



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Intelligent Information Retrieval System

Bola Abimbola*

¹Universidad de Oviedo, Escuela Universitaria Jovellanos, Francisco Tomas y Valiente 33201 Gijon, Oviedo, 33006, Spain.
E-mail: UO285018@uniovi.es

Article Info

Volume 2, Issue 1, January 2022

Received : 22 May 2021

Accepted : 17 November 2021

Published : 18 January 2022

doi: [10.51483/IJAIML.2.1.2022.71-74](https://doi.org/10.51483/IJAIML.2.1.2022.71-74)

Abstract

An Intelligent Information Retrieval (IIR) is a machine learning system used by users to retrieve the massive volume of information available through internet. IIR offers personalization and efficiency in the current internet development allowing users to examine and acquire appropriate information. The paper will focus on the keyword searching issues of the current digital library system. The paper will analyze a model that solves the issue using metadata case-based and concept-based approach. The objective is to evaluate a set of categories and concepts in the domain field that demonstrates some relations between them to elaborate how poor quality of information retrieval in digital library system can be solved. The paper further argued that the already developed domain-specific ontology can be efficient for query advancement. Many researchers have used semantic retrieval technology using concepts to solve problems that lack semantics in traditional retrieval technology. Using concepts in ontology enhances search results

Keywords: Digital library system, Ontology, Query expansion, Intelligent information retrieval

© 2022 Bola Abimbola. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

1. Introduction

The information volume available through the internet has grown significantly over the past few years. People can now access and share a substantial amount of information. Nevertheless, the massive volume and unstructured nature of information worldwide have made it challenging for users to examine and acquire appropriate information. Academic and industrial researchers are now concentrating on Information Retrieval (IR) techniques to solve such problems. These methods focus on mathematical algorithms and models for retrieval. Users did not acquire a semantic description of the information they required in the traditional IR systems. Discovering more relevant information requires the use of Intelligent Information Retrieval (IIR) systems.

A digital library system is an integration of different resources, such as, collecting, classifying, storing, protecting, searching and retrieving information (Aruleba *et al.*, 2017). The current system has keyword searching issues. The proposal is to chose a model that solves the issue using metadata case base and ontology (concept-based approach). It involves pinpointing domain ideas in the user's query and expanding those concepts to produce more suggestions and

* Corresponding author: Bola Abimbola, Universidad de Oviedo, Escuela Universitaria Jovellanos, Francisco Tomas y Valiente 33201 Gijon, Oviedo, 33006, Spain. E-mail: UO285018@uniovi.es

enhance results. The aim is to bring in the ontology concept, employing its benefits of standard concept and generative semantics. Domain-specific ontology boosts IR from traditional technologies that relied on a keyword to technologies that depend on concept or knowledge and replace the traditional keyword matching with semantics matching retrieval process (Tehseen, 2018). The first approach is applying query expansion methods using a concept-based approach. The second approach is establishing a case-based connection evaluation for metadata IR using the Case-Based Reasoning (CBR) approach. Results demonstrate enhancements over traditional techniques.

1.1. Intelligent Information Retrieval System

According to Baoxian (2018), employment of semantic retrieval technology will enhance information quality retrieval. The work suggests a technological framework that will help users retrieve information by providing them with relevant case-based and semantic retrieval documents. Users will key in their query terms in their normal language and the system conduct analysis on them. Joby (2020) argues that comparing ideological query representation against the conceptual representations database aid in choosing the match within a close range. Joby (2020) further contends that users can use natural language or a relevant document or Boolean query to search-related documents.

The suggested sets of concepts for geographical-related information retrieval included a mixture of quantitative and qualitative geometric data, including sparse coordinate and topological relations information representing the geographical place's footprints. Geographical categories classified places and then connected to non-geographical cases grouped by conceptual ranking. According to Afuan *et al.* (2019), the goal was to combine Euclidean and hierarchical distance analysis to establish a geographical distance evaluation. Another proposal was on query advancement approach that employed advanced dependent topologies to bring queries nearer to document collection characteristics and user's preferences (Afuan *et al.*, 2019). It aimed at linking each concepts' case and class with a feature vector to modify these ideas to the terminology and document collection utilized. The concepts and their related feature vectors processed results after identifying user's query on the search engine (Afuan *et al.*, 2019).

According to Tehseen (2018), concepts established in the Spirit web project were employed to support document retrieval that was geographically suitable to users' request, where query advancement used domain and spatial concepts. A concept network deduced from concepts from the original query words served as a knowledge base for modifying query advancement (Afuan *et al.*, 2019). The conceptual query advancement quality relied on the quality of concept network. The purpose of a concept network was to match original query words which resulted in the development of other concepts and queries terms.

While most concepts categories used the WordNet as a controlled vocabulary to advance the query, the approach suggested by Afuan *et al.* (2019) combines the advantages of statistical methods and concept use. For query advancement, field concepts were utilized as controlled terminology. The primary presupposition is that users create a query concurrently illustrating an issue they are trying to address. The CBR approach uses other related cases to addresses an information retrieval request (Lin, 2020). A solution is acquired by providing several links related to the user's query. A case-based approach contains a set of data that defines and provide information about other data. The model by Afuan *et al.* (2019) offers various intelligent query advancement and processing benefits.

1.2. Ontology

Traditional IR systems recovered information without defining any user's specific field of interest. As such, the system provided massive data that was not important to the user. On-Piu Chan (2020) demonstrates how to employ concepts effectively included in various multi-disciplinary fields comprising of different terminologies aiming goal to improve the browsing outcome quality for extensive search systems.

Query searching using concepts is a promising and unique approach in the retrieval process. Users do not need to know the documents implementation; their focus is on the conceptual searching level. Domain concept is helpful for query advancement by increasing the input terms with the appropriate domain ontologies. WordNet adds meronyms, homonyms, and synonyms to the index terms, making the indexing stage effective (On-Piu Chan, 2020). There exist two problems in utilizing the concept-based approach. The first problem is using keywords to extract semantic concepts. Its main problem is determining relevant concepts that identify documents and determine the language used in user queries. It is essential to avoid matching and connecting inappropriate concepts and disregard appropriate concepts. The second problem is document indexing. Field ontology is established, and property and concept relationship in the professional field is described. The model structure proposed by On-Piu Chan (2020) has seven steps including:

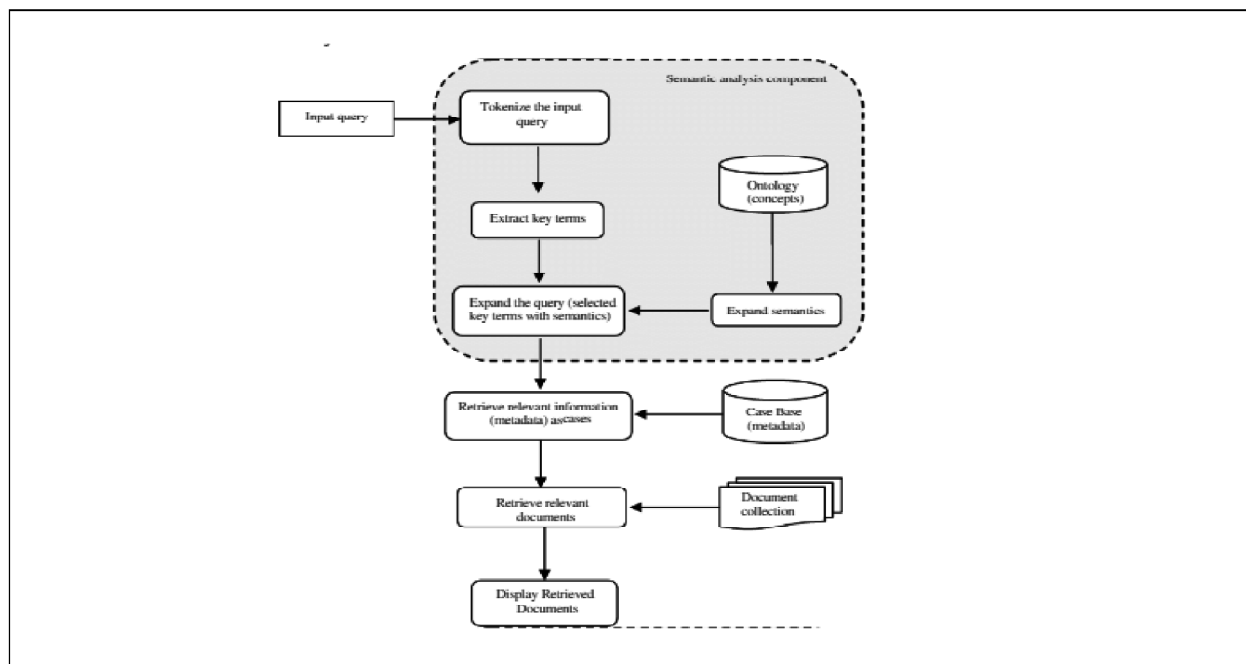


Figure 1: Suggested System Architecture

- Step One: Verify the professional specialty and ontology category.
- Step Two: Evaluating the likelihood of re-utilizing current ontology
- Step Three: Highlight essential ontology terms.
- Step Four: Define categories and classes ranking system.
- Step Five: Define classes property.
- Step Six: Define property aspects.
- Step Seven: Develop cases.

1.3. Case-Based Module

According to Anaissi *et al.* (2017), a concept-based method utilizes concepts from a specific domain and CBR approach with various metadata containing relevant documents defines a case. A case base is act as a piece of document information to examine the query process and retrieve information from appropriate documents in the digital library (Anaissi *et al.*, 2017). It aimed at improving concept-based information retrieval by integrating domain ontology, case-based reasoning process, and traditional information retrieval process (Lin, 2020). The model proposed by Lin (2020) employs concepts to expand queries and integrates textual and case-based closeness to recover a set of information for relevant documents to give users several document recommendation options. The steps are as follows:

- Step One: Matching a new case using other cases in the case base.
- Step Two: Recover the closest matching case from the past cases' library.
- Step Three: Reuse the recovered case to address the existing issue.
- Step Four: Re-evaluate and modify the suggested solution if needed.
- Step Five: Rename the final result as a new case.

Case similarity evaluation is done for the subject and author attributes. Devi and Gandhi (2020) Statistical IR methods used the Apache Lucene to measure Title attribute content. In Lucene, the Boolean technique and vector space model were used to determine the relevancy of a particular document to the user's request.

2. Semantic Analysis

Semantic analysis is the process of comparing syntactic structures of users' search terms. It expands the category words semantics structure and then retrieves the information to the user interface. According to Ahmed (2017), finding specific

concept matching is easy, and the challenging part is to match unsettled related concepts using knowledge repository. It provides concepts' details and their connections with other concepts and involves two steps: tokenizing user query and extracting vital domain words from tokenized terms (Ahmed, 2017). Irrelevant concepts are removed, and suitable ones are linked to documents for query creation. The processes are performed automatically without user feedback or intervention. Queries are generated using relevant and appropriate ontology terms using knowledge encrypted in concepts. The component involves three different ways of query expansion: interactive, manual, and automatic. Interactive and manual query needs user engagement. On the other hand, automatic query expansion involves adding extra phrases or terms to boost the retrieval execution without the user's interference.

3. Conclusion and Recommendation

The report has demonstrated that the already developed domain-specific ontology can be efficient for query advancement. Many researchers have used semantic retrieval technology using concepts. It helps to solve problems that lack semantics in traditional retrieval technology. Using concepts in ontology enhances search results. Expanding a query aims at minimizing document or query mismatch by adding related phrases and terms to the relevant documents set.

Nevertheless, query expansion contains some built-in dangers. Thesaurus has been used in information retrieval to identify the linguistic entities and synonymous expressions semantically the same. A query drift can occur due to query ambiguity providing information that is irrelevant to the user. For instance, the term windows could mean Microsoft Windows Operating System (OS) or the actual house windows. The system should employ domain concepts to solve the problem. Not every tokenized term should be set for expansion. Query expansion process should replace the terms in the domain concepts with the original user terms and their related domain concepts.

References

- Ahmed, M. (2017). [Semantic Based Intelligent Information Retrieval Through Data Mining and Ontology](#). *Int. J. Comput. Sci. Eng.*, 5(10), 210-217.
- Afuan, L., Ashari, A., and Suyanto, Y. (2019). [A Study: Query Expansion Methods in Information Retrieval](#), *J. Phys. Conf. Ser.*, 1367(1), 012001. DOI:10.1088/1742-6596/1367/1/012001
- Anaissi, A., Goyal, M., Catchpoole, D.R., Braytee, A., and Kennedy, P.J. (2017). [Case-Based Retrieval Framework for Gene Expression Data](#). *Cancer Inform.*, 14, 21-31. <https://doi.org/10.4137/CIN.S22371>
- Aruleba, K.D., Akomolafe, D.T., and Afeni, B. (2017). [A Full Text Retrieval System in a Digital Library Environment](#). *Intell. Inf. Manag.*, 8(1), 1-8.
- Baoxian, J. (2018). [Application of Intelligent Information Retrieval for Big Data Oriented Brain Science](#). *Advances in Engineering Research*, 66. DOI: <https://doi.org/10.2991/amcce-18.2018.156>
- Devi, M.U., and Gandhi, G.M. (2020). [Scalable Information Retrieval System in Semantic Web by Query Expansion and Ontological-Based LSA Ranking Similarity Measurement](#). *Int. J. Adv. Intell. Paradig.*, 17(1/2), 44.
- Joby, D. (2020). [Expedient Information Retrieval System for Web Pages Using the Natural Language Modeling](#). *Journal of Artificial Intelligence and Capsule Networks*, 2(2), 100-110. DOI: <https://doi.org/10.36548/jaicn.2020.2.003>
- Lin, K.S. (2020). [A Case-Based Reasoning System for Interior Design Using a New Cosine Similarity Retrieval Algorithm](#). *J. Inf. Telecommun.*, 4(1), 91-104.
- On-Piu Chan, J. (2020). [Digital Transformation in the Era of Big Data and Cloud Computing](#). *Int. J. Intell. Inf. Syst.*, 9(3), 16.
- Tehseen, R. (2018). [Semantic Information Retrieval: A Survey](#). *J. Inf. Technol. Softw. Eng.*, 8(4). DOI:10.4172/2165-7866.1000241.

Cite this article as: Bola Abimbola (2022). [Intelligent Information Retrieval System](#). *International Journal of Artificial Intelligence and Machine Learning*, 2(1), 71-74. doi: 10.51483/IJAIML.2.1.2022.71-74.