Evolution of Information Retrieval System: Critical Review of Multimedia Information Retrieval System Based On Content, Context, and Concept

Ridwan Andi Kambau
Faculty of Computer Science
University of Indonesia
ridwan.andi@ui.ac.id

Zainal Arifin Hasibuan
Faculty of Computer Science
University of Indonesia
zhasibua@cs.ui.ac.id

Abstract— **In recent years the explosive growth of information affects the flood of information. The amount of information must be followed by the development of the effective Information Retrieval System (IRS) so that the information will be easily accessible and useful for the user. The source of Information contains various media format, beside text there is also image, audio, and video that called multimedia. A large number of multimedia information rise the Multimedia Information Retrieval System (MIRS). Most of MIRS today is monolithic or only using one media format like Google[1] for text search, tineye[2] for image search, youtube[3] for video search or 4shared[4] for music and audio search. There is a need of information in any kind of media, not only retrieve the document in text format, but also retrieve the document in an image, audio and video format at once from any kind media format of the query. This study reviews the evolution of IRS, regress from text-based to concept-based MIRS. Unified Multimedia Indexing technique is discussed along with Concept-based MIRS. This critical review concludes that the evolution of IRS follows three paces: content-based, context-based and concept-based. Each pace takes on indexing system and retrieval techniques to optimize information retrieved. The challenge is how to come up with a retrieval technique that can process unified MIRS in order to retrieve optimally the relevant document.**

Keywords—information retrieval, multimedia information retrieval, content-based MIR, context-based MIR, concept-based MIR

I. INTRODUCTION

Development of IRS is strongly influenced by the growth of data and information. The exponential growth of data must be balanced with reliable data search technique like IRS. Evolution of IRS begins from text based search or text-based information retrieval that using a keyword as a query. In the early stage of IRS Development is known as classic information retrieval, IRS is divided among three models, Set Theory model, Vector model, and Probabilistic model [1]. Each model had its characteristics. The main characteristic of the set-theory model represents document and queries through sets of keywords. Similarities are derived from the set-theoretic operation on those set. Boolean, Extended Boolean [2] and Fuzzy [3] model are included in the set-theoretic model. Vector model arises to enhance Boolean model [4] problems that have not rank and exact match or binary weight. Vector model uses term weighting to rank retrieved document and performing a partial match. This model represents documents and queries usually as vectors, matrices or tuples. The similarity of the query vector and document vector is represented as a scalar value. Generalized Vector Space [5], Latent Semantic Indexing (LSI), and Neural Network [6] model are three models that included in the Vector model. The Probabilistic model treats the process of document retrieval as probabilistic inference. Similarities are computed as probabilities that a document is relevant for given query. BM25, Divergence from Randomness, Language Model [7], Bayesian Network Model [8] and Latent Dirichlet Allocation are an example of the Probabilistic model.

Another important thing in classic IRS is IRS Evaluation that measures how well the system meets the information need of the user. Precision and Recall are the most popular retrieval evaluation [9]. Another IRS evaluation like Mean Average Precision and F-Measure also widely used to measure the IRS [1]. Measuring IRS not only requires IRS evaluation, but also need Reference Collection like TREC (Text Retrieval Conference) Collection [10], Reuter Collection [11], INEX collection etc [12].

In text-based IRS, there is the annotation that performs searching multimedia data automatically using text query, like image annotation using SVM (Support Vector Machine) [13], video annotation [14], and audio annotation [15]. All of the system search label or tag of image, video or audio, but the system could not read the content of image, video or audio because label could not represent the content of multimedia data. The Content-based MIRS offers a solution for this problem.

[1] www.google.com
[2] www.tineye.com
[3] www.youtube.com
[4] www.4shared.com

Multimedia data grows quickly and MIRS is regarded as one of the extensive research issues in IRS area of research. The availability of a large amount of data, includes multimedia data, requires an effective MIRS to find relevant, accurate and completeness of multimedia data. IRS that extracting feature content of image, video, and audio are called content-based MIRS is a solution of the text-based retrieval that could not read the content of multimedia data. Content-based MIR is not an exact or partial matching like text-based IRS but a similarity matching that means the system perform matching process between the multimedia query and multimedia document in the database based on similarity features of the multimedia content [16].

Accurate and relevant information is not only dependent on the set of query or content from multimedia data but also determined by the context (user, time, location, document, environment, event, and so forth) [17]. Context based MIR improve effectivity of content based MIR, especially in the accuracy of the multimedia document retrieved, by adding context to the retrieval technique.

Many MIRS search result are based on the occurrence of query or based on a feature of content from multimedia data can not find a relevant document that does not mention query terms explicitly, especially when a user only entering very short queries, this shortcoming can be improved by incorporating human knowledge and concept detector in the MIRS. It is called concept-based MIRS [18].

MIRS that have been explained above and exist today only using one media like Flickr[5] and Google Image[6] for Image Search, Youtube, and Vuclip[7] for Video Search and for Music and Audio Search there are 4shared or Findsounds[8]. Multimedia data, including text, image, video, and audio can come from anywhere or any resource that has no relation to one another, but potentially interrelated. So it is possible if the user needs information from any kind of data from the variety of resources at one-time searching. But today it is still difficult for MIRS to retrieve all media at once.

In the case that almost same with one time searching to get any kind of data, some user still need more, they need multimedia data retrieved in semantic concept, it means data is not only limited by terms query explicitly (syntactic) but also including the meaning of query or the intent behind the query (semantic) [19]. Today it is still the problem of MIRS.

The remainder of this paper is organized as follows, Section 2, provide information about IRS evolution. Section 3 describes evolutions of MIRS based on content, context, and concept. Section 4 explains Critical Review and Section 5 is about the challenge and future work of MIRS.

[5] www.flickr.com
[6] image.google.com
[7] www.vuclip.com
[8] www.findsounds.com

II. Information Retrieval System

The existence of IRS can not be separated from the flood of data. IRS evolves and constantly improves.. The weakness of old IRS will be rectified in the new one. The first and simplest IRS is the Boolean model [4], that is stand long enough in its time as a search system. The Boolean model that included in the Set-Theoretic model using binary index term weight, it predicts the result only relevant or non-relevant, there is no ranking, which might lead to the retrieval of too few or too many documents. Vector model overcomes the shortcoming of the Boolean model with term weighting with considering how important this term for describing a document. The most popular term weighting is tf-idf (term frequency-inverse document frequency) based on frequency level [20].

Because of too many models in IRS, so we have to select some models as a representation of all IRS model. As a foundation of many IRS, and until now its technology still in used, text-based IRS will be discussed first. Text-based IRS with keyword has started with index term technique that using the term as a reference for indexing. Term Indexing [21] that perform indexing automatically was one of the early IRS, but the system had a very high computing cost and can not recognize synonymy and polysemy words. The issue of polysemy and synonymy is researched [22] with Latent Semantic Indexing (LSI). IRS with LSI had used Bag of Words (BoW) concept that could reduce computational cost and recognize some synonymy and polysemy words, but in the experiment, many synonymy and polysemy are not detected. This weakness is overcome by probabilistic Latent Semantic Indexing (pLSI) [23] that could improve ability to recognize the words that have multiple meanings (polysemy). The next step of IRS development using three layers of Bayesian probability technique that are called Latent Dirichlet Allocation (LDA) [24] is used to increase the effectivity of IRS, particularly to handle synonymy and polysemy problems. However, LDA can not realize difficulties of semantic knowledge problems. The improvement of LDA is Tag-LDA that could fix semantic knowledge problems with using corpus and lexical database [25]. The use of lexical database or ontology and corpus become the latest trend in text-based IRS and emerging the new IRS is called Concept-based IRS. One of the early concept-based text retrievals [26] is with Explicit Semantic Analysis (ESA). Concept-based text retrieval needs many resources and has to develop document corpus and BoW and Concept Detector. Further development of text-based retrieval followed concept-based retrieval system.

III. Multimedia Information Retrieval System

The main issue in MIRS was how to bridge the “Semantic Gap” or how to translate the easily computable low-level content-based media features to high-level concepts or terms which would be intuitive to the user [16].

Like IRS, MIRS also evolved constantly improve themselves. In this paper, the development of MIRS is divided into three major parts, Content-based MIRS, Context-Based MIRS and Concept-Based IRS.

Content-based MIRS focus on feature-based similarity over image, video, and audio. Extracting image features like color, shape and texture, [27] segmenting video (key frame or shot boundary) and extracting video feature like image feature plus motion feature [28] and Audio features consist of acoustic features (loudness, spectrum, pitch, bandwidth and spectrum) and semantic features (timbre, rhythm, events and instrument) [29]. Content-based MIRS match the multimedia query and multimedia document in the databases based on similarity features of multimedia data to produced relevant and accurate retrieved document [16].

Information also influenced by context or moment when performing a search. Capturing and integrating contextual information in the retrieval process can increase the search performance and reducing the ambiguity of information. [30] Context-based MIR combines the technique of search, query awareness, and user context into a single framework in order to provide the most appropriate response to their information need. Context affects all aspect of MIRS like how they interact with the system, what type of response they expect from a system and how they make the decision about the information object they retrieve. To many contexts, but based on [17] context can be a user, device, time, location, document, environment and event.

Content-based MIR and context based MIR are still inaccurate and incomplete when different keywords are used to describe the same concept in the document and in the query. Concept-based MIRS have attempted to solve this problem with using corpus and thesauri or human world knowledge. [26] With the knowledge base, retrieved document not only refer to query term explicitly but also refer to semantic meaning. Besides that, there is corpus-based with concept detector as a trainer. Effectivity of Concept-based MIRS is better than Content and Context based MIRS, but it requires too many resources like knowledge base from ontology mapping or lexical database and corpus. [31]

A. Content-based Multimedia Information Retrieval System

The fundamental problem is how to enable or improve multimedia retrieval using content-based methods that are necessary when text annotation is non-existent or incomplete. Content-based methods use the visual and audio content.

The initial evolution of MIRS was the development of Content-based MIR that consists of Content-based Image Retrieval (CBIR), Content-based Video Retrieval (CBVR) and Content-based Audio Retrieval (CBAR). The fundamental problem in this system was how to enable or improve multimedia retrieval using the content-based method.

1) Content-based Image Retrieval (CBIR)

Content-based image retrieval is a technique which uses visual content to search images from large-scale image database according to users' interest. [32]

One of the early CBIR was developed by IBM with QBIC project [27]. QBIC was a simple CBIR that using color, shape and texture features to recognize 1000 picture (any object) with R-Tree variation indexing. To evaluate this system was using Precision-Recall and Similarity measure matched image query and image in the database. The use of the global feature like GIST representation [33] increases the match quality between image query and image document in the database and optimizing the trade-off between memory usage and precision. Scale Invariant Feature Transform (SIFT) was Local Feature for Image that using key point to detect the visual similarity of another image. SIFT Descriptor [34] make image invariant in rotation and scale. It helps the acceleration of similarity image matching process. Like a SIFT, Speed-up Robust Feature (SURF) was a local feature for an image that using key point, but SURF have more invariant component, beside rotation and scale, there is the angle, blurring, and noise. SURF [35] had better performance than SIFT even they use the same concept.

Besides using the visual descriptor like SIFT and SURF, some CBIR utilizes learning algorithm to increase performance or to rank retrieved image like Learning to Rank CBIR [36]. CBIR also exploited Deep Learning with using Deep Auto-Encoder [37] for reconstructing the image and the label (bag of words) as a representation of image caption. The last approach of CBIR in this research using CENTRIS (CENsus Transform HISTogram), plus color and texture feature [38] were proving integrates three features could enhance the retrieval performance, but three kinds of similarity can not change self-adaptively which needs to improve.

2) Content-Based Video Retrieval (CBVR)

Content-based video retrieval (CBVR) systems analyze visual video content and generate appropriate data required to summarize and retrieve content from large video databases [39].

Content-based Video Retrieval (CBVR) was most complicated MIRS if we compare with CBIR and CBAR, too many components of this system, but research in this field wide open. First research in CBVR from [40] with Mining Temporal Pattern (MTP) Generation and indexed by Fast Pattern Index Tree. This system can deal with high dimension and visual feature problems. One of the machine learning algorithm, Support Vector Machine (SVM) Classification was used CBVR to create effective video retrieval [28], but the result of evaluation was low accuracy and precision. Another Video Retrieval Project that [41] called LivRE (Lucene Image Video Retrieval) utilizing combination of image and video retrieval algorithm in web-base. The modular characteristics cause easily to use it. Some CBVR used Deep Learning, one of them was Supervised Recurrent Hashing (SRH) for Large Scale Video Retrieval [42] using Convolutional Neural Network and Long Term Memory Network and comparing with Long Short-Term Memory Network (LSTMN). Based on comparison LSTMN was proven SRH performance had better then LSTMN.

3) Content-Based Audio Retrieval (CBAR)

Given any audio piece, we can instantly tell the type of audio (e.g., human voice, music or noise), speed (fast or slow), the mood (happy, sad, relaxing etc.), and determine its similarity to another piece of audio. This is the technique of content-based audio retrieval

Unlike CBIR and CBVR, Content-based Audio Retrieval using signal and frequency as the feature.

Actually, CBAR was divided into three areas, music, sound and speech, but for this research, we only used music and sound. Many research in CBAR, but we only use five papers to represent CBAR. Initial paper [29] about Hierarchical System in CBAR where 1500 pieces of sound are extracted with Mel-frequency cepstral coefficient (MFCC) and tested by Hidden Markov Model (HMM) and Gaussian Mixture Model, the result was Accuracy rate of a coarse feature about 90% and Perceptual Feature about 80%. Fingerprinting was audio detection because can track similar audio from audio database accurately. Single Value Decomposition included Discrete Fourier Transform (DFT) and Discrete Cosine Transform (DCT) is the algorithm that [43] created. Audio Fingerprinting also used Spectral Flux for Audio Retrieval, Its algorithm using Low Pass Filter and Fourier Transform and this algorithm better than another tested algorithm like Philips Algorithm. Like in CBIR and CBVR, we could use Deep Learning to improve the performance of audio retrieval. CBAR was using Deep Convolutional Neural Network (D-CNN) [44] significantly outperforming traditional BoW representation for audio retrieval. Another technique of CBAR was codebook-based [45], that was tested and compared with Query by Tag and Query by Example and the result audio retrieval that utilize codebook outperforms.

B. *Context-Based Multimedia Information Retrieval System*

Contextual Retrieval is defined as 'combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for user's information need'. [46]

Research on the contextual information retrieval field had proven that the state when the user conducts a search had a perceptible effect on the user's search behavior. The search context may include several dimensions such as time, location, user, current task etc. In MIRS field, it had taken a very important part of research aim to improve the relevance of the search result.

Here, some of Context-based MIR with context user, time & location, document and environment & event. In MIR with Context Document [47], contain two part in one system, the first part was CBIR and another was a context document. In CBIR using HSV Color and Gabor Filter while context document using index term and LSI Algorithm. The result was Combination text and image retrieval outperforms from single information retrieval. MIR with Context User was very popular than another context, social media often use this MIR with user context. [48]. With cluster algorithm, this MIRS with user context was better than a naïve model. CBIR with Context Time & Location often used in the gadget or device with an assortment of features this MIRS [49], can improve retrieval image & context location performance, with reducing computational cost for checking location. MIR with Context Event using Hapori Search as sample paper to test its performance, compared with Mobile Bing Local and the result was the performance of Hapori Search. Evaluation using precision-recall denoted Hapori Search had good performance.

C. *Concept-Based Multimedia Information Retrieval System*

Content-based retrieval is difficult to describe its semantic visual features or semantic audio features. Concept-based MIR has attempted to tackle these difficulties by using manually built thesauri or by extracting latent word relationship and concept from the corpus. For multimedia data, it needs classifier to build concept detector model by gathering a large pool of multimedia data and using machine learning to select training set and testing set so that we catch the semantic visual feature or semantic audio feature in concept terms. Concept based MIR was divided in Concept-based Image Retrieval, Concept-based Audio Retrieval and Concept-based Video Retrieval.

1) *Concept-based Image Retrieval (CpBIR)*

Concept-based Image Retrieval (CpBIR) aim at enabling indexing and subsequent retrieval of images based on concepts that are automatically detected from visual content of images, as well as from any accompanying metadata. Example of concepts include image scene elements ("sky", "sea"), action ("person running", "smiling face") or object ("car", "flower"). The use of concepts allows textual queries on non-annotated image collection. The paper [50] described Concept-based image retrieval with training weight computed from tags, it means every image in the database had tag and weight. To collect image concept using concept detector that built from training and testing data in the learning process. This MIRS is a highly effective method for ranking candidate training images was outlined, that uses existing image tags, a reference corpus, and WordNet to assign scores with respect to a concept. Artificial Neural Network (ANN) based distributed processing architecture for semantic image retrieval [51] can retrieve image quickly and detect image as a concept. The use of knowledge domain like WordNet and ImageNet to capture concept from visual features was researched by Feng and Bhanu [52] with the contribution to the literature on context-based co-occurrence pattern in computer vision where co-occurrences of concept used as contextual cues for improved concept inference.

2) *Concept-based Video Retrieval (CpBVR)*

Concept-based Video Retrieval is one of the video search techniques that automatically detected concept. The concept derived from the combination of the knowledge-based and corpus-based semantically. The semantic concepts are managed by National Institute of Standards and Technology (NIST). For the evaluation of video retrieval, TREC Video Retrieval Evaluation (TRECVID) dataset is utilized as well. [31]

Like CpBIR, CpBVR applies same technique, but still need an addition in motion features. [53] research about Concept-based Video Retrieval utilize unified 12 kinds of feature to reduce its computational complexity. The concept co-occurrence matrix and several assistant methods (B&W detection, audio detection, and motion detection) are suggested to enhance the performance of the video retrieval system. To bridge semantic gap, concept-based video retrieval has been considered as a feasible alternative technique for video search. In order to retrieve a desirable video shot, a query should be defined based on users' needs. In spite of the fact that query can be on object, motion, texture, color and so on, queries which are expressed in terms of semantic concepts are more intuitive and realistic for end users. Therefore, a concept-based video retrieval model based on the combination of the knowledge-based and corpus-based semantic word similarity measures is proposed with respect to bridging the semantic gap and supporting semantic queries. [31]. Another model of concept-based video retrieval unsupervised zero-shot retrieval where no training videos are provided: In this work, we introduce a new method for automatically identifying relevant concepts given a text query using the Markov Random Field (MRF) retrieval framework. We use source expansion to build rich textual representations of semantic video concepts from large external sources such as the web [54].

3) Concept-based Audio Retrieval (CpBAR)

Slightly different from CpBIR and CpBVR, CpBAR is using audio signal and audio feature to find sound or music, but the overall concept almost same. CpBAR exploits sound effect corpus and using machine learning to create audio concepts [55] with using Gaussian Mixture Model and Kullback Leibler divergence this system can improve its query from the sample of audio and increasing the performance quantitatively and qualitatively. Some CpBAR has used Deep Learning like Audio retrieval using semantic similarity. Hierarchical-Deep Neural Network H-DNN is main algorithm to build this system. [56]. HDNN significantly outperforms Gaussian Mixture Model (GMM) and Neural Network baselines. Another audio retrieval was Music Retrieval that using Support Vector Machine as its algorithm [57]. This system was efficient to retrieve music and simple to implement.

The illustration of evolution Text-Based IRS, Content-Based MIRS, Context-Based MIRS and Concept-Based MIRS can be shown in Fig.1. The Picture of MIRS shows the evolution process and their Techniques.

IV. Critical Review

The sequences of Text-based IRS, Content-based MIRS, Context-Based MIR, and Concept-based MIR (See. Fig.1) refer to the maturity level or complexity of the IRS techniques. Maturity in a term when text-based IRS like LDA is not able to detect the multimedia data, it is covered by content-based MIRS with SIFT and SURF technology. Context-based MIRS improves effectiveness by increasing the accuracy of content-based MIRS, it can be seen in HAPORI Search that adds event context in the search result. Concept-based MIR enhances the weakness of previous IRS, primarily in the meaning of semantic query whether query text, image, video, and audio. Further, increase the retrieved document relevance in multimedia format. For examples ESA for text, Concept-based Image Retrieval with Training Weight for Image, Concept-based Video Combination of Semantic Similarity for Video Search and Audio Information Retrieval Using Semantic Similarity for Audio.

So does the Concept-based MIRS is the best? The answer is uncertain yet, if more complex of the IRS the answer probably yes, because each IRS has a specific purpose. The text-based IRS is very popular as a search engine, content-based MIRS is very useful for a medical care. Context-based MIRS can increase the accuracy and relevance and more focus to find information need based on their context and Concept-based MIRS with a complete structure for research or to handle the complex problems of searching all kind of format data.

There are two problems need to be analyzed in the evolution of MIR, namely;

- Constraint of monolithic multimedia information retrieval
- Less relevant to the retrieved document from existing MIR.

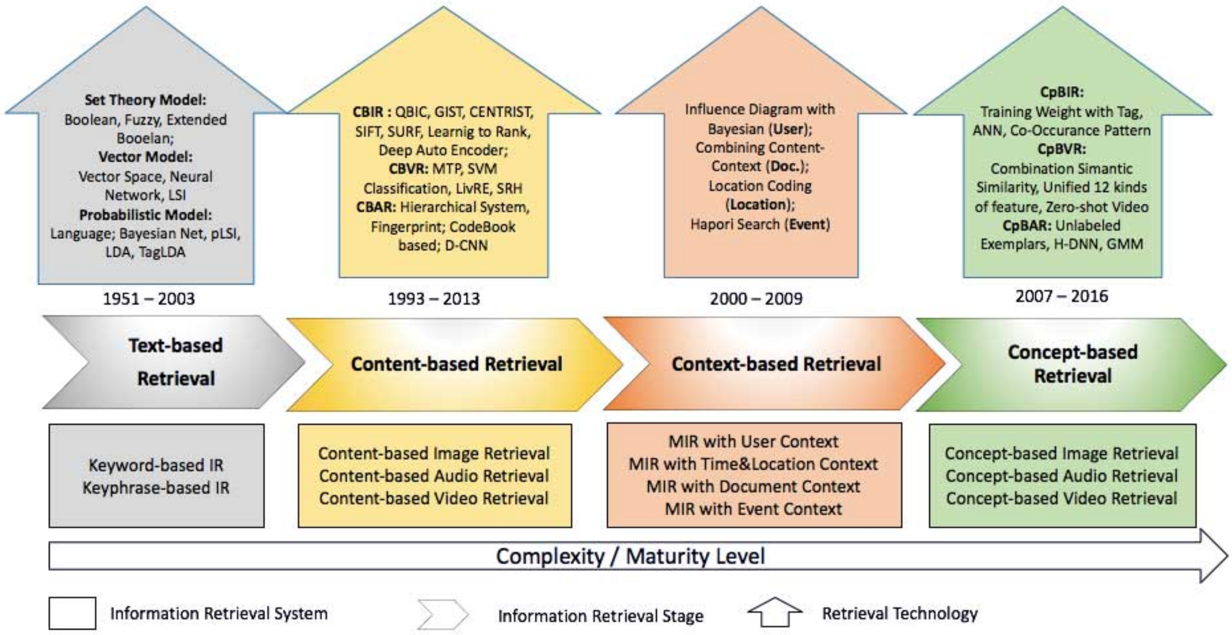

Fig.1. Multimedia Information Retrieval Evolution Stage with their Techniques

Multimedia data with text, image, video and audio data can come from anywhere and different resources that probably have not related one another, but the content has a potency to related even though in different media or from a different resource. For examples, the user needs one topic of information from anywhere and from any kind of media in text collection, image collection, video collection and audio collection. And then when they search the desired topic they will get information about the topic in the text, image, audio and video format at once. It is difficult for monolithic MIR like the Content-based MIR, the Context-based MIR, and the Concept-based MIR solve this problem. We need another multimedia information retrieval system that can unify all media format like A Multidimensional Approach of Content-based MIR [59].

Beside that, in the almost same case, the user need retrieved document in all media format and highly relevant to the topic. It uses semantic concept with knowledge-based or corpus-based like dictionary or ontology. Semantic concept means when the data not only limited by the similarity of the query or content of the query (syntactic) but also the meaning of the query or the intent behind the query or same concept that associated with the query (semantic). This technique will tackle the less relevant problem in the MIR.

V. Challenge & Future Works

Content-based MIRS, Context-Based MIRS, and Concept-Based MIRS are monolithic information retrieval system. It means each of MIRS is only perform one searching technique with one media format. It is difficult if the user needs one topic of information from anywhere and any kind of media format and to be taken at once. To handle this problem, it is required the unified multimedia retrieval technique or unified multimedia indexing that may retrieved document in all any kind of media.

In another case, besides needing unified multimedia retrieval technique the user also needs information that associated with the same concept of the multimedia query to get highly relevant to the topic. For this case, a concept-based MIRS with knowledge-based or corpus is needed.

For the future work, we need to develop a framework that unified MIRS based on content or based on concept with unified multimedia indexing technique to find or to catch any information from any kind of media format and not only retrieved multimedia document base on the similarity of the query, but also to find information with the same concept or semantic information from the query. There is a research attempt to design Unified MIRS in content-based [59], but it still need improvement to find information that related conceptually.

Today the development of MIRS will face many challenges for the future, besides of monolithic MIRS constraints, we also need to expand knowledge domain or create a specific or global ontology to enrich concept domain for the highly relevant retrieved document. In addition, for image, audio, and video we should be built multimedia concept detector using a large number multimedia dataset in order to capture the concept and relation between the concept that contained in the visual and audio features. This is for extended concept-based MIRS.

VI. Conclusion

This critical review concludes that the evolution of IRS follows three paces: content-based, context-based and concept-based MIRS as the research issues that raise the monolithic IRS problem and less relevance problem of existing MIRS.

Unified Concept-based MIRS is the proposed solution to solve monolithic MIRS problem with unified multimedia indexing technique and Less relevance of retrieved document from concept-based MIRS.

References

[1] R. B. Yates and B. R. Neto, Modern Information Retrieval, The Concept and Technology Behind Search, 2nd, England: Addison Wesley, 2011.

[2] G. Salton, E. Fox and H. Wu, "Extended Boolean Information Retrieval," in *Communication of the ACM*, 1983.

[3] Y. Ogawa, T. Morita and K. Kobayashi, "A Fuzzy Document Retrieval System using Keyword Connection Matrix and a Learning Method," in *Fuzzy Sets & Systems*, 1991.

[4] S. Warwick, W. Frakes and R. Baeza, "Boolean Operations," *Prentice Hall,* vol. Information Retrieval, pp. 264-292, 1992.

[5] M. Wong, W. Ziarko and N. Wong, "Generalized Vector Space Model in Information Retrieval," in *Eight ACM SIGIR Conference on Research & Develoment in Information Retrieval*, New York, 1985.

[6] R. Wilkinson and P. Hingston , "Using the Cosine Measure in Neural Network for Document Retrieval," in *14th ACM SIGIR Conference on Research & Development in Information Retrieval*, Chicago, USA, 1991.

[7] C. Zhai, "Statistical Language Model for Information Retrieval," *Synthesis Lectures of Human Language Technologies, Morgan Claypool,* 2008.

[8] H. Turtle and B. W. Croft, "Evaluation of Inference Network Based Retrieval Model," *ACM Transaction on Information Systems,* vol. 9, no. 3, pp. 187-222, 1991.

[9] C. Claverdon, "The Cranfield Test on Index Language Device," in *ASLIB Proceedings*, 1967.

[10] E. M. Voorhees and D. K. Harman, "TREC: Experiment and Evaluation in Information Retrieval," *The MIR Press,* 2005.

[11] D. L. David, "http://www.daviddlewis.com/resources/testcollections/," [Online].

[12] I. T. Developer, "http://inex.mmci.uni-saarland.de/," [Online].

[13] C. Cusano, G. Ciocca and R. Schettini, "Image annotation using SVM," *Proc. SPIE 5304, Internet Imaging V, 330 (,* 12 2003.

[14] V. Lavrenko, S. Feng and R. Manmatha, "Statistical Model for Automatic Video Annotation and Retrieval," in *IEEE International Conference Acoustics, Speech, and Signal Processing*, Montreal, 2004.

[15] D. Turnbull, L. Barrington, D. Torres and G. Lanckriet, "Semantic Annotation and Retrieval of Music and Sound Effects," in *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 2008.

[16] M. S. Lew, N. Sebe, C. Djeraba and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges," in *ACM Transaction on Multimedia Computing, Communication and Application*, 2006.

[17] D. Boughareb and N. Farah, "Context in Information Retrieval," in *International Conference on Control, Decision and Information Technologies (CoDIT)*, 2014.

[18] Z. A. Hasibuan, A. Kurniawan and R. Budiarto, "Multi-Format Concept-Based Information Retrieval Using Data Grid," in *Journal of Advanced Computing and Application*, 2010.

[19] A. D. Lin, P. J. Graydon and et. al., "Concept-based Search and Retrieval System," in *Google Patent*, 2004.

[20] G. Salton and C. S. Yang, "On the Specification of Term Values in Automatic Indexing," *Journal of Documentation,* vol. 29, pp. 351-371, 1973.

[21] G. Salton and C. Buckley, "Term Weighting Approaches in Automatic Text Retrieval," *Information Processing and Management,* p. 322, 1988.

[22] S. Deewester, S. T. Dumais, G. W. Furnas and et al, "Indexing by Latent Semantic Analysis," *Journal of American Society for Information Science,* p. 391, 1990.

[23] T. Hofmann, "Probabilistic Latent Semantic Indexing," in *Proceedings of the 22nd Annual International ACM SIGIR Conference* , 1999.

[24] D. M. Blei, A. Y. Ng and I. M. Jordan, "Latent Dirichlet Allocation," *Journal of Machine Learning,* pp. 993-1022, 2003.

[25] F. LI, H. SHEN and T. HE, "Tag-Topic Model for Semantic Knowledge Acquisition from Blogs," in *7th International Conference NLP-KE*, Tokushima, Japan, 2011.

[26] O. Egozi, S. Markovitch and E. Gabrilovich, "Concept-based Information Retrieval Using Explicit Semantic Analysis," in *ACM Transactions on Information Systems*, 2011.

[27] W. Niblack, R. Barber and et al, "QBIC project: querying images by content, using color, texture, and shape," in *Spie digital Library*, 1993.

[28] G. Nagaraja, R. Murthy and T. Deepak, "Content Based Video Retrieval Using Support Vector Machine Classification," in *Applied and Theoretical Computing and Communication Technology*, 2015.

[29] T. Zhang and C. J. Kuo, "Hierarchical System for Content-based Audio Classi cation and Retrieval," in *Proceedings Spiedigital Library*, 1998.

[30] L. Tamine-Lechani, M. Boughanem and M. Daoud, "Evaluation of Contextual Information Retrieval Effectiveness: Overview of Issues and Research," *Knowledge Information System,* vol. 24, pp. 1-34, 2010.

[31] S. Memar, L. S. Affendey, N. Mustapha and M. R. Ekteva, "Concept-based Video Retrieval Model Based On The Combination of Semantic Similarity Measures," in *13th International Conference on Intelligent Systems Design and Applications (ISDA)*, Malaysia, 2013.

[32] F. Long, H. Zhang and D. D. Feng, "Fundamental of Content-based Image Retrieval," *Multimedia Information Retrieval,* 2006.

[33] M. Douze, H. Jegou, H. Sandhawalia, L. Amsaleg and C. Schmid, "Evaluation of GIST descriptors for web-scale image search," in *Proceedings of the ACM International Conference on Image and Video Retrieval*, Santorini, 2008.

[34] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* pp. 1-28, 2004.

[35] A. Alfanindya, N. Hashim and C. Eswaran, "Content Based Image Retrieval And Classification Using Speeded-Up Robust Features (SURF) and Grouped Bag-of-Visual-Words (GBoVW)," in *2013 International Conference on Technology, Informatics, Management, Engineering & Environment (TIME-E 2013)*, Bandung, 2013.

[36] F. F. Faria, A. Veloso, H. M. Almeida and et al, "Learning to Rank for Content-Based Image Retrieval," in *10th Proceedings of the international conference on Multimedia information retrieval,* Philadelphia, 2010.

[37] A. Krizhevsky and G. E. Hinton, "Using Very Deep Autoencoder for Content Based Image Retrieval," in *European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, Bruges, 2011.

[38] G. Lin Shen and X. Jun Wu, "Content Based Image Retrieval by Combining Color, Texture and CENTRIST," in *Constantinides International Workshop on Signal Processing*, London, 2013.

[39] N. Sebe, M. Lew and A. Smeulders, "Video Retrieval and Summarization," *Computer Vision and Image Undertanding,* pp. 141-146, 2003.

[40] J.-H. Su, Y.-T. Huang and V. S. Tseng, "Efficient Content-based Video Retrieval by Mining Temporal Patterns," in *Proceedings of the 9th International Workshop on Multimedia Data Mining*, Las Vegas, 2008.

[41] G. d. Olieveira Barra, M. Lux and X. Giro i-Nieto, "Large Scale Content-Based Video Retrieval with LIvRE," in *14th International Workshop on Content-Based Multimedia Indexing (CBMI)*, 2016.

[42] Y. Gu, C. Ma and J. Yang, "Supervised Recurrent Hashing for Large Scale Video Retrieval," in *Proceedings of ACM on Multimedia Conference*, Amsterdam, 2016.

[43] P. Cano, E. Battle, T. Kalker and J. Haitsma, "A Review of Audio Fingerprinting," in *Journal of VLSI Signal Processing*, Netheland, 2005.

[44] A. v. de Oord, S. Dieleman and B. Schrauwen, "Deep content-based music recommendation," in *Advances in Neural Information Processing Systems 26*, 2013.

[45] Y. Vaizman, B. McFee and G. Lanckriet, "Codebook-Based Audio Feature Representation for Music Information Retrieval," in *IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, 2014.

[46] J. Allan, "Challenges in Information Retrieval ang Language Modelling," *Report of a workshop held at the Center fo Intelligent Information Retrieval,* 2002.

[47] T. Westerveld, "Image Retrieval: Content versus Context," in *Proceeding Content-Based Multimedia Information Access*, Paris, 2000.

[48] W. Zemirli, L. Tamine-Lechani and M. Boughanem, "A Personalized Retrieval Model based on Influence Diagrams," in *Conference on Modeling and Using Context* , Roskilde, 2007.

[49] S. S. Tsai, D. Chen, G. Takacs, V. Chandrasekhar and et al, "Location Coding for Mobile Image Retrieval," in *Proceedings of the 5th International ICST Mobile Multimedia Communications Conference*, London, 2009.

[50] V. Papapanagiotou, C. Diou and A. Delopoulos, "Improving Concept-Based Image Retrieval with Training Weights Computed from Tags," in *ACM Transactions on Multimedia Computing, Communications, and Applications*, New York, 2015.

[51] S. J. Karande and V. Maral, "ANN based Semantic Content Based Image Retrieval with Distributed Processing," in *International Conference on Contemporary Computing and Informatics (IC3I),* , Mysore, 2014 .

[52] L. Feng and B. Bhanu, "Semantic Concept Co-Occurrence Patterns for Image Annotation and Retrieval," in *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 2016.

[53] X. Guo, Z. Zao, Y. B. Chen and A. Cai, "AN IMPROVED SYSTEM FOR CONCEPT-BASED VIDEO RETRIEVAL," in *3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC)*, 2012.

[54] J. Dalton, J. Allan and P. Mirajkar, "Zero-Shot Video Retrieval Using Content and Concepts," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, San Fransisco, 2013.

[55] L. Barrington, A. Chan, D. Turnbull and G. Lanckriet, "AUDIO INFORMATION RETRIEVAL USING SEMANTIC SIMILARITY," in *IEEE International Conference on Acoustics, Speech and Signal Processing* , Hawaii, 2007.

[56] M. Ravanelli, B. Elizalde, K. Ni and G. Friedland, "AUDIO CONCEPT CLASSIFICATION WITH HIERARCHICAL DEEP NEURAL NETWORKS," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO),* , Lisbon, 2014.

[57] P. Keng and Y.-H. Yang, "Music Annotation and Retrieval using Unlabeled Exemplars: Correlation and Sparse Codes," in *IEEE SIGNAL PROCESSING LETTERS*, 2015.

[58] A. Rubhasy, Z. A. Hasibuan, I. Budi and G. P. Mindara, "A Multidimensional Approach in Content-based Multimedia Information Retrieval System," in *International Conference on Computer Engineering and Applications (ICCEA 2011)*, 2011.

This page intentionally left blank