Digital and Smart Libraries Researches

O R I G I N A L   A R T I C L E

Identifying Effective Factors in Information Retrieval from Linked Databases in Virtual Libraries

Zahra Dayyan[1], Safiyeh Tahmasebi Limooni[2*], Seyed Ali Asghar Razavi[3]

1. Ph.D. Student, Department of Knowledge and Information Science, Bab.C., Islamic Azad University, Babol, Iran.
2. Department of Knowledge and Information Science, Bab.C, Islamic Azad University, Babol, Iran.
3. Department of Knowledge and Information Science, Bab.C, Islamic Azad University, Babol, Iran.

*Correspondence
Safiyeh Tahmasebi Limooni
E–mail:sa.tahmasebi2@iaviac.ir

A B S T R A C T

Introduction
Semantic web and linked data technologies offer transformative potential for knowledge organization in digital libraries, enabling machine-readable connections that enhance discoverability and interoperability. Virtual libraries, as major digital repositories, are ideal for implementing these technologies. However, their practical adoption, especially in contexts like Iran, faces significant challenges. These include data quality issues, metadata inconsistencies, vocabulary mismatches, structural heterogeneities, and language-specific barriers. For Persian-language resources, unique orthographic features and native subject systems further complicate semantic linking and increase error rates. Additionally, gaps exist between the sophistication of linked data systems and the skill levels of both end-users and information professionals, while user interfaces often fail to support the exploratory search that linked data enables. Prior research lacks a systematic, holistic investigation of factors affecting information retrieval from linked databases in Iranian virtual libraries. Consequently, this study aimed to identify these critical factors and develop a comprehensive conceptual model illustrating their dynamic interrelationships to provide a strategic framework for both theory and practice.

Methodology
This study employed a qualitative methodology based on the grounded theory approach by Glaser and Strauss. This interpretive, inductive paradigm was chosen for its suitability in exploring a complex phenomenon where pre-existing theory is limited, allowing theory to emerge directly from empirical data. The target population consisted of experts with substantial knowledge and experience in Information Science, Information Technology, and Computer Science. Participants were selected via purposive and theoretical sampling, with strict inclusion

criteria: a minimum doctoral degree and at least five years of direct experience with linked data, digital libraries, or semantic information retrieval. Data were collected through 15 in-depth, semi-structured interviews, continuing until theoretical saturation was achieved after the twelfth interview, with three additional confirmatory interviews. Interviews focused on participants' experiences, perceptions, and insights regarding enablers and barriers in linked data information retrieval. Data analysis followed the grounded theory stages of open, axial, and selective coding. Open coding involved line-by-line analysis to identify initial concepts. Axial coding refined these concepts into broader categories and subcategories, exploring their relationships. Selective coding integrated all categories into a coherent theoretical model. To ensure trustworthiness, validation techniques included member checking, peer review, detailed methodological documentation, and test-retest procedures for coding consistency.

Findings

The rigorous analysis of the qualitative data yielded a rich and nuanced understanding of the ecosystem surrounding information retrieval from linked databases. The findings culminated in the identification of four principal, interdependent categories of factors that collectively determine the efficacy and success of the retrieval process. These categories form the pillars of the proposed conceptual model, emphasizing that optimal performance requires their synergistic integration rather than isolated optimization:

Technical Factors

This dimension covers the essential infrastructure and system capabilities needed to store, process, and query linked data effectively and dependably. It consists of four key components. First, Infrastructure Stability and Performance involves critical requirements like high system availability, fast response times, strong processing power for complex queries, scalability for increasing data, and reliable backup systems. Second, Query Efficiency and Optimization focuses on the system's internal sophistication, including advanced indexing, query optimizers for efficient searches, caching for common queries, and detailed performance analysis. Third, Quality of Links to External Data is central to linked data's value, dealing with the breadth, reliability, and maintenance of connections to outside datasets—encompassing link quantity, authority, consistency, and upkeep. Fourth, Advanced Semantic Capabilities refers to intelligent system features, such as reasoning engines for drawing conclusions, support for custom inference rules, navigation of ontology structures during search, and tools like fuzzy text search.

Content Factors

This dimension concerns the inherent quality, structure, and currency of the data within the linked database. High-quality content is vital for any effective retrieval system and includes four aspects. Data Accuracy and Correctness is the basic requirement that data be error-free, factually correct, validated before entry, and supported by user error-reporting and duplicate minimization. Semantic Richness and Depth goes beyond accuracy to involve a rich data model, featuring diverse entity relationships, detailed ontologies, deep classification systems, thorough descriptions, and varied data types like geographic or temporal information. Controlled Vocabularies involve systematic terminology management using standards such as SKOS, including synonym networks, multilingual support, selection guidance, and links to external ontologies for better interoperability.

Human and Organizational Factors

This category emphasizes the crucial role of people, their skills, and the institutional environment that supports them. Technology and content alone are insufficient without skilled human involvement and organizational backing, divided into four elements. User's Linked Data Literacy involves the user's basic understanding of linked data concepts like entities and RDF, ability to read data visualizations, skill in forming simple queries, and knowledge of ontologies. User's Information and Subject Literacy covers broader information skills, such as clearly defining information needs, effectively using keywords, critically evaluating results, and having sufficient topic knowledge. Active Role of Librarians sees librarians shifting to key mediators, requiring deep awareness of linked data capabilities, ability to help users craft queries, leadership in training programs, and bridging user needs with technical complexity.

User Interface and User Experience (UI/UX) Factors

This dimension focuses on the design and functionality of the system's front-end, which directly handles all human-computer interaction. A powerful backend becomes inaccessible without an intuitive and supportive interface, defined by four characteristics. User-Friendly Design follows principles of clarity, with logical layout and navigation, responsiveness across devices, familiar visual elements, and accessibility standards for inclusivity. Advanced Search Features go beyond basic search boxes to offer tools like graphical query builders that simplify complexity, faceted filtering options, and capabilities to save, reload, and share search strategies. Results Visualization and Display takes advantage of linked data's graph nature by providing visual representations of results and their connections, alongside standard list and table views, plus mapping for spatial data and explanations for inferred results. Support and Guidance includes integrated help systems like context-sensitive assistance, pre-made query examples, chatbots or FAQs for quick solutions, and easy feedback and error-reporting channels.

Discussion and Conclusion

The study demonstrates that effective information retrieval from linked databases is not merely a technical challenge but a complex, multi-dimensional socio-technical phenomenon. Success depends on the harmonious and dynamic integration of four inextricably linked pillars: Technical Robustness, Content Credibility, Human-Organizational Competence, and User-Centric Design. The proposed model posits that these dimensions interact systemically; weakness in any one fundamentally undermines the entire system. For example, an advanced semantic reasoner (Technical) is useless with inaccurate data (Content), and perfectly structured data (Content) is inaccessible without user literacy (Human) or a clear interface (UI/UX).