Journal of Theoretical and Applied Information Technology
30th September 2019. Vol.97. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

UNIFIED CONCEPT-BASED MULTIMEDIA INFORMATION RETRIEVAL SYSTEM USING DEEP LEARNING PROCESS WITH ONTOLOGY

RIDWAN ANDI KAMBAU , M.OCTAVIANO PRATAMA
Faculty of Computer Science, University of Indonesia, Depok, Indonesia
E-mail: ridwan.andi@ui.ac.id www.com, m.octavioano61@ui.ac.id

ABSTRACT

The amount of digital data is growing at a staggering pace and mostly in the form of text. The growth of data, including multimedia (text, images, video, and audio) poses challenges in developing Multimedia Information Retrieval Systems (MIRS). Today, MIRS uses one or two media as a query input, for example, Google using text and image. There are comprehensive information needs in multimedia, including video and audio media as query input that can increase the amount and variety of information in retrieval result. Also, it is difficult to organize the relationship between the query input and the retrieval result in the same context or semantically related. This paper proposes a Unified Concept-based MIRS using deep learning with Ontology to tackle these problems. There are three main processes of this research; the first is Indexing Process which consist of collecting multimedia data, creating the multimedia dataset, extracting multimedia features, identifying and classifying objects and media format with Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), then storing object indexed as the concepts. The second is Query Processing which consists of inputting multimedia query or object to identifying and classifying object becomes a concept, and then the third is Retrieval and Rank Process of concept or object indexed and concept of query input. The ontology organizes the relationship between the concepts. Resource of this research uses the Cultural Heritage domain. The Unified Concept-based MIRS shows the capability of the system to extract features of four media (text, image, audio, and video) as concept representation to identify the multimedia object from query input and retrieve the object in four types of media at once. The retrieval results performance with deep learning increase about 30%-40% than support vector machine technique, and the usage of ontology increases the retrieval relevant results about 30 % is compared with MIRS without Ontology.

Keywords: *Information Retrieval System, Multimedia Information Retrieval System, Ontology, Content-based MIRS, Context-based MIRS and Concept-based MIRS.*

1. INTRODUCTION

The growth of digital data exponentially in the last two decades followed by various media or multimedia influences the development of the Information Retrieval System (IRS). IRS become more accurate, complete and relevant along with data growth of multimedia. The complexity of the data and their structure then the comprehensive information need of user in various media affect the changing retrieval techniques. The retrieval results are more complete if using various media rather than only using media text. Further with concept-based information retrieval, the multimedia format of data can be related with another data in different media even though the input query is explicitly different, but they have the same meaning and related in the same context with the input query.

The initial phase of IRS starts from the development of the classical IRS then followed by Text-based information retrieval dominance moves towards multimedia information retrieval where media not only text but also image, video, and audio that are called Multimedia Information Retrieval System (MIRS). Initially, MIRS uses annotation techniques to search multimedia information using text queries automatically, such as image annotations using the Support Vector Machine [1], automatic video annotations [2], and annotations audio and music [3]. The MIRS looks for labels or tags on the metadata and titles of images, video, and audio, but it cannot recognize the content or the

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

features of multimedia data. To identify the representation of multimedia data content or features is required Content-based MIR that comparing and matching the characteristic of the multimedia data to find similarity between the multimedia query and multimedia data collection.

Content-based Image Retrieval (CBIR) [4], Content-based Video Retrieval(CBVR)[5], and Content-based Audio Retrieval (CBAR) [6] extract each low-level features from images, video, and audio to perform similarity measure between the features of a multimedia query and the multimedia data collection in the repository. To increase accuracy and decrease the polysemy or synonymy problems of the MIR is used context. Contexts such as users, times, locations, documents, environments, or events when associated with MIRS will separate the query based on the context is called Context-based MIR. The weakness of Content-based and Context-based MIR cannot understand and recognize the meaning of the query. This problem is solved by Concept-based MIRS that using the thesaurus or extracting latent word relationships and obtaining concepts from the corpus or mapping the ontology.

Content-based and Concept-based MIR which is discussed above is mono-modal information retrieval or only applying one media while there is comprehensive information need to unify text, image, video, and audio format in one MIRS or multi-modal of IRS. Today, Search Engine like Google and Bing allow text and image media for their queries and the search results present text, image, and video formats. Several research perform multi-modal media in MIR such as unify text-based and content-based image retrieval [7], [8], [9], unify text-based and content-based video retrieval [10], [11], [12], unify text-based and content-based audio retrieval [13], [14], or unify text, image, audio, and video retrieval [15], [16].

Development of machine learning and deep learning enrich extraction features and classification techniques of multimedia data or object including text, image, audio, and video features, especially in content-based information retrieval. A machine learning technique like Support Vector Machines (SVM) can be applied in text retrieval [17], content-based image retrieval [18], content-based audio retrieval [19], and content-based video retrieval[20]. Besides the SVM technique, Naïve Bayes[21] and Ensemble [22] model is also applied in Text Retrieval. Deep Learning improves Machine Learning capability with several techniques such as Deep Belief Network (DBN), Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN). The techniques are applied in MIRS in the deep learning for Text Retrieval use RNN [23], for Content-based Image Retrieval [24], Audio Retrieval [25] and Video Retrieval [26] can use CNN with several adjustments.

Concept-based Information Retrieval exploits knowledge representation with the concept terms, Latent Semantic Analysis, Thesaurus (e.g.,WordNet), and Ontology Mapping are samples of knowledge representation [27]. Concept-based Information Retrieval using Ontology not only search based on keyword but also retrieve information related to the keyword or multimedia query. Ontology maps and structures information or keyword to the concept or context that can be easy to understand by a human [28].

In the context of cultural heritage a few research on Concept based MIR using a standard ontology, for instance, CIDOC-Concept Reference Model (CRM) [29] and Europana Model [30]. The models are a reference for cultural heritage MIRS for examples REACH Project [31], CultureSampo[32], or Cantabria [33]. The uniqueness object of cultural heritage such as heterogeneous media and various object, and then relatedness with another object are suitable with concept-based MIRS using an ontology.

This paper proposes the Unified Concept-based MIRS using an Ontology with deep learning techniques try to answer these questions. How to perform unified multimedia indexing that can identify query input in four types of media and retrieve object in four types of media at once ? and How to organize and structure object in the formal concept representation and build relationship among concept ?. The cultural heritage object is used as research domain refer to CIDOC-CRM structure.

The methodology is started with data collection and creating multimedia dataset for six object in three ethnics which is trained in CNN and RNN architecture to develop Multimedia Classifier Model in Unified Indexing Process. The Unified indexing is embedded in MIRS architecture to become the unified concept-based MIRS. For structuring the object is applied ontology refer to CIDOC-CRM. After processing then perform the evaluation of the system to obtain the results of experiment.

The contribution of this paper is threefold. First, we present Unified MIRS that integrating all features of media text, image, video, & audio. The second, multimedia object is identified and classified to recognize text image, video, and audio as a concept representation with CNN and RNN techniques, and the third, Ontology design of Cultural Heritage

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

object that provide object structured refer to CIDOC-CRM that easily to organize and develop.

This paper is organized as follow, section 2 to provide related works that contain the main reference of this paper. Section 3 describes the proposed unified concept-based MIRS using deep learning with ontology. Section 4 is the methodology and implementation. Section 5 is results and discussion and Section 6 is concluded the paper.

2. RELATED WORKS

The development of Unified Concept-based MIRS using deep learning with ontology requires some techniques not only multimedia information retrieval system but also ontology mapping for building relationship between the concepts. Deep learning technique, in this case, using CNN for extracting feature, identification and classification of image, video, audio and RNN for extracting feature, identification and classification of text and retrieval to support unified indexing technique of the system. Cultural heritage domain is used as dataset and data collection in this system. IRS, MIRS, Multi-modal MIRS, Object Classification with Deep Learning, Ontology, and cultural heritage will be discussed here.

2.1 Information Retrieval System (IRS)

Information Retrieval System (IRS) emerged because of the rapid growth of digital data. IRS evolve continuously and keep improving. The results of the new IRS today is the improvement to the old IRS. Modern IRS is started with Classic IRS that consisting of three main models; Set Theory Model, Vector Space Model, and Probabilistic Model. Besides these models, Classic IRS also introduce Term Weighting and Term Frequency-Inverse Document Frequency (TF-IDF) [34] for rank retrieval which is still in use today as retrieval techniques.

Modern IRS is the continous development of classic IRS such as Extended Boolean Model [35] and Fuzzy Set Model [36] in Set Theory Model. The extension of the Vector Space Model is Generalized Vector [37], Latent Semantic Indexing (LSI) [38]., and Neural Network Model [39]. The probability of distribution is the basic principle of the Probabilistic Model that is used by Language Model [40]., Divergence from Randomness, and Bayesian Network Model [41].

Modern IRS generally is formed from three main process; indexing process to acces and relate easily with original document , query processing manage input, rank and retrieval process manage rank retrieval results (see Fig.1).

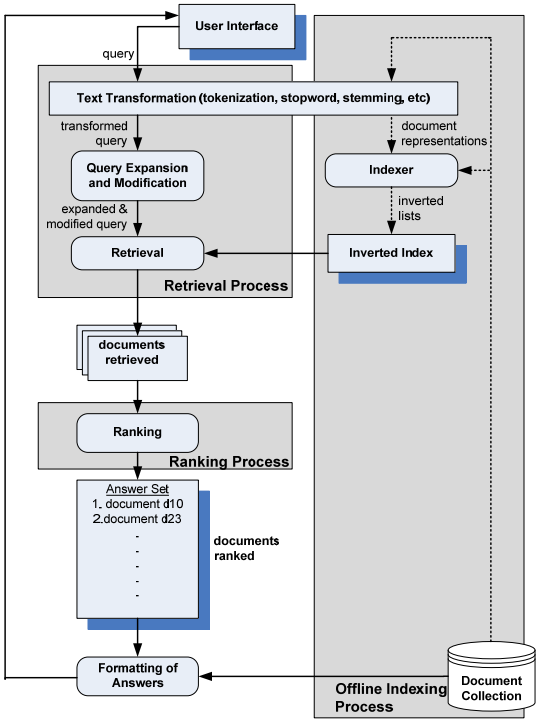

Fig.1.Information Retrieval System;Indexing, Querying, Retrieval and Ranking [42]

2.2 Multimedia Information Retrieval System (MIRS)

The MIRS aim to extract multimedia data (text, image, video, and audio) and to retrieve multimedia data related to the information need of the user and rank them according to a degree of relevance or relation with regard to the user query. The problem in MIRS is "Semantic Gap" when the distance between computable low-level features content and high-level concepts that is understood by the human. Bridging the gap is the main issue of MIR [43]

Content-based MIRS and Concept-based MIRS are parts of MIRS that will be discussed in the related works. Content-based MIR is divided into Content-based Image Retrieval, Content-based Video Retrieval, and Content-based Audio Retrieval. Concept-based MIR is divided into Concept-based Text Retrieval, Concept-based Image Retrieval, Concept-based Video Retrieval, and Concept-based Audio Retrieval.

Content-based MIRS is the task of retrieving multimedia media object based on their

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

content with using the features of the object like; image with color, texture, and shape features; audio with acoustic and semantic features; video with the combination between the feature of an image, motion feature and feature of audio.

Content-based Image Retrieval (CBIR) is one of the active research fields where there are various techniques for extracting colors such as Color Co-occurrence Matrix, Color Correlogram, Color Histogram and Dominant Color Description. The feature of texture such as Gabor Filter, Wavelet Transform, Steerable Pyramid and Tamura Texture. For the feature of the shape such as Fourier Descriptor [4].

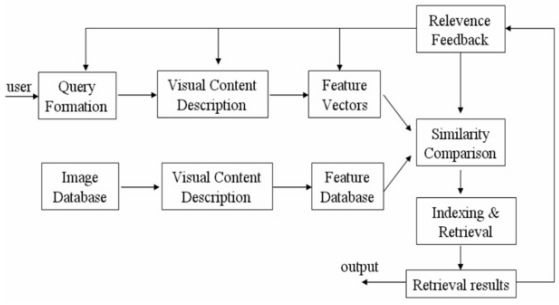

Fig.2. Content-based Image Retrieval [4]

The aim of Content-based Audio Retrieval (CBAR) is to find desired audio document based-on their acoustic features. The acoustic features are contained Loudness, Spectrum Power, Pitch, Spectrum, Bandwidth, and Brightness. The process is almost same with another Content-based retrieval but is more specific with acoustic features. For similarity measure, it is need learning approach to match audio query or text query with the audio collection[44].

Content-based Video Retrieval (CBVR) is a process of retrieving the desired video from a large collection of video with using motion features, audio features, color histogram, motion histogram, text feature, and features are extracted from face and object existing in its frames. The process flow of CBVR is started from video segmentation that differentiating video component like a frame, shot or scene and then there are classified. Video features are extracted for each component and stored in video features database. On the other hand, the query video is segmented before their extract. The result of video extraction is matched in the similarity measure and the result or video output is desired video [5].

Concept-based MIRS has a different approach with Content and Context-based MIR, It exploits knowledge representation with the concept terms. The knowledge representation is in the form of manual thesauri, automatic corpus or ontology mapping. The result of retrieval is dependent on how big the corpus or how good the structured of ontology.

Concept-based Text Retrieval with Explicit Semantic Analysis using Wikipedia as a corpus and evaluating with several TREC dataset. It tries to solve the conventional keyword-based retrieval problems like polysemy and semantic problems. The keyword-based retrieval is not able to recognize two different words in the same concept. The corpus or manually thesauri is the answer with performs extracting latent word relation and concept from the corpus. [45]

Concept-based Image Retrieval is enabled by ontology mapping that offers to bridge the low-level image features to high-level ontology concept. The ontology with a specific domain is structured by the concept will manage human knowledge. It requires huge data image to improve its accuracy and relevance. Learning approach and ontology is applied to create the concept from content or image features like color, shape, texture and put the concept as a label. For the ranking of image search result is calculated as a sum of matched ontology concepts with reference to user query[46].

Concept-based Audio Retrieval is learned from the training data of audio examples and text caption using Gaussian Mixture Model (GMM) distribution for modeling algorithm that will produce the concept. Semantic features map the audio track as the concept and rank list will be handled by with semantic similarity measure based on ranking database track [47].

To bridge semantic gap Concept-based Video Retrieval has been considered as a feasible alternative technique for video search. Video query can be frame, object, motion, color, texture which are expressed in terms of semantic concept. Colombia374 has automatic concept detector maps the video query onto the concept. Colombia374 has 374 concepts with its score are extracted from LSCOM Ontology and annotating the concept of TRECVID2005 [48].

2.3 Multi-modal MIRS

Text-based, Content-based, Context-based, and Concept-based MIRS which is explained above is mono-modal MIRS or only applying one media in the retrieval process. Along with information growth and its various media there are comprehensive information need of user to unify text, image, video, and audio format in one MIRS or multi-modal IRS. Today, Search Engine like Google and Bing allow

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

text and image format for their queries and for the search results present media text, image, and video.

Several research perform multi-modal format media in MIRS with unifying a few media, there are bi-modal such as unifying text-based and content-based image retrieval, unifying text-based and content-based video retrieval, or unifying text-based and content-based audio retrieval; tri-modal to unifying content based text, image, video retrieval and content-based text, audio, video retrieval [49] or quad-modal unifying text, image, audio, and video retrieval at once [15], [16].

Unifying text-based retrieval and content-based image retrieval or bi-modal MIRS implement image mining [7] and find image in medical resource database [8] that applying image annotation and image processing techniques (extracting features of image) to retrieve images from the collection. Another technique using text and image similarity matching [9] for image retrieval. Unifying text-based and content-based video retrieval have been implemented for video search using text and semantic retrieval utilize text extraction from video subtitle [10], for video retrieval using text metadata and speech [11] and for video retrieval with early fusion approach [12]. Unifying text-based and content-based audio retrieval is implemented in digital music search using Vector Space Model and represent the audio as "bag of audio-word" [13] and also is implemented in incorporating audio-based similarity to retrieve and index music pieces [14].

Unifying three format media or tri-modal in a MIRS are implemented for Image and Video search using hybrid approach for text extraction in text block on image or Video [50] and also implemented for lecture video retrieval using audio and text transcript using part-of speech tagging and key framework for audio extraction [49].

Quad-modal is unifying four formats media in a MIRS to fuse text, image, audio, and video in a system that can process four format media. The implementation for Content-based MIR is increasing various media that can be inputted and retrieved in a MIRS to satisfy information need of the user with multidimensional approach [15] or Indexing and Retrieval Techniques for Text, Image, Audio, and Video using text and audio-visual features [16].

2.4 Deep Learning for Classification in MIRS

The classification is processing to perform categorization a set of the object using some basic features to describe it. There are some classifier like Logistic Regression (LS), Support Vector Machine (SVM), Naive Bayes, and Neural Network. For the simple recognizing pattern or simple classification using LS and SVM is good enough, but for the complex classification, it is required deep neural network using more layer or Deep Learning.

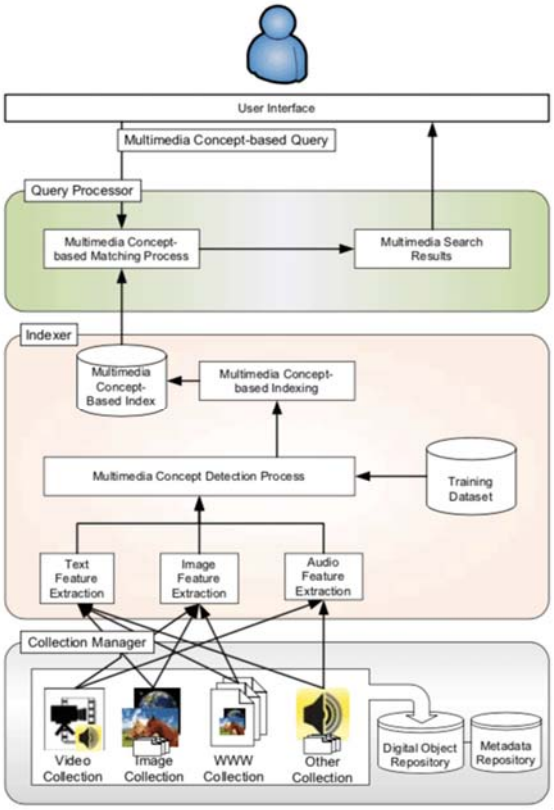

Fig.3. A Multidimensional Approach in Content-based MIRS.[15]

Deep learning allows computational models that are composed of multiple processing layers to learn features representations of data with multiple levels of abstraction. The visual object recognition, object detection, speech recognition and much other domains is dramatically improved by deep learning. The complex structure in large data sets is handled by deep learning by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. For processing data image, video, speech, and audio, the convolutional net is a breakthrough method and with outperforms result for classification and recognition. And then for the sequential data such as text and audio, the recurrent net is the great solution [51]. Convolutional Neural Network Architecture is the breakthrough of image classification. Simple form architecture of CNN can be seen on. Fig.4.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

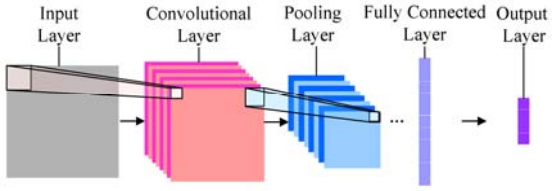

Fig.4. Simple Form of Convolutional Neural Network

CNN has proven capable of image classification with outperform result. The experiment with classifying 1.2 million high-resolution images into 1000 different classes achieved error rate 37.5 % and 17 % which is better than the previous state of the art for image classification. For training, implementation is used 60 million parameters, and 6500 neurons, consist of five convolutional layers, some of which are followed by max-pooling layer and fully-connected layer with a final 1000-way softmax. Regularization method is called 'dropout' to reduce overfitting in the fully-connected layer that proved is very effective [52]. For Video Classification we also use CNN techniques with frame manipulation. The video is treated as a still image on its frames. Frames (1 or 2 frames) form video is classified with image classification or CNN in this case.

CNN for Image classification perform very well and show promise for audio classification. With various CNN architectures to classify the soundtrack of dataset 70 M training Video with 30,871 video-level labels to extract the audio features. This experiment examines fully-connected Deep Neural Network, AlexNet, VGG, Inception, and RestNet to increase audio classifier capability. The result is outperform if compared with *Audio Set* Acoustic Event Detection (AED) classification task [25].

2.5 Ontology

An ontology is an explicit, formal specification of a shared conceptualization. The term is borrowed from philosophy, where an Ontology is a systematic account of Existence. For Artificial Intelligence (AI) systems, what 'exist' is that which can be represented [53]. The meaning of conceptualization is the abstract model such as domain, identified relevant concepts, and relations. The explicit means all concept must be defined, formal means machine understandable and shared means consensus about ontology.

Ontology-based IRS model provides better search capability is compared with a full text-based search. Ontology maps keyword to the concept and create relation to the keyword in the same context that can be easy to understand by a human. Improvement of Vector Space model IRS is significant when using ontology-based knowledge representation [54]. The ontology might be seen as concepts/terms and relation between concept/term, besides that ontology as the representation of knowledge and provide a formal structure for describing the concept and its relationship, ontology also make a ranking as result of semantic similarity matching [55].

Ontology is a knowledge representation in a domain, there are many ontology standards map the real word. In the cultural heritage context is recognized CIDOC-Conceptual Reference Model (CRM) and Europana Model as the ontology standard. The standards become a reference in the cultural heritage field because the data structure gives benefit in development of data structure of information.

2.6 Cultural Heritage Search

The cultural heritage as the things pre- served by the memory institutions, i.e. museums, sites and monuments records ("SMR"), archives and libraries. Cultural heritage can be tangible like movable cultural heritage (paintings, sculptures, coins, manuscripts), immovable cultural heritage (monuments, archaeological sites, and so on), and underwater cultural heritage (shipwrecks, underwater ruins, and cities) or intangible like oral traditions, performing arts, and rituals.

The special characteristic of cultural heritage is various format media (text, image, video, and audio) and rich in semantic terms. The collection items have their history in text format, have a picture, a video, and sometimes also have an audio format. The items of cultural heritage are related in many ways to their environment, the society, and to other items. The resource of cultural heritage is managed by UNESCO.

Based on UNESCO classification there are Cultural Heritage, Natural History and combining both of them, but we use terms "Cultural Heritage" to descript them. The Cultural Heritage is comprised of two categories, tangible such as clothing, books, monument, building and other artifacts and intangible languages, social value, tradition, artistic expression and another aspect human activity. For Natural History refers to the elements of biodiversity (including flora and fauna) and geodiversity (including mineralogical, geomorphological, paleontological, etc.). And the combination cultural heritage and natural history have characteristic both

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

of them.

Initially, CultureSampo [32] presented prototype system for integrating cultural heritage content in Finland. It intends to search and combine easily heterogenous cultural heritage related content. CultureSampo provides a various way for searching cultural heritage, such as historical event-based and geographical map-based including mapping the metadata schema with ontology. Cantabria's Cultural Heritage [33] is another cultural heritage platform on the national level for the region of Cantabria, Spain. It used ontology mapping for integration of cultural heritage data and relating the content. In Europe, there is Europana Data Model (EDM) [56] that is a shared ontology infrastructure to unify data from a different institution. Korean Cultural Heritage Data Model (KCHDM) [57] almost same with another cultural heritage platform is built an ontology-based data model to create a semantic link between separated institution. All these platforms are based on a wide and formally defined ontology, the CIDOC Conceptual Reference Model (CIDOC-CRM) [58]. It provides definitions and formal structure for describing the implicit and explicit concepts and relationship used in cultural heritage.

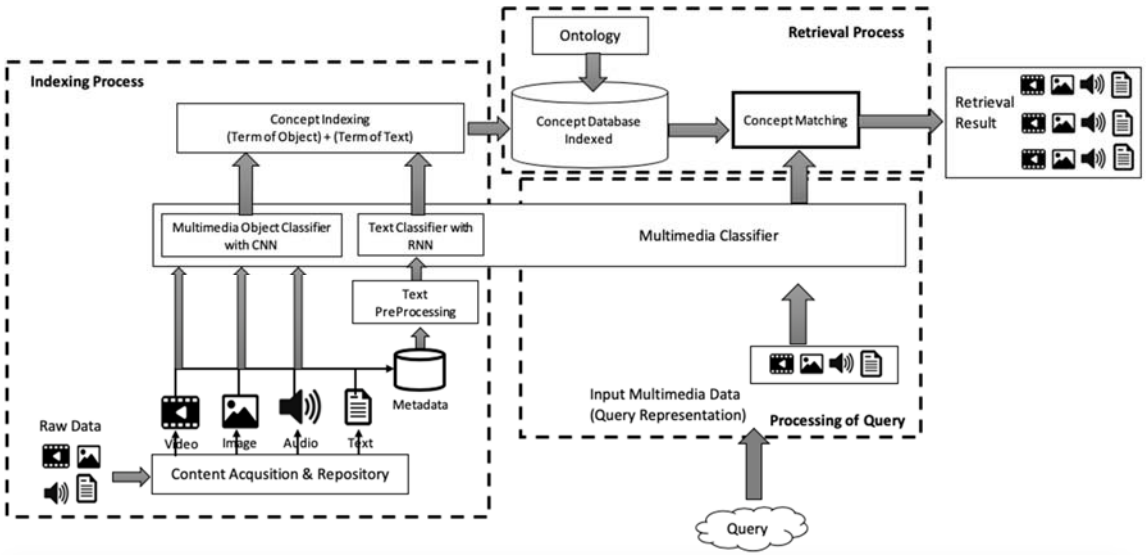

Figure 5. Unified Concept-based Multimedia Information Retrieval using Ontology

3. THE ARCHITECTURE OF UNIFIED CONCEPT-BASED MIRS USING DEEP LEARNING WITH ONTOLOGY

Unified Concept-based MIRS using Ontology is proposed to create multi-modal MIRS to tackle the limitation of media format problems. The system could apply input query with text, image, video or audio format and presenting retrieval result with text, image, audio, and video format in the same page. The retrieval results of MIRS are more complete because applying four media formats (multi-modal) if we compare with mono-modal MIRS. Using concept-based to solve the semantic problems or creating relationship between object and using ontology to create map or formal structure of Indonesia's cultural heritage domain referes to CIDOC Conceptual Reference Model (CIDOC-CRM).

The architecture design of Unified Concept-based MIRS using Ontolgy comprises of three main parts; Unified Indexing Process consists Content Acquisition and Repository, Multimedia Classifier Model, and Concept Indexing; the second, Processing of Query consists Multimedia Query Input and also Multimedia Classifier Model, the third, Retrieval and Rank Process consists Ontology, Concept Database Indexed and Eucledian Distance as matching process plus Rank List Result. (Figure. 5).

Unified Indexing Process performs information processing to speed up that searches. The key of indexing process in this paper is the Multimedia Classifier Model that is developed using Indonesian ethnic multimedia dataset. This model using deep learning techniques CNN and RNN to identify and classify multimedia object become the concept that easily understood by the human. From Content Acquisition and repositories the multimedia

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

objects (text, image, audio, video) pass the Multimedia Classifier Model to index the multimedia object as the concept then is stored in Concept Database Index.

In **Processing of Query**, a multimedia query passes the multimedia classifier model to be recognize as a concept. The concept from the multimedia query is compared with the concept in Concept Database Index in **Retrieval and Rank Process**. The ontology maps the concept and manage the relationship between concept to generate rank list result.

4. THE METHODOLOGY AND IMPLEMENTATION

Several steps to develop the system is initialized with creating a multimedia dataset as resources which is built the model. The model or algorithm that is created from training, validating, and testing the multimedia dataset establish the model and then the model is applied in the MIR system. Developing concept-based MIR need reference or ontology maps to create a relationship between concept and ontology is implemented in the MIR system. Implementing the system using a multimedia query with "tongkonan" concept is necessary to know the performance of the system.

Table 1. Cultural Heritage Multimedia Object Dataset

Activity	Method/Tool	Results
Creating Image Dataset Data training 70%, Data Testing 20% Data Validation 10%	Limit image size to 100x100 px, then reduce again to 70x70 px with Python Programming Language	Indonesia's Cultural Heritage Image Dataset 600 image dataset for six object (class)
Creating Video Dataset Data training 70%, Data Testing 20% Data Validation 10%	Limit image size to 100x100 px, then reduce again to 70x70 px with Python Programming Language (Same with Image Dataset)	Indonesia's Cultural Heritage Image Dataset 600 image dataset for six object (class)
Creating Audio Dataset Data training 70%, Data Testing 20% Data Validation 10%	Converting Audio signal to spectrogram, Cutting the audio in 10 second for dataset	Indonesia's Cultural Heritage Audio Dataset 30 kinds of 5 minutes Audio for three object (class)
Creating Text Dataset Data training 70%, Data Testing 20% Data Validation 10%	Word and Metadata Processing with Python Programming Language	Indonesia's Cultural Heritage Text Dataset, 600 short text and metadata for six object (class)

4.1. Creating Multimedia Dataset

The multimedia dataset consists of four media formats, image datasets, video datasets, audio datasets, and video datasets. Dataset use six objects as follows Tongkonan and Rambusolo are part of Toraja ethnic, Rumah Bolon and Tortor Dance are part of Batak ethnic, and then Pura Uluwatu and Pendet Dance are part of Bali Ethnic. The activity, method, and result for creating dataset are shown below (Table. 1).

4.2. Training CNN and RNN for Building Multimedia Classifier Model

Multimedia Classifier Model is an algorithm formed by multimedia dataset which functions to extract multimedia features and performs classification. It uses deep learning with CNN techniques to extract features and classify image, video, and audio then RNN technique to extract features and classify text. Figure 6. shows CNN Architecture for Image and Video Extracting Features and Classification and Figure 7. shows CNN Architecture for Audio Extracting Features and Classification, then Text Extracting Features and Classification is used RNN Architecture in Figure 8.

Image datasets are inputted using the Mini Batch Gradient Descent64 technique (input 64 images at once). Feature Learning and Classification are activity on the training phase of image and video extracting features and classification. In the first convolutional layer, feature learning applies convolution with filter or kernel to extract image feature, and max-pooling reduce the size of the image. ReLu is an activation function that passes through the layer or goes to the next layer. While image representation vector becomes smaller, it is formed flatten vector. The flatten vector pass the fully connected and softmax to classify values. In this case values, 0,1,2,3,4,5 are the representation of tongkonan, rambusolo, rumah bolon, tari tortor, pura uluwatu, and tari pendet that predetermined. If

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

image is tongkonan.jpg the CNN output is 0 that means the concept of tongkonan. Every single epoch (feed-forward and backpropagation) process is performed image data validation and after CNN convergence is done image data testing before becoming CNN model.

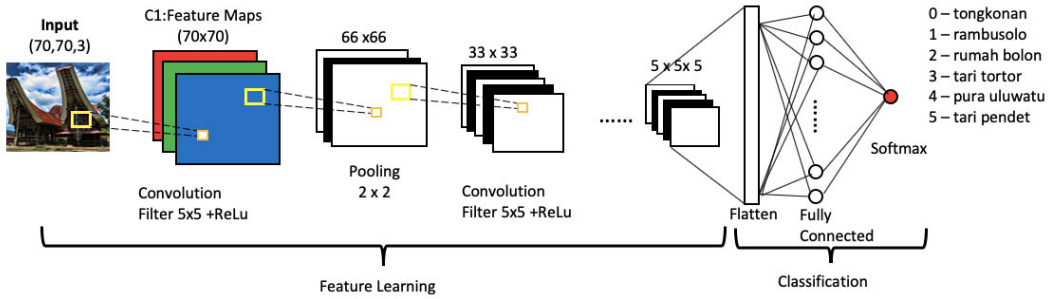

Figure 6. CNN Architecture for Image and Video Extracting Features and Classification

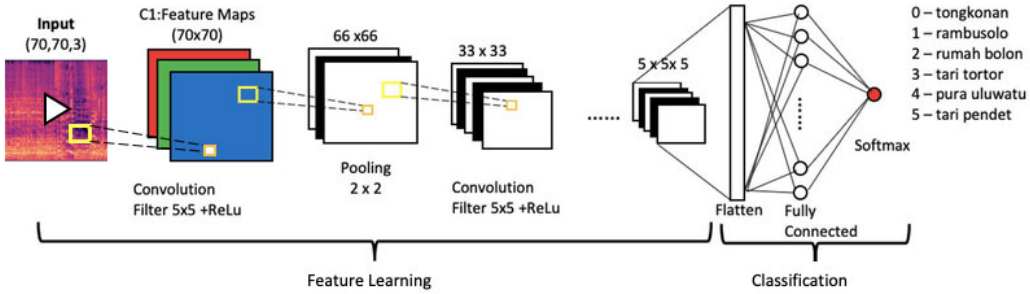

Figure 7. CNN Architecture for Audio Extracting Features and Classification

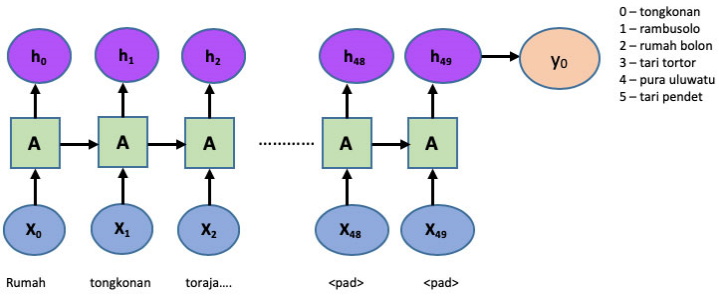

Figure 8. RNN Architecture for Text Features and Classification

Process of Audio Extracting Features and Classification using CNN is almost same with image and video (Figure 7). The difference is only in the audio data that is converted to the spectrogram and cut to the 10-second audio spectrogram. The audio data is processed until convergence then performs validation and testing with audio data spectrogram.

Text Features and Classification using RNN is different from Image, Audio, and Video extraction and classification. Text dataset as a short narration of objects is inputted in RNN Architecture with mini batch gradient 64 technique. After text preprocessing continues to change text dataset with the word embedding become a set of vectors (word2vec) before pass through to the RNN layers. The uutput of the previous layer is used by the next layer until the last layer (50 layers). The output of the last layer refers to the predetermined value

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

0,1,2,3,4,5 that are the representation of tongkonan, rambusolo, rumah bolon, tari tortor, pura uluwatu, tari pendet. Before becoming a model the architecture is validated per epoch and tested with text dataset after training and validating.

After training, validating and testing CNN and RNN for each format media with the multimedia dataset, the model is formed. In the MIR context CNN Model for Image and Video Indexing and RNN Model for Text Indexing. Combination of Image Indexing Model, Video Indexing Model, Audio Indexing Model, and Text Indexing Model is called Unified Indexing Model or Multimedia Classifier Model (Figure 9).

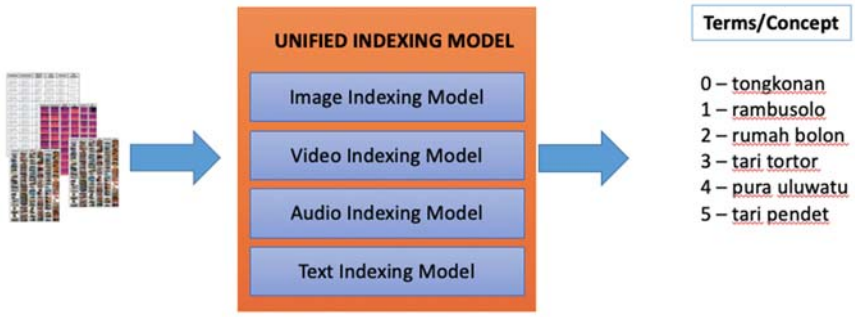

Figure 9. Unified Indexing or Multimedia Classifier Model

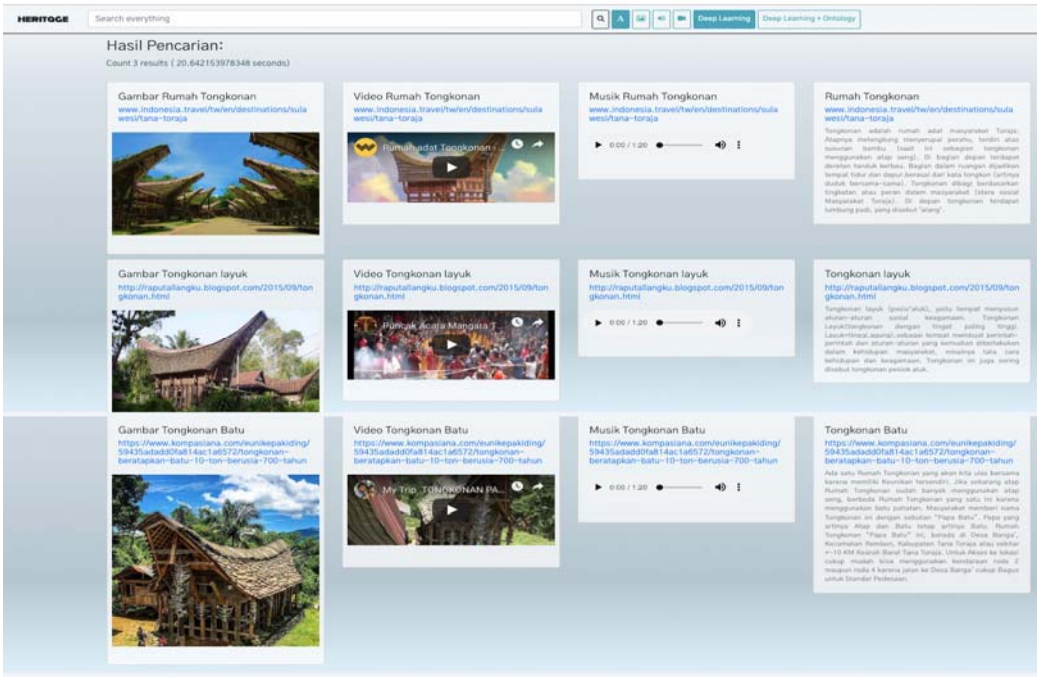

Figure 10. Retrieval Result of Unified MIR in four media format with "tongkonan" query.

4.3. Applying Multimedia Classifier Model in MIRS

The Multimedia Classifier Model is part of the Indexing Process that performs learning feature extraction and classification of multimedia data. Applying or testing the work MIR need multimedia data collection that is different from the dataset. The data acquisition and repository store 41 images (type: jpg), 41 short audios duration (type: mp3), 41 short videos (type: mp4), and 41 short text (type: doc). The MIR only search six concepts (tongkonan, rambusolo, rumah bolon, tari tortor, pura uluwatu,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and tari pendet) from three ethnic (Toraja, Batak, and Bali). The usage of six concepts refers to the object that has been classified in CNN and RNN model.

Input query 'tongkonan' is inputted in the search field for the experiment and multimedia classifier model recognize the query as a concept of 'tongkonan' or the traditional house of Toraja ethnic. The 'tongkonan' concept is compared with the multimedia database indexed to find similarity using Euclidean disctance. The similarity between concept query and concept in multimedia database indexed generate rank list results and in this case the output is information about tongkonan in four format text, image, audio, and video.

The retrieval result on this experiment generate information about "tongkonan" or all information that is shown have Tongkonan terms (Figure 10.). It means the retrieval result only produce information that explicitly same with the query while the user needs information that related to the query (semantic relation) even it is not the same with query explicitly. The answer of the problem is concept-based MIR is the ontology that is discussed in the next section.

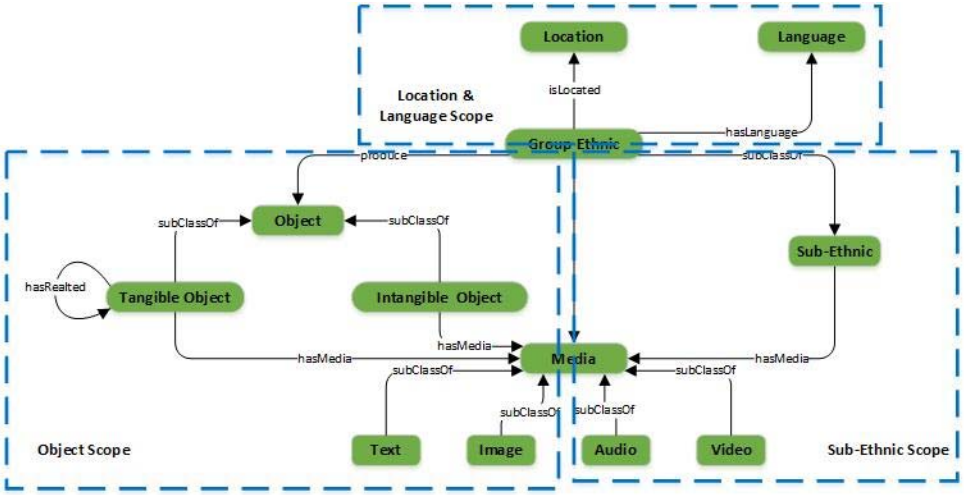

Figure 11. Main Ontology of Indonesia's Cultural Heritage with Ethnic Domain

4.4. Constructing Ontology

Ontology design in this paper manages Indonesia's Cultural Heritage domain with an ethnic group is the center of domain. The ontology refers to Information structure of UNESCO Cultural Heritage Categories and ontology standard of cultural heritage is CIDOC-CRM. The ethnic group as a class has relation with Location and Language class, Object class. The Group ethnic has sub-ethnic class, and the class of Object contain Tangible and Intangible subclass. All class in main ontology related to the four media formats that are text, image, audio, and video. (Figure 11.). Table 2. shows classes, properties or relation and data type of main Ontology. It designs with simple ontology development [64] techniques.

Besides main Ontology, there are three ethnic classes (green) with the instance (yellow) or concept representation. Fig. 12. depicts classes and the instance of Toraja Ethnic The green box is a object class, the yellow box is an instance, the blue box is a blank node, and the white box is a literal value.

Table 2. Classes and Instance of Indonesian Ethnic

Class	Instance
Ethnic	Batak, Bali, Toraja
Sub-Ethnic	Karo, Simalungun, Bali Aga, Bali Majapahit, Rongkong, Kaili
Location	Sumatera Utara, Pulau Bali, Tana Toraja
Language	Batak, Bali, Toraja
Tangible Object	Rumah Bolon, Pura Uluwatu, Tongkonan
Intangible Object	Ngaben, Tari Tortor, Rambusolo
Media	Teks Citra Audio Video

Three ethnics (Toraja, Bali, and Batak) are the representation of ethnics in Indonesia with related class in the cultural heritage fields, such as sub-ethnic, location, language, tangible and intangible objects, and media. Each class has instances related to their ethnic.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

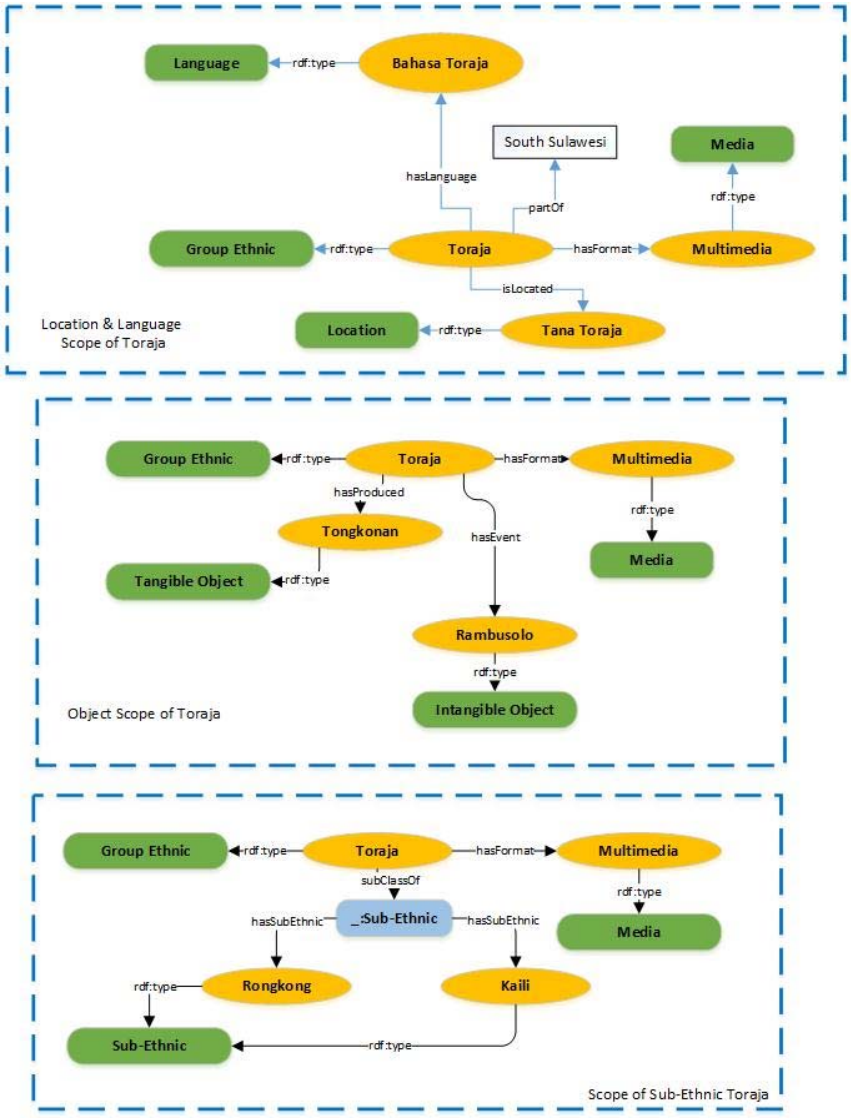

Figure 12. Concept representation or Instance of Toraja Ethnic

4.5. Implementation

MIR has limitless in lack of semantic relation that is improved with ontology. The ontology creates information structure and maps the multimedia object or class based-on their ontology design. MIRS using Ontology no need similarity because the relation between objects (classes) and instances is managed by an ontology. In the experiment shows the retrieval results using ontology more complete and various. (Figure 13.)

5. THE RESULT AND DISCUSSION

Two importants part of the system, first is Unified Multimedia Indexing Process with feature extraction, identification and classification using CNN and RNN techniques and the second is Concept-based MIR with or without Ontology. Both of them will be discussed below.

5.1. Feature Extraction and Classification with Deep Learning

The implementation of Multimedia Classifier Model as an important part on the Unified Multimedia Indexing Process need high accuracy of object identification and classification. It uses some measuring techniques like Accuracy, Precision-Recall, and F1-Score. The measurement shows the performance classification of the model that show in Table 3.

Journal of Theoretical and Applied Information Technology
30th September 2019. Vol.97. No 18

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

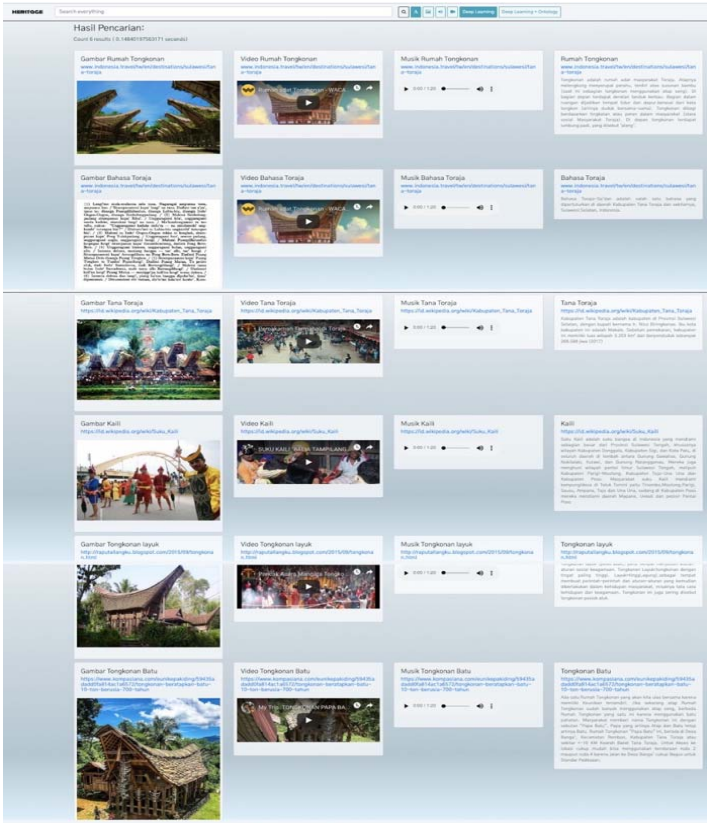

Figure 13. Retrieval Results of Unified Concept-based MIR using Ontology

The implementation also tests multimedia (text, image, audio, and video) classification apply machine (shallow) learning for comparison with deep learning using Support Vector Machine (SVM). The measurement of SVM for multimedia classification in table 4. The comparison between deep learning and machine learning techniques show deep learning with CNN and RNN techniques is outperform about 30%-40% than machine learning with SVM technique.

Table 3. Evaluation of Image, Audio, Video, and Text Classification with CNN and RNN

Tribes Object	Image Classification					Audio Classification				
	A	P	R	F1	Size	A	P	R	F1	Size
Rumah Bolon	0.66	0.65	0.65	0.68	20	0.53	0.54	0.52	0.57	20
Tari Tortor	0.62	0.54	0.68	0.64	20	0.52	0.50	0.54	0.58	20
Uluwatu	0.60	0.59	0.65	0.56	20	0.49	0.46	0.59	0.49	20
Tari Pendet	0.59	0.60	0.58	0.58	20	0.53	0.55	0.58	0.54	20
Tongkonan	0.64	0.66	0.70	0.65	20	0.56	0.55	0.50	0.56	20
Rambusolo	0.65	0.62	0.68	0.67	20	0.54	0.52	0.58	0.52	20

Tribes Object	Video Classification					Text Classification				
	A	P	R	F1	Size	A	P	R	F1	Size
Rumah Bolon	0.58	0.65	0.65	0.58	20	0.76	0.75	0.75	0.75	20
Tari Tortor	0.52	0.54	0.56	0.54	20	0.74	0.75	0.75	0.75	20
Uluwatu	0.58	0.68	0.66	0.67	20	0.77	0.78	0.78	0.78	20
Tari Pendet	0.58	0.52	0.54	0.55	20	0.76	0.76	0.79	0.78	20
Tongkonan	0.62	0.60	0.64	0.62	20	0.75	0.76	0.76	0.76	20
Rambusolo	0.60	0.58	0.62	0.62	20	0.73	0.74	0.75	0.75	20

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 4. Evaluation of Image, Audio, Video, and Text Classification with SVM

Tribes Object	Image Classification					Audio Classification				
	A	P	R	F1	Size	A	P	R	F1	Size
Rumah Bolon	0.20	0.22	0.21	0.22	20	0.18	0.19	0.18	0.19	20
Tari Tortor	0.16	0.19	0.17	0.18	20	0.15	0.17	0.17	0.16	20
Uluwatu	0.23	0.21	0.21	0.22	20	0.20	0.19	0.18	0.19	20
Tari Pendet	0.24	0.20	0.22	0.21	20	0.19	0.20	0.19	0.18	20
Tongkonan	0.15	0.17	0.15	0.15	20	0.16	0.16	0.17	0.16	20
Rambusolo	0.19	0.19	0.19	0.20	20	0.17	0.18	0.18	0.17	20

Tribes Object	Video Classification					Text Classification				
	A	P	R	F1	Size	A	P	R	F1	Size
Rumah Bolon	0.19	0.20	0.19	0.19	20	0.37	0.38	0.40	0.39	20
Tari Tortor	0.20	0.20	0.18	0.19	20	0.40	0.38	0.37	0.39	20
Uluwatu	0.18	0.16	0.16	0.17	20	0.35	0.35	0.35	0.35	20
Tari Pendet	0.21	0.19	0.19	0.20	20	0.36	0.36	0.36	0.36	20
Tongkonan	0.22	020	0.18	0.20	20	0.35	0.37	0.40	0.38	20
Rambusolo	0.21	0.18	0.18	0.18	20	0.32	0.39	0.35	0.36	20

5.2. Concept-based MIR with or Without Ontology

The implementation present Concept-based MIR using ontology is outperform then without ontology. The retrieval result of Concept-based using ontology shows various of information not only in media formats but also the semantic relationship between query concept and multimedia data collection while the MIR without ontology present information containing the query explicitly. The table below (Table 5.) shows the Precision-Recall Measurement of Concept-based MIR with and without Ontology. As we can see the Precision-Recall Concept-based MIR using Ontology outperforms about 30% the Concept based MIR without Ontology.

Some limitation in this research such as the concept represent four media from several existing media and represent as a term not a feature of object. Also, only using specific domain in cultural heritage on the Indonesian ethnics not for all cultural heritage.

Table 5. Precision-Recall Comparison Concept-based MIR with and without Ontology

Terms (Object)	with Ontology		without Ontology	
	Precision	Recall	Precision	Recall
Tongkonan	33%	31%	67%	61%
Rambusolo	33%	31%	67%	61%
Rumah Bolon	33%	31%	67%	61%
Tari Tortor	33%	31%	67%	61%
Pura Uluwatu	33%	27%	67%	53%
Tari Pendet	33%	27%	67%	53%

6. CONCLUSION

The success of deep learning techniques affects the improvement of feature representation including multimedia features extraction, identification and classification in the MIRS field. The paper aims to unify multi-modal MIRS with four format media or multimedia to increase the various type of retrieval results of information. The deep learning techniques such as CNN and RNN help the MIRS to unify multimedia feature extraction and classification in the Indexing Process. Predetermined objects are specified before training, validation, and testing the multimedia dataset. The Accuracy, Precision-Recall, and F1 of CNN and RNN affect the ability of the model to carry out the classification. The performance of CNN and RNN is outperformed than SVM about 30-40%. It means the capability of the system is feasible to identify and classify object.

For the retrieval process is utilized Precision-Recall to measure the retrieval performance. The comparison between Concept-based MIR without or with Ontology is about 30%

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

where the Concept-based MIR with ontology is better than without Ontology.

For future works Unified concept-based MIRS use more than four media, included animation, graph, x-rays, etc. It can develop with another deep learning techniques to increase the accuracy of identifying object.

In the cultural heritage context, the system can be implemented in cultural heritage education, preservation and tourism with more object from many ethnics.

REFERENCES:

[1] C. Cusano, M. Bicocca, and V. Bicocca, "Image annotation using SVM," *Proc. SPIE*, no. 1, pp. 330–338, 2003.

[2] V. Lavrenko, S. L. Feng, and R. Manmatha, "Statistical models for automatic video annotation and retrieval," *2004 IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 3, 2004.

[3] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, "Semantic annotation and retrieval of music and sound effects," *IEEE Trans. Audio, Speech Lang. Process.*, vol. 16, no. 2, pp. 467–476, 2008.

[4] D. Long, Fuhui; Zhang, Hongjiang; Feng, "Fundamental of Content-based Image Retrieval," no. 1, pp. 15–27, 2009.

[5] A. Ansari and M. H. Mohammed, "Content based Video Retrieval Systems - Methods , Techniques , Trends and Challenges," *Int. J. Comput. Appl.*, vol. 112, no. 7, pp. 13–22, 2015.

[6] G. Chechik, E. Ie, M. Rehn, S. Bengio, and D. Lyon, "Large-scale content-based audio retrieval from text queries," *Proceeding of the 1st ACM international conference on Multimedia information retrieval*. pp. 105–112, 2008.

[7] K. Thirunavu, "A Survey on Text and Content Based Image Retrieval System for Image Mining," *Int. J. Eng. Res. Technol.*, vol. 3, Issue 3, no. March 2014, 2015.

[8] W. Li and B. Li, "Combining Text and Content Based Image Retrieval on Medical Resource Database," *3rd Int. Conf. Multimed. Technol.*, pp. 1771–1783, 2013.

[9] B. Dinakaran, A. Jonnalagadda, and A. K. Cherukuri, "Interactive Image Retrieval Using Text and Image Content," *SCYBERNETICS Inf. Technol. Bulg. Acad. Sci.*, vol. 10, no. January, 2010.

[10] V. K. Bante and A. N. Bhute, "A SURVEY ON TEXT BASED VIDEO RETRIEVAL USING SEMANTIC," *Int. J. Adv. Comput. Eng. Netw.*, vol. 3, Issue-4, no. 4, pp. 56–63, 2015.

[11] N. Radha, "Video retrieval using speech and text in video," *Int. Conf. Inven. Comput. Technol.*, vol. 2, pp. 1–6, 2016.

[12] H. Xu, K. He, L. Sigal, S. Sclaroff, and K. Saenko, "Text-to-Clip Video Retrieval with Early Fusion and Re-Captioning," *arXiv Prepr. arXiv1804.05113*, 2018.

[13] M. Riley and E. Heinen, "A TEXT RETRIEVAL APPROACH TO CONTENT-BASED AUDIO RETRIEVAL," *Int. Soc. Music Inf. Retr.*, pp. 295–300, 2008.

[14] P. Knees, T. Pohle, M. Schedl, D. Schnitzer, K. Seyerlehner, and G. Widmer, "AUGMENTING TEXT-BASED MUSIC RETRIEVAL WITH AUDIO SIMILARITY," *Int. Soc. Music Inf. Retr.*, pp. 6–11, 2009.

[15] I. Budi, Z. A. Hasibuan, A. Rubhasy, and G. P. Mindara, "A Multidimensional Approach in Content-based Multimedia Information Retrieval System," *Int. Conf. Comput. Eng. Appl.*, 2011.

[16] M. N. S. Zope and A. Gadekar, "Techniques for Text , Image , Audio and Video Indexing and Retrieval," *Int. J. Emerg. Trends Technol. Comput. Sci.*, vol. 4, no. 5, pp. 123–126, 2015.

[17] A. Basu, C. Watters, and M. Shepherd, "Support Vector Machines for Text Categorization," *cornel Comput. Sci.*, pp. 1–7, 2002.

[18] K. Sugamya, A. Vinayababu, and S. Pabboju, "A CBIR CLASSIFICATION USING SUPPORT VECTOR MACHINES," *Int. Conf. Adv. Hum. Mach. Interact.*, pp. 1–6, 2016.

[19] G. Guo and S. Z. Li, "Content-Based Audio Classification and Retrieval by Support Vector Machines," vol. 14, no. 1, pp. 209–215, 2003.

[20] D. Nagaraja, G S, Murthy, Rajashekara, "Content based Video Retrieval Using Support Vector Machine Classification," vol. 217, pp. 821–827, 2015.

[21] A. Mccallum and K. Nigam, "A Comparison of Event Models for Naive Bayes Text Classification," *AAAI-98 Work. "Learning Text Categ. Springer*, 1998.

[22] A. C. Alhadi, A. Deraman, M. Masila, and A. Jalil, "An Ensemble Similarity Model for Short Text Retrieval An Ensemble Similarity Model for Short," *Citeseer*, no. March 2018, 2017.

[23] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Convolutional Neural Networks for Text Classification," *Twenty-Ninth AAAI Conf. Artif. Intell.*, pp. 2267–2273, 2015.

[24] A. Krizhevsky and G. Hinton, "Using Very Deep Autoencoders for Content-Based Image Retrieval," *Proc. Eur. Symp. Artif. Neural Networks*, pp. 1–7, 2011.

[25] S. Hershey *et al.*, "CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION," *IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 131–135, 2017.

[26] A. Karpathy, T. Leung, and at all, "Large-scale Video Classification with Convolutional Neural Networks," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 1725–1732, 2014.

[27] O. Egozi, S. Markovitch, and E. Gabrilovich, "Concept-based information retrieval using explicit semantic analysis," *ACM Trans. Inf. …*, vol. 0, no. 0, pp. 1–38, 2011.

[28] D. Vallet, M. Fernández, and P. Castells, "An Ontology-Based Information Retrieval Model," *Semant. Web Res. Appl.*, pp. 455–470, 2005.

[29] Icom-Cidoc, "CIDOC-CRM: the CIDOC Conceptual Reference Model ," vol. 4, 2006.

[30] A. I. (Ed), "Europana Data Model Primer, Europana Technical Document," 2013.

[31] C. Doulaverakis, Y. Kompatsiaris, and M. G. Strintzis, "Ontology-based access to multimedia cultural heritage collections-The REACH project," *Comput. as a Tool, 2005. EUROCON 2005. Int. Conf.*, vol. 1, pp. 151–154, 2005.

[32] E. Hyv, J. Takala, K. Puputti, and H. Kuittinen, "CultureSampo : A National Publication System of Cultural Heritage on the Semantic Web 2 . 0 1 A Basis for Cultural Heritage on the Semantic Web," *Springer-Verlag Berlin Heidelb.*, pp. 851–856, 2009.

[33] F. Hernandez, L. Rodrigo, J. Contreras, and F. Carbone, "BUILDING A CULTURAL HERITAGE ONTOLOGY FOR CANTABRIA," *Annu. Conf. CIDOC*, pp. 1–14, 2008.

[34] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5. pp. 513–523, 1988.

[35] G. Salton, E. A. Fox, and H. Wu, "Extended Boolean information retrieval," *Commun. ACM*, vol. 26, no. 11, pp. 1022–1036, 1983.

[36] Y. Ogawa, T. Morita, and K. Kobayashi, "A fuzzy document retrieval system using the keyword connection matrix and a learning method," *Fuzzy Sets Syst.*, vol. 39, no. 2, pp. 163–179, Jan. 1991.

[37] S. K. M. Wong, W. Ziarko, and P. C. N. Wong, "Generalized vector spaces model in information retrieval," *Proc. 8th Annu. Int. ACM SIGIR Conf. Res. Dev. Inf. Retr. - SIGIR '85*, pp. 18–25, 1985.

[38] T. K. Landauer *et al.*, "An Introduction to Latent Semantic Analysis," *Discourse Process.*, vol. 25, no. 2–3, pp. 259–284, 1998.

[39] R. Wilkinson and P. Hingston, "Using the cosine measure in a neural network for document retrieval," *Dl.Acm.Org*, pp. 202–210, 1991.

[40] C. Zhai, "Statistical Language Models for Information Retrieval A Critical Review," *Found. Trends® Inf. Retr.*, vol. 2, no. 3, pp. 137–213, 2007.

[41] H. A. W. B. C. Turtle, W. B. Croft, and H. A. W. B. C. Turtle, "Evaluation of an Inference Retrieval Model," *ACM Trans. Inf. Syst.*, vol. 9, no. 3, pp. 187–222, 1991.

[42] R. Baeza-Yates and B. Ribeiro-Neto, "Modern information retrieval," vol. 463, p. 513, 1999.

[43] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain, "Content-based multimedia information retrieval: State of the art and challenges," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 2, no. 1, pp. 1–19, 2006.

[44] G. Chechik, E. Ie, M. Rehn, S. Bengio, and D. Lyon, "Large-scale content-based audio retrieval from text queries," *Proceeding 1st ACM Int. Conf. Multimed. Inf. Retr.*, pp. 105–112, 2008.

[45] O. Egozi, S. Markovitch, and E. Gabrilovich, "Concept-Based Information Retrieval Using Explicit Semantic Analysis," *ACM Trans. Inf. Syst.*, vol. 29, no. 2, pp. 1–34, 2011.

[46] U. Manzoor and M. A. Balubaid, "Semantic Image Retrieval : An Ontology Based Approach," vol. 4, no. 4, pp. 1–8, 2015.

[47] L. Barrington, A. Chan, D. Turnbull, and G. Lanckriet, "Audio Information Retrieval Using Semantic Similarity," *Tracks A J. Artist. Writings*, vol. 2, pp. 2–5, 2007.

[48] S. Memar, L. S. Affendey, N. Mustapha, and M. Ektefa, "Concept-based video retrieval model based on the combination of semantic similarity measures," *2013 13th Int. Conf. Intellient Syst. Des. Appl.*, pp. 64–68, 2013.

[49] S. V. S. Associate, "Lecture Video Retrieval Using Audio and Text Transcripts," *Int. J. Adv.*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Res. Basic Eng. Sci. Technol., vol. 3, no. 39, pp. 48–52, 2017.

[50] C. Misra and S. Sural, "Content Based Image and Video Retrieval Using Embedded Text," *7th Asian Conf. Comput. Vis.*, no. January 2006, 2006.

[51] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[52] A. Krizhevsky, G. E. Hinton, and I. Sutskever, "ImageNet Classification with Deep Convolutional Neural Networks," *Commun. ACM*, vol. 60, no. 6, pp. 1–9, 2017.

[53] T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *Int. Work. Form. Ontol.*, 1993.

[54] D. Vallet, M. Fernández, and P. Castells, "An Ontology-Based Information Retrieval Model," *Semant. Web Res. Appl. Lect. Notes Comput. Sci.*, pp. 455–470, 2005.

[55] S. A. Elavarasi and K. Menaga, "Ontology Based Semantic Similarity Measure Using Concept Weighting," pp. 15–20, 2014.

[56] M. Doerr, S. Gradmann, S. Hennicke, and A. Isaac, "The Europeana Data Model (EDM)," *76th Ifla Gen. Conf. Assem.*, pp. 1–12, 2010.

[57] H. Kim, J. Kim, and S. Kim, "Towards a semantic data infrastructure for heterogeneous Cultural Heritage data - challenges of Korean Cultural ... Towards a Semantic Data Infrastructure for Heterogeneous Cultural Heritage Data - Challenges of Korean Cultural Heritage Data Model (KCHDM," no. February 2017, 2015.

[58] C. Main *et al.*, "Definition of the CIDOC Conceptual Reference Model Documentation Standards Group ," *ICOM/CIDOC CRM Spec. Interes. Gr.*, 2003.