

Leveraging Graph Databases for Dynamic Relationship Management and Real-World Data Representation

Venkata Naga Sai Kiran Challa

USA

ABSTRACT

This paper explores the advantages of using graph databases as the primary mode of data storage to represent complex, real-world relationships. Unlike traditional relational databases, graph databases naturally handle dynamic relationships and evolving schemas, making them ideal for applications with deeply nested and frequently changing data. We delve into the fundamental principles of graph databases, including index-free adjacency, native graph storage, and specialized query languages like Cypher and Gremlin. Additionally, we present a practical implementation of a social media platform's data model in both graph and relational databases, highlighting the simplicity and efficiency of graph databases in managing complex relationships. Through a case study involving advertisement recommendations based on user behavior, we demonstrate the superior capabilities of graph databases in handling dynamic relationship updates and providing insights via graph analytics. This paper aims to provide a comprehensive understanding of the practical applications and benefits of graph databases in modern data management.

*Corresponding author

Venkata Naga Sai Kiran Challa, USA.

Received: December 05, 2021; **Accepted:** December 08, 2021; **Published:** December 18, 2021

Keywords: Graph Databases, Relational Databases, Dynamic Relationships, Real-World Data Representation, Neo4j, Cypher, Graph Analytics, Social Media Data Model, Advertisement Recommendations, Machine Learning Integration

Introduction

From the technical-scientific revolution following the Second World War, the large volume of information generated in the growing number of areas of knowledge began to demand a higher level of informational organization, as information must be ordered, structured or materially fixed, making it if a document, otherwise it will remain amorphous and unusable. It can be said that the advances that occurred from the 1950s to the present day were relevant and marked the development both in the form of storage and representation and in the retrieval of information.

Thus, users' needs become the central point of studies in the area, since information retrieval is the main objective of the entire area of the Information Organization. As a result, librarians began to face new challenges with changes in the conceptualization and way of delivering and assessing library services, therefore having to assume several roles, in addition to what was already foreseen.

The storage of information, previously done only on the hard drives of computers with large processing capacity, now takes place in a different way, so that there is the possibility of accessing files, data and applications anywhere and at any time, using them, whether both a computer and mobile devices, as long as there is an Internet connection, through storage in so-called "clouds".

Within the scope of the processes that involve the organization and

retrieval of information, in the context of Library and Information Science (BCI), the materiality of these processes is considered in the sphere of Information Retrieval Systems (SRI). For Lancaster (1978), the main function of an SRI is to act as an interface between a particular population of users and the universe of information resources in printed or other form. It is in this environment that the cataloging, indexing and classification processes support the organization and retrieval of information from various information supports. The product of these activities is the creation of catalogs of a given physical collection, or of a digital library database, or even of an online catalog and digital repositories.

The effective performance of SRI depends on the quality with which the organization of information is carried out, which will imply a retrieval of the most relevant and pertinent information, since these procedures are conditioned to each other, creating a bridge between the input and output of the information. In this context, attention must be paid to the exhaustiveness and specificity variables when carrying out the indexing, as well as to the level of recall and precision that the system proposes to serve its users. This will influence the system's information retrieval capacity, resulting in relevant and pertinent documents for the user. Relevance consists of the degree of similarity between terms that make up user search expressions and their occurrence in documents in the collection or in indexing terms. Relevance is the relationship that exists between the information obtained in a search that responds to the user's need or demand for information, that is, information that is useful to the user.

In this article, the role of information organization and retrieval within the scope of an Information Retrieval System (SRI)

is discussed, with the aim of evaluating the impact of the exhaustiveness and specificity variables and recall and precision measures, as well as the concepts of relevance and pertinence in SRIs. As a contribution, a flow proposal for a user-centered Information Storage and Retrieval System is presented, bringing together the various aspects presented.

From such discussions, it was considered to bring, in a didactic way, very important concepts for professional practice and that could assist in procedures related to the elaboration of indexing policies, fundamental procedures for the training of professionals who are dedicated to information processing. In this sense, a flow is presented in which the different aspects of variables and measurements were brought together within a proposal for an Information Storage and Retrieval System (SARI).

In the next sections, the methodology used to carry out this study is first described, and the organization and information retrieval as they are being considered within the scope of this study is presented. Subsequently, the SRI and their flows are described, and the processes to help users achieve relevance and pertinence in their searches; Soon after, such discussions are brought up and consolidated through the proposal of a SARI and, finally, the final considerations are presented.

Methodology

This study is characterized as descriptive and exploratory, with the aim of understanding, through a flowchart, how the processes of organization and retrieval of information (ORI) impact on the variables exhaustiveness and specificity and, consequently, on the recall and precision measures, with the aim of achieving greater relevance and pertinence in the results of user queries within the scope of an SRI.

To study how these variables and measures are highlighted in the structures of a flowchart of SRI processes in the context of ORI, it was proposed, firstly, to prepare a narrative review of the literature, since the research is conducted by more open questions, to: (1) map the main SRI flowcharts present in the literature and (2) collect initial inputs to support the theoretical-methodological proposal of an SRI that encompasses all these elements. These criteria are justified because the flowcharts that represent the processes of an SRI do not explicitly place these variables and measures in this specific context. In the case in question, five flowcharts were selected based on the following criteria: (1) the seminal authors in the area and (2) those most cited in the literature.

To this end, an exploratory search was carried out, without temporal definition, on Google Scholar, and in five specialized databases: Library and Information Science Abstracts (LISA), Information Science & Technology Abstracts (ISTA), Library, Information Science & Technology Abstracts with Full Text (LISTA), Scopus and Web of Science. These bases were chosen considering their relevance and relationship with the area and sub-area of knowledge delimited for this review, using the following search expressions, presented in Table 1.

Table 1: Expressions Used to Search the Literature

Acronym	Search expressions
E1	"Information Retrieval System" OR "SRI"
E2	Information Retrieval Systems" OR jrIR system"
E3	"Indexing Policy" AND "Information Organization" AND "Information Retrieval"
E4	"Indexing Policy" AND "Organization of Information" AND "Information Retrieval"
E5	"Exhaustiveness" AND "Specificity" AND "Recall" AND "Precision" AND "Relevance" AND "Relevance"
E6	"Exhaustivity and Specificity" AND "Precision and Recall" AND "Relevance and Pertinence"

Source: Prepared by the author.

Initially, 48 documents were selected from those recovered, using three criteria: (1) documents that dealt with the information retrieval system and indexing policy; (2) documents that presented studies on the exhaustiveness and specificity variables, on recall and precision measures, and on relevance and pertinence within the scope of the SRI; (3) documents that had these terms in the title or keywords. Of these 48 selected documents, 33 documents were used that specifically deal with the topic, among which are the five flowcharts used to support a new structure that would contextualize the position of each of these variables and measures within a proposal for a System Storage and Retrieval System (SARI), as presented in section 6 of this article.

Organization and Retrieval of Information

In this section, the concepts of information organization and information retrieval are presented, which are considered relevant to the context of this article.

Organization of Information

The area of Information Organization (IO) comprises all studies related to the processes and instruments used in the organization of information resources of any nature, with the aim of enabling the meeting of the information needs of a given community of users.

Dahlberg (2006) defines information organization as the ordering of objects in order to create a link between the object of an area and its own activity. Novellino (1996) states that the information representation process is characterized, mainly, by replacing the descriptive and thematic content of a document with an abbreviated description, which will be stored for later retrieval.

Thus, OI aims to represent, store and retrieve information. According to Barreto, the objective of the information organization process is to enable and facilitate access to information, which, in turn, has the competence and intention of producing knowledge [1]. To this extent, it is understood that the input of inconsistent data will result in the output of inconsistent data. In this process, cognitive mechanisms are activated that influence both the input and output of the information retrieval system, because they are dependent on the way we use our mind to carry out abstractions.

In the context of Library Science and Information Science, the representation of information is carried out through the cataloging, indexing and classification processes in an SRI, as illustrated in Figure 1.

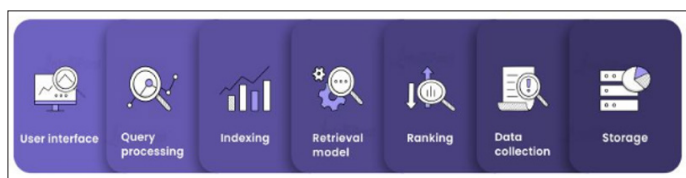


Figure 1: Architecture of Information Retrieval System

In the cataloging process, also known as descriptive representation, a bibliographic item is described with the aim of making it unique among the others in a given collection, allowing it to be identified, located and represented in catalogs. Mey (1995, p. 5) considers that “cataloging is the study, preparation and organization of coded messages, based on existing items or items that can be included in one or several collections, in order to allow intersection between the messages contained in the items and users’ internal messages.”

Indexing is another important representation process that occurs within an SRI, in which the indexer is expected to read the document and distinguish between relevant and peripheral information to better represent it for subsequent retrieval. According to the ISO 5963-1985 standard, indexing is seen as “[...] the representation of the content of documents by means of special symbols, whether taken from the original text or chosen from an information or indexing” [2]. This process is carried out in two stages: the first is analyzing the document to identify its informational content; the second, the translation of concepts in terms of an indexing language, using knowledge organization systems, such as thesauruses and bibliographic classification systems.

While cataloging describes the physical characteristics of a bibliographic item and indexing is concerned with issues involving the intellectual content of the document, classification, as a process, involves the orderly and systematic assignment of each entity to just one class within a system of mutually exclusive, non-overlapping classes based on similarities and differences. According to Tristão et al (2004, p. 163), “classification is a mental process through which we can distinguish things, beings or thoughts based on their similarities or differences.” This is a fundamental activity of the human mind that processes ideas and distinguishes them based on common characteristics. According to Lima (2021), other concepts can still be accepted for the word classification:

Depending on the point of view, classification is considered a discipline, but it can also be the product that results from the act of classifying and, simultaneously, it is the tool used to carry out the classification process (LIMA, 2021).

In the area of Library Science and Information Science, at least four conceptualizations can be attributed to the word classification: classification has been studied as a discipline, as a process of grouping and ordering knowledge, as the product of the grouping and ordering process, and as an information representation instrument.

Information Retrieval

Information Retrieval (IR) is an original area of Computer Science (CC), and the expression was attributed to the American engineer Calvin Mooers, in 1951, who defined it, at the time, as a process that “[...] encompasses the intellectual aspects of describing information and its specificities for the search, in addition to any

systems, techniques or machines used to perform the operation”, within the scope of Library Science and Information Science, states that this conception of information retrieval brought by Mooers was centered on the construction of systems, however, from the end of 1970, it expanded to a user-centered approach, taking into account users' mental models [3-4].

For Ferneda (2003), the IR process consists of “identifying, in the set of documents (corpus) of a system, that meet the user's need for information”. Thus, IR is considered an important operation in an Information Retrieval System, which aims to relate the user's search with the items stored in the database, through a set of interconnected elements of information record processing routines, aiming to meet the information needs of a community of users. However, some authors define IR from different approaches.

For information retrieval involves intellectual aspects of information description and search specifications, in addition to the systems, techniques and equipment that are applied to carry out the entire process [4]. While Baeza-Yates and Ribeiro-Neto point out that [5]:

Information retrieval deals with the representation, storage, organization and access of information items such as documents, web pages, online catalogs, structured and semi-structured records, multimedia objects. The representation and organization of information items must provide users with ease of access to the information of interest.

Salton considers IR as an area of research that is concerned with the structure, analysis, organization, storage, retrieval and search of information [6]. In turn, Lancaster (1993) considers it as a process of searching a collection of documents in order to identify those texts that deal with a certain subject.

According to Rowley (1994, p. 113), IR is the “process of locating documents and items of information that have been stored”. For the author, the process is made up of three elements: consultation, comparison and result. The query is the user's question transformed into a search strategy; comparison is the action of checking whether the question asked matches the stored items; and the result is the list or items that match the user's search (ROWLEY, 2002). However, the information retrieval process depends greatly on the steps carried out at input (cataloging, indexing, classification) and storage, which directly impact the searches carried out in an Information Retrieval System.

Information Recovery System (SRI): An Analysis from the Flows For Rowley (2002), SRIs and computers were almost used as synonyms, however, before the emergence of any computer and information technology itself, paper-based card and file systems already existed. The SRI is an integral part of a communication system and/or information system. These systems allow users to search for information in a collection of documents (or other sources of information) through queries generally formatted as a set of metadata, and obtain it in a way that meets their needs with relevance and pertinence. But SRIs deal with at least two different problems regarding users' information needs or queries: (1) they must distinguish and identify the relevant information related to the query, and (2) they must get the answer quickly.

Salton and McGill (p. xi) define an SRI as “a system that deals with the representation, storage, organization and access to items of information, and this can be in a physical or digital collection [7].”

The authors consider it as an interconnected set of information record processing routines, with their own purposes and criteria, aiming to meet the information needs of a community of users. For Silva, Santos and Ferneda (2013, p. 29), information retrieval systems “have the function of representing the content of corpus documents and presenting them to the user in a way that allows them to quickly select items that satisfy fully or partially meet your need for information [...]”.

For Lancaster (p. 1), every SRI is composed of two subsystems: the information input (input) and the information output (output), represented in the form of a cycle, carried out by basically three stages: representation, storage and recovery, characterized as a continuous and feedback process [8].

When documents are selected to form a collection, whether physical or digital, they initially go through an organization taking into account the needs of the community to be served. This organization is carried out by the processes of cataloging, indexing, classification and summarization. Therefore, the SRI encompasses everything from organizational processes to the retrieval of information by the user.

There are several proposals in the literature to represent the flow of these subsystems of information input, representation and retrieval in an SRI. In this article, we will present some of these schemes, based on the literature, with the aim of verifying the structure and processes of their subsystems. For this, the following schemes were selected.

Lancaster, as early as 1978, in his book Information Retrieval Systems, identifies the elements that make up the SRI [9]. The subsystems identified by him are: selection and acquisition; indexing; vocabulary; search; user-system interaction (question negotiation); “match”¹. By calling all these elements subsystems, he brings to the fore the concept of interrelationship, in which there is a relationship between users, collections, vocabularies, indexing and retrieval. This means they affect each other. The overall view of an SRI presented by Lancaster, as we will see later, is a view of the whole, in which it is of fundamental importance to understand the behavior of each element so that an SRI can achieve its function, that is, enable consistent recoveries to its users.

For a greater understanding of this interrelationship, and pointing out that all actions always start from the identification of the user, as we will see later, we can present the following questions: what is your area of activity (subject), what activity do you carry out (teacher, student, researcher)?; what means of communication do you use?; what language do you use (scientific, technical, not specialist)?; and so on. The answers will provide elements to determine principles for forming the collection, for treating the document, for determining its classification and even actions that, apparently, are unrelated, such as the number of documents that a user can take on loan and the allowed period [10].

Based on this conception, Lancaster proposes two schemes in his publications (1986, 1993) [8]. In the first scheme, in his book Vocabulary Control for Information Retrieval (1986, p. 3), the author presents the components of an SRI. As can be seen, in its flow proposal, the representation of a document focused only on the indexing and cataloging processes [8]. The author did not insert details of the cataloging elements, leaving the indexing steps at the same level of understanding. After this organization, the documents are stored in a printed or digital database, in which

searches can be carried out to meet user requests, as can be seen in Figure 2.

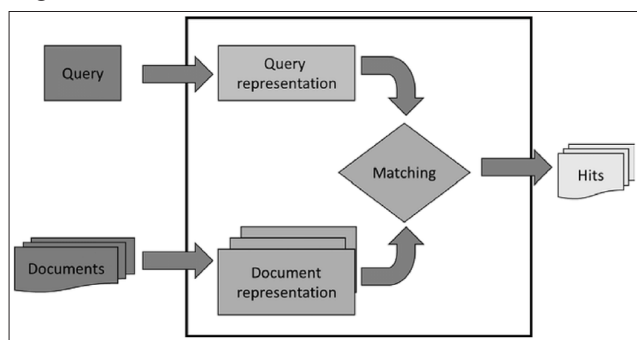


Figure 2: The Components of an Information Retrieval System

The author considers that cataloging has two important elements that must be considered: (1) the physical description of a document and (2) the choice of the access point to represent it. Regarding indexing, the author considers it as an intellectual process composed of two stages: conceptual analysis and translation, that is, the subject of a document is analyzed and, subsequently, a controlled vocabulary is used to standardize the representation of this subject. However, in his flow proposal, the author considers the conceptual analysis and translation stages within the indexing and cataloging processes, despite having differentiated them at the time of their descriptions.

Lancaster, in his book Indexation and Summaries: theory and practice (1993, p. 2), brings another, more expanded proposal, in which he also considers the preparation of indexes and summaries, as can be seen in Figure 3. These Modifications are also considered due to technological advances at the time.

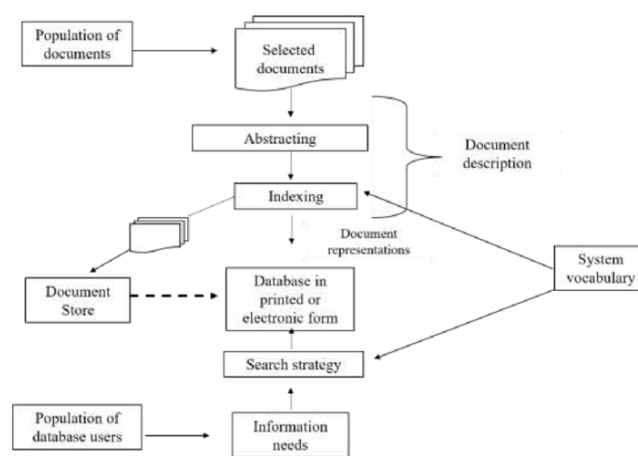


Figure 3: Role of Creating Indexes and Summaries in the Broader Framework of Information Retrieval

As can be seen, in this scheme, Lancaster precedes the writing of abstracts with indexing as a way of helping to condense the representation of the document's content, and considers them as the process of document description. On the other hand, the author points out the need to plan the search strategy according to users' needs. In both subsystems, the author suggests the use of controlled vocabulary to standardize the terminology used by the author with that of the user in their search. These terms assigned by the indexer become access points used to retrieve bibliographic items. A common feature in these schemes are the subject analysis and translation processes, which can occur both at the input, that

is, in the representation, and at the output, which is the retrieval of information. In addition to these aspects, this 1993 flow also includes the inclusion of the specification and differentiation of databases in printed or electronic form.

It appears that Lancaster flows alert us to the decision of which aspects of a document will be represented in an SRI, as well as what level of specificity or exhaustiveness will be attributed to the descriptors and which relate to the set of decisions adopted by the policy of SRI indexing. This decision permeates the type of vocabulary adopted in the System, as both, the indexing policy and vocabulary updating, go hand in hand.

Soergel adds another aspect to be observed in SRI, the issue involving information storage, in which information items need to be processed, searched, retrieved and disseminated to various user communities [11]. In this sense, Soergel considers that an Information Storage and Retrieval System (Information Storage and Retrieval - ISAR System) is a subsystem of an information system as a whole [11]. The author presents an Information Storage and Retrieval System structure, in which he emphasizes not only the information input and information output subsystems, but also the storage subsystem, as can be seen in Figure 4.

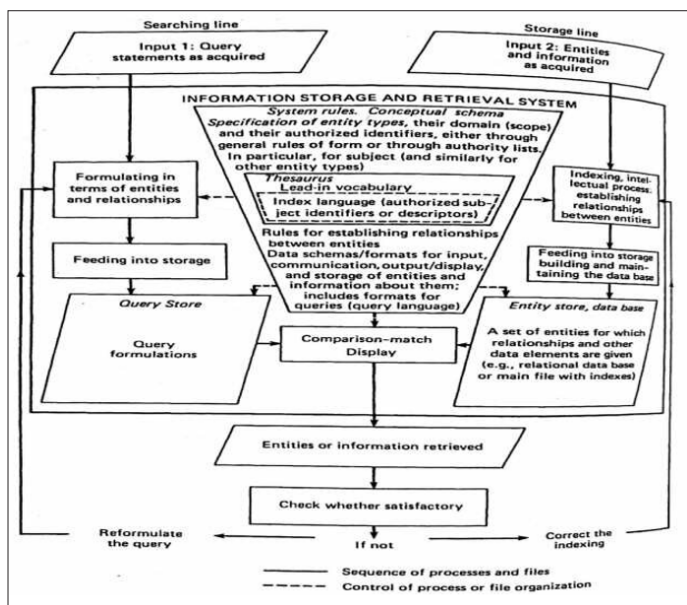


Figure 4: The Structure of the Information Storage and Retrieval System

Unlike the schemes presented by Lancaster (1986, 1993), Soergel begins by first presenting the information output system, which the author calls Search Line, in which users' profiles are studied and searches are formulated [8]. starting from terms and relationships to reach the results represented at level 1. At the other extreme, there is an information input system, with documents and data, in which the information representation processes take place based on the indexing processes (descriptive and subject), so that bibliographic items can be stored and made available to meet the needs of your community. The author adds to this flow a stage of comparing the results, when the relevance of the recovered items is assessed in relation to the search strategy carried out, and the potential relevance in the occurrence in documents in the collection.

Another important point in this scheme presented by Soergel is the suggestion to be made, based on the potential relevance

or not of the analysis of the documents, to redo the search or to correct the indexing process carried out for that document [11]. This analysis, when carried out, contributes to maintaining the efficiency and accuracy of the system.

Chu presents the main components of an SRI, which he calls the Information Representation and Retrieval (RRI) process [12]. For the author, these components are: the database, the search engine, the language and the interface, as illustrated in Figure 5.

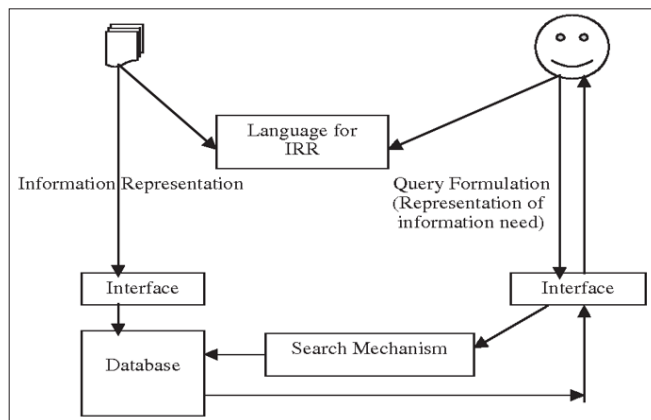


Figure 5: Information Representation and Retrieval Process

In this flow, we see that the process of representing information appears with emphasis instead of naming the processes according to the previous schemes. The author also presents the languages and the information retrieval process at the same level, and highlights the interface for accessing the database and the interface for formulating the user's search. Another important point, highlighted by Chu, is the role of the information professional in representing information using controlled vocabulary, as discrepancies may occur during this process, which can cause problems in returning searches, with low relevance and pertinence [12].

Chu points out that three problems can occur in processes within the scope of an SRI: (1) the documentary typology, which may not be in a format that can be archived; (2) the difficulty of finding a term that exactly represents the content of the document with the thesaurus descriptor; and (3) the inconsistency of information representation, which can occur when more than one indexer performs the representation processes [12].

On the other hand, when formulating the search, the user has to deal with the search in natural language which, in most cases, does not coincide with the descriptors used in the controlled vocabulary. In this case, the author suggests that the search may be more successful if there is standardization between decision-making and the instruments used in the representation and retrieval of information [12].

The last scheme presented in this work is that of Vickery and Vickery [13]. In this scheme, the authors present the flow based on 10 processes (1) indexing, (2) storage, (3) search strategy formulation, (4) results matching, (5) selection of relevant documents, (6) document retrieval, (7) document location, (8) information retrieved, (9) evaluation of results. If the results are satisfactory, this stage ends. If the results were not satisfactory, (10) reformulation of the search is carried out, as shown in Figure 6.

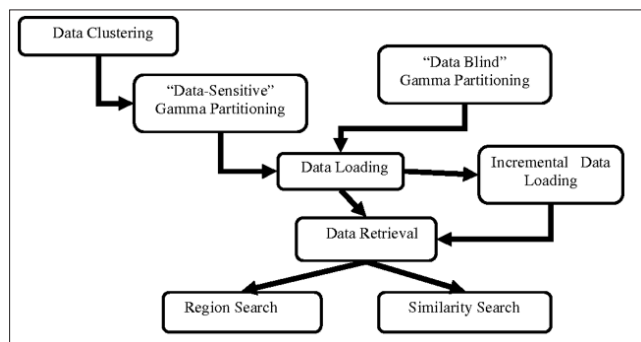


Figure 6: Information Storage and Retrieval

It is noted that, when entering information, in this scheme, the authors emphasize only the indexing process, without mentioning the other processes, moving directly to storage and, therefore, describing the information retrieval activities in more detail., which can currently be done through a computer interface. Furthermore, unlike the previously mentioned systems, this one does not present the issue of controlled vocabulary as an element of the system.

In general, what can be observed in these schemes presented is that there is no standardization in the flow of activities presented by the authors. All flows have input, storage, and retrieval stages; some specify the processes, and others do not. However, they all aim to improve the effectiveness and efficiency of retrieval, as these measures directly impact the relevance and pertinence of search results carried out by users. Therefore, it can be inferred that the quality of the information at the input of an SRI determines the quality of the information at the output.

Graph Databases

Graph databases are increasingly becoming the preferred method for data storage due to their ability to represent real-world relationships directly and intuitively. Traditional relational databases (RDBMS) work well for simple, static use cases but can become cumbersome when dealing with complex, dynamic relationships. This paper explores the fundamental principles of graph databases, their advantages over relational databases, and provides practical examples and use cases to demonstrate their effectiveness.

Graph databases manage and query graph-structured data using various strategies. One key concept is index-free adjacency, where each node directly references its adjacent nodes without using an index. This allows for rapid traversal of relationships, which is particularly beneficial for queries that require exploring connected data, such as finding friends of friends in a social network. Unlike other database systems that fit graph data into relational or other models, graph databases are designed to store graph data structures directly, optimizing performance and simplifying the data model. Another important strategy is graph partitioning, which involves dividing a graph into smaller, more manageable subgraphs. Each subgraph can be processed independently, enhancing scalability and performance in distributed systems.

Graph databases use specialized query languages like Cypher, Gremlin, and SPARQL to retrieve complex, interconnected data efficiently. They also support various path-finding algorithms, such as Dijkstra's algorithm, A*, and BFS/DFS, which are crucial for applications requiring shortest path calculations, route optimization, and network analysis. Furthermore, graph

databases facilitate the analysis of large-scale graph data to extract insights, identify patterns, and detect anomalies. This capability is invaluable in fields like social network analysis, fraud detection, and recommendation systems. To optimize performance, graph databases employ caching strategies such as frequent traversals, hot spots, LRU (Least Recently Used), and prefetching. They also ensure data integrity and enable distributed graph processing through consistency and concurrency control features, which are useful for fast updates and quicker analytics, particularly in machine learning applications.

To illustrate the advantages of graph databases, consider a social media platform where users interact in various ways. In a graph database schema, nodes represent users, posts, comments, ads, and ad preferences. Relationships between these nodes capture interactions such as follows, created, commented, liked, preferred, recommended, and notified. For example, creating nodes and relationships in a graph database involves commands to define users, posts, comments, ads, and ad preferences, and then establish connections among them.

Here is an example of creating nodes for users and posts and establishing relationships between them in Cypher:

```

CREATE (u1:User {userID: 'U001', name: 'Alice', email: 'alice@example.com', createdAt: '2021-01-01', version: 1, timestamp: timestamp()});
CREATE (u2:User {userID: 'U002', name: 'Bob', email: 'bob@example.com', createdAt: '2021-01-02', version: 1, timestamp: timestamp()});
CREATE (p1:Post {postID: 'P001', content: 'Hello World!', createdAt: '2021-01-10', version: 1, timestamp: timestamp()});
CREATE (p2:Post {postID: 'P002', content: 'Graph Databases are cool!', createdAt: '2021-01-15', version: 1, timestamp: timestamp()});
CREATE (c1:Comment {commentID: 'C001', content: 'Indeed!', createdAt: '2021-01-20', version: 1, timestamp: timestamp()});
CREATE (c2:Comment {commentID: 'C002', content: 'Absolutely!', createdAt: '2021-01-25', version: 1, timestamp: timestamp()});
CREATE (a1:Ad {adID: 'A001', content: 'Buy Shoes', createdAt: '2021-01-10', version: 1, timestamp: timestamp()});
CREATE (a2:Ad {adID: 'A002', content: 'Read Books', createdAt: '2021-01-15', version: 1, timestamp: timestamp()});
CREATE (ap1:AdPreference {prefID: 'AP001', createdAt: '2021-02-01', version: 1, timestamp: timestamp()});
CREATE (ap2:AdPreference {prefID: 'AP002', createdAt: '2021-02-01', version: 1, timestamp: timestamp()});
MATCH (u:User {userID: 'U001'}), (p:Post {postID: 'P001'})
CREATE (u)-[:CREATED {since: '2021-01-10', version: 1, timestamp: timestamp()}]->(p);
MATCH (u:User {userID: 'U002'}), (p:Post {postID: 'P002'})
CREATE (u)-[:CREATED {since: '2021-01-15', version: 1, timestamp: timestamp()}]->(p);
MATCH (u:User {userID: 'U001'}), (c:Comment {commentID: 'C001'})
CREATE (u)-[:COMMENTED {since: '2021-01-20', version: 1, timestamp: timestamp()}]->(c);
MATCH (u:User {userID: 'U002'}), (c:Comment {commentID: 'C002'})
CREATE (u)-[:COMMENTED {since: '2021-01-25', version: 1, timestamp: timestamp()}]->(c);
MATCH (p:Post {postID: 'P001'}), (c:Comment {commentID: 'C001'})
CREATE (p)-[:HAS_COMMENT {since: '2021-01-20', version: 1, timestamp: timestamp()}]->(c);
MATCH (p:Post {postID: 'P002'}), (c:Comment {commentID:
  
```

```
'C002'}) CREATE (p)-[:HAS_COMMENT {since: '2021-01-25', version: 1, timestamp: timestamp()}]->(c);
MATCH (u:User {userID: 'U001'}), (p:Post {postID: 'P002'})
CREATE (u)-[:LIKED {since: '2021-01-25', version: 1, timestamp: timestamp()}]->(p);
MATCH (u:User {userID: 'U002'}), (p:Post {postID: 'P001'})
CREATE (u)-[:LIKED {since: '2021-01-25', version: 1, timestamp: timestamp()}]->(p);
MATCH (u:User {userID: 'U001'}), (ap:AdPreference {prefID: 'AP001'})
CREATE (u)-[:PREFERS {since: '2021-02-01', version: 1, timestamp: timestamp()}]->(ap);
MATCH (u:User {userID: 'U002'}), (ap:AdPreference {prefID: 'AP002'})
CREATE (u)-[:PREFERS {since: '2021-02-01', version: 1, timestamp: timestamp()}]->(ap);
MATCH (ap:AdPreference {prefID: 'AP001'}), (a:Ad {adID: 'A001'})
CREATE (ap)-[:RECOMMENDED {since: '2021-02-01', version: 1, timestamp: timestamp()}]->(a);
MATCH (ap:AdPreference {prefID: 'AP002'}), (a:Ad {adID: 'A002'})
CREATE (ap)-[:RECOMMENDED {since: '2021-02-01', version: 1, timestamp: timestamp()}]->(a);
```

Updating preferences based on machine learning analytics can be straightforward in a graph database. For instance, if new analytics recommend a different ad for a user, a new ad preference node can be created and connected to the user node. Here's an example:

```
CREATE (ap3:AdPreference {prefID: 'AP003', createdAt: '2021-03-01', version: 1, timestamp: timestamp()});
MATCH (u:User {userID: 'U001'}), (ap:AdPreference {prefID: 'AP003'})
CREATE (u)-[:PREFERS {since: '2021-03-01', version: 1, timestamp: timestamp()}]->(ap);
```

Final Considerations

The processing and retrieval of information are fundamental steps for understanding the complex nature of information. Such processes have, in addition to a series of steps, a range of interrelated aspects, such as: the perspective of the informational context, the role of the user within this context, the role of the information professional and his systemic vision, the question the type of knowledge production and its materialization in documents, the type of controlled vocabulary adopted, the added technological issues and, consequently, the way of treating and retrieving information. If indexing is an activity that occurs at the data input stage, and retrieval at the output, in a systemic view, the input affects the output. Therefore, discussions involving the factors, measures and variables that must be considered in an indexing policy are of fundamental importance.

In this article, we consider highlighting these issues by showing how in the informational flows discussed, these aspects can be aggregated into a consistent whole, embodying what is called an Information Storage and Retrieval System (SARI). This systemic vision in which all the elements of a SARI are interrelated and in which the user is the center of attention, in our conception, is fundamental for the training of professionals capable of building Systems that aim to provide quality treatment in their representations. and which will result in the retrieval of the most relevant and pertinent information.

Thus, we highlight, as a final contribution of this study, the importance of a systemic view, in which all elements of an Information Storage and Retrieval System (SARI) are related, with the user as the main element; as well as activities that are considered important for training professionals capable of building

consistent Systems [14-19].

References

1. Barreto AAA (2002) condição da informação. São Paulo em Perspectiva 3: 67-74.
2. International Standard Organization ISO (1985) 5963-1985 - Documentation: methods for examining documents, determining their subjects, and selecting indexing terms. Suíça: ISO 1985.
3. Mooers CN (1951) Zatoncoding applied to mechanical organization of knowledge. American Documentation 2: 20-32.
4. Saracevic T (1999) Information science. Journal of the American Society for Information Science 50: 1051-1063.
5. Baeza-Yates R, Ribeiro-Neto B (2011) Modern information retrieval New York: Addison Wesley.
6. Salton G (1968) Automatic information organization and retrieval. New York: McGraw- Hill.
7. Salton G, McGill JM (1983) Introduction to modern information retrieval New York: McGraw-Hill.
8. Lancaster FW (1986) Vocabulary control for information retrieval 2. ed. Arlington: Information Resources Press 270.
9. Lancaster FW (1978) Information retrieval systems 2. ed. New York: Wiley.
10. Gomes HE, Campos MLA (1988) Indexing policy. SESC, 1998. Teaching material presented in the Training Course in the area of indexing.
11. Soergel D (1985) Organizing Information: principles of database and retrieval systems. California: Academic Press.
12. Chu H (2005) Information representation and retrieval in the digital age Melford: Information Today, Inc.
13. Vickery BC, Vickery A (2004) Information science in theory and practice 3. ed. rev. aum. Munique: KG Saur 400.
14. Cooper WSA (1971) definition of relevance for information retrieval. Information Storage and Retrieval 1: 19-37.
15. Cuadra CA, Katter RV (1967) Experimental studies of relevance judgements Santa Monica: Systems Development Corporation 3.
16. Fosket DJA (1972) Note on the concept of "relevance". Information Storage and Retrieval 2: 77-78.
17. Ingwersen P, Järvelin K (2005) The turn: Integration of information seeking and retrieval in context. Dordrecht: Springer 448.
18. Kemp DA (1974) Relevance, pertinence and information systems development. Information Storage and Retrieval 10: 37-47.
19. Saracevic T (1975) Relevance: A review of and a framework for the thinking on the notion in information science. Journal of the American Society for Information Science 26: 321-343.

Copyright: ©2021 Venkata Naga Sai Kiran Challa. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.