantic structures such as taxonomies, hierarchies, and knowledge graphs before being passed to the generative model (Jiang et al. 2025). Studies consistently show that *ontology1-guided* and *graph-based* approaches substantially outperform purely vector-based retrieval, particularly in disambiguation, multi-hop reasoning, and domain-specific tasks (Cruz, Tavares, and Belo 2025; Jiang et al. 2025).

Gajderowicz, Bhardwa and Fox (2026) review the intersection of *ontological structures* and RAG, showing how *ontologies* support query disambiguation, validation, and structured retrieval. It is within this landscape that DDC’s hierarchical structure acquires renewed relevance — not as a replacement for computational retrieval, but as a pre-existing, human-curated semantic framework capable of constraining and orienting retrieval in ways that purely statistical methods cannot fully replicate. No system, however, eliminates these problems entirely: structured semantic scaffolding reduces but does not eliminate ambiguity, drift, and context loss, and its effectiveness depends on the quality and granularity of the underlying knowledge organization.

These structural limitations suggest that contemporary information retrieval may benefit from complementary knowledge organization systems specifically designed to provide stable disciplinary contexts and transparent semantic organization.

DDC: A Tool That Remains Valuable

While other tools contribute significantly to information retrieval, they also exhibit inherent limitations; DDC, by contrast, should be interpreted not merely as a classificatory tool, but as an epistemological infrastructure, one that situates ideas within a coherent framework of knowledge. Unlike retrieval systems optimized for engagement metrics, DDC was designed with a different goal: to represent the relationships between fields of knowledge, not simply to highlight the most popular documents or produce results that convey a sense of certainty grounded in plausibility rather than verification. This distinction—often overlooked in discussions of digital discovery—becomes increasingly relevant in an environment saturated with AI-generated content and opaque ranking systems.

To illustrate this point, the main advantages of a DDC-based document retrieval system such as WebDeweySearch² can be examined.

WebDeweySearch systems—available in different linguistic versions—provide users with direct access to the full hierarchical structure of the Dewey Decimal Classification. Through these platforms, users can navigate the classification tree across all levels, from general classes to highly specific subdivisions.

In addition to hierarchical browsing, users can perform searches based on the verbal expressions associated with DDC classes, including both class headings and terms from the Relative Index, thereby identifying the most relevant classification numbers for their information needs.

¹ In this sense, “ontology” must be understood as a technical term referring to a formal, machine-readable representation of knowledge within a specific domain.

² There are currently four versions of the WebDeweySearch system, corresponding to the German, Swedish, and Norwegian editions of the DDC, and, starting in 2026, the Italian edition.

Once the appropriate class numbers have been identified, users can use them as structured search keys within one or more library catalogs (OPACs), querying either a specific class number or expanding the search to include all subordinate classes within the hierarchy. This may enable both precise and comprehensive retrieval of relevant documents from bibliographic records enriched with DDC notations, typically corresponding to curated and professionally developed collections. In this way, WebDeweySearch functions not only as a classificatory interface, but as an intermediary layer connecting the semantic organization of knowledge with the retrieval of verified bibliographic resources, and can be understood as a fully-fledged information retrieval tool in its own right.

Records retrieved through WebDeweySearch correspond to verified documents, often curated within thoughtfully assembled collections. While it is acknowledged that library collections inevitably reflect institutional choices and potential imbalances, they nevertheless can be understood, in most cases, as curated corpora,³ distinct from the more open and less curated information environment of the web.

Before beginning to outline the advantages of using the DDC as a knowledge organization system for searching for information sources, it is important to highlight a significant limitation of its use in the real world: the DDC is used almost exclusively to index documents in the form of books or, at most, pamphlets, while articles or other types of contributions contained within books are excluded. This limitation should be explicitly acknowledged when considering the arguments that follow. To address the challenges of searching for these other types of documents, the analysis should have been extended to other classifications, such as the JEL classification,⁴ the Mathematics Subject Classification (MSC),⁵ or other indexing tools—subject headings or thesauri—such as MESH,⁶ which share some of the DDC’s strengths, lack others, and possess features not found in the DDC. This extension would have gone beyond the scope of this work and will therefore be left for possible future developments.

1. Serendipity: the serendipitous discovery, or the “shelf effect”

In a digital environment, users tend to retrieve exactly what they search for—or rather, what retrieval systems return based on algorithmic or statistical models—and little else. This is not a flaw in keyword-based systems; it is their design logic. But it results in a measurable epistemic cost: the suppression of serendipitous discovery (Pievani 2021).

DDC, on the other hand, is based on the principle of placement—that is, the systematic positioning of works related to interconnected topics within a contiguous classification space (Mitchell and Vizine-Goetz 2009). This is not random but reflects Melvil Dewey’s assumption that knowledge is

³ By “curated corpora” we refer to collections developed through professional selection, cataloging, and acquisition processes, which may nonetheless reflect institutional priorities and limitations. The most significant exception is represented by legal deposit library collections, which comprise the entire bibliographic output of a country or region; consequently, their growth is not driven by the libraries’ own acquisition policies.

⁴ JEL Classification System / EconLit Subject Descriptors. <https://www.aeaweb.org/econlit/jelCodes.php?view=jel>

⁵ 2020 Mathematics Subject Classification. <https://mathscinet.ams.org/msc/msc2020.html>

⁶ MESH <https://www.ncbi.nlm.nih.gov/mesh/>

fundamentally relational rather than atomic (Wiegand 1996). A book does not simply belong to a single subject; it belongs to a system of contiguous ideas, a system based on the disciplinary mode through which humans produce knowledge.

In practice, whether consulting a physical shelf or a virtual one via WebDeweySearch or another class-navigation tool, a user examining the class “Economics of Climate Change” (333.72) will find nearby the class and documents on “Renewable Energy Resources” (333.79). This adjacency is not an algorithmic suggestion based on co-citation or the purchasing behaviours of a more or less large number of customers, but is structural, determined by the intellectual architecture of the classification itself. It is less susceptible to manipulation, personalization, or commercial prioritization. Its structure is determined through the collective decisions of distributed expert bodies with international participation, with diverse backgrounds and interests, and is subject to the oversight of a professional community capable of identifying and correcting both specific errors and systemic shortcomings.

A useful example can be found by searching for the term ‘dog’ in the Italian WebDeweySearch. This returns 190 results. Most of these correspond to names of dog breeds, but one also finds the main number for general works on dogs (636.7) falling within the field of animal husbandry, all specific treatments and behaviours of dogs, topics relating to dogs falling within the field of zoology (599.772), canine veterinary medicine (636.7089), the use of dogs in sports (Hunting with dogs, 799.234) and recreational activities (Dog racing, 798.8), police dog services (363.2) and also dogs in literature (e.g. 809.933629772). Upon selecting one of these 190 subjects, the user is presented with the corresponding number within the subject hierarchy, with the option to explore higher and lower levels, as well as related classifications. Whenever the user decides that a number may be of interest to them, they can access an OPAC – in the Italian case, that of the National Library Service – to view the works indexed under the chosen number.

For information professionals, this aspect is of considerable relevance and goes far beyond the mere search for documents and information. In the development of collections, subject classification reveals gaps and development needs: if a library holds resources under call number 333.72 but few documents under 333.79, the shelf—whether physical or virtual—communicates this imbalance. Classification transforms consultation into a sort of audit of the collection.

While research on digital information behaviour has extensively analyzed how users encounter information serendipitously and how digital environments can be designed to foster or inhibit this phenomenon (Makri and Blandford 2012a; 2012b; Makri et al. 2014), the specific role of hierarchical subject classifications as engines of structured serendipity remains one of the least theorized. As research on information behavior consistently demonstrates, users, even professional users, often cannot articulate verbally what they need until they find it. Subject classification creates the conditions for this encounter; keyword search, by its very nature, does not systematically enable such encounters in the same structured manner.

2. A universal language for organizing knowledge

DDC’s use of Arabic numerals is not merely a matter of practicality, but a deliberate design choice for interoperability that has proven extraordinarily effective over the course of a century and a half

(Wiegand 1996). In an era of multilingual digital collections and data sharing among institutions, DDC's numerical notation—thanks above all to its global reach—can provide a level of document organization independent of natural languages, alphabets, syntax, and morphology.

Thanks to this feature and the availability of more than 30 translations in addition to the English edition, DDC can be considered as a globally shared classification framework that could (and should) be applied to other formats, even if current library practice limits it to monographs. A researcher in Tokyo, a student in Nairobi, and a cataloger in Paris all work with the shared understanding that numbers in the 500 class indicate the natural sciences. No translation is required at the classification level, even when the bibliographic records themselves are in different languages (Panzer 2008).

More recently, DDC classifications have also become relatively stable identifiers within interconnected data environments. Building on the Dewey Linked Data initiative, OCLC has progressively developed persistent URIs for DDC numbers, enabling their use as globally identifiable semantic resources. Although the implementation of this infrastructure is still evolving, the hierarchical structure and international diffusion of DDC already support semantic interoperability across different information systems (Mitchell 2013; OCLC 2024; 2025).

As libraries adopt RDF-based metadata frameworks and publish their catalogs as linked open datasets, Dewey URIs are intended to serve as stable, globally resolvable identifiers, capable of connecting bibliographic resources across institutions that share the same semantic framework. A query built around a Dewey URI can, in principle, retrieve relevant resources from any institution that has mapped its metadata to the same ontological framework, regardless of the language in which those records were created.

This positions DDC not merely as a cataloging tool, but as a semantic interoperability infrastructure. In a landscape where FRBR, BIBFRAME, and schema.org compete for dominance in bibliographic description, the structure of DDC's classifications can offer a relatively stable reference point for alignment across different institutions for sharing data in a semantically interoperable way—a function that any keyword-based system can hardly replicate, given the barriers that natural languages create for defining one-to-one correspondences.

3. Hierarchical Logic vs. Keyword-Based Retrieval

Keyword retrieval is typically organized as a flat result set, even when supported by underlying semantic or relational models. A search engine indexes terms and returns documents ranked by relevance signals, but it does not represent the logical relationships between these terms (Svenonius 2000). Homonyms, polysemy, and specialized jargon introduce noise that users struggle to manage. The hierarchical structure of DDC addresses this problem at the architectural level—we might say natively. Classification proceeds from the general to the specific through a strictly hierarchical logic:

- 500 – Natural Sciences
- 510 – Mathematics
- 516 – Geometry
- 516.3 – Analytical Geometry

Each level represents a progressively more specific subdivision of the one above it, meaning that navigating the hierarchy allows users to refine their search with greater precision, without requiring them to know specific terminology in advance. A student unfamiliar with the term “analytical geometry” can still find relevant material by starting at level 510 and moving logically down through the structure.

This has direct implications for library practice during reference consultations. When a user asks a vague or poorly formulated question, DDC hierarchy can provide a shared framework for the conversation: the librarian can probe to understand where to place the user’s question within the knowledge tree; they can then move up or down the chain of classifications, engaging in a collaborative process of intellectual clarification based on a solid, shared structure, transforming a retrieval operation into a genuine learning experience.

Furthermore, hierarchical classification can support systematic literature reviews in ways that keyword searches often cannot fully replicate. A researcher requiring comprehensive coverage of a domain can define it using its DDC range and be confident that the resulting set of books reflects an intellectual boundary based on sound principles, rather than an arbitrary keyword match.

4. Countering the Fragmentation of Information

A major transformation in the contemporary information environment is the transition from document-level retrieval to the extraction of information fragments. AI-based search systems increasingly return extracted passages, summaries, or generated responses, rather than complete documents situated within a disciplinary context. The result is a form of decontextualized or weakly contextualized knowledge: excerpts separated from the interpretive contexts that actually give them meaning, excerpts that create confusion and misinformation (Floridi 2020; Bender et al. 2021; Carr 2025).

DDC operates according to the opposite logic: each classification number is simultaneously an index for retrieving one or more documents and a disciplinary indicator. When a cataloger assigns a DDC notation, they do not merely organize a physical or digital object, but can be interpreted as making an epistemological statement about the placement of this work within the structure of human knowledge.

Consider a work on trains. Assigned to category 385 (Rail Transport), it is contextualized within the economic and social history of mobility. Assigned to category 625.1 (Railway Engineering), it belongs to the tradition of applied mechanics and infrastructure design. The same information object takes on significantly different intellectual interpretations depending on its classification, and that meaning is made explicit, stable, and searchable in a way that no keyword tag or presence in the form of an excerpt within a text generated by GenAI can reliably enable.

For information professionals engaged in high-quality metadata work (Jelani A. and Ashkar 2025), this underscores the importance of subject analysis as an intellectual act, not merely a bureaucratic one. With the proliferation of automated cataloging tools, the risk of classification drift—numbers assigned through pattern matching rather than conceptual judgment—is real. The hierarchical rigor of DDC can offer a standard against which such drift can be detected and corrected.

5. WebDewey: A Living System

The perception of DDC as a static and obsolete system is not supported by empirical evidence. Over the past 15 years, the transition to WebDewey has produced a continuously updated and dynamically managed classification infrastructure, capable of responding to emerging fields with a speed and specificity that previous print editions could not achieve. The current DDC evolves rapidly, but always on the basis of a system whose “engine” is grounded in human intellectual judgment and the consensus of bodies such as the EPC, EDUG, and ultimately a community of professional users (OCLC 2026).

The tables of the current DDC include classifications dedicated to Artificial Intelligence and its various branches (006.3), as well as many class numbers based on its application to numerous disciplinary fields. Studies on LGBTQ+ topics or Sustainable Development—areas that were not previously represented or were underrepresented in previous editions—can now be indexed with greater precision. The editorial process, overseen by the Editorial Policy Committee and supported by a global network in which EDUG plays an essential role, involves ongoing consultation with subject specialists, ensuring that new numbers reflect genuine disciplinary consensus rather than terminological trends.

A classification system that cannot accommodate new fields of knowledge forces catalogers to rely on makeshift solutions: overly generic numbers, inconsistent local practices, or reliance on subject headings to convey the specificity that a static classification cannot provide. WebDewey’s continuous updating model reduces this friction and maintains the system’s ability to serve as a reliable map of current knowledge.

The tension between the stability of classifications and the continuous updating of semantic content could be one of the key elements of the contemporary DDC, allowing the system to maintain structural consistency while adapting to the evolution of disciplinary domains.

Finally, for institutions participating in cooperative cataloging, WebDewey updates directly impact the quality of shared records. When DDC is updated, these updates are proposed for propagation across the cooperative network: a systemic advantage that a locally managed keyword taxonomy or subject index would struggle to replicate.

6. DDC as a cognitive map, not just a retrieval tool

Keyword searches and AI-powered chatbots provide answers. DDC seeks to provide guidance within the disciplinary framework of knowledge.

When a user receives a list of results from a keyword query, they get documents, but they do not understand how these relate to one another or to the broader field of study. The logic of the search engine is invisible, and the placement of suggested documents within the disciplinary system remains opaque. The user has no map, only a list. Even more cryptic are the suggestions from a chatbot that uses GenAI to generate responses.

DDC, by contrast, functions as a cognitive cartography of knowledge (Miksa 1998). When a researcher places a topic under heading 301 (Sociology) rather than 150 (Psychology), they are not merely finding a shelf location, but are learning something about the disciplinary boundaries of their inquiry. Which academic tradition addresses this problem? What methodological assumptions arise from it? What language do the authors of the document use?

The classification number is informative in and of itself, regardless of the specific document it refers to. (Bowker and Star 2000)

This is particularly valuable for novice researchers, doctoral students, and even undergraduate students in advanced courses who are entering an unfamiliar field of study. A keyword search for “identity” yields an indistinct heterogeneous set of results ranging from philosophy to psychology, from sociology to computer science and marketing. A DDC-based search forces the user to first identify where the meaning of the word “identity” they are interested in fits, and in doing so, they learn something about the structure of the discipline they are entering.

In an era where AI-powered chatbots provide confident, fluid answers that conceal the epistemological assumptions underlying them, DDC’s transparency regarding disciplinary boundaries can function as a corrective mechanism.

7. Resistance to misinformation and low-quality content

Keyword search indexes the web as it is, including disinformation, SEO-optimized noise, and AI-generated filler content. The ranking signals that determine what appears at the top of results (backlinks, engagement, publication date) do not necessarily correlate with intellectual quality or academic reliability (Noble 2018; Gillespie 2010).

DDC-based retrieval typically operates within a different quality ecosystem, since it is generally embedded in professionally curated collections and cataloging practices. To receive a DDC number, as we have already mentioned, a work must undergo a document selection process and professional cataloging, typically performed by qualified library staff who apply the DDC schema and established quality standards. This process functions as an implicit, albeit not infallible, quality filter: DDC operates within library collections, which themselves are curated through acquisition. The effectiveness of this filter depends on the quality of the cataloging work itself: inconsistent or automated classification practices can introduce errors and reduce the reliability of DDC-based retrieval, meaning that the system’s structural strengths are only fully realized when supported by rigorous professional practice.

Nevertheless, relying on selected documentary sources does not make DDC-indexed collections free from bias: the development of collections has its own biases and gaps. It does mean, however, that a DDC-based search operates within a curated corpus rather than an open, unmediated one. For users seeking academic material or high-quality content, this distinction is certainly relevant. With the proliferation of AI-generated content and the growing difficulty of distinguishing it, at first glance, from content created by humans, the value of retrieval systems based on professionally developed collections can only increase. Library catalogs and the collections they index can offer a form of structural mitigation against content pollution that keyword-based search algorithms may currently struggle to match consistently.

8. Support for Interdisciplinary Research

One of the most significant challenges in the contemporary research environment is navigating interdisciplinary territory (Thompson Klein 2021; Thompson Klein and Philipp 2023): topics that

lie at the intersection of two or more established fields and may therefore be classified under multiple sections within the Dewey Decimal Classification system.

This apparent complexity actually can be an advantage. A researcher working at the intersection of environmental law and economics can consciously consult both section 344.046 (Environmental Law) and section 333.7 (Economics of Natural Resources), using the classification system as a tool to map the disciplinary topology of their research problem. Furthermore, DDC's subject index offers a series of cross-references between different disciplinary treatments of the same subject or between related subjects, which can facilitate the mapping process.

This approach is difficult to replicate through keyword-based search alone, because keywords do not encode disciplinary affiliation, but only lexical content.

DDC's auxiliary tables, which allow classifiers to specify geographical, linguistic, temporal, and formal aspects, can also help enhance the system's internal relationality, enabling the addition of analytical dimensions that highlight facets present across a variety of subjects. A work classified as 305.4209492 is not simply "concerning women" (305.42), but specifically concerns women in the Netherlands (09492), within a historical context. This precision is encoded in the number itself, invisible to keyword searches but immediately readable by an expert user. The presence of both concepts—women and the Netherlands—across multiple DDC classes can offer users the opportunity to explore the classification beyond its hierarchical structure.

For systematic reviews and scoping studies, where comprehensive and reproducible subject coverage is a methodological requirement, DDC-based search strategies offer a level of reproducibility and transparency that keyword searches cannot enable. Two researchers performing the same DDC-based search in the same catalog are likely to obtain broadly comparable sets of results when operating within the same catalog and classification framework.

9. Independence from Changes in Language and Terminology

Academic and professional terminology is subject to continuous evolution; terms fall out of use, acquire new connotations, or are replaced by preferred alternatives for political, cultural, or scientific reasons. Keyword-based searches are inherently sensitive to these shifts: a search strategy relying solely on contemporary jargon may overlook decades of literature published using terms that were preferred in the past or that remain confined to specific scientific circles.

DDC exhibits a structural resistance to these terminological changes because, while verbal expressions are updated, the class numbers remain constant. Furthermore, terms are rarely eliminated unless they are considered entirely obsolete or culturally unacceptable. Because classification numbers are assigned based on conceptual essence rather than surface-level vocabulary, a work once indexed under the label of "racial relations" and a modern study on "structural racism" can both be anchored to the same notation (305.8), even if they share no common keywords.

A similar, and perhaps even more significant, problem arises when different cultures adopt divergent approaches to the same concept. A poignant example is the treatment of "race": the English DDC identifies the notation 305.809 with the index term "White race," whereas the Italian DDC—in different national implementations of DDC, terminology and index terms may reflect local linguistic and cultural conventions—utilizes the expression "Popolazione bianca". Despite

these linguistic and cultural differences, works discussing the same group of human beings are unified under the same classification number.

The DDC number can thus serve as a conceptual anchor across time, culture, and context, regardless of how the associated terminology evolves. This has profound implications for retrospective bibliographic research—a requirement in historical and legal studies—where a DDC-based strategy can retrieve older literature that keyword searches using current terminology would bypass. Ultimately, this framework helps overcome the lack of parallelism between different linguistic domains, providing information professionals with a strong argument for DDC classification as a vital, stable complement to keyword retrieval.

10. Accountability and Transparency in Knowledge Organization

AI-based chatbots raise some epistemological challenges: they generate responses without revealing their sources, training data, classification logic, or confidence level. The user receives an answer but cannot verify the process that produced it. Unlike traditional information retrieval systems, generative models do not enable a direct linkage between generated textual strings that may resemble citations and existing documents, unless explicitly supported by external retrieval mechanisms. These are not merely practical inconveniences, but raise structural challenges for the informed and critical use of information which, evidently, create ethical and political problems that go far beyond simple bibliographic research. (Floridi 2023; Pollicino and Dunn 2024; Matsieli and Mutula 2025).

DDC, by contrast, is a relatively transparent and well-documented system. Its tables are published and maintained in publicly accessible form. Its editorial decisions are recorded and traceable. Its hierarchical logic can be systematically examined, critiqued, and taught. When a classification number is assigned, the reasoning behind it is, in principle, reconstructible by trained users or catalogers.

This transparency supports information literacy education in a way that opaque artificial intelligence systems cannot. Teaching a user how DDC works provides a transferable mental model of knowledge organization. Teaching a user how a specific search algorithm works is, in most cases, impossible: the algorithm is proprietary, constantly evolving, and deliberately obscured. Even in the case of AI programs based on machine learning and deep neural networks, only usage strategies can be taught, but the actual processes executed by the model - stochastic processes - remain largely opaque due to the system's architectural design.

As the profession increasingly grapples with issues of algorithmic accountability in information retrieval, DDC's longstanding commitment to a documented classification logic—based on sound principles and publicly available—stands as a model worth considering and further developing within library and information science communities.

DDC has not been and is not free from problems—including American-centrism, Western epistemological bias, and other culturally situated limitations—but unlike a proprietary (black box) algorithm, the biases of DDC can be identified, discussed, and corrected publicly precisely because DDC can be interpreted as a “transparent and open to criticism” system.

These critical issues should not be underestimated, but rather considered as elements of any his-

torically situated system that requires continuous critical review and updating based on an awareness of the culturally determined context of every system for interpreting reality.

A summary table is presented below comparing document searches based on keyword systems or generative AI models with those conducted using the DDC through digital platforms such as WebDeweySearch or WebDewey itself.

<i>Functionality</i>	<i>AI-Powered Search (GenAI)</i>	<i>Keyword Search (IR)</i>	<i>Dewey Decimal Classification (DDC)</i>
Primary objective	Generation of plausible, syntactically correct text	Retrieval of specific items/terms	Mapping of a disciplinary domain
Underlying logic	Statistical likelihood / Probabilistic inference	Relevance and popularity algorithms	Intellectual and hierarchical structure
Search methods	Prompt-dependent; statistically driven	Query-dependent; lexical matching	Structure-dependent; hierarchical navigation
Disciplinary context	Absent or weakly inferred	Absent or user-defined	Explicit, stable, and foundational
Interoperability	Platform-specific (Proprietary)	Platform-specific (Opaque)	Globally standardized (Universal Schema)
Resistance to manipulation	Low: Susceptible to prompt injection, training bias and sponsored inference (ad-sustained bias)	Low: Susceptible to SEO and commercial ad-ranking	Relatively high: editorially controlled at the system level; at the application level it may vary.
Support for a systematic literature review	Limited: High risk of “hallucinated” citations	Limited: Dependent on terminology and user expertise	Strong: May support more systematic disciplinary coverage within DDC-indexed collections.

Keyword Search – AI-powered search vs. DDC: A Comparative Analysis.

A Possible Extension: DDC in AI Contexts

Regarding the issues of reliability, transparency, and accountability posed by AI systems, the DDC could act not only as a critical alternative but also as a powerful technological ally for artificial intelligence itself, especially within Retrieval-Augmented Generation (RAG) architectures (Friedrich, Akbari, and Fürstenau 2026). The integration of Dewey notations into document metadata allows for the retrieval of information to be anchored to a curated and stable semantic structure. Instead of relying exclusively on statistical or vector proximity - which in certain configurations can be prone to thematic drift or hallucinations - an advanced system can use the DDC hierarchy to filter or weight retrieved text fragments, supporting the generative model in drawing primarily from coherent disciplinary contexts.

In this scenario, classification ceases to be limited to a passive organizational role and becomes instead a dynamic contextual filter for information retrieval. By mapping a query to a specific DDC class before generation begins, the system can reduce its exposure to irrelevant or weakly related data clusters that happen to share superficial lexical similarities with the prompt. This “classification-aware” AI would not only provide responses, but could also explicitly state its epistemological

routing (e.g., “*This answer is primarily grounded within Engineering [620] rather than Economics [330]*”), restoring the disciplinary context that current GenAI platforms typically flatten. In this framework, the DDC does not eliminate the inherent fragmentation of AI-based retrieval systems, but it can help to structure and semantically link that fragmentation, mitigating the risks of de-contextualization.

This interpretation is consistent with a long tradition in Knowledge Organization research that understands classification not simply as a mechanism for document retrieval, but as an instrument for representing relationships among concepts and for shaping the intelligibility of knowledge domains (Miksa 1998; Svenonius 2000; Bowker and Star 2000).

From this perspective, the DDC may be examined as a semantic and epistemic structure whose significance extends far beyond mere bibliographic control. As a hierarchical system of notational distinctions, the DDC contributes to structuring domains of knowledge as retrievable and navigable conceptual spaces; it does not simply describe subjects, but establishes explicit relationships among them, providing a framework within which documentary resources can be interpreted. When such a structure is integrated into a retrieval-augmented architecture, the classification system becomes an active structure for organizing retrieval pathways: a set of rules governing not what is known, but under what conditions something may be retrieved, recombined, and made meaningful within a given epistemic context.

Recent developments in retrieval-augmented generation increasingly move beyond purely vector-based retrieval toward architectures that incorporate structured semantic scaffolding, such as hierarchies, ontologies, and knowledge graphs (Jiang et al. 2025). In hierarchical RAG approaches, multi-level indexing structures are explicitly used to control retrieval granularity and preserve semantic organization across different levels of abstraction, from coarse thematic clusters to fine-grained document segments (Huang et al. 2025). Similarly, graph-based retrieval systems introduce hierarchical or community-based indexing schemes in which documents and entities are organized into layered structures that guide retrieval and reduce search complexity (Wang et al. 2026). Other approaches explicitly reconstruct retrieval as a tree- or graph-traversal process, preserving structural relations during search and enabling multi-level semantic navigation (Wei et al. 2026).

Taken together with the structural failure points identified in standard RAG implementations (Barnett et al. 2024), these developments indicate a clear convergence toward retrieval systems in which structured taxonomic or ontological representations function not as passive metadata, but as active constraints on retrieval, query expansion, and semantic routing (Jiang et al. 2025; Gajderowicz, Bhardwaj, and Fox 2026).

Within this landscape, the Dewey Decimal Classification can be understood as a pre-existing, globally standardized, and human-curated ontology whose structural virtues function in ways analogous to the routing mechanisms now emerging in GraphRAG and ontology-guided frameworks. The notational hierarchy acts as an explicit routing mechanism across conceptual fields, while scope notes, relational cross-references (“see” and “see also”), and Relative Index entries serve as local boundary conditions that stabilize meaning, resolve lexical ambiguity, and anchor concepts within a robust associative network.

Consequently, the DDC offers foundational design principles for future retrieval architectures, navigating a pre-segmented interpretive space where meaning is rigorously co-produced by vector similarity and taxonomic hierarchy.

Discussion

In an information ecosystem increasingly characterized by the hyper-personalization of algorithmic rankings, the deceptive fluency of generative AI, and the erosion of document integrity via decontextualized fragments, the Dewey Decimal Classification emerges as a vital, structured counter-model. While contemporary search environments prioritize engagement metrics and probabilistic plausibility, the DDC preserves explicit relationships between intellectual domains through a transparent, hierarchical architecture that safeguards disciplinary context.

The arguments developed throughout this paper converge on a central premise: the DDC does not merely organize physical or digital objects; it organizes knowledge itself. This distinction becomes paramount at a historical juncture where the boundary between information retrieval and text generation is increasingly blurred. Keyword search isolates individual terms, and generative AI synthesizes plausible prose; the DDC, conversely, situates information within an inspectable, relational framework that actively fosters user learning and critical navigation. These modalities are fundamentally complementary, and the objective of this analysis has been to illuminate the unique, irreplaceable value of classification systems at a moment when they risk being overshadowed by statistical computational paradigms.

To maintain analytical rigor, some structural objections to this position must be explicitly acknowledged. First, the operational scope of the DDC is inherently constrained by its traditional focus on monographic literature, leaving journal articles, conference proceedings, and digital-born grey literature largely unindexed by its main notations.⁷ While this is a genuine limitation, it points toward the need for future integrations with complementary indexing tools rather than an abandonment of the model. Second, the DDC carries its own well-documented historical and ideological biases, including Western epistemic frameworks and Anglo-American centrality. However, unlike proprietary “black-box” ranking algorithms whose biases remain hidden and commercially protected, the DDC operates as an open, transparent, and publicly documented system whose shortcomings can be systematically identified, debated, and corrected by an international community of professionals. Third, practical deployment reveals usability barriers, as DDC-based retrieval demands a level of structural familiarity that contemporary users rarely possess, often requiring the professional mediation of librarians. This challenge underscores the critical importance of investing in intuitive, user-facing interfaces—such as WebDeweySearch—that make hierarchical navigation seamlessly accessible to non-experts, while reaffirming the role of information literacy programs as an indispensable alternative to passive reliance on AI chatbots. Finally, it must be noted that the institutional strengths of the DDC do not emerge in isolation, but rather from its systemic interaction with professionally curated library collections and the robust cataloging environments in which it is traditionally embedded.

Ultimately, the proposal to deploy the DDC as a semantic scaffolding layer within RAG architectures remains speculative. Whether this specific classification scheme constitutes the optimal structure for technical implementation in AI workflows is an open empirical question requiring systematic prototyping, benchmarking, and evaluation. This proposition is advanced here as a

⁷ Integration with Linked Open Data (LOD) is helping, to some extent, to extend its scope to other types of digital resources as well.

programmatic research proposal rather than a finalized empirical result. Given the theoretical and argumentative nature of this position paper, its validity rests upon the logical coherence of its arguments and the integration of current literature. It aims to contribute to a necessary and ongoing dialogue within Library and Information Science (LIS) regarding the enduring relevance of structured knowledge organization in a machine-driven age.

Conclusion

This paper has re-examined the Dewey Decimal Classification not as an obsolete legacy system for physical shelf arrangement, but as a sophisticated, dynamic infrastructure for knowledge organization. Within contemporary library environments, document interpretation, classification operations, and access notations are structurally intertwined. Through its rigorous network of hierarchical notations, scope notes, and cross-references, the DDC reveals an intricate semantic architecture that maps information spaces through explicit conceptual distinctions and disciplinary relationships.

The central finding of this study is therefore not a proposal to transform the DDC into a tool within the computational system, nor is it a claim that classification schemes and information retrieval systems are equivalent. Rather, the proposed analysis suggests that systems such as the DDC represent modes of organising knowledge in which the meaning of each notation is defined through an explicit and historical hierarchical structure, rather than through statistical analysis alone. This mode remains relevant even in a context dominated by search algorithms and GenAI models.

From this perspective, the significance of the DDC extends beyond its traditional role as a bibliographic classification system used for the physical organisation of materials in libraries; rather, it lies in its ability to highlight the processes that remain central to any form of mediation of recorded knowledge: the segmentation and restructuring of the space of meanings, the management of ambiguity, and the interpretation of documents. These operations are bound to reappear in the new digital context as well, albeit often without explicit reflection capable of drawing on the insights gained from 150 years of bibliographic classification history.

At the same time, this work is limited to a theoretical argument developed through conceptual reconstruction and structural analogy. Consequently, the theses put forward here should be understood as heuristic contributions aimed at reaffirming certain fundamental characteristics of knowledge organisation systems.

It is necessary to acknowledge that the DDC remains historically and culturally situated within specific institutional and editorial frameworks. Any contemporary reinterpretation must therefore carefully navigate the risk of abstracting the system from its original operational functions simply to emphasize structural features that align with new computational AI paradigms.

Future research should expand upon these theoretical foundations in several distinct directions. First, a deeper exploration into the epistemological foundations of bibliographic classifications is required to reinforce their structural integrity, ensuring that their core strengths are preserved amid ongoing technological disruptions. Second, empirical investigations should systematically evaluate how the structural properties of the DDC can optimize the design of novel document discovery platforms. Increased awareness of taxonomic benefits should inspire retrieval systems

that, in tandem with advanced AI models, highlight hierarchical relationships rather than generating flat, uncontextualized results. Third, further scholarly attention must be dedicated to the broader epistemological implications of introducing explicit classification constraints into probabilistic and opaque information environments. Finally, the professional community must engage in a rigorous reflection on the role of the human indexer. Document classification is fundamentally an act of intellectual interpretation rather than a mechanical transcription—a vital human mediation that risks being marginalized due to immediate economic pressures rather than structural or epistemological obsolescence.

References*

- Afane, Mohamed, Emaan Hariri, Derek Ouyang, and Daniel E. Ho. 2026. *Benchmarking Legal RAG: The Promise and Limits of AI Statutory Surveys*. In *CS&law '26: ACM Symposium on Computer Science and Law, March 03–05, 2026*, Berkeley, CA; New York, NY: Stanford Digital Horizons Observatory. https://dho.stanford.edu/wp-content/uploads/Stat_Surveys.pdf.
- Algaba, Andres, Vincent Holst, Floriano Tori, Melika Mobin, Brecht Verbeke, Sylvia Wenmackers, and Vincent Ginis. 2025. "How Deep Do Large Language Models Internalize Scientific Literature and Citation Practices?." *arXiv preprint*. <https://arxiv.org/html/2504.02767v1>.
- Bender, Emily M., Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?." *Association for Computing Machinery*. <https://dl.acm.org/doi/10.1145/3442188.3445922>.
- Barnett, Scott, Stefanus Kurniawan, Srikanth Thudumu, Zach Brannelly, and Mohamed Abdelrazek. 2024. "Seven Failure Points When Engineering a Retrieval Augmented Generation System." In *Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering – Software Engineering for AI (CAIN 2024)*, 194–99. <https://dl.acm.org/doi/10.1145/3644815.3644945>.
- Boateng, Frank. 2025. "The Transformative Potential of Generative AI in Academic Library Access Services." *Information Services & Use* 45 (2).
- Bowker, Geoffrey C., and Susan Leigh Star. 2000. *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.
- Carr, Nicholas. 2025. *Superbloom: How Technologies of Connection Tear Us Apart*. New York: W. W. Norton & Company.
- Cruz, Tiago da, Bernardo Tavares, and Francisco Belo. 2025. "Ontology Learning and Knowledge Graph Construction: A Comparison of Approaches and Their Impact on RAG Performance." *arXiv preprint*. <https://doi.org/10.48550/arXiv.2511.05991>.
- Dahl, Matthew, Varun Magesh, Mirac Suzgun, and Daniel E. Ho. 2024. "Large Legal Fictions: Profiling Legal Hallucinations in Large Language Models." *Journal of Legal Analysis* 16 (1): 64–93. <https://doi.org/10.48550/arXiv.2401.01301>.
- Floridi, Luciano. 2020. *Pensare l'infosfera: La filosofia come design concettuale*. Milano: Raffaello Cortina Editore.
- Floridi, Luciano. 2023. *The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities*. Oxford: Oxford University Press.
- Friedrich, Tilman, Karl Akbari, Daniel Fürstenau. 2026. "Data Governance Practices for Generative AI Powered Organizational Knowledge Management Systems Using Retrieval Augmented Generation." In *Proceedings of the 59th Hawaii International Conference on System Sciences*. <https://scholarspace.manoa.hawaii.edu/items/b7fce874-a2a4-49ee-9180-d9bf5bd987d5>.

* Last date of access to the websites: June 30, 2026.

- Gajderowicz, Bart, Aviral Bhardwaj, and Mark Fox. 2026. "RAG and Ontologies for Information Retrieval: A Literature Review." In *Proceedings of the Joint Ontology Workshops (JOWO 2025), ONTOLLM Workshop, Catania, Italy*. <https://ceur-ws.org/Vol-4176/ontollm-2.pdf>.
- Gillespie, Tarleton. 2010. "The Politics of 'Platforms'." *New Media & Society* 12 (3): 347–64.
- Huang, Haoyu, Yongfeng Huang, Junjie Yang, Zhenyu Pan, Yongqiang Chen, Kaili Ma, Hongzhi Chen, and James Cheng. 2025. "Retrieval-Augmented Generation with Hierarchical Knowledge." *arXiv preprint*. <https://doi.org/10.48550/arXiv.2503.10150>.
- Jelani A., Sayed Mohamed, and Kairali Ashkar. 2025. "AI in Libraries: Assessing ChatGPT's Capabilities for Dewey Decimal Classification Systems." In *Conference: Building sustainable libraries for connecting communities: Librarianship, technology and collaboration*, 59–70. Tamilsuvadi Publication.
- Jiang, Pengcheng et al. 2025. "A Survey on Retrieval and Structuring Augmented Generation with Large Language Models." In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 2025)*, Toronto, Canada. arXiv:2509.10697.
- Linardon, Jake, Hannah K Jarman, Zoe McClure, Cleo Anderson, Claudia Liu, and Mariel Messer. 2025. "Influence of Topic Familiarity and Prompt Specificity on Citation Fabrication in Mental Health Research Using Large Language Models: Experimental Study." *JMIR Mental Health* 12. <https://mental.jmir.org/2025/1/e80371>.
- Lissack, Michael, and Brenden Meagher. 2024. "LLM's Embody Sloppy Science: Reliance on Pattern Matching Alone is Slodderwetenshap." SSRN. <https://ssrn.com/abstract=4949825>.
- Makri, Stephann and Ann Blandford. 2012a. "Coming across information serendipitously - Part 1. A process model." *Journal of Documentation*, 68 (5): 684-705. <https://doi.org/10.1108/00220411211256030>.
- Makri, Stephann and Ann Blandford. 2012b. "Coming across information serendipitously - Part 2: A classification framework." *Journal of Documentation*, 68 (5): 706-24. <https://doi.org/10.1108/00220411211256049>.
- Makri, S., Ann Blandford, Mel Woods, Sarah Sharples, and Deborah Maxwell. 2014. "Making 'my own luck': How people perceive, and change their behaviour toward, serendipity." *Journal of the Association for Information Science and Technology (JASIST)*, 65 (11): 2181-94. <https://doi.org/10.1002/asi.23200>.
- Matsieli, Molefi, and Stephen Mutula. 2025. "Generative AI and the Information Society: Ethical Reflections from Libraries." *Information* 16 (9): 771.
- Merton, Robert K. 1968. "The Matthew Effect in Science." *Science* 159 (3810): 56–63.
- Merton, Robert K. 1988. "The Matthew Effect in Science, II: Cumulative Advantage and the Symbolism of Intellectual Property." *Isis* 79 (4): 606–23.
- Miksa, Francis L. 1998. *The DDC, the Universe of Knowledge, and the Post-modern Library*. Albany, NY: Forest Press.

- Mitchell, Joan S., and Diane Vizine-Goetz. 2009. "Dewey Decimal Classification." In *Encyclopedia of Library and Information Science*, 3rd ed., edited by Marcia J. Bates and Mary Niles Maack. Boca Raton, FL: CRC Press. <http://www.oclc.org/research/publications/library/2009/mitchell-dvg-elis.pdf>.
- Mitchell, Joan S. 2013. "Dewey linked data: Making connections with old friends and new acquaintances." *JLIS.it* 4 (1): 177-99. <https://www.jlis.it/index.php/jlis/article/view/263>.
- Mitchell, Melanie. 2022. *L'intelligenza artificiale: Una guida per esseri umani pensanti*. Torino: Einaudi.
- Niu, Cheng, Yuanhao Wu, Juno Zhu, Siliang Xu, KaShun Shum, Randy Zhong, Juntong Song, and Tong Zhang. 2024. "RAGTruth: A Hallucination Corpus for Developing Trustworthy Retrieval-Augmented Language Models." In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024)*, 10862–78. <https://aclanthology.org/2024.acl-long.585/>.
- Noble, Safiya Umoja. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.
- OCLC. 2024. OCLC builds on its linked data strategy by creating uniform resource identifiers for Dewey Decimal Classification numbers. <https://www.oclc.org/en/news/releases/2024/20240815-uniform-resource-identifiers-for-ddc-numbers.html>.
- OCLC. 2025. 2024: A Year of Accelerating Linked Data. <https://www.oclc.org/en/news/announcements/2025/2024-accelerating-linked-data.html>.
- OCLC. 2026. *Organize Your Materials with the World's Most Widely Used Library Classification System*. <https://www.oclc.org/en/dewey.html>.
- Panzer, Michael. 2008. "From Heteroglossia to Polyglottism: Multilingual MelvilClass." Presentation at Dewey Breakfast/Update, ALA Annual Meeting, Anaheim, July 3. http://www.oclc.org/dewey/news/conferences/multilingual_melvil_class.pdf.
- Pasquale, Frank. 2015. *The Black Box Society*. Cambridge, MA: Harvard University Press.
- Perc, Matjaž. 2025. "Counterfeit Judgments in Large Language Models." *Proceedings of the National Academy of Sciences* 122 (48).
- Pievani, Telmo. 2021. *Serendipità: L'inatteso nella scienza*. Milano: Raffaello Cortina Editore.
- Pollicino, Oreste, and Paula Dunn. 2024. *Intelligenza artificiale e democrazia: Opportunità e rischi di disinformazione e discriminazione*. Milano: Bocconi University Press.
- Santosa, Faizhal Arif. 2025. "Artificial Intelligence in Library Studies: A Textual Analysis." *JLIS.it* 16 (1): 61–71.
- Svenonius, Elaine. 2000. *The Intellectual Foundation of Information Organization*. Cambridge, MA: MIT Press.
- Thompson Klein, Julie. 2021. *Beyond Interdisciplinarity: Boundary Work, Communication, and Collaboration*. Oxford: Oxford University Press.
- Thompson Klein, Julie, and Thorsten Philipp. 2023. "Interdisciplinarity." In *Handbook Transdisciplinary Learning*, edited by Thorsten Philipp and Tobias Schmohl, 195–204. Bielefeld: transcript Verlag. <http://doi.org/10.1515/9783839463475-021>.

Wang, Shu, Yixiang Fang, Yingli Zhou, Xilin Liu, and Yuchi Ma. 2026. “ArchRAG: Attributed Community-based Hierarchical Retrieval-Augmented Generation”. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 40 (19): 15868–76. <https://dl.acm.org/doi/10.1609/aaai.v40i19.38619>.

Wei, Chunyu, Huaiyu Qin, Siyuan He, Yunhai Wang, and Yueguo Chen. 2026. “T-Retriever: Tree-based Hierarchical Retrieval Augmented Generation for Textual Graphs.” In *Proceedings of the AAAI Conference on Artificial Intelligence*, 40 (19): 15924–32. <https://dl.acm.org/doi/10.1609/aaai.v40i19.38625>.

Wiegand, Wayne A. 1996. *Irrepressible Reformer: A Biography of Melvil Dewey*. Chicago: American Library Association.