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

IPFJRO: THE DEVELOPMENT OF IMAGE PROCESSING ONTOLOGY

[1]OOI JESSIE, [2]MANSOOR ABDULLATEEF ABDULGABBER, [3]SIAU CHIUN LIEW

Universiti Malaysia Pahang, Faculty of Computer Systems and Software Engineering, Malaysia

Email: jss_ooi@hotmail.com, hakmansoor@ump.edu.my

ABSTRACT

IPFJRO is an image processing-based ontology develop specifically for an ontology-based keywords search system. Ontology ease the human computer communication by presenting the concept or information in a way that can be understand by both machine and human without any misunderstanding. At the same time, it allows the information reuse across different applications. However, despite its advantages, development of an ontology is still a lengthy and complicated process. Hence, not every domain has its own complete ontology developed and ready to use. In this study, we had developed a domain ontology. We had select Image Processing as the domain of our ontology. This ontology was based on the terminologies from research articles and image processing text book. The development of this ontology followed the guidelines in Ontology Development 101. We used Protégé as the development tool. This study focused on the methodology, design and development of the ontology. There's a total of 63 classes and 831 individuals in IPFJRO.

Keywords: *Ontology, Ontology development, Knowledge Representation, Image Processing, Query*

1. INTRODUCTION

Image processing is a method of processing selected images based on mathematical operations by using any form of signal processing [1]. It is a huge research area where a new researcher will be overwhelmed by it if they do not possess the minimum amount of the basic domain knowledge. To discover any information, they needed, most of the user will depends on the search engines due to the huge amount of information available on the internet and the easy access to the said information. However, when the user does not have the appropriate keywords for their search session, their effort may be fruitless. Vocabulary mismatch is one of the most common problem exists especially in search engines especially when dealing with a specific technical domain [2]. This problem arises when an item or a situation was described using different terms by different users. This situation is often encounter by the novice user.

To encounter the problem of the vocabulary mismatched, controlled vocabularies are needed [3, 4]. To resolve this problem, ontology has been developed. Ontology is a collection of formal, machine-process-able and human interpretable representation of the entities, and the relations among those entities, within a defined application domain [5] . The development of the ontology allows the users to "speak the same language", thus they can connect freely despite the ambiguous used terms [6]. Moreover, ontology also able to unite the disparate data from different sources [7]. With the implementation of the ontology, the users not only able to communicate with the machine precisely, they are also able to reuse the data effortlessly. Due to the nature of the ontology that able to communicate despite the ambiguous term, ontology can perform well in a domain specific application. In addition to that, in a field such as image processing that has huge branches in their research area, standardising the terms and area exist would hugely improve the learning curve of a new researcher or students.

While there are several ontologies available online, based on our findings, none of it is image processing based. With the motivation to contribute in the knowledge sharing and reuse in image processing, we had proposed an ontology model to model the image processing knowledge area.

There are 3 main challenges that we identified in this ontology development.

(1) To design a reusable framework that can be applied easily by other related fields.
(2) To develop an ontology that can cover most of the topic in the selected domain.
(3) To develop a domain ontology that can be reuse in different situations or applications.

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

1.1 Research Objectives

This study focused on connecting the gap between different Image Processing topics and create a clear connection between each topic. To achieve this objective, we had developed a domain ontology based on image processing (IPFJRO). This ontology was established based on 3 different elements, "Class", "Individuals" and "Relationship". The classes represent different research field in the image processing and their general methods while the individuals represent applicable area, algorithm and other related information for every classes.

The rest of the study will be structured as below: (2) The discussion of the past literature of ontology development. (3) Discussed the tools used for ontology development. (4) Showcase the ontology development design. (5) Display the methodology for the ontology development. (6) Discuss the result of the ontology development and (7) Summary and future work.

2. ONTOLOGY

Ontology is commonly used in the semantic web and bioinformatics. One of the reasons for the ontology development is the ability of sharing knowledge. Developing ontology is a lengthy and iterative process. However, a well build ontology associate terms with the concepts and relations in the ontology and devise syntax for encoding knowledge in terms of the concepts and relations [8]. It clarified the structure of the knowledge exist within the domain and present them in a structured way. This allows the knowledge to be shared within those who have similar needs in that domain. Another reason for building ontology is their ability to be reused by different situation. This is because of the flexibility of the data structure that can be adapt and changed without much difficulties [9].

Ontology often applied in the biomedical application and semantic web due to its ability to provide a controlled vocabulary and ability for knowledge sharing. In a biomedical field, Gene ontology can be considered as one of the most significant ontology development project. The main purpose of creating this ontology is to create well-structured, controlled, dynamic vocabularies that can be applied across several domains in molecular and cellular biology [15]. Creation of the gene ontology is based on these three organism database: FlyBase [16], Mouse Genome Informatics (MGI) [17] and the Saccharomyces Genome Database (SGD) [18]. Wordnet ontology is a well-known common and biggest open-source ontology that focus on semantic relationship between words [19].

The ability to exchange knowledge within different domains or applications is one of the most important features for an ontology, thus, there are 3 impotent aspect that need to be satisfied. First, ontology must be highly intuitive to the user. Second, ontology must have a well-defined formal semantics with established reasoning properties in terms of completeness, correctness, and efficiency. Last, the ontology must be able to create a proper linkage with the existing web language such as XML and RDF to ensure its interoperability [10]

There is not only one way to develop ontology [11]. Several ontology development methods had been proposed over the years such as Ontology Development 101 [12], and Methondology [13]. However, no matter what the ontology development method is, it should contain these elements [14]:

(1) Classes that represent the concept. These classes should be arranged in a hierarchy of superclass and subclass.

(2) Relations between each concept.

(3) Attribute that will describe the features of the concepts.

(4) Formal axioms to model the sentences that are always true.

(5) Function that represent the special case of the relation.

(6) Instances that represent the elements or individuals in the ontology.

3. TOOLS: PROTÉGÉ

In this study, Protégé has been selected as the development tools for the ontology. It is a free, open source ontology editor and a framework to build the ontology. Protégé was developed by Stanford Centre for Biomedical Informatics Research at the Stanford University School of Medicine [20]. Protégé is supported by a strong community of academician and corporate users with frequent update and support for reasoning tools, which is one of the reasons Protégé has been selected as the development tools for the ontology. Another reason for us to select Protégé as the development tools is due to

Journal of Theoretical and Applied Information Technology
15th May 2018. Vol.96. No 09

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

its full support of the OWL2 Web Ontology Language and RDF specification from the World Wide Web Consortium. Figure 1 shows the interface of Protégé.

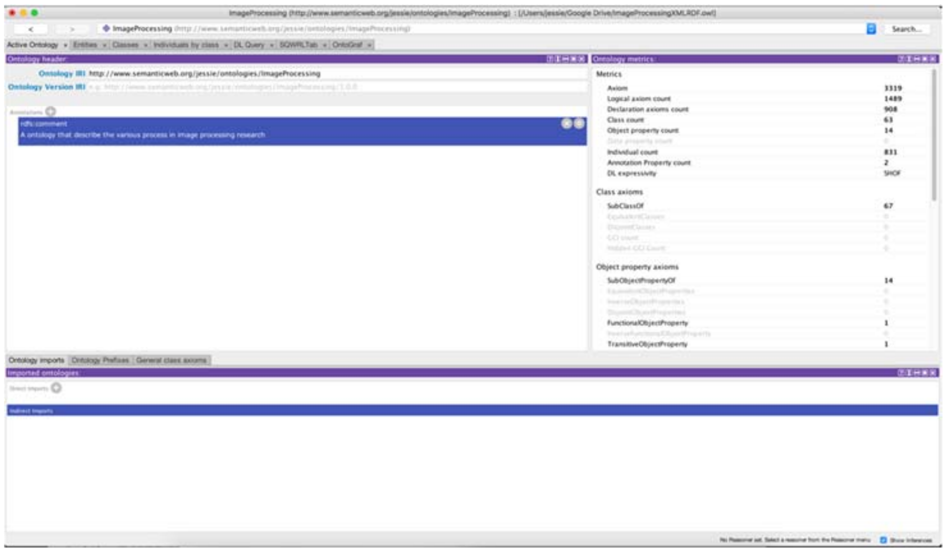

Figure 1 Protégé Interface

4. ONTOLOGY DESIGN

Given that our ontology is a task-based ontology, hence the design of the proposed ontology will revolve around the designated task. The major concern while developing this ontology is how it can provide the most suitable answer for its given question. Therefore, to fit the motive of our ontology, we've decided that, the development of this ontology based on the terminology that are commonly found in a research field will be most suitable for our cause.

The ontology framework can be categorised into 3 major sections: the main category which is the selected domain and categorisation. This will become the superclass of the ontology. The next section is the subclass that based on research fields in image processing. In this section, classes can be categorised based on their attributes such as the process or methods of every parent class. This section can be designed according to the understanding of the ontology developer and how it suits the usage of the ontology or it's applications. The last section is the related information such as the application area, algorithm or related field of the selected sub categories that will become the instances. The framework of the designed ontology is shown in Figure 2 below.

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

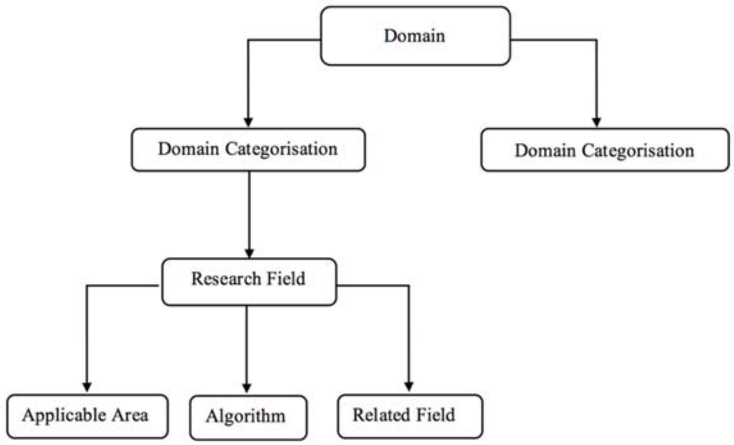

Figure 2 IPFJRO framework

5. METHODOLOGY

The development of the ontology is based on the Ontology 101 guideline [12] and it was created in OWL2 with RDF schema. There are 2 major process in our ontology development. (1) Knowledge acquisition. (2) Formalised the ontology based on our design. These processes can be implemented based on the Ontology 101 guideline.

According on the guideline, the ontology development can be categorised into 7 different steps:

(1) Determine the scope and domain of the ontology.

The main purpose of this ontology is to support an ontology-based keyword search system that will improve keywords selection in the information seeking process for research and journal paper. Hence, this ontology should be able to provide keywords suggestion based on the selected domain. The ontology will also provide the users knowledge that needed to answer the question given by the junior researcher. The ontology also required to provide a clear view of the relationship between different image processing techniques and topics.

(2) Consider reusing the existing ontologies.

After the determination of the scope and domain of the ontology, the possibility of reusing other ontology should be considered as one of the options. Since, building ontology from stretch is a long and exhausting process where it requires a lot of time and manpower to build one. Hence, by reusing the ontology will help to reduce a lot of time and effort in the research. The reusability of an ontology has been determined by the structured of the ontology, their architecture and the scope of the ontology. However, based on our searching effort, at least within our knowledge, there's no suitable domain ontology that can be applied into our designed ontology.

(3) Create a list of important terms of the selected domain.

One of the major steps in developing IPFJRO is the selection of a list of important terms. These terms will be selected as the classes of the ontology which will affect the entire architecture of the ontology. This process was carried out simultaneously with step 4 which is the development of the class hierarchy.

After determining the most important terms, we've started the process of selecting the terms, which are the algorithm, related field and applicable area of each important term. Journal articles and research papers are the main sources for selecting the important terms in our ontology. Articles or papers will be search based on the predetermined terms. The method for the terms collection in this study has been referred to the term selection methods in the corpus dependent knowledge model in query expansion. If the terms belong to the same article, then it will be assumed that these terms belong in the same category.

Every terminology has been categorised based on research papers and these research papers were selected according to the keywords provide in it. If the important terms are part of the keywords in the research paper, it will be assumed the rest of the keywords provided in that research paper can be categorised together.

The collections of the terms are based on the keywords from the selected research papers. For the building of this ontology, around 250 journals and conference papers has been collected. To make sure the basic knowledge of image processing was covered, to the glossaries in the book "Digital Image Processing" has been used as additional material [1]. These terms will be categorised accordingly based on our classes.

(4) Define the classes and the classes hierarchy.

A set of terms will first be determined as the framework of the ontology. These terms will eventually become the classes of our proposed ontology. To fulfil the purpose of the ontology development, we had design a class hierarchy that are able to connect different area of image processing together. These classes are based on different type of image processing and their research field. Moreover, these classes will become the guideline for collecting the terms that will make up our ontology. The selection of these terms were referred from the "Digital Image Processing" by Gonzalez and Woods [1]. In the original guideline of Ontology 101, the important terms will have to be created before defining the classes and classes hierarchy. However, in this study we felt that creating a list of important terms and define the classes and its hierarchy is equally important. Hence, should be carried out simultaneously without a clear definition of which process come first. Figure 3 shows the example of the classes in IPFJRO.

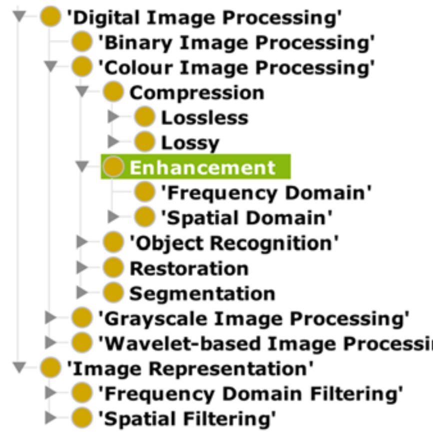

Figure 3 IPFJRO Class example.

(5) Define the properties of the classes.

At this stage, relationship between each class in the ontology will be defined. The relationships between the classes are defined based on the connection between the classes and subclasses. Such as, if the technique can be categorised into several types of process, or the different type of techniques that can be applied to certain types of the images. In general, the first tier can be considered as the general type of the domain knowledge. It will then be categorised further based on the process of each classes. These classes will then have a series of child classes based on their respective methods. Table 1 shows the example of the relationship between classes in the image processing ontology while Figure 4 shows the view of the properties in Protégé.

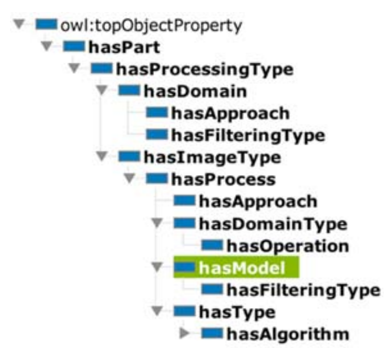

Figure 4 IPFJRO Relationship

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

Superclass	**Subclass**	**Relationship**	**Remarks**
Digital Image Processing	Binary Image Processing	HasImageType	Digital Image Processing has different image format.
	Colour Image Processing		
	Wavelet-based Image Processing		
	Grayscale Image Processing		
Image Representation	Spatial Domain	HasDomain	Image Representation has 2 different domains.
	Frequency Domain		
Spatial Domain	Convolution	HasFilteringType	Convolution and Correlation are filtering techniques of Spatial Domain.
	Correlation		
Frequency Domain	Discrete Fourier Transform	hasTechnique	Frequency Domain has 6 different techniques.
	Discrete Cosine Transform		
	Walsh function		
	Hadamard Transform		
	Slant Transform		
	Karhunen Loeve Transform		
Grayscale Image Processing	Enhancement	HasProcess	Grayscale Image Processing and Colour Image Processing have 5 different processes.
Colour Image Processing	Restoration		
	Compression		
	Object Recognition		
	Segmentation		

Table 1 Example of relationship in IPFJRO

(6) Define the facets of the slots.

In our ontology, the only facets we've included is the definition for the classes. Table 2 shows part of the top tier class's definitions that has been decided in the domain ontology and these definitions will be included as the facets of the ontology for future use.

Table 2 Example of definition for top tier classes

Class	Definition
Frequency Domain	-Image value at image position F represents the amount that the intensity values in image is vary over a specific distance related to F. -Changes in image position correspond to changes in the spatial frequency.
	-Explicit periodic relationships in the spatial domain.
Binary image processing	-Only two possible values for each pixel. -Typically, black and white though any two colours can be used. -Bi-tonal, bi-level, two-level, monochrome, monochromatic, Bitmap. -Pixel is stored as a single bit. -Result of certain operations such as segmentation, thresholding, and dithering.
Enhancement	-Improving the quality of a digitally stored image. -Process of adjusting digital images.

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

	-Manipulating the image with software. -Results are suitable for display or further image analysis.
Image Restoration	-Taking a corrupt/noisy image and estimating the clean, original image. -Motion blur, noise and camera mis-focus. -Reverse the process that blurred the image -Point Spread Function (PSF). Reduce noise and recover resolution loss.

(7) Create individuals for every classes.

In the last step of the ontology creation, important terms that have been collected in step 3 will be categorised based on the classes that had been defined in step 4. All the categorised terms will be considered as the individuals for their respective classes. These individuals will act as the extended knowledge of the classes. Figure 5 shows part of the individuals for "Binary Image Processing" in Protégé.

Figure 5 Individuals for Binary Image Processing

(8) Verification

The last part of the ontology development is the verification of the ontology for their consistency and inconsistency. This process is not included in the original guideline of Ontology Development 101. However, we've felt that it is appropriate to evaluate and verify the developed ontology is without contradiction. A reasoner is a program that deduces the logical consequences from a set of explicitly asserted facts or axioms and usually provides support for reasoning task such as classification, debugging and querying [21]. Hence the reasoner can verify if there are any logical contradiction exist in the ontology.

To verify the ontology, we had used the FaCT++ reasoner. FaCT++ is a reasoner that implements the tableaux decision procedure. It is able to validate any inconsistence classes during the development of the ontology. It incorporates several performances enhancing optimisation such as absorption, model merging, ordering heuristics and taxonomic classification. Hence, could eliminate any contradiction exist in our ontology. Based on the implementation of FaCT++ onto the developed ontology, there isn't any contradiction exists.

Figure 6 shows that there are no indications of any inconsistency of the classes in the ontology.

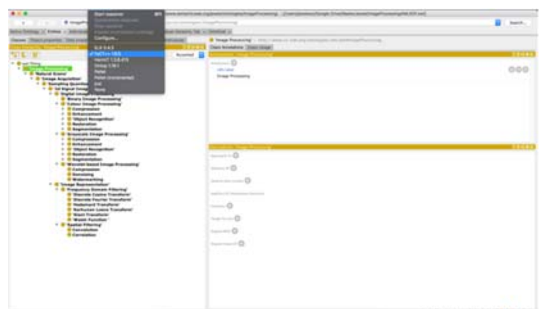

Figure 6 FaCT++ Reasoner

6. RESULT

The development of the ontology is a process of developing an ontology that will unify the research field and the related terms in image processing domain. The ontology was developed based on controlled vocabularies, terminologies and "Digital Image Processing" textbook. In this study, we're able to create a total of 63 classes and 831 individuals in IPFJRO.

The categorization of the classes was according to the level of detail into the image processing research level. In other words, the superclass of IPFJRO start with Image Processing and then go into different sub categories of image processing fields. More than 800 controlled terminologies have been chosen as the sources of IPFJRO. These terminologies have been categories based on their connection to the defined classes. These concepts exist in IPFJRO

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

was defined manually. The core top-level of IPFJRO is the important concept in image processing.

The superclass of the ontology is the main keywords of the selected domain. For example, in this study, the main keyword of our selected domain is Image Processing. It will then be categorised based on the categorisation of the domain type. These categorisations will be the first layer of the ontology classes. The first categorisation of IPFJRO are based on the different type of the image processing.

After the first tier of the categorisation, classes will then be categorised again based on different process carried out on each domain type. The subclasses of IPFJRO consist 63 different classes. Every subclass will be categorised into more subclasses based on different type of categorisation. For example, "Wavelet Image Processing" has two different processes, "Compression" and "De-noising". These processes will become the child class of "Wavelet Image Processing". The last tier of the main structure of the ontology is the methods of each process. In this tier, different methods applicable to each process will be selected as the classes. "Deterministic" is one of the method for "Restoration". Hence, it has become one of the child class of "Restoration".

In every classes, it consists of 2 different attribute, names and concept definition. Names of the classes shows the clear naming of their respective classes. On the other hand, concept definition displayed a clear definition of each classes, and a short introduction about the classes.

The individuals of the image processing type were a series of algorithm, related field and applicable area of each research field. The terms for that will be consider as the individuals were collected from a series of research articles and image processing textbooks. Part of our developed ontology is shown in Figure 6.

Due to the time constrain this is only a small fraction of what could be included in this ontology. The terminology of the ontology can be broadened significantly. Moreover, the development of the ontology is a huge and lengthy process, the knowledge in our domain are mostly mastered in different group of researchers. Hence, the process of knowledge acquisition can be handle by different research groups. Figure 7 has shown the visualisation of our image processing ontology.

Aforementioned, development of an ontology is a lengthy and iterative process. Hence, not every domain has their own specific domain ontology. Based on our knowledge, there is not any Image Processing domain ontology available right now. While every ontology development largely depends on the author view on the said subject, the development of IPFJRO shows one way of how a domain-based ontology can be developed when it is largely based on academic materials and provided an opportunity for knowledge reuse in the future.

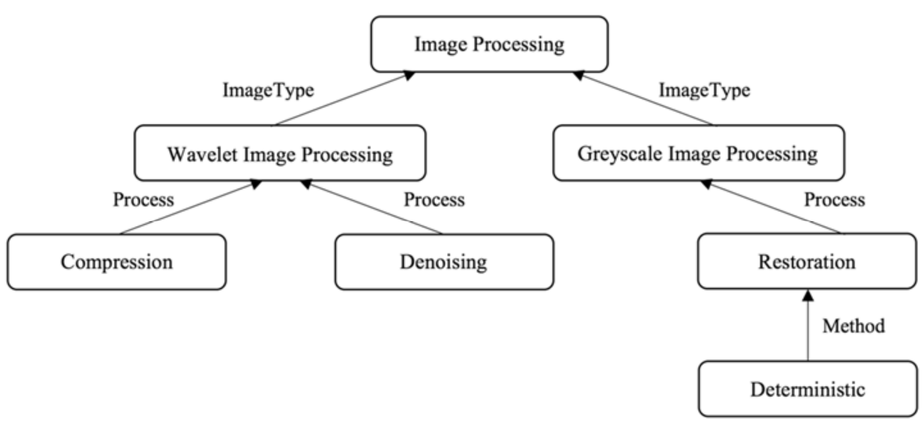

Figure 7 IPFJRO Ontology Architecture

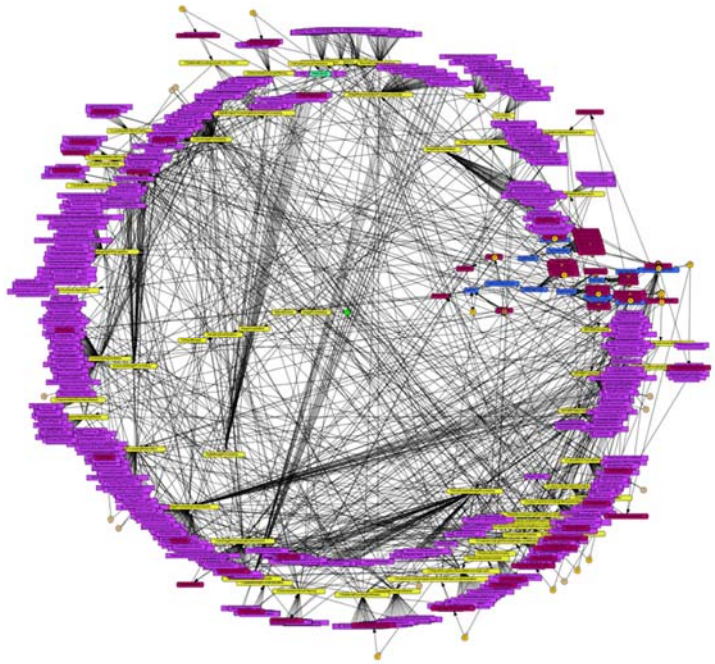

Figure 8 Radical view of IPFJRO

7. SUMMARY AND FUTURE WORK

The objective of developing this image processing ontology is to provide an overview in image processing topics and provide a better understanding about the relationship between each concept. Our developed IPFJRO consists of 800+ terminologies covering part of the image processing glossary. The developed ontology provides a library of terminologies that allows the user to explore the relationship between each term. Moreover, it can be integrated in different systems or area such as knowledge management system or recommendation system. Hence, encourage the knowledge reuse. While the ontology proved a draft overview of the image processing research field, it can be improved by improving the terminologies in the ontology, expand the ontology to have a better coverage in image processing and lastly collaboration with several experts to ensure it covers every expect of image processing.

ACKNOWLEDGMENTS
This research is financially supported by Universiti Malaysia Pahang (UMP) Research Grant No GRS150375

REFERENCES:

[1] Gonzalez RC, Woods RE. Digital Image Processing: Pearson/Prentice Hall; 2008.
[2] Martins F, Haslhofer B, Magalhães J, editors. Query expansion using open web-based skos vocabularies. ACM SIGIR Workshop on Health Search and Discovery: Helping Users and Advancing Medicine; 2013.
[3] Smith B. The logic of biological classification and the foundations of biomedical ontology. Logic, Methodology and Philosophy of Science Proceedings of the 12th International Conference2005.
[4] Shekarpour S, Marx E, Auer S, Sheth A. RQUERY: Rewriting Natural Language Queries on Knowledge Graphs to Alleviate the Vocabulary Mismatch Problem. 2017.
[5] Rubin DL, Lewis SE, Mungall CJ, Misra S, Westerfield M, Ashburner M, et al. National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. Omics : a journal of integrative biology. 2006;10(2):185-98.
[6] Mortensen JM, Minty EP, Januszyk M, Sweeney TE, Rector AL, Noy NF, et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a

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

study of SNOMED CT. Journal of the American Medical Informatics Association : JAMIA. 2015;22(3):640-8.

[7] Mortensen JM, Horridge M, Musen MA, Noy NF. Modest use of ontology design patterns in a repository of biomedical ontologies. Proceedings of the 3rd International Conference on Ontology Patterns - Volume 929; Boston. 2887728: CEUR-WS.org; 2012. p. 37-48.

[8] Chandrasekaran B, Josephson JR, Benjamins VR. What are ontologies, and why do we need them? IEEE Intelligent Systems and their applications. 1999;14(1):20-6.

[9] Jing D, Yang H, Shi M, Zhu W, editors. Developing a research ideas creation system through reusing knowledge bases for ontology construction. Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual; 2015: IEEE.

[10] Fensel D, Van Harmelen F, Horrocks I, McGuinness DL, Patel-Schneider PF. OIL: An ontology infrastructure for the semantic web. IEEE intelligent systems. 2001;16(2):38-45.

[11] Brusa G, Caliusco ML, Chiotti O, editors. A process for building a domain ontology: an experience in developing a government budgetary ontology. Proceedings of the second Australasian workshop on Advances in ontologies-Volume 72; 2006: Australian Computer Society, Inc.

[12] Noy NF, McGuinness DL. Ontology development 101: A guide to creating your first ontology. Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880, Stanford, CA; 2001.

[13] Fernández-López M, Gómez-Pérez A, Juristo N. Methontology: from ontological art towards ontological engineering. 1997.

[14] Bermejo J. A simplified guide to create an ontology. The Autonomous Systems Laboratory. 2007.

[15] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.

[16] Gramates LS, Marygold SJ, Santos Gd, Urbano J-M, Antonazzo G, Matthews BB, et al. FlyBase at 25: looking to the future. Nucleic acids research. 2017;45(D1):D663-D71.

[17] Eppig JT, Smith C, Blake JA, Ringwald M, Kadin JA, Richardson JE, et al. Mouse Genome Informatics (MGI): Resources for Mining Mouse Genetic, Genomic, and Biological Data in Support of Primary and Translational Research. 2017.

[18] Costanzo MC, Engel SR, Wong ED, Lloyd P, Karra K, Chan ET, et al. Saccharomyces genome database provides new regulation data. Nucleic acids research. 2014;42(D1):D717-D25.

[19] Miller GA. WordNet: a lexical database for English. Communications of the ACM. 1995;38(11):39-41.

[20] Protege About: Stanford Center for Biomedical Informatics Research; 2016 [Available from: http://protege.stanford.edu/about.php.

[21] Dentler K, Cornet R, Ten Teije A, De Keizer N. Comparison of reasoners for large ontologies in the OWL 2 EL profile. Semantic Web. 2011;2(2):71-87.