International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

ENHANCING WEB INFORMATION RETRIEVAL EFFICIENCY THROUGH ONTOLOGY-BASED SEMANTIC TECHNIQUES: A COMPARATIVE REVIEW

Dr. C.N.RAVI, LINU POULOSE, PREMA MANI,REKHIL M KUMAR
Department of Computer Science and Engineering,
Indira Gandhi Institute of Engineering and Technology
Nellikuzhi P.O, Kothamangalam, Ernakulam (Dist) Pincode 686691

Abstract

The vast expanse of the Internet harbors an extensive array of information, catering to the needs of numerous users who regularly access, research, and utilize its resources. Characterized by its multilingual nature and rapid growth, the Web encompasses diverse data types, ranging from unstructured to semi-structured formats such as websites, texts, journals, and files. Extracting pertinent information from this vast and varied pool has proven to be a daunting and laborious task. Traditional methods of data retrieval, reliant on simple keyword-based systems, often encounter challenges stemming from statistical limitations, leading to semantic inconsistencies and contextual ambiguities. Consequently, there arises a pressing need for systematic organization of this massive volume of data to facilitate efficient analysis and meet users' information requirements in relevant contexts. Ontologies have emerged as a prominent solution within the realm of the semantic Web, enabling the structured organization of disparate information and significantly enhancing the efficacy of information retrieval methodologies. Ontological information retrieval systems operate by retrieving files based on the semantic relationships between search queries and searchable data. This paper critically evaluates contemporary techniques for ontology-based information extraction across various data formats, including text, interactive media, and multilingual content. Furthermore, the study undertakes a comparative analysis to categorize and assess the most notable advancements in search and retrieval techniques, highlighting their respective strengths and weaknesses.

Keywords: *Web Information Retrieval, Ontology, Semantics, Multimedia Information Retrieval.*

INTRODUCTION

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

The quest for efficient information storage, retrieval, and extraction has been ingrained in human endeavors since the advent of written language. Initially confined to academic libraries, the concept of information extraction soon expanded with the anticipation of widespread adoption of machine-based initiatives worldwide. In the early 1970s, the emergence of devices tailored for specific user groups such as medical professionals, educational institutions, and governmental bodies marked the beginning of prompt information extraction. With the proliferation of innovative technologies and the rise of social media platforms like LinkedIn, Facebook, Instagram, Twitter, Snapchat, WhatsApp, and TikTok, users began sharing vast amounts of information in various formats, including text, visuals, sound recordings, and multimedia, collectively referred to as "multimedia data." This surge in communication technology, coupled with the increasing number of users and devices, led to exponential growth in data across domains such as eGovernment, eLearning, eBusiness, and eCommerce. Consequently, there arose a need for a versatile information retrieval system capable of accommodating diverse data sources and user needs, replacing the multitude of specialized processes initially tailored for specific user groups. However, information retrieval has become increasingly challenging and time-consuming due to the sheer volume and diversity of data.

To address the challenges, numerous multipurpose search engines were developed, alongside generic search engines catering to a broader user base. However, these search engines often fell short in understanding user queries in context and providing relevant responses. Instead, they tended to return a plethora of irrelevant search results, making information retrieval a frustrating experience for users. User queries, often expressed in natural language, may differ significantly from the terms used in indexed datasets, further complicating the retrieval process. For example, a user seeking advice on diabetes prevention measures may struggle to articulate their information needs effectively amidst the complexity of natural language.

In this context, the most relevant information can often be found within files containing specific terms such as "practitioner" or its synonyms, like "qualified physician." Thus, it becomes crucial to recognize that both professionals and doctors are part of the same conceptual framework to effectively retrieve such documents. To tackle this challenge, various methods have been developed over time to enhance users' ability to express effective queries using theoretical knowledge. One commonly employed approach is the integration of a dictionary component within Information Retrieval (IR) mechanisms, which captures semantic relationships across different domain ontologies. Additionally, the incorporation of content understanding features into IR systems has become prevalent, marking a shift from simple keyword-based approaches to thematic initiatives due to their effectiveness. Moreover, the concept of domain-specific knowledge integration, where information is linked to relevant data in a machine-readable format, has been proposed by researchers. Ontologies play a crucial role in this approach, facilitating the incorporation of machine-understandable semantic information. Semantic languages like RDF, OWL, RDFS, and

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

SPARQL are employed in the Semantic Web to support ontology-based approaches, demonstrating the efficacy of theoretical data modeling, such as ontologies, in retrieving pertinent information.

This paper addresses state-of-the-art ontology-based conceptual information retrieval methods, including "Text-based IR," "Multimedia IR," and "Cross-lingual IR." The author aims to compare and categorize the most contemporary methods used in these ontology-based information retrieval tasks. The importance of the Semantic Web in natural language evolution and its impact on information processing efficiency are briefly discussed in the initial segment. Furthermore, contemporary knowledge representation models like conceptual maps and frameworks are explored to enhance the effectiveness of semantically related information retrieval. Subsequently, various methods recommended for scriptural information extraction, followed by interactive media and cross-lingual processing, are outlined. The effectiveness of these techniques in written text, interactive media, and cross-lingual pattern recognition is then compared. Finally, future directions for annotation information extraction are discussed in the concluding section of the study.

II. IMPORTANCE OF SEMANTICS IN INFORMATION RETRIEVAL

The ultimate objective of speech recognition processing is to comprehend and convey information articulated in a particular language. Semantic understanding plays a crucial role in effectively managing vast amounts of information within meaningful contexts. Data semantics not only reveal the actual content of the message but also elucidate the framework of words utilized in the subject matter. Incorporating appropriate semantic and syntactic classification systems into queries has been found to significantly enhance question classification efficiency. Researchers have employed various information sources to determine the "semantic-relatedness" between words. Additionally, studies have explored different WordNet-based conceptual similarity metrics and their applications in web-based information extraction.

As text information proliferates rapidly, word mismatches become increasingly prevalent. Ambiguity in language often leads to a single word indicating multiple concepts. Consequently, information semantics have emerged as a crucial element in information retrieval, enhancing the search process and yielding more comprehensive results. Semantic feature extraction augments a user's query by extending its semantic content, thereby improving the effectiveness of system queries in terms of information gain. Semantic data enhancement transcends conventional data and information extraction methods, leveraging semantic meaning to aid in data processing.

Pioneering research in the field has laid the groundwork for semantic-based information extraction. Researchers have developed Semantic Information Retrieval (SIR) frameworks designed to interpret questions expressed in dense English language. Implemented within the Speech impediment computer program, SIR demonstrates the capability to comprehend semantic information. The program's functionality is based on an underlying structure that establishes

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

connections between parsed components within a given communication using word associations and property lists. Semantic information is retrieved from input inquiries using a configuration function, and the system evaluates inquiries to determine their forms and process them accordingly. For declarative inquiries, the structure populates the model with relevant information, while for questions, the framework either provides answers or identifies reasons for their absence in the model parameters. SIR effectively addresses semantic uncertainties in search queries and adapts the model framework to conserve computer memory. Semantic analysis has revolutionized Information Retrieval (IR) and is prevalent across various fields, including the semantic Web. Ontologies, which form the foundation of semantics, are constructed using semantic features. Numerous researchers have explored semantic contexts within the realm of the semantic Web, utilizing it as a tool for information retrieval. For instance, in one study [9], researchers developed a prototype allowing users to annotate their queries with semantic information extracted from existing conceptual frameworks. The integration of annotated semantic features resulted in increased precision compared to conventional text retrieval methods. Similarly, in another study [20], the semantic Web was leveraged for extracting knowledge and information from biological and medical patents. The aim was to utilize web applications to enable personalized information extraction from current semantic data sources. Furthermore, efforts have been made to automate the process of semantic annotation, facilitating observation of semantics within the semantic Web [36]. Additionally, researchers have employed unsupervised information processing to generate seed documents, leading to the creation of systems like the Sem-Tag System, which automatically categorizes large corpora with semantic information [13,14]. Moreover, ongoing research focuses on the development of deep learning ontologies [9, 21, 40].

Semantic systems, such as semantic networks, serve as prominent means of representing complex knowledge structures. These systems visualize concepts and their relationships using vertices and edges, effectively conveying semantic relationships between ideas. Semantic-based information extraction has extensively utilized these semantic networks. In-depth articles have been published on thematic mapping, highlighting the efficacy of semantic channels and maps in condensing vast amounts of information into a compact space while elucidating associations between semantic information, familiar concepts, and documents [30]. Additionally, novel systems like the "Grant System" [34,47] have been developed to aid in identifying funding mechanisms through resource constraints dissemination activation. These systems have shown enhanced user engagement and prediction accuracy values compared to conventional Convolutional Neural Network (CNN) systems. Authors have also proposed fragmented mentoring inference engines based on participant networks, distributing file indices across the system based on paper semantics determined by latent semantic archiving (LSA). This approach reduces search costs by locating all semantically similar indices in the same network location [47].

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

III. THE ROLE OF ONTOLOGIES IN INFORMATION RETRIEVAL

Ontologies serve as hierarchical structures comprising machine-readable, interpretable, and processable data [43]. Supercomputing ontologies encompass highly heterogeneous groupings or categories formed by domain-specific terms, attributes, and interconnections. They find application in sentiment analysis, information retrieval, term annotation, text categorization, and information retrieval frameworks. The following outlines ontology's significance in semantic information extraction:

A. Query Expansion Semantic information resembling statements found in domain-specific ontologies is utilized to expand user queries [7].

B. Resolution of Term Disambiguation Interpretive paradigms related to the same concept are disambiguated [37].

C. Document Classification Ontological concepts are employed to classify documents and aid in query expansion [16].

D. Enhanced IR Model Incorporating ontology into established IR models results in customized and improved data retrieval models due to semantic classification effects [1, 17, 33].

Ontology has been integral to various information retrieval projects for numerous reasons, as elaborated below:

- Semantic Digital Library (SDL): SDL leverages ontology as a knowledge repository. Subject matter and metadata, encompassing all files, are integrated into the ontology to facilitate rapid information retrieval.
- Crime News Retrieval (CNR): CNR employs ontology to foster semantic-based information retrieval, modeling named entity recognition [32].
- Multi-Modality Ontology-Based Image Retrieval (MMOBIR): MMOBIR has postulated several ontologies, including text-based ontology, graphic ontology, and website ontology, to characterize images in terms of textual, visual, and contextual embeddings. Furthermore, MMOBIR demonstrates how these ontologies can be integrated into DBpedia, an open-sourced knowledge database, to facilitate in-depth exploration.

IV. THE ROLE OF ONTOLOGIES IN INFORMATION RETRIEVAL

In various contexts, ontology provides a robust and structured representation of data, making it processable, shareable, and reusable [3,4]. For instance, researchers utilized an ontology model to generate audio media information, surpassing conventional keyword-based information extraction

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

approaches [44]. Similarly, in [52], the author utilized ontology as a knowledge provider to enhance the efficacy of a single recognition system. Additionally, ontology was employed to improve the effectiveness of a constitutional search engine [51], benefiting users by suggesting synonyms. Ontology has been integrated into the semantic IR process to dynamically classify web content [19].

Similarly, user queries may encompass terms from multiple domains, complicating the selection of an accurate domain ontology for the search strategy. In [15], researchers proposed a filter feature selection methodology to address this challenge. Semantic terms extracted by combining multiple semantically similar ontologies were incorporated into every user query [1]. When no semantic classification was available, the WordNet ontology supplemented the user query with semantic aspects [3,5].

V. TYPES OF MULTILINGUAL INFORMATION RETRIEVAL

A. Text-based Information Retrieval

The essential semantic relatedness between query terms and corpus poses challenges for ontology-based content retrieval in research [39]. The vector space model, also known as the vector representation, depicts records and queries as vectors, measuring their similarity using trigonometric measures. While other clustering methods like TF-IDF vectors and BM25 are employed in vector spatially IR, they have not been applied to Ontological-IR systems. Moreover, ontologies play a crucial role in extracting constructs from files and queries. Researchers proposed an ontology-based IR approach that extracts concepts from queries [39]. Their model posits that key query aspects are already evident in established ontologies like WordNet. Using ontology, a set of conceptual frameworks for each document was extracted, and the retrieved concepts were compared to query concepts, with a scoring system estimating document ranks. They evaluated the integrity of the best system using 1239 MEDLINE documents with the keyword Cystic Fibrosis. In [24], a similar strategy was introduced for Arabic context learning, comparing query and document similarity using a cosine similarity metric. A generic vector space model combining entities (NEs) and search terms was suggested for semantic information retrieval, representing static or semantic data across various media for communication in Wh-queries [8]. They utilized ontology to expand query semantics and classify the entire collection with ontological constructs to retrieve semantically connected documents. This model outperformed other keyword-based approaches in their experiments. Authors proposed ontology-based IR, relying on a foundational taxonomy generated from Contextual Information or lexical data systems [33]. They incorporated an ontological stack mapping component into the core taxonomy and devised a weighting scheme for vector space to address vocabulary mismatches and minimize vector dimensions in IR.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

B. Retrieval of Multimedia Information

In the contemporary era, exploring, indexing, storing, and retrieving audio, video, and image content, either individually or in combination, is a common practice. With the proliferation of social media networks like Facebook and Twitter, the collection of multimedia data sets has surged. Multimedia retrieval is crucial due to the increasing use of media-sharing platforms and the growing volume of multimedia content on the internet. However, characterizing visual features and utilizing them effectively for content examination and retrieval pose significant challenges. Multimedia retrieval techniques have been developed to retrieve descriptive features from images and audio data, tracing them to high-level query functionalities. Nonetheless, the exponential growth of web content has made computing and managing vast amounts of multimedia content time-consuming. MIR encompasses a wide range of research topics, methodologies, and data types, including sound, visuals, images, motion graphics, video files, rich text, HTML tags, and combinations thereof [2,25]. The volume of datasets, such as images, songs, and video content, has surged over time, leading to an increasing demand for effective audio-visual search and retrieval methods for web users. To enhance the human-centered nature of the MIR framework, system responses must be precise to meet user expectations. Users employ various MIR processes, such as Google for video and image search and Altavista for audio files, to find multimedia assets. Additionally, there are numerous MIR seminars and conferences, including ACM and SIGMM. MIR programs primarily serve two user needs: exploration and navigation of the Web through media summaries [18]. Two main techniques are employed to address these needs: feature-based and classification methods. Classification methods have gained popularity in recent years due to their ability to assert semantic information about media, which is crucial for information retrieval. As technology advances rapidly, researchers are increasingly focusing on information strategies. Moreover, non-textual information is becoming more prevalent than textual data, suggesting that it will soon become a common method of data sharing. Given these trends, a comprehensive examination of state-of-the-art retrieval strategies for non-textual audio-visual data is necessary. While information extraction systems like internet search engines are well-established and widely available, interactive media IR systems are less developed and less accessible.

C. Text-based Image Retrieval (TBIR)

TBIR processes are commonly used in web applications for image retrieval [32]. This approach utilizes text associated with the image, such as a file name, web address, or annotation, to describe the image's content. When a user enters a textual query, various techniques are employed to address the issue of polysemy before extracting keywords. The query is then tagged using these keywords, and the dataset is searched for images that match the annotated queries, with the most relevant images returned as results.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

Figure 1 illustrates the structure of feature extraction for text-based images. An ontology-driven IR framework with a soccer context application was proposed [29]. The focus was primarily on three challenges in information retrieval: scalability, retrieval efficiency, and functionality. To enhance retrieval efficiency, they incorporated inference, rules, and a URL information applicator. A semantic indexing framework based on Apache Lucene was suggested to enhance keyword-based search capability by providing retrieved and inferred information using ontology. They employed a searchable model to address functional ambiguity issues. Thanks to the efficient semantic indexing model, this approach covers various aspects of metadata. Figure 2 displays the architecture of feature extraction for content-based images.

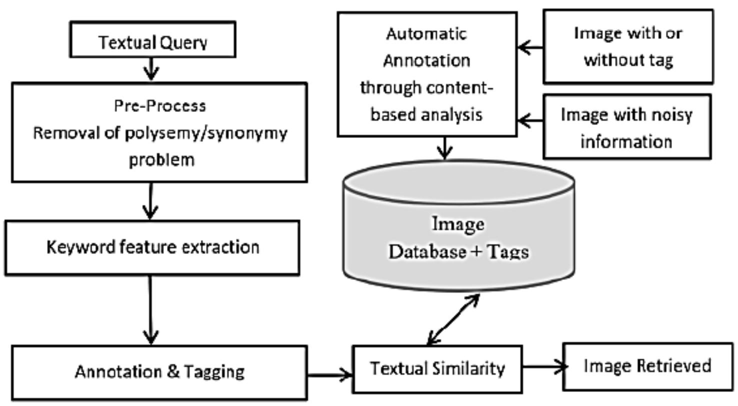

Figure 1. The architecture of text-based image retrieval feature extraction

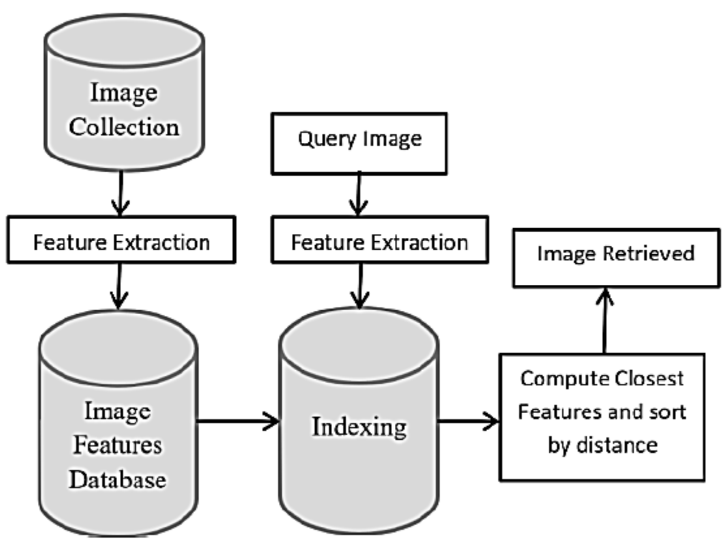

Figure 2. The architecture of content-based image retrieval feature extraction

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

Digital data is expanding rapidly due to remarkable advancements in digital devices that capture it. This includes documents, images, audio, and video files, among others. Video retrieval involves extracting information from a video dataset based on user requirements. Effective data retrieval methods are essential to manage such extensive databases.

Text-based video retrieval methods involve retrieving videos based on textual elements present in them, such as subtitles, performer names, and event descriptions [31]. These methods analyze textual components like characters, phrases, and frames within video sequences to extract features. These keyframes need to be annotated based on the textual data they contain. Optical Character Recognition (OCR) tools are commonly used to extract text data from videos. Categories are then employed to extract data, which can serve as search terms for indexing. Figure 3 illustrates the architecture of content-based video information retrieval.

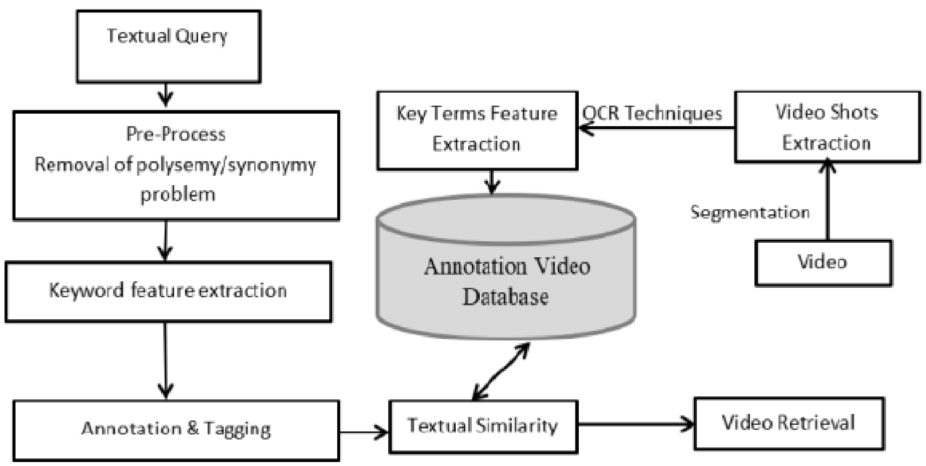

Figure 3. The architecture of content-based video information retrieval

Text-based video retrieval methods utilizing ontologies are crucial for searching an artist's discography, identifying musical genres, and pinpointing event locations mentioned in videos. These techniques are also beneficial for educational video libraries used in online learning applications. However, a limitation of these methods lies in selecting the most suitable Optical Character Recognition (OCR) platform, and they do not fully assess videos semantically. Consequently, there has been a shift in focus among researchers towards analyzing videos beyond just their textual content. Figure 4 illustrates the architecture of content-based video retrieval.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com
ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395
Volume 9, Issue 12 – August 2021-22 - Pages 16-29

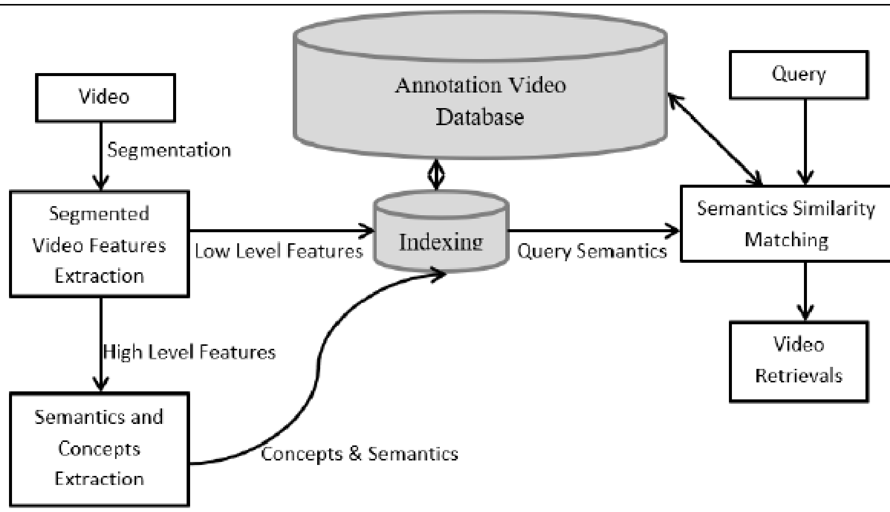

Figure 4. The architecture of content-based video retrieval

V DISCUSSION

In this section, the performance of various frameworks for text, multilingual, and interactive media retrieval is highlighted and compared based on expertise and evaluation methods. In a study by [7], a scheme for text-based information extraction was developed, focusing on context-based semantic retrieval to enhance search engine accuracy. RDF frequent patterns were utilized for source information containment, with query ideas matched against established RDF triples instead of keywords. This approach emphasizes the combined effect of concepts and the resemblance of their relationships, leading to improved search accuracy. The comparison between recall and precision in indexing and retrieval specific keywords, simple semantics, and semantic neighborhoods revealed significant correlations, with semantic IR showing higher values. Additionally, [20] proposed an ontology-based IR approach, which notably improved IR efficacy by incorporating a fuzzy ontology generated automatically from documents. Evaluations demonstrated an increase in IR efficiency across various techniques. Similarly, [19] recommended a three-layer semantic-based archiving approach, achieving higher accuracy compared to regular index-based IR. Moreover, [43] tested an ontology-based IR technique on datasets, outperforming the standard approach on all metrics. Another study by [44] introduced a semantic query expansion technique, which outperformed existing strategies, particularly in terms of F-score and P@20. Additionally, [47] presented a key-phrase semantic retrieval model, demonstrating improved performance through semantic archiving, ontology-based retrieval, and deductive reasoning. Finally, [33] introduced an ontology-based semantic information augmented platform for cloud computing, achieving high F-scores for both single and multi-topic queries.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

VI CONCLUSION

Semantic-based data retrieval faces numerous challenges, including the scarcity of semantic knowledge forms, evaluation benchmarks, data sources, effective IR techniques, and the evolving nature of the domain. Similarly, multimedia content retrieval encounters obstacles such as semantic discrepancies between user query keywords and attributes of interactive media resources. Another significant constraint in interactive media IR is the lack of extensive datasets for indexing classifiers needed to accurately analyze multimedia features. Additionally, cross-lingual information extraction lacks essential resources like corpora, ontologies, and idiomatic expressions for various languages, such as Arabic. Moreover, cross-lingual feature extraction struggles with knowledge discovery challenges, posing a significant hurdle for researchers and practitioners alike. To advance in the field of information extraction, substantial research is needed in translation software, instant ontology acquisition from unstructured data, and semantic data interpretation and extraction.

REFERENCES

1. Ali, A., Alfaycz, F., & Alquhayz, H. (2018). Semantic Similarity Measures Between Words: A Brief Survey. *Sci. Int. (Lahore), 30*(6), 907-914.
2. Ali, A., & Alourani, A. (2021). An Investigation of Cloud Computing and E-Learning for Educational Advancement. *International Journal of Computer Science and Network Security, 21*(11), 216-222.
3. Ali, A., & Ahmad, I. (2011). Concept-based information retrieval approaches on the web: A brief survey. *IJAIR, 3*(6), 14-18.
4. Ali, A., & Ahmad, I. (2012). A Novel Approach for Information Retrieval on the Web. *International Journal of Advance and Innovative Research, 1*(6), 20-26.
5. Ali, A., & Ahmad, I. (2011). Information Retrieval Issues on the World Wide Web. *International Journal of Computer Technology and Applications, 2*(6), 1951-1955.
6. AlKhunzain, A., & Khan, R. (2021). The Use of M-Learning: A Perspective of Learners' Perceptions on M-Blackboard Learn. *International Journal of Interactive Mobile Technologies (iJIM), 15*(2), 4–17.
7. Bendersky, M., Metzler, D., & Croft, W. B. (2012). Effective query formulation with multiple information sources. Paper presented at the Proceedings of the fifth ACM international conference on Web search and data mining.
8. Cao, T. H., Ngo, V. M., Hong, D. T., & Quan, T. T. (2008). A Named-Entity-Based Multi-Vector Space Model for Semantic Document Clustering. In *Proc. of the 1st Pacific-Asia Workshop on Web Mining and Web-Based Application (WMWA'2008, May 20, Osaka, Japan)*.
9. Casteleiro, M. A., Prieto, M. J. F., Demetriou, G., Maroto, N., Read, W. J., Maseda-Fernandez, D., ... & Stevens, R. (2016). Ontology Learning with Deep Learning: a Case Study on Patient Safety Using PubMed. Paper presented at the SWAT4LS.
10. Charron, B., Hirate, Y., Purcell, D., & Rezk, M. (2016). Extracting semantic information for e-commerce. Paper presented at the International Semantic Web Conference.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

11. Chicco, D., Sadowski, P., & Baldi, P. (2014). Deep autoencoder neural networks for gene ontology annotation predictions. Paper presented at the Proceedings of the 5th ACM conference on bioinformatics, computational biology, and health informatics.
12. Corcoglioniti, F., Dragoni, M., & Rospocher, M. (2016). Knowledge extraction for information retrieval. Paper presented at the European Semantic Web Conference.
13. Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., ... & Tomlin, J. A. (2003). SemTag and Seeker: Bootstrapping the semantic web via automated semantic annotation. Paper presented at the Proceedings of the 12th international conference on World Wide Web.
14. Dingli, A., Ciravegna, F., & Wilks, Y. (2003). Automatic semantic annotation using unsupervised information extraction and integration.
15. Egozi, O., Markovitch, S., & Gabrilovich, E. (2011). Concept-based information retrieval using explicit semantic analysis. *ACM Transactions on Information Systems (TOIS), 29*(2), 1-34.
16. Elhadad, M. K., Badran, K. M., & Salama, G. I. (2018). A novel approach for ontology-based feature vector generation for web text document classification. *International Journal of Software Innovation (IJSI), 6*(1), 1-10.
17. Ensan, F., & Bagheri, E. (2017). Document retrieval model through semantic linking. Paper presented at the Proceedings of the tenth ACM international conference on web search and data mining.
18. Ezzikouri, H., Madani, Y., Erritali, M., & Oukessou, M. (2019). A new approach for calculating semantic similarity between words using WordNet and set theory. *Procedia Computer Science, 151*, 1261-1265.
19. Gupta, S., & Garg, D. (2011). Comparison of semantic and syntactic information retrieval system on the basis of precision and recall. *Int J Data Eng, 2*(3), 93-101.
20. Hersh, W. (2021). *Information retrieval Biomedical Informatics*. Springer, 755-794.
21. Hohenecker, P., & Lukasiewicz, T. (2017). Deep learning for ontology reasoning. *arXiv preprint arXiv:1705.10342*.
22. Jain, V., & Singh, M. (2013). Ontology based information retrieval in semantic web: A survey. *International Journal of Information Technology and Computer Science (IJITCS), 5*(10), 62.
23. Johnston, B., & Webber, S. (2005). As we may think: Information literacy as a discipline for the information age. *Research strategies, 20*(3), 108-121.
24. Kaloub, A. I. (2013). Automatic ontology-based document annotation for Arabic information retrieval.
25. Khan, R. M. I., Ali, A., Alourani, A., Kumar, T., & Shahbaz, M. (2022). An Investigation of the Educational Challenges During COVID-19: A Case Study of Saudi Students' Experience. *An Investigation of the Educational Challenges During COVID-19: A Case Study of Saudi Students' Experience, 11*(1), 353-363.
26. Khan, R. M. I., Radzuan, N. R. M., Farooqi, S., & Khan, M. S. (2021). Learners' Perceptions on WhatsApp Integration as a Learning Tool to Develop EFL Vocabulary for Speaking Skill. *International Journal of Language Education, 5*(2), 1-14.

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105
JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

27. Khan, R. M. I., Radzuan, N. R. M., Shahbaz, M., & Ibrahim, A. H. (2018). EFL Instructors' Perceptions on the Integration and Implementation of MALL in EFL Classes. *International Journal of Language Education and Applied Linguistics*, 39-50.
28. Kohonen, T. (2013). Essentials of the self-organizing map. *Neural networks, 37*, 52-65.
29. Krishnan, K., Krishnan, R., & Muthumari, A. (2017). A semantic-based ontology mapping–information retrieval for mobile learning resources. *International Journal of Computers and Applications, 39*(3), 169-178.
30. Lin, X. (1997). Map displays for information retrieval. *Journal of the American Society for information Science, 48*(1), 40-54.
31. Liu, Y., Albanie, S., Nagrani, A., & Zisserman, A. (2019). Use what you have: Video retrieval using representations from collaborative experts. *arXiv preprint arXiv:1907.13487*.
32. Liu, Y., Huang, Y., Zhang, S., Zhang, D., & Ling, N. (2017). Integrating object ontology and region semantic template for crime scene investigation image retrieval. Paper presented at the *2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*.
33. Meštrović, A. (2017). Collaboration networks analysis: Combining structural and keyword-based approaches. Paper presented at the Semanitic Keyword-based Search on Structured Data Sources.
34. Michalke, V. N., & Hartig, K. (2016). Explanation Retrieval in Semantic Networks.
35. Moreno, L., & Martinez, P. (2013). Overlapping factors in search engine optimization and web accessibility. *Online Information Review*.
36. Murthy, V. N., Maji, S., & Manmatha, R. (2015). Automatic image annotation using deep learning representations. Paper presented at the *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*.
37. Pal, A. R., & Saha, D. (2015). Word sense disambiguation: A survey. *arXiv preprint arXiv:1508.01346*.
38. Panchenko, A., Loukachevitch, N., Ustalov, D., Paperno, D., Meyer, C., & Konstantinova, N. (2018). Russe: The first workshop on russian semantic similarity. *arXiv preprint arXiv:1803.05820*.
39. Paralic, J., & Kostial, I. (2003). Ontology-based information retrieval. In Proceedings of the 14th International Conference on Information and Intelligent systems (IIS 2003), Varazdin, Croatia, pp. 23-28.
40. Petrucci, G., Ghidini, C., & Rospocher, M. (2016). Ontology learning in the deep. Paper presented at the European Knowledge Acquisition Workshop.
41. Purcell, K., Rainie, L., & Brenner, J. (2012). Search engine use 2012.
42. Ramli, F., Noah, S. A., & Kurniawan, T. B. (2016). Ontology-based information retrieval for historical documents. Paper presented at the *2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP)*.
43. Sanfilippo, E. M., Belkadi, F., & Bernard, A. (2019). Ontology-based knowledge representation for additive manufacturing. *Computers in Industry, 109*, 182-194.
44. Selvalakshmi, B., & Subramaniam, M. (2019). Intelligent ontology based semantic information retrieval using feature selection and classification. *Cluster Computing, 22*(5), 12871-12881.

International Journal on Recent Researches in Science, Engineering & Technology (IJRRSET)
A Journal Established in early 2000 as National journal and upgraded to International journal in 2013 and is in existence for the last 10 years. It is run by Retired Professors from NIT, Trichy. Journal Indexed in JIR, DIIF and SJIF.
Available online at: www.jrrset.com

ISSN (Print) : 2347-6729
ISSN (Online) : 2348-3105

JIR IF : 2.54
SJIF IF : 4.334
Cosmos: 5.395

Volume 9, Issue 12 – August 2021-22 - Pages 16-29

45. Shahbaz, M., & Khan, R. M. I. (2017). Use of mobile immersion in foreign language teaching to enhance target language vocabulary learning. *MIER Journal of Educational Studies Trends and Practices*, 66-82.
46. Vijayarajan, V., Dinakaran, M., Tejaswin, P., & Lohani, M. (2016). A generic framework for ontology-based information retrieval and image retrieval in web data. *Human-centric Computing and Information Sciences*, *6*(1), 1-30.
47. Yang, Y. (2012). Semi-structured semantic overlay for information retrieval in self-organizing networks. Paper presented at the Proceedings of the 21st International Conference on World Wide Web.
48. Ye, J., Dasiopoulou, S., Stevenson, G., Meditskos, G., Kontopoulos, E., Kompatsiaris, I., & Dobson, S. (2015). Semantic web technologies in pervasive computing: A survey and research roadmap. *Pervasive and Mobile Computing*, *23*, 1-25.
49. Ye, X., Shen, H., Ma, X., Bunescu, R., & Liu, C. (2016). From word embeddings to document similarities for improved information retrieval in software engineering. Paper presented at the Proceedings of the 38th international conference on software engineering.
50. Yu, H., Mine, T., & Amamiya, M. (2005). An architecture for personal semantic web information retrieval system. Paper presented at the Special interest tracks and posters of the 14th international conference on World Wide Web.
51. Zhang, N., Pu, Y.-F., & Wang, P. (2015). An ontology-based approach for chinese legal information retrieval. Paper presented at the Proc. CENet.
52. Zomahoun, D. (2013). Collaborative Semantic Annotation of Images: Ontology-Based Model. Paper presented at the Signal & Image Processing: An International Journal (SIPIJ).