Abstract—The semantic gap problem existing in Content Based Image Retrieval (CBIR) remains a challenge due to the failure of unification between low-level image features and high-level human perception. User studies revealed that users tend to seek information using high-level semantics description rather than using quantitative approach. Therefore, ontology as a semantic technology might be a solution since the ability of ontology can be employed to define the high-level concepts. The objectives of this paper are to describe the implementation of the multi-modality ontology using Protégé as semantic image annotation tool and to show how the desired images are retrieved using text-based image query interface (TBIQI) without having to learn the ontology query languages i.e. SPARQL Protocol and RDF Query Language (SPARQL).

Keywords—semantic-based image retrieval; multi-modality ontology; SPARQL; text-based image query interface; semantic gap

I. INTRODUCTION
The factor of technologies advancement for internet, ubiquitous electronic devices and computer nowadays makes the amount of digital images collection grows rapidly. Therefore, without appropriate tool it will make image search and retrieval challenging. According to Rui et al. [1] we cannot retrieve or make use of the image information unless it is well organized and structured so as to allow effective searching, browsing and retrieval of images.

The early research in image retrieval (IR) area began with textual-based image retrieval (TBIR) approach [2]. In TBIR, user need to enter the textual-based query or related keyword to search the images and it will return all possible images which match between the images description and the query. However, the mismatched problem always occurred when there is incorrect description about images. For example, you may find human picture while you are searching for herbal medicinal plant (HMP) e.g. “tongkat ali”. Therefore the major disadvantage of TBIR is polysemy problem where the same word can be used for different images.

The content-based image retrieval (CBIR) research then replace the TBIR research based on extracting and indexing of low-level image feature such as colour, shape and texture [3] automatically using machine e.g. computer. To perform the image query in CBIR, the user needs to have sample in hands or sketch of an image. However the problem arise when the user don’t have image in hand. They could not implement the query and this is become one of the disadvantages in CBIR. A major weakness of CBIR, it fails to mimic how human understand image. Hence, an issue of a semantic gap exists. As revealed by Smeulders et al. [4] semantic gap is defined as “the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation”. Humans have a tendency to exploit high-level concepts, such as keywords and text descriptors to deduce images and measure their similarity [2], however CBIR unable to represent the semantic image in a qualitative manner. Thus, there is a gap between the way of human understanding the images qualitatively and the way of machine interprets the images quantitatively. Therefore the multi-modal ontology (MMO) which is a semantic technology is proposed in this paper to reduce the semantic gap between the visual features and the richness of human semantic. The MMO integrates the textual information which represents the high-level concepts and visual feature which represents the low-level feature.

The ability to provide an explicit domain-oriented semantics in terms of defined high-level concepts for semantic image annotation in image retrieval is an advantage of a semantic technology like ontology. The ontology are able to describe visual content using well-structured concepts and relationships that are not only machine-readable but also human readable and meaningful.

In this study, the HMP has been chosen as a case study since there is a significant demand of people around the world to consume it. World Health Organization (WHO) [5] stated that between 70% to 80% of many developed countries has used some form alternative or complementary medicine for their health care. Furthermore, 80% of the population some Asian and African country depend on the HMP as primary health care. These high percentages portray the crucial for the necessity of semantic descriptions for HMP images in order to cater for such diverse users’ information needs e.g. entrepreneur, researcher, government, public, pharmacologist, botanist, pharmacist and practitioners [6][7].

II. MULTI-MODALITY ONTOLOGY
The high-level architecture for the proposed MMO is depicted in Figure 1. It consists of four main components. The Semantic-Based Image Retrieval (SBIR) systems components are: a textual description and visual feature collection from corpora, an information extraction
component, a multi-modality ontology construction component and a semantic image retrieval module component. A detail explanation of MMO architecture can be found in Sulaiman et al. [8].

In this section, the methods for image annotation, taxonomy structure and describing the resources of MMO for HMP are described. We have created the MMO from 60 species and Table 1 depicts 10 of the overall HMP species.

![Figure 1. General architecture of MMO for SBIR](image)

There are more than 180 images collected from three sources which are Malaysian Herbal Monograph volume I [9], Malaysian Herbal Monograph volume II [10] and Malaysian Herbal Monograph volume III [11]. According to our finding, there are six plant morphologies that can be used as medication source which are a whole plant, leaf, flower, rhizome or stem, fruit or seed and root or tuber.

![Figure 2. Linking image to MMO](image)

### Table 1. Sample of 10 HMP Species

<table>
<thead>
<tr>
<th>No.</th>
<th>Vernacular Name (Malay)</th>
<th>Scientific Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Akar cerita, hempedu</td>
<td>Andrographis paniculata</td>
</tr>
<tr>
<td>2</td>
<td>Capa, cepa, sembong,</td>
<td>Blumea balsamifera DC.</td>
</tr>
<tr>
<td>3</td>
<td>Gelenggang besar</td>
<td>Cassia alata L.</td>
</tr>
<tr>
<td>4</td>
<td>Kunyit</td>
<td>Curcuma longa Auct.</td>
</tr>
<tr>
<td>5</td>
<td>Tutup bumi, tapak</td>
<td>Elephantopus scaber L.</td>
</tr>
<tr>
<td>6</td>
<td>Bunga cengkeh</td>
<td>Euginea caryophyllata</td>
</tr>
<tr>
<td>7</td>
<td>Tongkat ali, penawar</td>
<td>Eurycoma longifolia Jack</td>
</tr>
<tr>
<td>8</td>
<td>Gandarusa</td>
<td>Gendarussa vulgaris Nees</td>
</tr>
<tr>
<td>9</td>
<td>Bunga lawing</td>
<td>Illicium verum Hook</td>
</tr>
<tr>
<td>10</td>
<td>Cekur</td>
<td>Kaempferia galanga Linn.</td>
</tr>
</tbody>
</table>

### A. Semantic Image Annotation

There are a numbers of semantic image annotation tool that can be used. Image annotation can be defined as a process of mapping the unknown image to one of a known class [12]. Figure 2 shows the snapshot of semantic image annotation using Protégé. For each particular image inherits the class properties and relationships of its assigned class. Our approach is based on the idea of image annotation using MMO. The high-level concepts and low-level visual features are incorporated in order to improve the richness of metadata vocabulary so that can enhance the retrieval accuracy.

### B. Terminology Box

T-Box is an acronym of terminology box which refers to concept or class definition which comprises a set of terminology axioms. The combination of concept or class, object properties and data properties form the T-Box [13]. In HMP ontology domain, the knowledge of plant description, external morphology of herbal plant, traditional uses and visual features such as colour, texture and shape are modeled. All of these metadata are modeled based on monographs, domain expert experiences and plant identification terminology glossary book. An example of T-Box can be referred to Figure 3. In HMP ontology, there are 61 concepts consists of 2 superclasses and 59 subclasses have been identified to implement the MMO. We have categorized the owl:Thing into two major divisions of ontology which are TextualConcept class and VisualConcept class.
C. Class taxonomy structure

The TextualConcept class represents the textual description information of metadata while the VisualConcept class represents the visual feature information of metadata. Under the TextualConcept class there is only one class namely as Plant. Under Plant class there are seven subclasses which are Synonym, TraditionalUses, PlantDescription, Herb, Character, Definition and VernacularName classes. Under PlantDescription there are six subclasses and two subclasses for Character and VernacularName class respectively. The Leaf, Stem, Fruit, GeneralHabit, Root, Flower are subclasses of PlantDescription while Taste and Odour are subclasses of Character. English and Malay are subclasses of VernacularName. There are 19 subclasses under the lowest level of classes in this MMO. The first nine classes are Arrangement, Venation, LeafHabit, Attachment, LeafPart, Bases, Margin, Division and Apices which are subclasses of Leaf. The second four classes are StemHabit, StemForm, StemPart and StemType which are subclasses for Stem. The third three classes are FruitPart, FruitHabit and FruitType which are subclasses for Fruit. The forth two classes are FlowerPart and FlowerHabit which are subclasses for Flower and the last class is RootHabit which is subclasses for Root.

D. Web Ontology Language

Web Ontology Language (OWL) properties represent relationships between two individuals [14]. There are two main types of properties involved in implementing HMP ontology. Such properties are object properties and datatype properties. Object properties link between an individual to an individual. We have identified 46 object properties can be used in this MMO. Such properties are hasHerb, hasSynonym, hasLeafHabit, hasLeafPart, hasLeafMargin, hasLeafAttachment, hasFruitHabit, hasFruitHabit, hasDefinition and hasVernacularName.

E. Assertion Box

A-Box is an acronym of assertion box which refers to individual or instance definition. Combination of individual or instances and relationship form the A-Box. An individual is a particular realization “object” of class which representing a real entity in the real world [13]. There are 1114 individual or instances for this HMP ontology. Figure 3 depicts the correlation between T-Box and A-Box whereby Image_vol1_10_1 and ColourFlower_10 are example of individual or instances. The hasColourFlower is an object property and Image, Colour and ColourFlower are classes. We can see that an individual Image_vol1_10_1 has a relationship hasColourFlower with individual ColourFlower_10.

F. Resource Description Framework (RDF)

RDF is a framework for metadata that provides interoperability between applications that exchange machine understandable information on the web. RDF is not a language but a model for describing resources in a decentralized open-world assumption. RDF is a general method to decompose any type of knowledge into small pieces in the form of subject-predicate-object expressions. These expressions are known as triples in RDF terminology. Every RDF descriptions can also be represented as a directed labeled graph (a semantic network) that is equivalent to RDF statements [15]. Figure 2 depicts that four oval-shaped dotted lines parts which are class for Image and the individual or instances Image_vol1_10_1, ColourFlower_10 and c:\20 spesis vol 1\10 cekur (kaempferia galanga Linn.)\flower.jpg can be portraying as triples as depicted in Figure 4. The rdf:domain is used to infer subject membership while the rdfs:range is used to infer object membership. More than 3384 triples have been constructed in this MMO.

Once all the triple has been completed, then the RDF/XML code can be generated. Figure 5 and 6 show the example fragmentation of RDF/XML code for MMO triple in OWL format. More than 5346 lines of RDF/XML code has been generated for this MMO.
III. TEXT-BASED IMAGE QUERY INTERFACE

The usability of MMO need to be validated once the MMO has been constructed as a knowledge repository that stores metadata about high-level concepts and low-level features. To perform this task, we can pose queries to retrieve the related images. In Protégé there is a special function for ontology developer to validate their ontology by using SPARQL query panel. They need to enter the ontology language query syntax and then execute it to retrieve the required images. However, it is hard for end user to understand the SPARQL syntax before being able to perform the query. Therefore, in this section, we present the TBIQI as a convenient access mechanism for them. The main significant of using TBIQI is that the users able to retrieve the desired HMP images from MMO without having to master the SPARQL query knowledge.

A. SPARQL

The SPARQL which pronounced as “sparkle” is an acronym for SPARQL Protocol and RDF Query Language. SPARQL was adopted by World Wide Web Consortium (W3C) as the means to access and query ontologies built using RDF [16]. The SPARQL queries match graph pattern against the target graph of the query. The patterns are like RDF graph, but may contain named variables in place of some of the nodes (resources) or links/predicate (i.e. properties).

B. Image Retrieval Using TBIQI

The retrieval process of image using TBIQI can be referred from Figure 1. It begins with user entering any text-based query inside the text-box provided in the interface. The semantic image retrieval module then will process the input entered, execute the query construction and trigger the SPARQL ontology query language to retrieve a set of answers that satisfies the query expressions. In order to process the SPARQL query and to find all possible answers, we employed the Jena API embedded with ARQ query engine. In TBIQI, we used Liferay platform as an enterprise open source web based interface to receive the input, indexing the images and to display a set of matching images result to the user.

An example of text-based user query is “cough” and the result can be depicted in Figure 7. The result displays 5 HMP images that related to “cough” query.

In semantic image retrieval module, the SPARQL query will be generated automatically using TBIQI as shown in Figure 8.

The Keyword means the text-based input entered by user. The PREFIX means the abbreviations of Uniform Resource Identifier (URIs) which is a string of characters used to identify a name of a resource. The SELECT query expression identifying what information should be return from the query that are bound to the variable ?image, ?imageEntryId, ?x and ?label. The WHERE clause specifying what is the query pattern in underlying dataset. In Figure 8, we have a set of four graph patterns:

- ?x rdf:s:label ?label; is used to retrieve the object properties ?x that has label ?label
- ?image rdf:type plant:Image; is used to retrieve instance (image) which is type Image class
- ?image plant:imageEntryId ?imageEntryId; is used to retrieve instance (image) that has index identity number ?imageEntryId
- ?image ?prop ?x; is used to retrieve instance (image) that has object properties ?prop
The FILTER clause instructs the query processor to only pass along triples that meet a certain condition and normally it will be used with `regex()` function that checks for string matching a certain pattern. Accordingly, FILTER `regex(?label, "\bcough\b", "i")` is used to retrieve label that exactly contain the word “cough” only in MMO which is case insensitive mode (“i”).

C. Validation query result TBIQI with Protege

To validate the TBIQI query result, we manually entered the similar SPARQL query that automatically generated in TBIQI to SPARQL query panel in Protégé as shown in Figure 9. The result can be depicted in Figure 10.

IV. CONCLUSION AND FUTURE WORK

In this paper we have presented the implementation of MMO for HMP as a successful unification between low-level image features with high-level human perceptions in order to reduce the semantic gap in CBIR. In addition, we have also provided the TBIQI as a convenient access mechanism for end user to retrieve the images without having to learn SPARQL ontology language technically. To prove that our TBIQI can retrieve the image accurately, we benchmark the result with SPARQL query panel in Protégé. The comparison result between TBIQI and Protegé had shown that almost 100% similar.

For the future work, we plan to enhance the TBIQI to be a natural language image query interface (NLIQLI), where the user can enter then human-like query. Therefore, some improvements in semantic image retrieval module such as tokenizer and eliminating the stop words are needed.

ACKNOWLEDGMENT

The author would like to thank Universiti Teknologi MARA (UiTM) Malaysia and the Ministry of Education (MoE) Malaysia for providing the research grant (Grant No. RAGS/2012/UiTM/SG05/1).

REFERENCES


