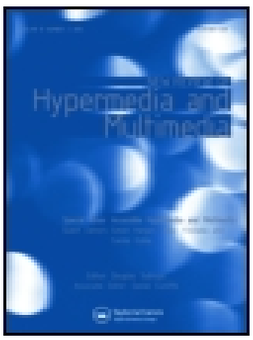


New Review of Hypermedia and Multimedia

ISSN: 1361-4568 (Print) 1740-7842 (Online) Journal homepage: https://www.tandfonline.com/loi/tham20

Multimedia context interpretation: a semantics-based cooperative indexing approach

Mohammed Maree

To cite this article: Mohammed Maree (2020): Multimedia context interpretation: a semantics-based cooperative indexing approach, New Review of Hypermedia and Multimedia, DOI: 10.1080/13614568.2020.1745904

To link to this article: https://doi.org/10.1080/13614568.2020.1745904

NEW REVIEW OF HYPERMEDIA AND MULTIMEDIA
https://doi.org/10.1080/13614568.2020.1745904

Taylor & Francis
Taylor & Francis Group

Multimedia context interpretation: a semantics-based cooperative indexing approach

Mohammed Maree

Department of Information Technology, Faculty of Engineering and Information Technology, Arab American University, Jenin, Palestine

ABSTRACT

The relative ineffectiveness of semantics-based multimedia indexing systems on the Web is caused by the semantic knowledge incompleteness and semantic heterogeneity problems. Nevertheless, the need to search multimedia documents with precision on the Web is persistently growing; pressing the demand for effective and efficient indexing strategies. In this article, we present an ontology-based multimedia indexing approach that cooperatively identifies the semantic and taxonomic relations that exist between annotation words that surround multimedia documents on Webpages. In this context, multiple ontologies are jointly employed for indexing each document. We construct inverted indexes in the form of semantic networks where nodes of each network are identified and added based on a majority-voting technique, while edges represent the semantic and taxonomic relations that hold between those nodes. To alleviate the heterogeneity between the resulting networks, we employ ontology merging algorithms to integrate them into consistent networks. We also utilise concept relatedness measures to enrich the networks with semantically-relevant entities that are not recognised by the used ontologies. To validate our proposal, we have developed a prototype system based on the proposed techniques. The produced results using real-world datasets demonstrate an improvement of the effectiveness against state-of-the-art baseline metrics.

ARTICLE HISTORY
Received 29 November 2019
Accepted 18 March 2020

KEYWORDS
Context interpretation; cooperative ontologies; semantic heterogeneity; semantic-relatedness measures; semantic networks

Introduction

The size of multimedia documents is constantly increasing on the Web (Shrivastav et al., 2017). Precise search within and across such a huge space of documents is a challenging task for current multimedia indexing approaches (Kroupi et al., 2016). For instance, in a general search context for different types of multimedia documents, searching for "*impala speed in jungle*" in a video search engine (e.g. Google, Bing, etc) results in a mix of videos about felines and automobiles.

CONTACT Mohammed Maree mohammed.maree@aaup.edu Department of Information Technology, Faculty of Engineering and Information Technology, Arab American University, P.O Box 240, 13 Zababdeh, Jenin, Palestine

Similarly, searching for the acronym "*C.A.T.*" in an image search engine, such as Google and Yahoo results in a mix of images about felines and computer games. Failing to precisely meet user information needs in such general search contexts over multimedia documents on the Web is because the underlying techniques employed by existing approaches suffer from a number of limitations that can be summarised and categorised as follows:

(1) For content-based multimedia indexing and retrieval approaches, the goal is to index multimedia documents based on their content (also referred to as descriptors or low-level features) such as histograms (Bhunia et al., 2019), patterns and textures (Pal et al., 2018; Wei et al., 2018), basic geometric shapes (Song et al., 2019) and metadata and audiovisual segments (Ishtiaq et al., 2018; Manocha et al., 2018). Once documents are indexed, a user can search for the desired multimedia documents using the *query-by-example* approach, wherein he/she needs to upload an example document, and the system matches the content of the uploaded examples to their closest set of documents. Although these indexing approaches proved to be effective in finding matches between user queries and their corresponding multimedia documents, they ignore the surrounding contextual information (in the form of annotation keywords, metadata or textual descriptions) of the documents in the matching process. Nevertheless, the incorporation of such information can play a twofold role. First, users are often reluctant to providing example queries and prefer to express their information needs either using keywords or natural-language-like queries. Second, the process of indexing textual content on the Web is more efficient, considering the time complexity required for indexing multimedia documents based on their low-level descriptors.
(2) For text-based multimedia indexing approaches, the constructed inverted indexes (a.k.a. inverted files) are based on the textual information (such as annotation keywords, captions, and textual content of Webpages) that surround the documents. This information is usually processed using various natural language processing (NLP) techniques to filter out irrelevant keywords and retain those that appear to be significant and contribute to the meaning of the text that describes each document. More details on this approach is provided in section 4. As stated in (Shrivastav et al., 2017), this approach has proved to outperform content-based indexing approaches as it facilitates users' access to their desired information in a relatively more efficient manner. However, the quality of this approach is still penalised by the semantic-gap problem; i.e. the mismatch between the terms used by users to express their information needs and those used by content creators who upload new multimedia documents on the Web.
(3) To tackle problems associated with the abovementioned approaches, researchers propose exploiting semantic resources and knowledge bases

during the indexing as well as retrieval tasks (Abdulmunem & Hato, 2018; Amato et al., 2016; Bendib & Laouar, 2018; Guo et al., 2018; Strezoski & Worring, 2018; Tani et al., 2017; Tulasi et al., 2016; Weese et al., 2018). These systems employ domain-specific ontologies such as Medical Subject Headings[1] (MeSH) which is used for indexing medical images and Iconclass[2] which is used by museums and art institutions for the description and retrieval of subjects represented in images (works of art, book illustrations, reproductions, photographs, etc.). For more examples on domain-specific ontologies that have been exploited for multimedia indexing, we refer the reader to the BioPortal[3] and DERI[4] ontology online repositories. On the other hand, other systems exploit generic ontologies that offer broader domain coverage such as WordNet (Podlesnaya & Podlesnyy, 2016; Rinaldi & Russo, 2018), OpenCyc (He et al., 2016; Hong et al., 2015) and YAGO (Iftene & Alexandra-Mihaela, 2016; Orlandi et al., 2018). Although the domain coverage of these ontologies is more comprehensive, the quality of generic-ontology based approaches is still penalised by the semantic knowledge incompleteness problem (Maree & Belkhatir, 2015). In addition, it is important to point out that — in many scenarios — each ontology returns different types of semantic relations that may hold between the same entities. This heterogeneity is caused by several factors such as the domain coverage and size of the ontology, ontology construction methodology (i.e. manual vs automatic), authors of the ontology and their perceptions on the respective domain and the syntax/ontology language used to define the ontological structure and its components. Because of these differences, it is impractical to utilise a single generic ontology for indexing multimedia documents based on their surrounding contextual information.

Starting from this position, we propose a new multimedia indexing approach incorporating several generic and domain-specific ontologies. In the context of our work, the term *ontology* is defined as a semantic resource that formally and explicitly defines entities (concepts, instances and relations that hold between them) in the domain of interest at both lexical and knowledge levels. Our aim here is to exploit semantics at search time, combining search with reasoning-based techniques to increase the precision of the retrieved results. Consequently, knowledge encoded in the used ontologies is processed, filtered and aggregated to provide cooperative decisions on the semantic relations that will be used in constructing the inverted indexes. The main goal in this context is to overcome the semantic knowledge incompleteness problem by

[1] http://purl.bioontology.org/ontology/MeSH/D008511/
[2] http://www.iconclass.nl/home/
[3] http://bioportal.bioontology.org/
[4] http://vocab.deri.ie/

having a wider domain coverage through using an integrated source of semantic and taxonomic information. As such, and unlike conventional approaches that use a single ontology for indexing purposes, a group of ontologies is exploited to identify and extract concepts and relations from the surrounding contextual information of multimedia documents on the Web. These concepts and their corresponding relations form the basis for constructing semantic networks that are further processed and merged to represent the semantic indexes for multimedia documents. For example, when we have the term "swine flu" among the list of indexing terms, we can enrich this term with semantically-relevant terms that can be obtained using WodNet and YAGO ontologies as depicted in Figure 1 below.

It is important to point out that despite the fact that both ontologies have recognised the term "swine flu", we can see that more synonyms can be further obtained (such as: agn43, ECK1993, JW1982, yzzX, agn, yeeQ) when exploiting domain-specific ontologies, such as the Ontology of Genes and Genomes which can be used for indexing images in the medical domain. Similarly, other domain-specific ontologies can be used for indexing multimedia documents that belong to the same domain of interest. For instance, for indexing medical images and videos, we can exploit domain-specific ontologies such as the MeSH ontology which targets the medical domain (Lipscomb, 2000). Even though we can broaden the domain coverage when using multiple ontologies, we acknowledge the fact of still having missing semantic information. This is because of two main reasons. First, when tackling problems of a considerable size, such as indexing multimedia documents on the Web, we need to have quality ontologies (i.e. ontologies that are precise and accurate in terms of the knowledge that is encoded in their structures) that cope with such size, and this is practically not possible. Second, new entities (concepts, instances, semantic relations) are constantly added to the domains of interest, requiring

(a) (b)

Figure 1. Semantic networks for the term "swine flu". (a) Semantic Network for the term "swine flu" using WordNet ontology. (b) Semantic Network for the term "swine flu" using YAGO ontology.

ontologies to be updated frequently; which is also not practical manually. To address this issue, we utilise statistical-based concept relatedness techniques to enrich the semantic indexes with semantically-relevant entities. We provide a detailed example in Section 4, where we present two document examples (images in our context) associated with their surrounding contextual information, and demonstrate how multiple ontologies can provide richer contribution for indexing multimedia documents. We also demonstrate the problem of semantic knowledge incompleteness and why, and how we employ statistical information about missing entities in the exploited ontologies.

The main contributions of our work are summarised as follows:

- Constructing semantics-based inverted indexes for indexing multimedia documents based on their surrounding contextual information. In the proposed approach, a group of domain-specific and generic ontologies are aggregated to discover the latent semantic dimensions that are hidden in the content of the textual information that surrounds multimedia documents.
- Merging heterogeneous semantic networks and further enriching them with semantically-relevant entities based on statistical-based concept-relatedness measures.

The rest of this paper is organised as follows. In section 2, we review the related work and identify the strengths and weaknesses of a set of approaches that are related to our work. Theoretical foundations of the proposed techniques are discussed in section 3. Section 4 provides an overall description of the architecture of our proposed system. Section 5 presents the experimental setup and the results produced by the proposed system. Section 6 discusses the conclusions and outlines the further extensions of our work.

Related work

In recent years, many approaches have been proposed to address the issue of designing precision-oriented multimedia indexing and retrieval approaches on the Web (Asim et al., 2019; Bhunia et al., 2019; Bracamonte et al., 2018; Budikova et al., 2018; De Oliveira Barra et al., 2016; Guo et al., 2018; Hu et al., 2018; Irtaza et al., 2015; Iyer et al., 2019; Ke et al., 2017; Kroupi et al., 2016; Manzoor et al., 2012; Mukhopadhyay & Sinha, 2019; Nazir et al., 2018; Orlandi et al., 2018; Sarwar et al., 2019; Shrivastava & Tyagi, 2015; Wang et al., 2010). These approaches can be classified under two broad categories. These are: content-based and semantics-based approaches. The goal of the approaches that fall under the first category is to extract descriptors that can be utilised in constructing indexes for multimedia documents. For example, the authors of (De Oliveira Barra et al., 2016) proposed employing Content-based Video Retrieval (CBVR) techniques to index and retrieve video shots

based on example queries given by the user. In (Ke et al., 2017), the authors combined multiple low-level descriptors extracted from example images that are submitted by users. Indexing and retrieval of relevant images was based on the extracted descriptors. The works proposed in (Kroupi et al., 2016) and (Bhunia et al., 2019), used a combination of low-level features extracted from the content of multimedia documents and keywords extracted from the surrounding textual content. Although content-based approaches have proved to be effective, they integrate descriptors within a loosely-coupled architecture, i.e. there is no explicit mapping mechanism between the extracted features and the surrounding textual information of the multimedia documents. Other systems such as (Irtaza et al., 2015) employed genetic algorithms with support vector machines and incorporated user feedbacks for image retrieval purposes to improve the retrieval accuracy. However, user feedback is highly subjective and engages the user in a tedious task involving several interaction loops.

More recent approaches propose to exploit ontologies to improve the retrieval performance by taking into account the semantic dimension of the documents' contextual information. In (Guo et al., 2018) and (Budikova et al., 2018) the authors proposed exploiting ontologies to reformulate user queries and rank the retrieved results based on the knowledge captured by the utilised ontologies. Another example is the system proposed in (Manzoor et al., 2012), which used a manually constructed ontology as the core component for semantic image annotation and retrieval. The used ontological model provided the terminology and concepts for characterising the image metadata. The system exploited the semantic relations defined between the concepts of the ontology to recommend other results that may be of interest to the user. Wang et al. (Wang et al., 2010) exploited a multi-modality ontology for image retrieval on the Web. The process of building the used ontology consisted of two main phases. In the first phase, a high-level textual ontology was automatically constructed based on the Wikipedia encyclopedia. In the second phase, visual words or concepts were built based on the low-level features of the images. These approaches solely depended on the domain coverage of the used ontology. This means that the effectiveness of such approaches is governed by the number of recognised entities (concept, and relations) by the used ontologies. To overcome this problem, we propose utilising multiple generic and domain-specific ontologies that cooperatively provide decisions on the initial semantic indexes that can be used to index multimedia documents. Initially produced indexes are then merged to construct coherent and non-heterogeneous indexes that can be further expanded and enriched with terms that are not encoded or captured by any of the exploited ontologies. In this context, and unlike conventional approaches, by coupling multiple ontologies as well as term-relatedness measures, we attempt to construct semantics-based indexes that provide a richer source of semantic relations about the surrounding contextual information of multimedia documents.

Theoretical foundations

In this section, we first present the theoretical foundations behind the proposed approach. We formally define the terms *Domain-specific Ontology*, *Semantic Network*, *Merging*, *Semantic Index Enrichment*, and *Normalized Retrieval Distance*. Next, we introduce the overall architecture of our proposed approach and highlight the main components of the developed prototype.

Formal definitions

Definition 1: A Domain-specific Ontology Ω is defined as a 4-tuple $\langle C, R, I, A\rangle$ where the components of Ω are:

- $C = \{(c_i),\ i \in [1, \text{Card}(C)]\}$ that represents the set of domain concepts of Ω. The concept hierarchy of Ω is a pair $(C,\ \leq)$, where $\leq$ is an order relation on $C \times C$. We call $c \in C$ the set of concepts, and $\leq$ the subsumption relation. This relation is a.k.a. $is - a$ relation, and when it holds between the concepts c_i and c_j, then c_j is considered as the parent of c_i. For instance, if $c_i =$ object-oriented programming language and $c_j =$ programming language, then we say that programming language is a parent concept for the concept object-oriented programming language.
- $R = \{(r_i),\ i \in [1, \text{Card}(R)]\}$ is used to represent the set of semantic relations that hold between $C \in \Omega$. In our work, the set $\{r_1, r_2, \ldots, r_n\} \in R$ is obtained using multiple domain-specific ontologies $\{\Omega_1, \Omega_2, \ldots, \Omega_n\}$ and generic ontologies $\{O_1, O_2, \ldots, O_n\}$.
- I is the set of instances or individuals that belong to each concept $c_i \in C$ in Ω.
- A is the set of axioms verifying:

$$A = \{(r_i, c_j, c_k)\} \text{s.t.}\ i \in [1, \text{Card}(R)],\ j, k \in [1, \text{Card}(C)],\ c_j, c_k \in C \text{ and } r_i \in R.$$

As we explained in section 1, the semantic knowledge incompleteness in current ontologies forms a major limiting factor toward their full adoption in practical application domains. Because of this issue, we also incorporate generic ontologies to cooperatively assist in the process of constructing semantic indexes from the surrounding contextual information of multimedia documents on the Web. A generic ontology O in this context captures knowledge that covers multiple heterogeneous domains, where the set of concepts $C = \{(c_i),\ i \in [1, \text{Card}(C)]\} \in O$ are defined to encode generic-domain knowledge. It is important to point that using a single generic ontology is not sufficient in practical settings. This is because there is no single generic ontology that provides coverage of all concepts in all domains. This illustrates the semantic knowledge incompleteness issue in current generic ontologies as well. This is also the reason behind our exploitation of multiple generic ontologies since we are tackling a large-scale problem where

the covered concepts need to be as broad as possible to meet the requirement of precise indexing of multimedia Web documents.

Using both types of ontologies, we construct semantic networks that represent the semantic indexes for the documents in the dataset. In this context, a semantic network is formally defined as:

Definition 2: A Semantic Network ζ can be defined as a triplet $\langle C_\zeta, R_\zeta, I_\zeta \rangle$ where:

- C_ζ represents concepts defined in the semantic network. These concepts represent the nodes of the constructed semantic network that are identified and added based on their definitions in Ω and O.
- R_ζ are the relations between C_ζ. The set R_ζ is obtained based on the exploited ontologies Ω and O.
- I_ζ represents the set of instances of the concepts in C_ζ. The instances are also attached to their parent concepts based on their definition in Ω and O.

It is essential to point out that due to the semantic heterogeneity problem in the exploited ontologies, we may face the same problem in the constructed networks. This is because we may find two or more ontologies that define the same entity differently. To address this issue, we utilise ontology merging techniques to merge heterogeneous networks into a single and coherent network. In this context, a merging technique is defined as:

Definition 3: Given two heterogeneous semantic networks ζ_1 and ζ_2, the Merging operation finds semantic correspondences between their concepts and produces a single merged network ζ_{merged} as output. In this context, semantic correspondences between ζ_1 and ζ_2 are 4-tuples $\langle c_{id}, c_i, c_j, r \rangle$ such that:

- c_{id} is a unique identifier of the identified correspondence.
- $c_i \in \zeta_1, c_j \in \zeta_2$ are corresponding concepts of the input networks ζ_1 and ζ_2.
- $r \in R$ is a semantic relation holding between both entities c_i and c_j.

We would like to point that, we re-use the merging steps proposed by Maree and Belkhatir in (Maree & Belkhatir, 2015) to accomplish the merging task. It is important to mention that other algorithms can be exploited for the purpose of merging heterogeneous networks that may be produced by different ontologies. For example, Shao et al. proposed RiMOM, which is a system for instance matching between heterogeneous ontologies (Shao et al., 2016). The proposed system utilised a weighted exponential function based similarity aggregation method to tackle the problem of unbalanced aligned predicate numbers among different instance pairs. However, parameters of the proposed method needed to be manually configured to tone the output to meet the desired

matching task. In addition, semantic relations were not incorporated as part of the alignment method in the proposed approach. ALIN, is another system that aimed at finding mappings between concepts, attributes or relationships of the ontologies (Da Silva et al., 2018). In the proposed system, attribute mappings were interactively submitted to domain experts for feedback on their quality. However, one of the main limitations of this approach is the need for human intervention (made by domain experts) to evaluate the quality of the produced mappings and give feedback accordingly.

Despite the use of various semantic resources, we practically demonstrate that they still suffer from semantic information deficiency. To handle this issue, we use statistical-based semantic-closeness measures to compute the strength of semantic relatedness between the missing concepts and the concepts of the merged semantic networks. Formally, we can characterise this technique as follows:

Definition 4: Semantic Index Enrichment: The enrichment algorithm takes the concepts that are not recognised by the used semantic resources $S_{\text{missing}} = \{c_1, c_2, c_3, \ldots, c_n\}$ and the network ζ_{merged} as input, and produces for each $c \in C$ in ζ_{merged} a set of $S(c) \subseteq S_{\text{missing}}$ where,

- $S(c)$ represents the proposed expansion candidates for c. A candidate $c \in S_{\text{missing}}$ can be a single-word or compound-word from S_{missing}. The proposed set $S(c)$ can be obtained using the normalised retrieval distance (NRD) algorithm. In this algorithm, we use a threshold value v based on Equation 1.

$$S(c, v) = \{c \in S_{\text{missing}} | \text{NRD}(c, c_\zeta) \geq v\} \tag{1}$$

Definition 5: Normalized Retrieval Distance (NRD): We developed the NRD algorithm to be a general form of the Normalized Google Distance (NGD). The NGD determines the semantic closeness between two terms as follows:
Given two terms C_{miss} and C_{in} the NRD between both terms can be measured using Equation 2 as follows:

$$\text{NRD}\,(C_{\text{miss}}, C_{\text{in}}) = \frac{\text{Max}\{\log f(C_{\text{miss}}),\ \log f(C_{\text{in}})\} - \log f(C_{\text{miss}}, C_{\text{in}})}{\log M - \text{Min}\{\log f(C_{\text{miss}}),\ \log f(C_{\text{in}})\}} \tag{2}$$

where:

- C_{miss} is an entity that is not defined in the ontology.
- C_{in} is an entity that exists in the ontology.
- $f(C_{\text{miss}})$ is the number of hits for the search entity C_{miss}.

- $f(C_{in})$ is the number of hits for the search entity C_{in}.
- $f(C_{miss}, C_{in})$ is the number of hits for the search entities C_{miss} and C_{in}.
- M is the number of indexed Webpages.

This calculation is necessary to examine how closely related two semantic concepts are by analysing pairwise co-occurrence frequencies. A distance of zero indicates that the two concepts always appear together. Formally, this is a measure for the symmetric conditional probability of co-occurrence of the semantic concepts C_{miss} and C_{in}. Given a document containing one of the concepts C_{miss} or C_{in}, NRD(C_{miss}, C_{in}) measures the probability of that document also containing the other concept.

System overview

In this section, we present a high-level overview of the architecture of the proposed system. As depicted in Figure 2, we propose a conceptual indexing framework to index multimedia documents (image, video and audio documents on generic Webpages) on the Web by employing knowledge represented in a combination of multiple domain-specific and generic ontologies.

We would like to point out that the focus of our approach is on the contextual information (also referred to as annotation or auxiliary texts) that surround multimedia documents (images with known image file extensions, videos with known video file extensions, audio segments with known audio file extensions) on generic

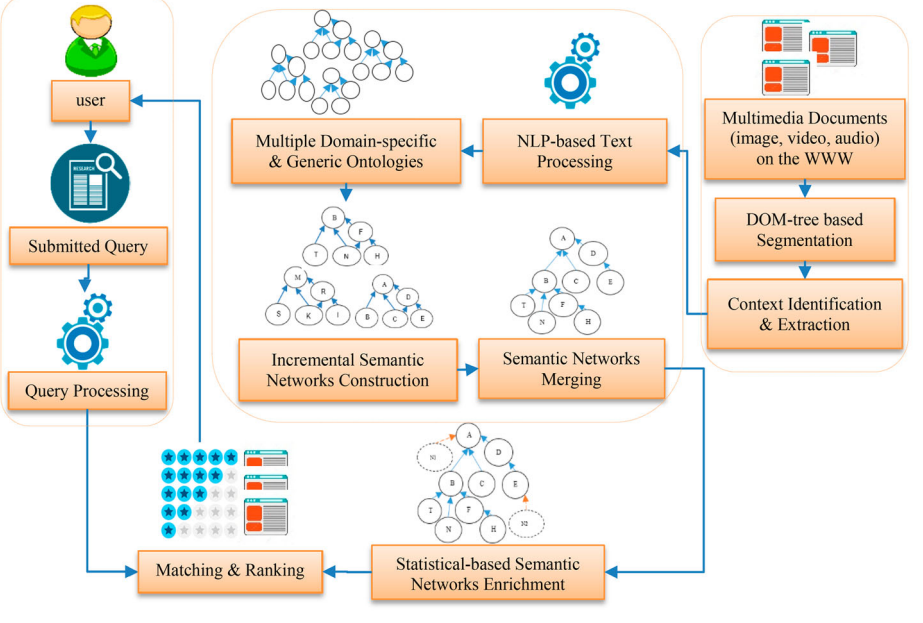


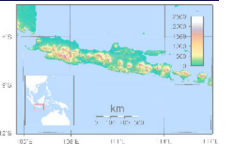
Figure 2. General architecture of the proposed system.

Webpages. As long as we can access the Document Object Model (DOM) for the Webpages where such documents reside, we can apply the proposed indexing scheme.

To begin the indexing process, we first utilise the DOM tree based segmentation techniques to segment and select the surrounding contextual information of multimedia documents. For more details on these techniques, we refer the reader to (Gulati & Yadav, 2019). Textual information extracted during this step are characterised by their noisy and sparse nature. Therefore, we filter out and refine the extracted information by employing several NLP steps, including stopwords removal, n-gram based text tokenization, de-pluralization, part-of-speech tagging and word sense disambiguation. After this step, semantic networks are incrementally constructed depending on the exploited ontologies. These networks represent the initially recommended semantic indexes for multimedia documents. To build the networks, each n-gram token is submitted to each of the used ontologies to find whether it is defined in it or not. As such, recognised tokens are represented as nodes in the semantic networks. An edge is inserted between two nodes if there is a semantic relation between them according to their definition in the used ontologies. Since we employ multiple domain-specific and generic ontologies, the number of produced semantic indexes may be zero, one, or more depending on the knowledge coverage of each ontology i.e. whether the submitted tokens are defined in the ontology or not. Additionally, there may be a conceptual and terminological difference between the produced semantic indexes. Therefore, we utilise the merging techniques proposed in (Maree & Belkhatir, 2015) to combine the different indexes into a single coherent semantic index. Consequently, each semantic index is created upon a cooperative decision made by the exploited ontologies. On the other hand, for unrecognised entities, we employ statistical-based semantic networks enrichment techniques. Namely, we utilise the NRD function to measure the semantic relatedness between the missing entities (those that are part of the surrounding contextual information of multimedia documents, but were not recognised by any of the exploited ontologies) and other entities that are represented by the nodes of the merged semantic networks. In this context, we enrich the merged semantic index by attaching concepts with strong semantic relatedness measures. On the other hand, we utilise the same NLP and enrichment techniques to process and index each query. In this context and unlike conventional approaches that use the bag-of-words model to expand users' queries, we construct an inverted index that stores query terms in addition to their semantically-relevant entities that are obtained from the exploited ontologies. In the next section, we provide a detailed description of the methods and techniques that we employ in our proposed framework.

To automatically extract multimedia documents and their relevant surrounding contextual information from the Web, we use the DOM Tree based Webpage segmentation algorithm proposed in (Sanoja & Gançarski, 2014).

As portrayed in Figure 3, the segmentation process is based on the DOM Tree structure of Webpages. Using this algorithm, multimedia documents are identified based on a three-phase segmentation process. These phases are page analysis, page understanding and page reconstruction. For more details on these phases, we refer the reader to the Black-o-Matic![5] Website where a free copy of the segmentation suite is publicly available. The results of this component are further filtered out and refined using several NLP steps as demonstrated by the next example. Given the two following segments (Seg_1 and Seg_2):

Segments	Multimedia Documents
Seg_1 *Java* (Indonesian: Jawa) is an island of Indonesia and the site of its capital city, Jakarta. Once the centre of powerful Hindu-Buddhist kingdoms, Islamic sultanates, and the core of the colonial Dutch East Indies, Java now plays	
Seg_2 *Java* is a general-purpose programming language that is class-based, object-oriented, and designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere" (WORA), meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. The syntax of Java is similar to C and C++, but it has fewer low-level facilities than either of them	

We first call the stopwords removal function using a pre-defined list that includes 480 stopwords such as: a, the, an, … etc. Then, the text tokenization algorithm tokenises the input text into n-gram tokens of lengths from 1 to 4. The results of these steps are shown in Table 1.

We would like to point that the exploited ontologies are utilised to identify n-gram tokens as part of the tokenization process. Our goal in this context is to produce meaningful n-gram tokens (also referred to as the set of correct tokens T_{correct}) that have a significant contribution to the meaning of the semantic index. For instance, tri-gram tokens such as "general-purpose programming language" and "dutch east indies" are recognised by either one or more of the exploited ontologies and are accordingly added to the set T_{correct}. Next, we use the Stanford CoreNLP[6] to classify the tokens of each T_{correct} into the grammatical category that they belong to. This is namely accomplished using the POS-Tagger module. Following this step, the semantic networks construction algorithm takes the set of correct tokens as input and produces semantic networks by automatically identifying the semantic relations that may hold between the tokens. The created networks are accordingly suggested as the initial semantic indexes for the multimedia documents. However, the probability that these networks are

[5] http://bom.ciens.ucv.ve/
[6] https://stanfordnlp.github.io/CoreNLP/

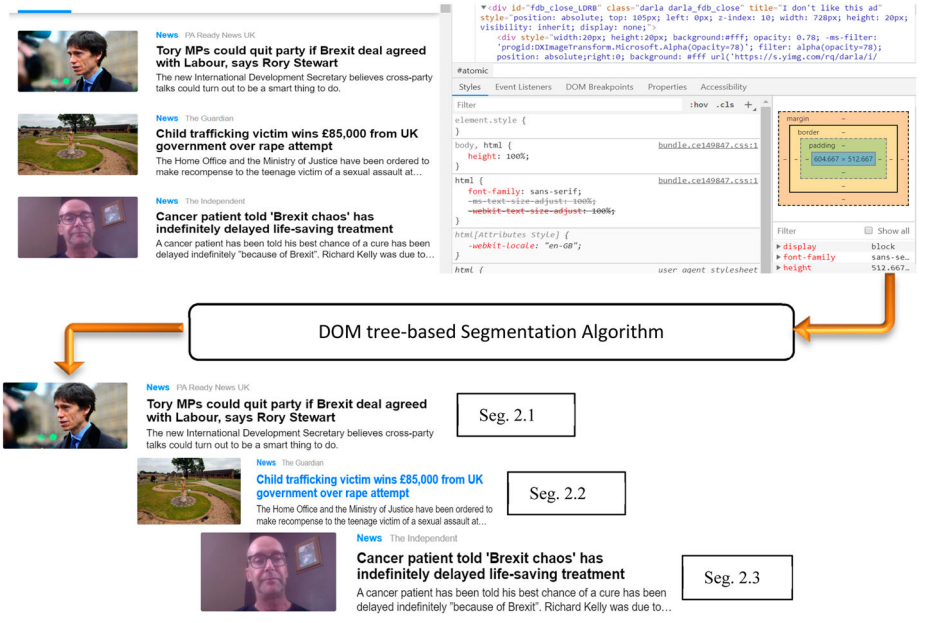


Figure 3. Identification and extraction of multimedia documents and their surrounding contextual information.

Table 1. Results of NLP Processing.

Stopwords Removal and Tokenization Steps
java - indonesian - jawa - island - indonesia - site - capital city - jakarta - centre - powerful - hindu-buddhist - kingdom - islamic - sultanate - core - colonial - dutch east indies - java - play - dominant - role - economic - political - life - Indonesia
java - general-purpose programming language - class-based - object-oriented - designed - implementation - dependency - possible - intended - application developer - write once run anywhere - wora - meaning - compiled - java code - run - platform - support - java - recompilation - syntax - java - c and - c++ - fewer - low-level - facility

semantically heterogeneous is high. For instance, as shown in Figures 4 and 5, more than one network are produced to index the Segments (Seg_1 and Seg_2).

As we can see in Figure 4, three semantic networks are proposed to index Seg_1. These initially proposed semantic indexes are obtained using the exploited ontologies. By carefully examining these networks, we can notice that there are conceptual and terminological differences across the suggested indexes. For example, some ontologies propose that the term "*Java*" is a meronym ("part of") of "*Indonesia*" as shown in Semantic Netowks$_{1,3}$, while others suggest that the same term should be considered as a holonym ("related to") of "*Indonesia*" as we can see in Semantic Network$_2$. Similarly, we can see that some ontologies consider "*Java*" and "*Dutch East Indies*" to be synonyms, while others consider the terms "*Java*" and "*Jawa*" as synonymous terms.

As we can see in Figure 5, two heterogeneous semantic networks are proposed to index Seg_2. Similar to the networks depicted in Figure 4, we can see another

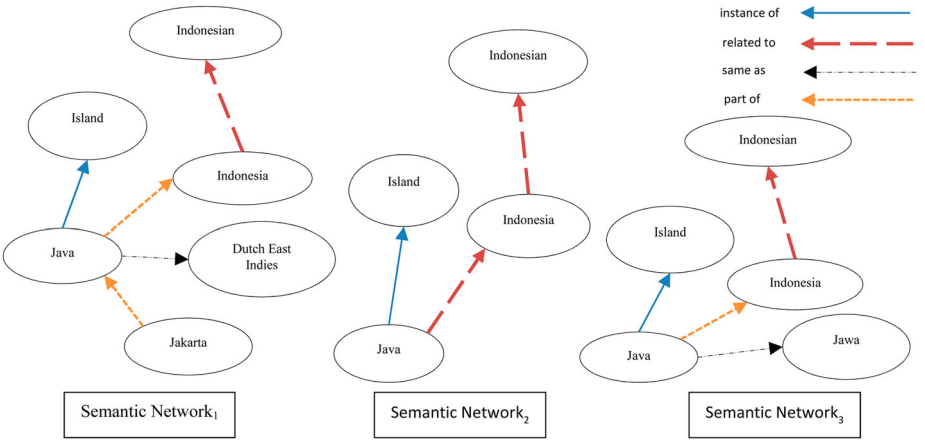


Figure 4. Heterogeneous semantic networks produced for indexing the image of *Seg*$_1$.

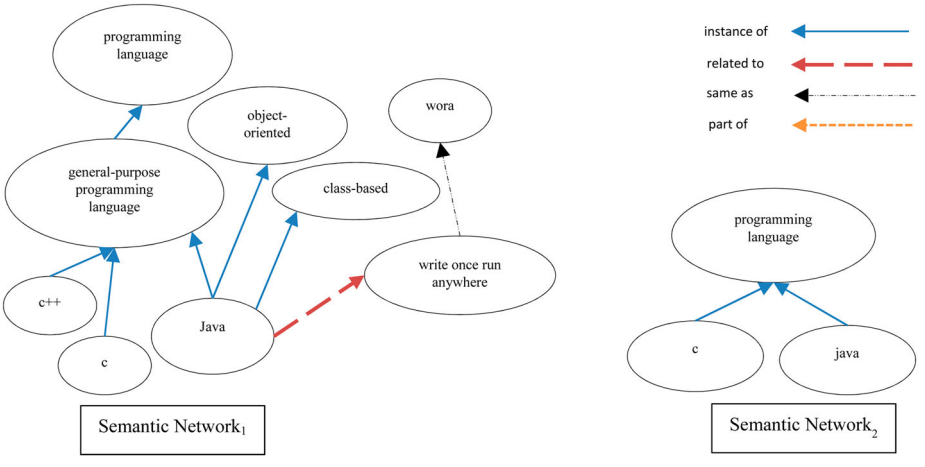


Figure 5. Heterogeneous semantic networks produced for indexing the image of *Seg*$_2$.

type of conceptual and terminological differences where some networks incorporate additional semantically-relevant terms that are missing in other networks. We can clearly see this effect when comparing the number of entities in Semantic Network$_1$ against those defined in Semantic Network$_2$. The main reason behind this heterogeneity is the lack of semantic knowledge in some of the exploited ontologies. This indeed demonstrates the need to (1) exploit more than one ontology and (2) merge the produced semantic indexes into a single coherent index.

To overcome the heterogeneity problem, we utilise the merging techniques proposed in (Maree & Belkhatir, 2015) to merge these networks into a single coherent network. The merging algorithm takes pairs of semantic networks (ζ_1 and ζ_2) as input, and finds all possible mapping elements between their

nodes, producing a single merged network ζ_{merged} as output. We formally define the mapping function as:

Definition 6: Mapping Element: Given two semantic networks ζ_1 and ζ_2, we compute the $N_1 \times N_2$ mapping elements $\langle Id_{ij}, n_{1i}, n_{2j}, R \rangle$ where:

- Id_{ij} is a unique identifier of the mapping element
- n_{1i} is a node in the first semantic network (n_{1i} ε ζ_1)
- n_{2j} is a node in the second semantic network (n_{2j} ε ζ_2)
- R includes relations such as Synonymy (≡), Disjointness (⊥), Hyponymy, etc. that hold between the nodes n_{1i} and n_{2j},i, $j(i = 1, \ldots, n_1)$, $((j = 1, \ldots, n_2))$

As shown in Figure 6, the initial semantic indexes for both contextual segments are merged using the above technique into single networks.

It is important to highlight that we use the majority voting algorithm proposed in (Maree & Belkhatir, 2015) for merging the semantic networks. The reason behind this step is because — as we can see in Figures 4 and 5 — the relations that hold between the same entities in two or more networks are not consistent. For example, the relation that holds between the terms "*java*" and "*indonesia*" is "*part of*" in both semantic networks 1 and 3, while the relation that connects the same terms in network 2 is "*related to*". Accordingly, using the majority voting technique, decisions on the relations that should link the terms in the merged networks are obtained based on the majority of the employed ontologies. It is worth noting that even though various domain-specific and generic ontologies are exploited, we still find entities that are either not recognised by any of the used ontologies or cannot be linked to other entities in the produced semantic networks. This happens because there are no semantic/taxonomic relations that hold between such entities and the

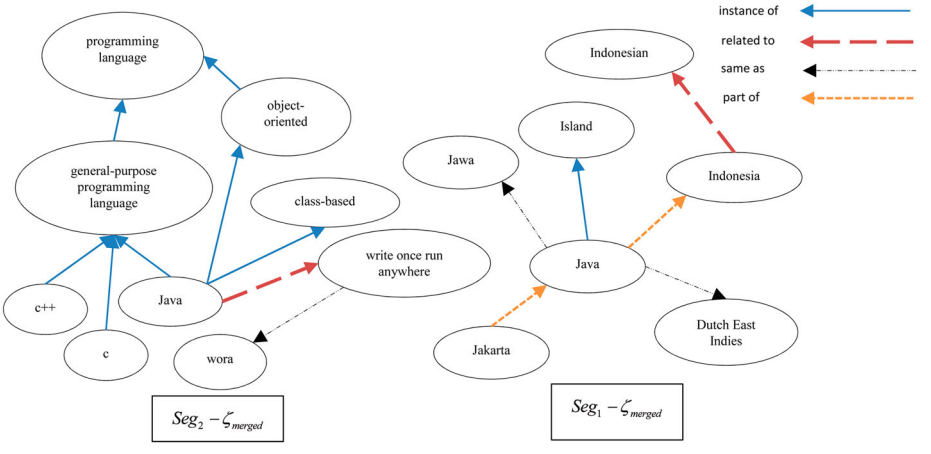


Figure 6. Merged semantic indexes for Seg_1 and Seg_2.

rest of entities that are included in the semantic networks. For instance, the words "platform" and "syntax" were not added to the initially proposed indexes. To address this problem, we propose utilising statistical-based concept-relatedness measures to find whether those entities can be proposed as potential enrichment candidates to the merged semantic networks. We formally define the network enrichment technique as:

Definition 7: Statistical-based concept-relatedness Enrichment: This technique takes a given set of terms $T = \{t_1, t_2, t_3, \ldots, t_n\}$ that are not recognised by the exploited ontologies or were not add to ζ_{merged} and produces for each $t \in T$ in ζ_{merged} a set $S(t) \subseteq T$ as output, where:

- $S(t)$ is the set of suggested enrichment candidates for t.

In this context, $S(t)$ is obtained using the Normalized Retrieval Distance (NRD) function proposed in (Cilibrasi & Vitanyi, 2007) and based on a threshold value v using Equation 3.

$$S(t, v) = \{w \in W | \text{NRD}(t, w) \leq v\} \tag{3}$$

where:

- $S(t, v)$ is the set of suggested enrichment candidates for each term t that is obtained based on the threshold value v. In the context of our work, we have manually configured the threshold value to be $v = 0.80$. As such, we only considered candidate expansion terms with all $NRD(t, w)$ values that were greater than or equal to v. In fact, we empirically found the current threshold value to be appropriate for index expansion purposes as it appeared to minimise the number of suggested enrichment candidates and exclude terms with weak co-occurrence degrees. We believe that further investigation of the various methods of setting an appropriate threshold value is still required in this domain.
- $w \in W$ represents entities that were not recognised by the exploited ontologies. A word w is then added to S if its $NRD(t, w)$ result is greater than or equals to v.

In Tables 2 and 3 below, we show the measures of semantic relatedness between a subset of the terms that belong to both the first and second contextual segments Seg_1 and Seg_2 and the term "*java*" from the merged semantic indexes $\text{Seg}_1 - \zeta_{\text{merged}}$ and $\text{Seg}_2 - \zeta_{\text{merged}}$.

As we see in Tables 2 and 3, different strengths of semantic relatedness exist between the terms from $T = \{t_1, t_2, t_3, \ldots, t_n\}$ and the term "*java*" from $Seg_1 - \zeta_{merged}$ and $Seg_2 - \zeta_{merged}$. In this context, we only consider a subset of enrichment candidates from $T = \{t_1, t_2, t_3, \ldots, t_n\}$ based on an automatic

Table 2. Sample measures of statistical-based semantic relatedness for the term "java" from *Seg$_1$-merged*.

Term 1	Term 2	Semantic Relatedness
colonial	java	0.37
site	java	0.42
capital city	java	0.34
centre	java	0.50
powerful	java	0.37
hindu-buddhist	java	0.10
kingdom	java	0.49
islamic	java	0.33
sultanate	java	0.42
core	java	0.40
colonial	java	0.37
play	java	0.00
dominant	java	0.41
role	java	0.41
economic	java	0.41
political	java	0.37
life	java	0.00

Table 3. Sample measures of statistical-based semantic relatedness for the term "java" from *Seg$_2$-merged*.

Term 1	Term 2	Semantic Relatedness
application developer	java	0.42
java code	java	0.98
compiled	java	0.41
run	java	0.41
platform	java	0.41
syntax	java	0.31
implementation	java	0.16
intended	java	0.24
support	java	0.37
meaning	java	0.40
designed	java	0.33
possible	java	0.33
dependency	java	0.28
recompilation	java	0.32
low-level	Java	0.45

threshold value v. This value is determined based on Equation 4:

$$v = t_{\max} - \frac{\sum_{i=1}^{n} t_i}{n} \tag{4}$$

where:

- v is the threshold value for selecting enrichment candidates.
- $t_{\max}$ is the maximum semantic relatedness measure.
- n is the number of semantic relatedness values.

As such, in our scenario, for $\mathrm{Seg}_1 - \zeta_{\mathrm{merged}}$ and $\mathrm{Seg}_2 - \zeta_{\mathrm{merged}}$, the suggested enrichment candidates are those that have a value of $v \geq 0.48$ and $v \geq 0.95$

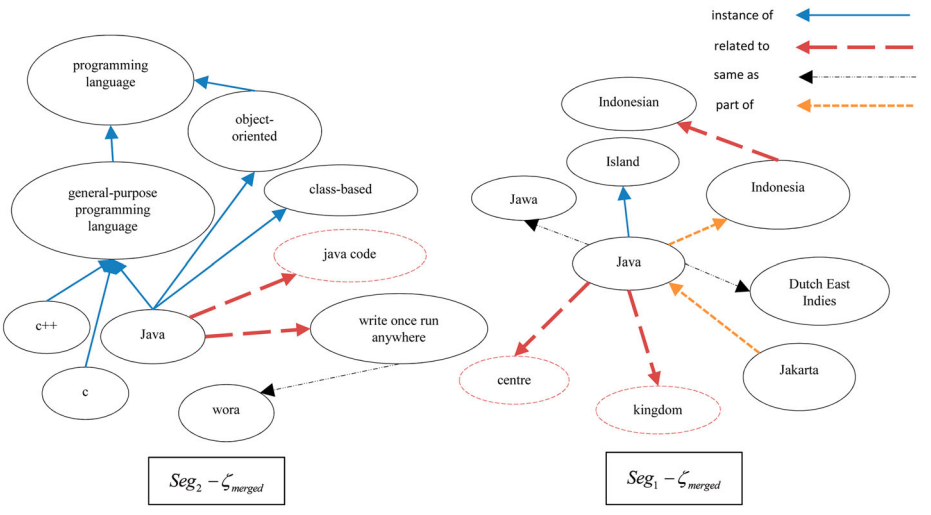


Figure 7. Attaching new entities to the merged semantic indexes for Seg_1 and Seg_2.

respectively. Accordingly the set $T = \{\text{centre, kingdom}\}$ is suggest for enriching $\text{Seg}_1 - \zeta_{\text{merged}}$, while the set $T = \{\text{java code}\}$ is suggested for enriching $\text{Seg}_2 - \zeta_{\text{merged}}$. We use the semantic relation "related to" to attach new entities to the merged indexes. As a result of the enrichment process, enrichment candidates from $T = \{t_1, t_2, t_3, \ldots, t_n\}$ are attached to ζ_{merged} and the hierarchical structure of ζ_{merged} is updated accordingly. Figure 7 depicts the result of enriching both semantic networks that belong to (Seg_1 and Seg_2).

Experimental results

In this section, we present and discuss the experiments that we have conducted to evaluate the quality of our proposed cooperative knowledge-based multimedia indexing approach. In these experiments, we exploit two publicly available ontologies, namely WordNet (Miller, 1995) and YAGO (Fabian et al., 2007), in addition to a set of domain-specific ontologies obtained from the online ontology repositories available at oor.net[7] and bioportal[8] websites. These domain-specific ontologies are: Ontology of Genes and Genomes[9] and Medical Subject Headings[10]. All solutions have been implemented using Java as the main programing language, and additional NLP and ontology processing software packages and APIs, such as Appache's JENA[11] API and Stanford's CoreNLP.[12]

[7] http://oor.net/
[8] http://bioportal.bioontology.org/
[9] http://bioportal.bioontology.org/ontologies/OGG/
[10] http://bioportal.bioontology.org/ontologies/MeSH/
[11] https://jena.apache.org/
[12] https://stanfordnlp.github.io/CoreNLP/

Experiments using single against multiple generic ontologies

Before we carry on to our experiments using the proposed cooperative ontology-based indexing approach, we conduct experiments using two generic ontologies to demonstrate the differences among them in terms of their role in improving the quality of the indexing process. This is an important step to validate our proposal of incorporating multiple ontologies that cooperatively assist in multimedia documents indexing. To do this, we have used three different datasets. The first dataset is a manually constructed dataset that comprises 18 queries and 300 randomly-selected Webpages that contain different types of multimedia documents (100 per multimedia type — image, video and audio segments). The second dataset (henceforth referred to as Dataset A) is a publicly-available dataset that has been used extensively by a variety of semantics-based information indexing and retrieval approaches. The dataset comprises 93 documents in the form of textual queries and 11,429 documents that are assigned with their corresponding relevance scores. The reason behind using this dataset is to practically demonstrate the effectiveness of utilising the proposed approach in indexing documents that contain textual information that can be treated as the surrounding contextual information of multimedia Web documents. In addition, since the authors of this dataset provide the relevance judgements, we can compute the precision of our system and evaluate the quality of the produced results when compared to the manual scores. It also assists in allowing researchers in the field to re-use the dataset (in addition to our results) to ensure the reproducibility of our experiments and for further research activities in this domain. The third dataset (henceforth referred to as Dataset B) is the CLEF iaprtc12[13] dataset. This dataset contains 41 folders with a total of 20,000 images associated with their contextual information in the form of .xml documents. Each document contains several xml tags to describe the image in terms of its title, description, notes about each image, and other relevant information. Figure 8 depicts an example of an image associated with its surrounding contextual information. We have defined 40 queries, in addition to their relevance judgements to evaluate the quality of the produced results when using baseline metrics against semantically-enhanced indexing strategies.

To evaluate the effectiveness of the proposed system in the same manner as reported in (Suchanek et al., 2009), we first start with the manually constructed dataset and compare our ground truth (i.e. manually-assigned relevance scores) to the corresponding scores that were automatically produced by the system. Figures 9 and 10 demonstrate the precision/recall differences among two experimental scenarios. First, in Figure 9, we start with a single generic ontology; that is WordNet and then we use another single generic ontology; that is YAGO to index the documents in the dataset. We experimentally demonstrate that when

[13]http://www-i6.informatik.rwth-aachen.de/imageclef/resources/iaprtc12.tgz/

<DOC>
<DOCNO>annotations/08/8991.eng</DOCNO>
<TITLE>New pupil in the School San Juan Apostol</TITLE>
<DESCRIPTION>a tourist is sitting with the other pupils at the wooden desks; some pupils are standing; there is a blue shelf with lots of stuff on it in the background;</DESCRIPTION>
<NOTES></NOTES>
<LOCATION>Arequipa, Peru</LOCATION>
<DATE>10 October 2003</DATE>
<IMAGE>images/08/8991.jpg</IMAGE>
<THUMBNAIL>thumbnails/08/8991.jpg</THUMBNAIL>
</DOC>

Figure 8. An example image with its surrounding contextual information from the CLEF iaprtc12 dataset.

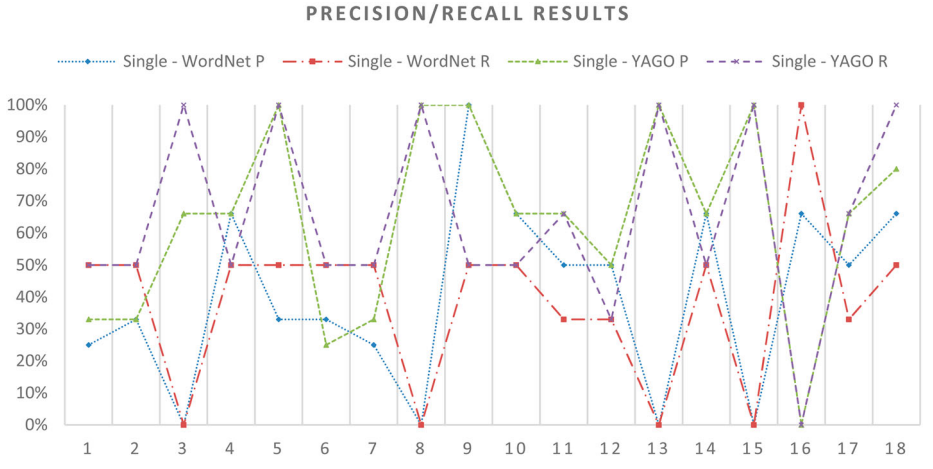


Figure 9. Precision/Recall results when using a single generic ontology.

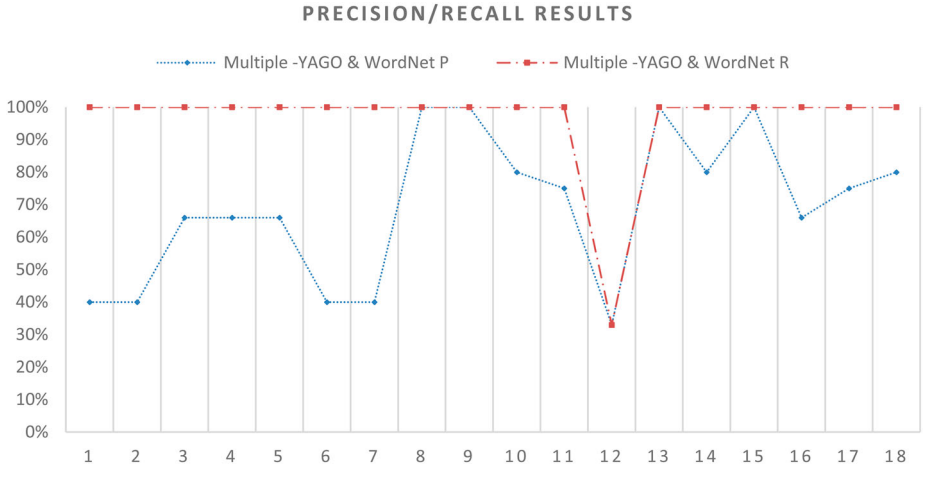


Figure 10. Multiple ontology-based (YAGO and WordNet) Precision/Recall measures.

using WordNet alone, the precision, as well as recall results, were not satisfactory due to the semantic knowledge incompleteness problem in this ontology. We would like to emphasise here the fact that WordNet was originally created for computational linguistic purposes, however, we have re-used knowledge encoded in this resource for information retrieval purposes since it provides a rich source of lexical-semantic information, including synonymous concepts, their hypernyms and hyponyms. Nevertheless, the domain coverage of WordNet is limited, namely considering the broadness and diversity of the terms used to describe the content of multimedia Web documents. Therefore, many terms were found to be missing in this ontology, demonstrating the necessity for exploiting additional ontologies to provide broader and more in-depth domain coverage. To further demonstrate this point, we used another generic ontology, that is YAGO for indexing the documents in the dataset. The main reason behind utilising another generic ontology is to demonstrate how the vast knowledge captured by this ontology can better assist in the semantic index construction process; enhancing the quality of the produced indexes.

As shown in Figure 9, we can see that for most of the queries, the precision has improved when using YAGO ontology. This is namely because this ontology provides broader and deeper domain coverage, and accordingly, it recognised most of the terms that are extracted from the surrounding contextual information. However, we can still find some terms that are also missing in YAGO. Therefore, as depicted in Figure 10, we utilised both ontologies for indexing purposes. As shown in this figure, the quality of the indexing process has improved against using single-based generic ontologies. This demonstrates that, on the one hand, multiple ontologies can cooperatively provide a wider domain coverage, and they can jointly assist in constructing the semantic indexes for each segment of contextual information on the other.

It is essential to point to the fact that despite the improvement achieved when utilising both ontologies, we faced the problem of inconsistent semantic relations that were returned by each of these ontologies. Therefore, we can see for some of the queries; the precision of the indexing process has degraded. This fact demonstrates the need for employing the majority voting technique (introduced in section 3) to resolve such inconsistencies among the utilised ontologies. The need for utilising this technique becomes crucial when incorporating domain-specific ontologies in the indexing process.

As depicted by the above figures, it is also important to highlight that for some results, the level of improvement on the precision and recall values was not satisfactory. This is because the used ontologies — despite their domain coverage — still suffer from semantic knowledge incompleteness problems. This problem can be mitigated by utilising the statistical-based concept of relatedness measures. These techniques can be used to further enrich the employed ontologies with additional terms that are not recognised in their current specifications. In the next section, we experimentally demonstrate the impact of

employing our proposed approach on the quality of four baseline indexing techniques.

Experiments using the semantically-enhanced model

In these experiments, we have used the second and third datasets (Dataset A and Dataset B) that are described in Section 5.A. By using these two datasets, our aim is to demonstrate the impact of enhancing the indexing process using semantic relations and semantically-relevant entities that are extracted from the employed ontologies. To validate the effectiveness of our proposal, we first start with the baseline results depicted in Figure 11. We have used four metrics to calculate the similarity between the queries and their corresponding contextual segments. Our aim in this context is to experimentally evaluate the accuracy of the proposed techniques in assigning relevance scores between query-document pairs, and measure their effectiveness in retrieving multimedia documents that satisfy users' information needs. The used metrics are: Cosine Similarity, Jaccard Similarity, JaroWinkler Similarity and SorensenDice Similarity. For

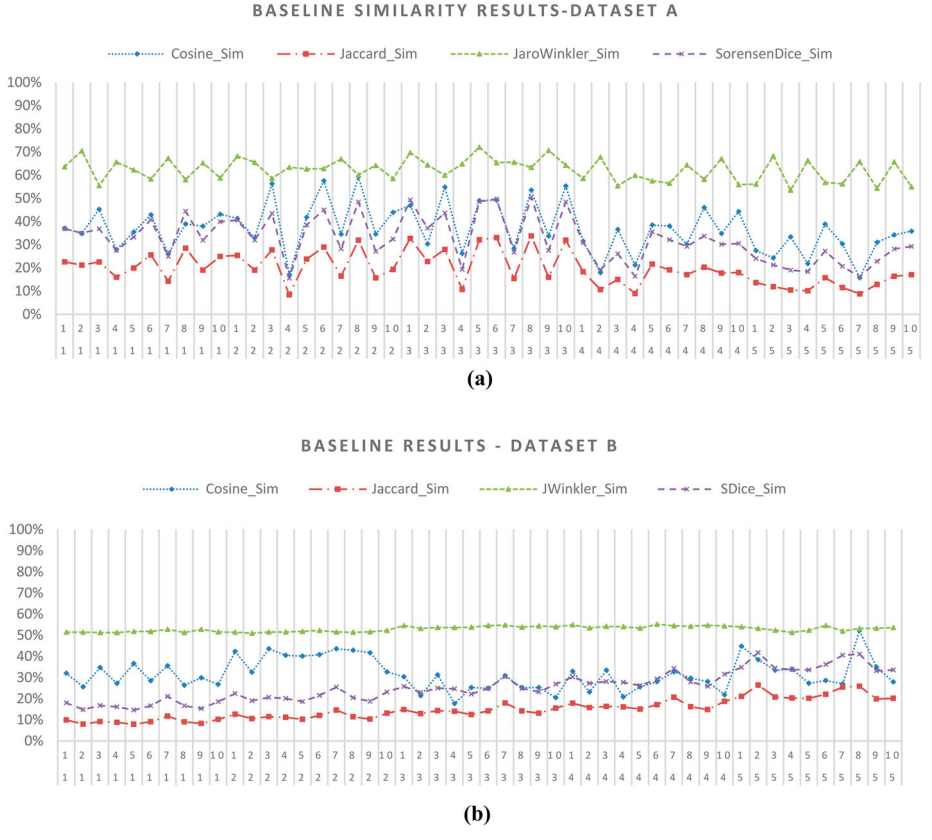


Figure 11. Baseline similarity results for both datasets using the four edit distance techniques. (a) Baseline Similarity Measures –Dataset A. (b) Baseline Similarity Measures –Dataset B.

more details on these measures, please refer to (Gomaa & Fahmy, 2013) where the authors provide a summary of a variety of edit distance measures including the ones that we use in our experiments. Before we proceed into discussing the results, we would like to point out there are two important parameters that have been incorporated to calculate similarity scores using each of the employed techniques. These are: (1) the calculation method (i.e. mathematical formula for calculating the similarity between terms in indexes and their correspondences in the queries) and (2) the number of shingles[14] that we assign to represent the character sequence of each term. Assigning different shingle values have empirically produced different similarity scores. In the context of our work, we explored using shingles of values 2, 4 and 6. We have empirically found that using 2-shingle character values produced the highest precision scores. When combining this number of shingles with the Jaro-Winkler's formula, we have obtained the highest precision scores.

For demonstration purposes, we have selected the first five queries and their associated contextual segments (10 segments per query) produced for both the first and second datasets.

As we can see in Figure 11, the returned results for both datasets (noting that the results produced by the Jaro-Winkler technique were flatter than those produced by the rest of the techniques) using the four edit distance measures are characterised by their low quality. This is because we ignore all semantic and taxonomic relations that exist between the terms of the used contextual segments. Accordingly, the elements (nodes and relations) of the constructed semantic indexes were at their minimal level leading to degrading the overall quality of the indexing process. Next, we attempt to enrich the constructed indexes with synonyms extracted from the employed ontologies in order to explore the impact of this enrichment procedure on the quality of the produced results. The results of this step are depicted in Figure 12. As we can see in this figure, there is an improvement on the precision of the indexing process using the four techniques. However, the level of improvement in this context is considered of little value, namely when compared to the experiments where we incorporate additional semantically-related terms such as hypernyms and hyponyms that are defined in the used ontologies.

As shown in Figure 12, when we added synonymous terms to the constructed indexes, the system retrieved additional semantically-related images and accordingly, the levels of precision have increased. For instance, for the query "*pupils at schools*", all images with "*pupils/students at school/schoolhouse*" were retrieved as well. After this step, we further enriched the constructed indexes with additional semantically-related terms such as hypernyms and hyponyms using the employed ontologies. To carry out this step, we have obtained broader/narrower terms from the used ontologies

[14]A Shingle is a sequences of n consecutive characters in a string.

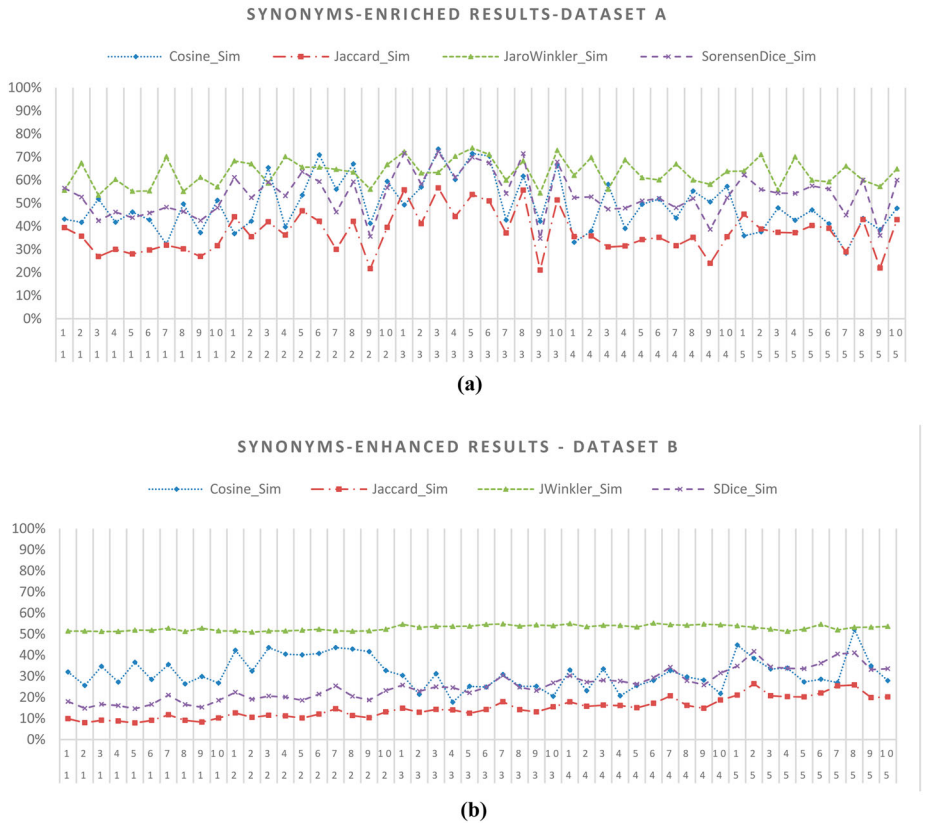


Figure 12. Similarity results after enriching the indexes with synonyms. (a) Synonyms-enhanced Similarity Measures – Dataset A. (b) Synonyms-enhanced Similarity Measures –Dataset B.

(both domain-specific and generic ontologies as long as this information is provided in their hierarchical structures) by moving one level up and one level down in the hierarchy of the ontology. For instance, for query terms such as "*school*", all documents containing "*educational institution*"; which is a hypernym to "*school*", and "*academy*" which is a hyponym of "*school*" were also retrieved among the set of relevant results. Figure 13 shows the results of carrying out this step.

As depicted in Figure 13, a higher level of improvement upon the overall quality of the indexing process has been achieved. This achievement supports our argument that multiple ontologies can cooperatively enhance baseline techniques. Adding semantically-relevant entities, in addition to synonymous terms, has proved to return additional relevant multimedia documents that correspond to the terms in the initial queries. To quantify the level of improvement achieved on the average precision when using the four based metrics against the synonyms-based and semantically-enhanced techniques, we have compared the total average precision between the three different approaches using both datasets as depicted in Figure 14.

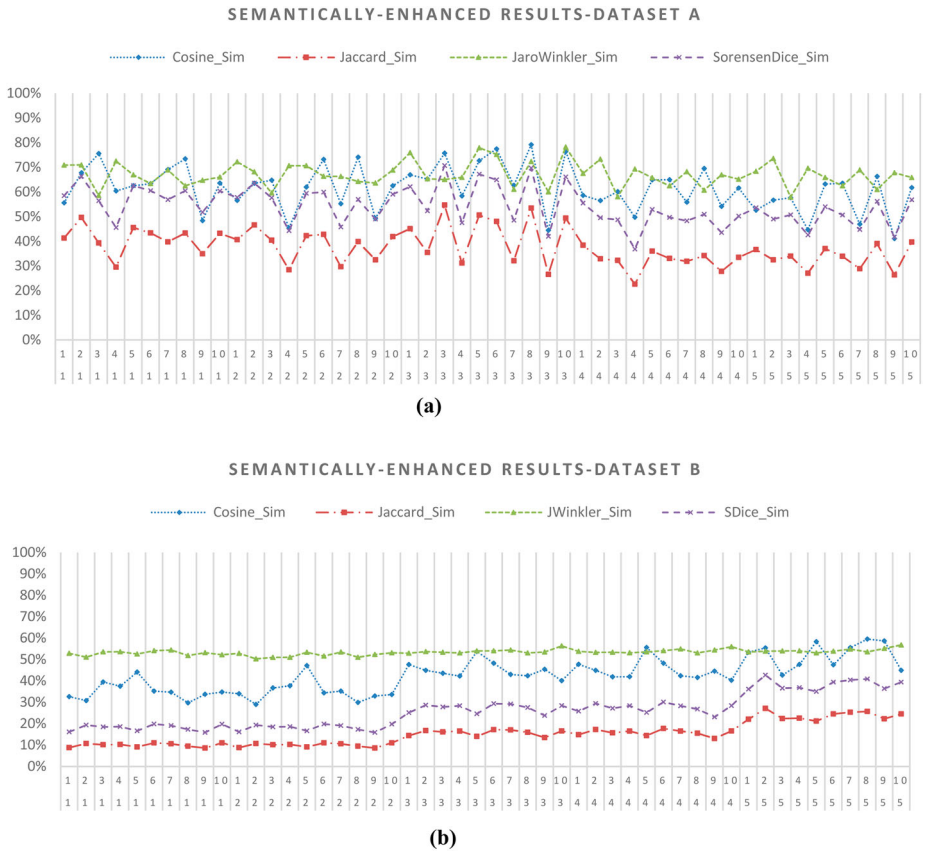


Figure 13. Similarity results after enriching the constructed indexes with semantically-related terms. (a) Semantically-enhanced Similarity Measures –Dataset A. (b) Semantically-enhanced Similarity Measures –Dataset B

As shown Figure 14, a higher level of improvement on the overall quality (precision values) of the baseline measures is achieved when applying our proposed techniques. For instance, for the first dataset (Dataset A), when enriching the Cosine metric with synonyms (acquired from the employed ontologies) the average precision increases by 12%, while it increases by 25% when exploiting additional semantic and taxonomic relations. The same applies to the rest of the baseline metrics but with different levels of improvement. For example, enriching the Jaccard metric with synonyms improved the overall quality of the Jaccard baseline by 17%, and by 18% when exploiting the semantically-enhanced model. The level of improvement on the JaroWinkler baseline metric was 1% and 5% respectively, while it was 21% and 22% when applying the semantics-based techniques on the SorensenDice metric.

Accordingly, we can conclude that, for the four metrics used in our study, the semantic as well as taxonomic entities obtained from multiple ontologies have improved their overall quality. A similar level of improvement has been also achieved when using the second dataset (Dataset B). It is important however

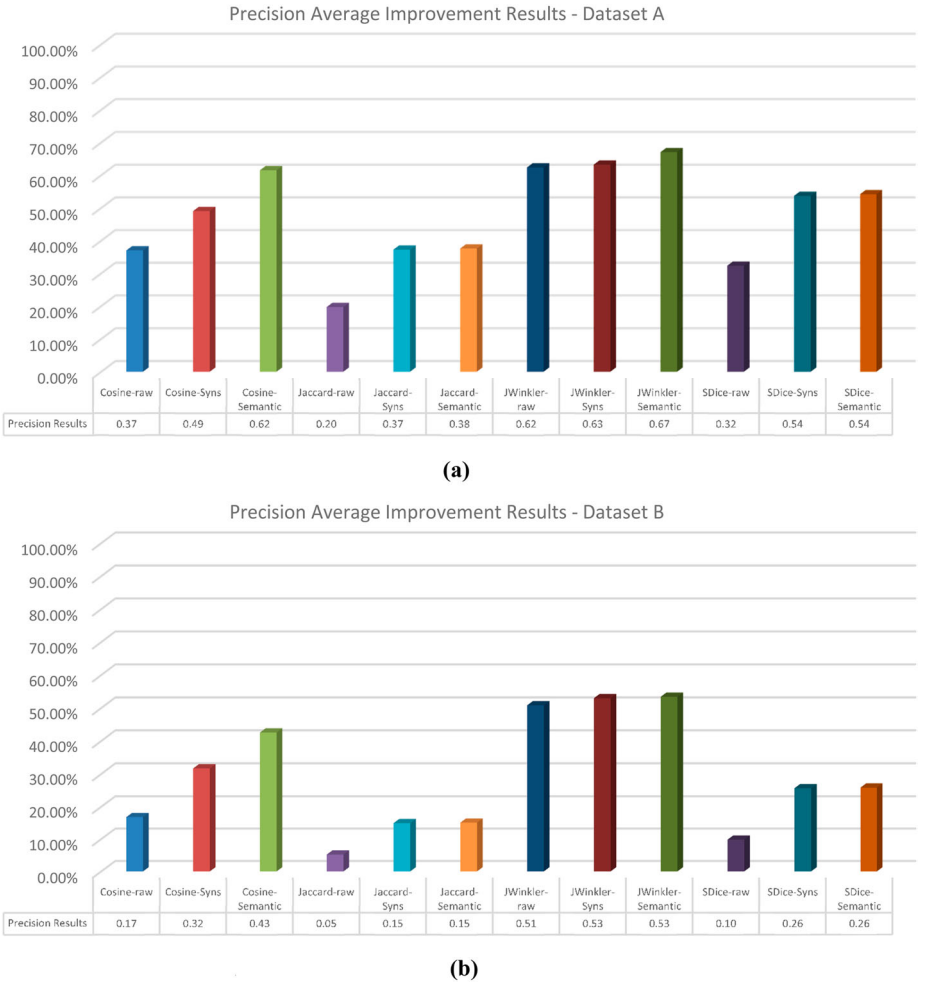


Figure 14. Average Improvement on the System's Precision Using Datasets A and B. (a) Average Improvement on the Precision – Dataset A. (b) Average Improvement on the Precision – Dataset B

to mention that, although the rate of improvement was similar for both Datasets A and B, the results were generally lower for Dataset B. We refer this to the fact that the level of contribution of the employed ontologies was higher for the first dataset. In particular, the topics of the annotation texts of Dataset B were about tourism-related images in specific locations, which were partially covered by the exploited ontologies. Also, the size of the annotation texts of this dataset was shorter compared to those in Dataset A, leading to a smaller number of enrichment and expansion candidates that were added to the constructed indexes. Accordingly, we conclude that despite the achieved improvement, our approach is still hindered by the fact that it heavily relies on the quality (depth and breadth of domain coverage, and precision of the semantic and taxonomic relations that

link various ontological entities) of the exploited generic and domain-specific ontologies.

As we have demonstrated earlier, when using a generic ontology such as WordNet alone, the level of improvement on the quality of the produced results was limited and not satisfactory. This is due to the semantic knowledge incompleteness problem in this ontology and in other similar generic ontologies. On the other hand, we argue that, even when utilising heavy-weight ontologies such as YAGO, which provides much wider domain coverage, we still face missing knowledge problems and also a lack of semantic relations that precisely describe and relate ontological entities. Therefore, exploiting domain-specific ontologies in our context can offer an important source of semantic information about the domain of interest.

However, since we are addressing the issue of indexing multimedia documents on the Web, the task of gathering and aligning generic and domain-specific ontologies in almost all domains is complex and time consuming. Therefore, this is among the limiting factors that hinders our current approach. We plan to address this problem in the future extension of our proposed system's prototype through proposing a new technique that automatically searches, collects and aligns multiple domain-specific ontologies in order to exploit them as part of the backbone of our proposed approach. This will of course open the door for further research issues such as: (1) measuring the efficiency of the proposed techniques, (2) evaluating the quality of the acquired ontologies, and (3) defining a mechanism for accessing ontological entities and prioritising the processes (i.e. to start with entities defined in domain-specific ontologies against those that are encoded in generic ontologies, in addition to identifying the hypernyms/hyponymy/synonymy levels that should be incorporated during the index construction phase) based on their precedence and impact on the quality of the produced results.

Conclusions and future work

In this paper, we discuss the issue of the relative ineffectiveness of multimedia indexing systems in the context of general search over multimedia documents on the Web. We demonstrate that the semantic knowledge incompleteness and semantic heterogeneity in existing ontologies are two main obstacles towards exploiting ontologies in practical application domains. To overcome these problems, we propose a hybrid ontology-based multimedia indexing approach that cooperatively identifies and extracts the semantic and taxonomic relations that exist between annotation words that surround multimedia documents on Webpages. To do this, we employ multiple generic and domain-specific ontologies to assist in constructing inverted semantic indexes that can be used for matching and ranking the retrieved multimedia documents. Accordingly, we constructed inverted indexes (in the form of semantic networks) where

nodes of each network are identified and added based on the voting results returned by the majority of the used ontologies. The semantic, as well as taxonomic relations that hold between those nodes, are used to connect related nodes in the networks. We also apply ontology merging algorithms to integrate the produced networks into consistent networks where each network represents a merged semantic index. To demonstrate the effectiveness of our proposal, we have developed a prototype system and used two datasets to test and evaluate the quality of the produced results. We have also compared the results produced by our system to four baseline metrics and demonstrated achieving promising results against conventional indexing approaches.

In future work, we plan to extend the current version of our system's prototype by incorporating additional ontologies. We plan to use other domain-specific datasets to study the impact of employing multiple ontologies on indexing multimedia documents that pertain to the investigated domains. Besides, we plan to study the performance/efficiency of the proposed approach when using heavy-weight ontologies that contain thousands to millions of entities, as we believe that using multiple large-scale ontologies will be a very costly process. On the other hand, we plan to explore state-of-the-art ontology merging techniques and investigate their impact on the quality of the merged indexes. We believe that studying the various approaches for merging heterogeneous semantic indexes can lead to further improvements on the overall's system effectiveness. Moreover, we plan to use additional large-scale datasets for testing the quality of the proposed techniques.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Mohammed Maree http://orcid.org/0000-0002-6114-4687

References

Abdulmunem, M. E., & Hato, E. (2018). Semantic based video retrieval system: Survey. *Iraqi Journal of Science*, *59*(2A), 739–753. https://doi.org/10.24996/ijs.2018.59.2A.12

Amato, F., Colace, F., Greco, L., Moscato, V., & Picariello, A. (2016). Semantic processing of multimedia data for e-government applications. *Journal of Visual Languages & Computing*, *32*, 35–41. https://doi.org/10.1016/j.jvlc.2015.10.012

Asim, M. N., Wasim, M., Khan, M. U. G., Mahmood, N., & Mahmood, W. (2019). The use of ontology in retrieval: A study on textual, multilingual, and multimedia retrieval. *IEEE Access*, *7*, 21662–21686. https://doi.org/10.1109/ACCESS.2019.2897849

Bendib, I., & M. R. Laouar. (2018). A semantic indexing approach of multimedia documents content based partial transcription. In *2nd international conference on natural language and speech processing (ICNLSP)* (pp. 1–6). IEEE.

Bhunia, A. K., Bhattacharyya, A., Banerjee, P., Roy, P. P., & Murala, S. (2019). A novel feature descriptor for image retrieval by combining modified color histogram and diagonally symmetric co-occurrence texture pattern. *Pattern Analysis and Applications*, *22*(4), 1517–1526. https://doi.org/10.1007/s10044-018-00771-2

Bracamonte, T., Bustos, B., Poblete, B., & Schreck, T. (2018). Extracting semantic knowledge from web context for multimedia ir: A taxonomy, survey and challenges. *Multimedia Tools and Applications*, *77*(11), 13853–13889. https://doi.org/10.1007/s11042-017-4997-y

Budikova, P., Batko, M., & Zezula, P. (2018). Conceptrank for search-based image annotation. *Multimedia Tools and Applications*, *77*(7), 8847–8882. https://doi.org/10.1007/s11042-017-4777-8

Cilibrasi, R. L., & Vitanyi, P. M. (2007). The google similarity distance. *IEEE Transactions on Knowledge and Data Engineering*, *19*(3), 370–383. https://doi.org/10.1109/TKDE.2007.48

Da Silva, J., K. Revoredo, F. Araujo Baião, & J. Euzenat. (2018, December 23). Interactive ontology matching: Using expert feedback to select attribute mappings. In *13th ISWC workshop on ontology matching*, Monterey, United States: No commercial editor, 25-36.

De Oliveira Barra, G., M. Lux, & X. Giro-I-Nieto. (2016, June 15–17). Large scale content-based video retrieval with livre. In *14th international workshop on content-based multimedia indexing (CBMI)* (pp. 1–4). IEEE.

Fabian, M., K. Gjergji, & W. Gerhard. (2007, May 8–12). Yago: A core of semantic knowledge unifying wordnet and wikipedia. In *16th international World Wide Web conference* (pp. 697–706). WWW.

Gomaa, W. H., & Fahmy, A. A. (2013). A survey of text similarity approaches. *International Journal of Computer Applications*, *68*(13), 13–18. https://doi.org/10.5120/11638-7118

Gulati, P., & Yadav, M. (2019). A novel approach for extracting pertinent keywords for web image annotation using semantic distance and Euclidean distance. In M. Hoda, N. Chauhan, S. Quadri, & P. Srivastava (Eds), *Advances in Intelligent systems and Computing* (pp. 173–183). Springer Singapore.

Guo, K., Liang, Z., Tang, Y., & Chi, T. (2018). Sor: An optimized semantic ontology retrieval algorithm for heterogeneous multimedia big data. *Journal of Computational Science*, *28*, 455–465. https://doi.org/10.1016/j.jocs.2017.02.005

He, Y., Li, Y., Lei, J., & Leung, C. H. (2016). A framework of query expansion for image retrieval based on knowledge base and concept similarity. *Neurocomputing*, *204*, 26–32. https://doi.org/10.1016/j.neucom.2015.11.102

Hong, R., Yang, Y., Wang, M., & Hua, X.-S. (2015). Learning visual semantic relationships for efficient visual retrieval. *IEEE Transactions on Big Data*, *1*(4), 152–161. https://doi.org/10.1109/TBDATA.2016.2515640

Hu, H., Wang, Y., Yang, L., Komlev, P., Huang, L., Chen, X. S., Huang, J., Wu, Y., Merchant, M., & Sacheti, A. (2018, August 19–23). Web-scale responsive visual search at bing. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 359–367). ACM.

Iftene, A., & Alexandra-Mihaela, B. (2016). Using semantic resources in image retrieval. *Procedia Computer Science*, *96*, 436–445. https://doi.org/10.1016/j.procs.2016.08.093

Irtaza, A., Jaffar, M. A., & Muhammad, M. S. (2015). Content based image retrieval in a web 3.0 environment. *Multimedia Tools and Applications*, *74*(14), 5055–5072. https://doi.org/10.1007/s11042-013-1679-2

Ishtiaq, F., Fonseca, Jr, B. J., Baum, K. L., Braskich, A. J., Emeott, S. P., Gandhi, B., Li, R., Smith, A. M., Needham, M. L., & Dellahy, I. O. (2018, February 6). Content based video content segmentation. *Google Patents*, 1–32. U.S. Patent No. 9,888,279.

Iyer, S., Chaturvedi, S., & Dash, T. (2019). Image captioning-based image search engine: An alternative to retrieval by metadata. In J. Bansal, K. Das, A. Nagar, K. Deep, & A. Ojha

(Eds), *Soft Computing for Problem Solving. Advances in Intelligent Systems and Computing* (Vol 817, pp. 181–191). Springer.

Ke, Q., Liu, M., & Li, Y. (2017, July 18). Content-based image search. *Google Patents*, 1–29. U.S. Patent 9,710,491.

Kroupi, E., Hanhart, P., Lee, J.-S., Rerabek, M., & Ebrahimi, T. (2016). Modeling immersive media experiences by sensing impact on subjects. *Multimedia Tools and Applications*, *75* (20), 12409–12429. https://doi.org/10.1007/s11042-015-2980-z

Lipscomb, C. E. (2000). Medical subject headings (mesh). *Bulletin of the Medical Library Association*, *88*(3), 265–266.

Manocha, P., Badlani, R., Kumar, A., Shah, A., Elizalde, B., & Raj, B. (2018, April 15–20). Content-based representations of audio using siamese neural networks. In *IEEE international conference on acoustics, speech and signal processing (ICASSP)*. (pp. 3136–3140). IEEE.

Manzoor, U., Ejaz, N., Akhtar, N., Umar, M., Khan, M. S., & Umar, H. (2012, December 10–12). Ontology based image retrieval. In *International conference for internet technology and secured transactions* (pp. 288–293).

Maree, M., & Belkhatir, M. (2015). Addressing semantic heterogeneity through multiple knowledge base assisted merging of domain-specific ontologies. *Knowledge-Based Systems*, *73*, 199–211. https://doi.org/10.1016/j.knosys.2014.10.001

Miller, G. A. (1995). Wordnet: A lexical database for English. *Communications of the ACM*, *38*(11), 39–41. https://doi.org/10.1145/219717.219748

Mukhopadhyay, D., & Sinha, S. (2019). Web-page indexing based on the prioritize ontology terms. In D. Mukhopadhyay (ed.), *Web searching and mining* (pp. 75–84). Springer.

Nazir, A., Ashraf, R., Hamdani, T., & Ali, N. (2018, March 3–4). Content based image retrieval system by using hsv color histogram, discrete wavelet transform and edge histogram descriptor. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1–6). IEEE.

Orlandi, F., Debattista, J., Hassan, I. A., Conran, C., Latifi, M., Nicholson, M., Salim, F. A., Turner, D., Conlan, O., & O'sullivan, D. (2018, November 26–29). Leveraging knowledge graphs of movies and their content for web-scale analysis. In *14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*. (pp. 609–616). IEEE.

Pal, N., Kilaru, A., Savaria, Y., & Lakhssassi, A. (2018). Hybrid features of tamura texture and shape-based image retrieval. In P. Sa, S. Bakshi, I. Hatzilygeroudis, & M. Sahoo (Eds), *Recent findings in intelligent computing techniques* (pp. 587–597). Springer.

Podlesnaya, A., & Podlesnyy, S. (2016). Deep learning based semantic video indexing and retrieval. In Y. Bi, S. Kapoor, & R. Bhatia (Eds), *Proceedings of SAI intelligent systems conference* (pp. 359–372). Springer.

Rinaldi, A. M., & Russo, C. (2018, December 10–13). User-centered information retrieval using semantic multimedia big data. In *IEEE international conference on big data (Big Data)* (pp. 2304–2313). IEEE.

Sanoja, A., & Gançarski, S. (2014, April 14–16). Block-o-matic: A web page segmentation framework. In *International conference on multimedia computing and systems (ICMCS)* (pp. 595–600). IEEE.

Sarwar, A., Mehmood, Z., Saba, T., Qazi, K. A., Adnan, A., & Jamal, H. (2019). A novel method for content-based image retrieval to improve the effectiveness of the bag-of-words model using a support vector machine. *Journal of Information Science*, *45*(1), 117–135. https://doi.org/10.1177/0165551518782825

Shao, C., Hu, L.-M., Li, J.-Z., Wang, Z.-C., Chung, T., & Xia, J.-B. (2016). Rimom-im: A novel iterative framework for instance matching. *Journal of Computer Science and Technology*, *31*(1), 185–197. https://doi.org/10.1007/s11390-016-1620-z

Shrivastav, S., Kumar, S., & Kumar, K. (2017). Towards an ontology based framework for searching multimedia contents on the web. *Multimedia Tools and Applications*, *76*(18), 18657–18686. https://doi.org/10.1007/s11042-017-4350-5

Shrivastava, N., & Tyagi, V. (2015). An efficient technique for retrieval of color images in large databases. *Computers & Electrical Engineering*, *46*, 314–327. https://doi.org/10.1016/j.compeleceng.2014.11.009

Song, Y., Lei, J., Peng, B., Zheng, K., Yang, B., & Jia, Y. (2019). Edge-guided cross-domain learning with shape regression for sketch-based image retrieval. *IEEE Access*, *7*, 32393–32399. https://doi.org/10.1109/ACCESS.2019.2903534

Strezoski, G., & Worring, M. (2018). Omniart: A large-scale artistic benchmark. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, *14*(4), 1–21. https://doi.org/10.1145/3273022

Suchanek, F. M., Sozio, M., & Weikum, G. (2009, April 20–24). Sofie: A self-organizing framework for information extraction. In *Proceedings of the 18th international conference on World Wide Web* (pp. 631–640). ACM.

Tani, M. Y. K., Ghomari, A., Youcef, L. D., Lablack, A., & Bilasco, I. M. (2017, July 18–20). An audio indexing and retrieval approach using a video surveillance ontology. In *Computing conference* (pp. 258–261). IEEE.

Tulasi, R. L., Rao, M. S., Usha, K., & Goudar, R. (2016). Ontology-based annotation for semantic multimedia retrieval. *Procedia Computer Science*, *92*, 148–154. https://doi.org/10.1016/j.procs.2016.07.339

Wang, H., Chia, L.-T., & Gao, S. (2010, March 29–31). Wikipedia-assisted concept thesaurus for better web media understanding. In *Proceedings of the international conference on multimedia information retrieval* (pp. 349–358).

Weese, J., Lehmann, H., Qian, Y., & Ten Kate, W. R. T. (2018, April 24). Accessing medical image databases using medically relevant terms. *Google Patents*, 1–14, U.S. Patent 9,953,040.

Wei, G., Cao, H., Ma, H., Qi, S., Qian, W., & Ma, Z. (2018). Content-based image retrieval for lung nodule classification using texture features and learned distance metric. *Journal of Medical Systems*, *42*(1), 1–7. https://doi.org/10.1007/s10916-017-0844-y