Binding Semantic to a Sketch Based Query Specification Tool

Shahrul Noah and Saiful Sabtu
Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Malaysia

Abstract: *In image retrieval systems, user information needs is expressed using multiple types of query. Unfortunately, due to user subjectivity perception to visual features and semantic depths of images, the conventional query submitted to the system encounter difficulties to identify user information need. The blooming of interest in semantic image retrieval requires current research direction to be more concerned into semantics. This paper describes our on-going research work in formulating a new query approach for image databases. The enabling technologies of the semantic web formed the building blocks of our query specification. Using the MRML as the communication protocol, XML in the form of SVG as the visual content description (sketch based) and RDF for binding the semantic meanings of images, this research provides an initial framework towards semantic based query formulation framework.*

Keywords: *Image retrieval system, semantic retrieval, multimedia retrieval markup language, semantic web*

Received March 19, 2007; accepted November 16, 2007

1. Introduction

The amount of multimedia data, particularly images have been increasing. On August 2005, Google alone had indexed almost 2.2 billion images on the web and that was an increased from 1.3 billion images reported a year before that. This vast amount is only considering those images available on the net, not including personal collections, pictures archives and photo CD-ROMs. As a result, there has been a great demand for solutions in the form of image retrieval systems or search engines that can organize and provide easier access to this massive volume of data.

Image retrieval solutions have been the hot pursuit of two research camps [5]; the computer vision group which believes that the best way of retrieving images, is through a content-based approach (via visible information such as color, texture and shape) and the database management group, whereby images are annotated and retrieved using the conventional text-based approach. However, issues regarding the semantic of multimedia data in image retrieval systems have particularly received little attention. With the advance semantic web technology and a language for describing two-dimensional graphics in XML called Scalable Vector Graphics (SVG), semantic annotation and querying of multimedia data can be significantly addressed.

In an image retrieval system, user information need is expressed using multiple types of query. Unfortunately, due to user subjectivity perception to visual features and semantic depths of images, the conventional query submitted to the system encounter difficulties to identify user information need. For example, query by sketch and query by example (image) do not represent semantic content of the targeted image. Query and matching are done using images visual features only. Meanwhile, query by text annotation faces the ambiguity of user description that poses difficulty for natural language processing techniques. Different mappings and lack of uniform correlation between system's annotated images and user's own constructed annotation in the query pose problem for image retrieval systems. The blooming of interest in semantic image retrieval requires current research direction in image retrieval system to be more concerned into semantics. Currently, vast querying techniques are used to accommodate user query to different retrieval systems. For querying to an image retrieval system based on text, the query item consists of metadata, keywords and textual annotation, which describe the image. For image retrieval system based on visual features, queries fall under the query by sketch and query by example types. Category browsing is used for both systems depending to the database size and the respective system classification abilities.

This paper presents our on going work in developing a semantic image search engine. Our focus at the moment is on the query specification tool. Therefore in this paper we only report on the progress of developing the semantic query specification tool. We proposed a query specification framework that will accommodate effective retrieval of images using query by sketch. We model a query format that is interoperable, reusable and semantically rich for submitting synthetic query image constructed using sketch. Technologies that form the building block of semantic web or emerged from the semantic web

vision such as XML, RDF, XSLT, MRML, SVG and DOM all contributed to our research work.

2. Problem Statement and Motivation

Image retrieval systems, as derived from its larger set, the multimedia information systems, inherit the same trait as a multidisciplinary field from computer science. The ever growing knowledge comprises elements from database management, signal processing, computer vision, natural language processing, networking and human computer interaction [20]. Most of the early image retrieval systems such as the QBIC [2], Photobook [19], VisualSEEK [21] and Blobworld [4] attempt to adapt human perceptual capabilities [23] by using various features extraction models and spatial information to represent images. Techniques from pattern recognition fields were also favored by many early approaches in image retrieval systems [1, 6]. The main notion was to analyze the image and extract perceptually salient visual features such as color, shape and texture, which will be numerically represented for indexing purpose. These image representations will be mapped to point in a multidimensional feature space where spatial access methods (SAMs) will be used to locate other points close to it. The general idea was that closer points meant higher similarity value from the human perception. The distance between points is known as similarity distance.

These systems, however, did not pay great attention on how human perceptions about images can be formulated in the form of query. As a result a number of sketch based approach in query formulation have been proposed [3, 7, 16, 22]. While the sketch-based query specification approached has the capacity of capturing human perception about images, they still lack the semantic elements in image retrieval. Enser *et al.* [10] proposed that the semantic based image retrieval can be achieved through the combination of textual and visual based methods.

Graham and Eakins [11] have pointed out that among the main issues in the development of image retrieval systems are the needs for systems to "understand" user requirements of images and the needs for query matching based on human similarity judgment. Query formulation is the main mechanism to express user needs but ironically it is one of the key problems in Information Retrieval (IR) field [15]. In IR, query is defined as an object of surrogating information need within its model [24]. Hofstede *et al.* [15] argued that ad-hoc queries burdened the user as it fails to manage the richness of information in the retrieval system. It overwhelmed the user, thus misleading them to be lost in conceptual space. According to Hearst [12, 13], to formulate a query, a user must select collections, metadata descriptions, or information sets against which query to be matched, and must specify words, phrases, descriptors, or other kinds of information that can be compared to or matched against the information in the collection. The majority of current approaches, however, only support their own native format of query and understand only a limited set of features use for image matching pre-defined by the developer. Users don't have the freedom to express their semantic needs of the required images and the capacity to providing image descriptions is very limited. Furthermore the available query format restricts the user in forwarding the constructed query to other available search services.

3. The Approach

The axiom of our approach circulates around the subject whereby to increase the understanding of user query submitted to the system. In addition to this, the query also has to be in a portable format to allow interoperable usage by different heterogeneous retrieval systems.

To make computers "understand" what we need, we have to equip them with adequate knowledge [18]. Therefore, the semantic meanings of images must be inherent to computers. It is inevitable that future image retrieval systems are required to incorporate multiple query understanding abilities in order to cater the increasing and diverse information need by users. The semantic web vision fits nicely these requirements. The semantic web vision is about creating smart data for machines to process, and the XML technology is the cornerstone of the semantic web [8]. Therefore, the enabling technologies of the semantic web (particularly XML) have been the building blocks of our proposed research work. Our approach, however, does not try to provide a generic universal query language for all image retrieval systems. The research focus is to generate a simple query document that is portable enough to be used against two most generic image retrieval systems as illustrated in figure.

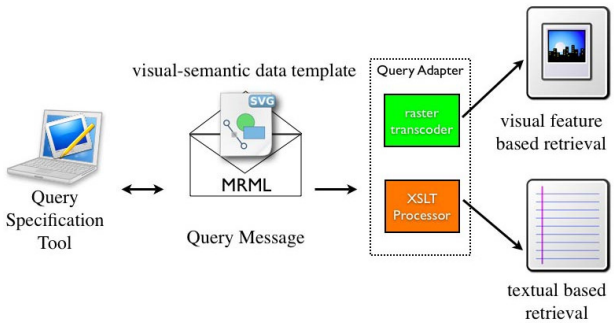

Figure 1. Proposed framework for querying to multiple retrieval systems.

We chose to apply freedom of expression as basis of the query. Thus query by sketch is selected as a foundation of the query format. Sketching allows the user to project the targeted image. This "mental image" is then drawn to an abstract piece of canvas. User sketch will provide content-based features like color,

shape texture and spatial information. User will also be supplied with a set of tool to group components and explains the meaning of the drawing by adding semantic binding which is almost similar to textual annotation. The proposed framework is further described in the following section.

3.1. The Query Specification Format

As shown in Figure, the query itself comprises of an XML document, formatted in SVG syntax. It is wrapped in a Multimedia Retrieval Markup Language (MRML) message queue. A query adapter is needed to act as a cartridge to adapt to different retrieval systems, either textual based system or visual based system. We defined a textual based image retrieval system as a system which used text-based query, such as metadata information, keyword or annotation as a mean to infer user requirements of images. Conversely, visual based retrieval system refers to any system that accepts images in any format (such as PNG, JPEG, GIF, *etc*) as a query item. Since the user query is in SVG format, it can be handled by the query adapter for submission to these different heterogeneous databases. SVG retains the structural integrity of the user's drawing, along side with any data binding placed on each of the object drawn. SVG provides the mean to transfer rich information to databases, provided that the drawing and notation followed some form of structured scheme.

This query format can be applied to any compliant image databases, be it a raster or a vector image collection, as long as it provides a query specification tool using sketch as an input. Instead of searching for the right images using a query seed, as implemented in query-by-example approach, it is easier for the user to describe the required image to the system by drawing it, and possibly embed semantic values to it for those systems that have the capability to process such request. Such an approach of querying is potentially significant as the number of sketch-based tools for Content Based Image Retrieval (CBIR) is currently increasing; as well as the availability of drawing input devices such as mouse and tablets for microcomputers; and stylus for PDAs and smart phones. The success of this approach mainly resides within the implementation and adaptation of the current XML technology (SVG) into CBIR field. With the advent and wide acceptance of XML technology, the Web community gets another boost when SVG Specification (SVG - a description language for image based upon XML) was introduced. It features all the advantages XML has to offer and it renders images of the Internet with sizes smaller that any format with standardize color palette to ensure consistent color rendering across output devices. While providing such quality, we must not forget it is actually text based, making any XML parser able to read it. Furthermore, industry support is wide as it conforms to the standard by W3C.

3.2. Adding Semantic with Data Binding Method

Queries submitted by the user will consist of strokes corresponding to the silhouette of the object and the color composition the user is looking for. In a simple sketch-based query, whereby shape as the main visual feature, queries are considered as objects constructed using combinations of lines, curves, polygons *etc*. Therefore, by considering the visual nature of such queries, vector based format are the most suitable representation to express these queries [9]. Currently, there are several choices for such picture format. Among others are Adobe Illustrator (AI), CorelDraw (CDR), Computer Graphics Metafile (CGM), DXF (CAD Software) and SVG. SVG is a W3C recommendation for describing image by using XML. Since among the aim of this research is to assist in terms of interoperability among systems and bridging the semantic gap, XML based approach seems to offer some promising solutions. Hence, with every stroke the user created, it will be generated to SVG-like form regardless of how the original sketch is being drawn. Semantic binding tags will be made available for user who wants to describe more on the content meaning. For an even more detail on the non-visual content, Dublin core based metadata tagging can be added.

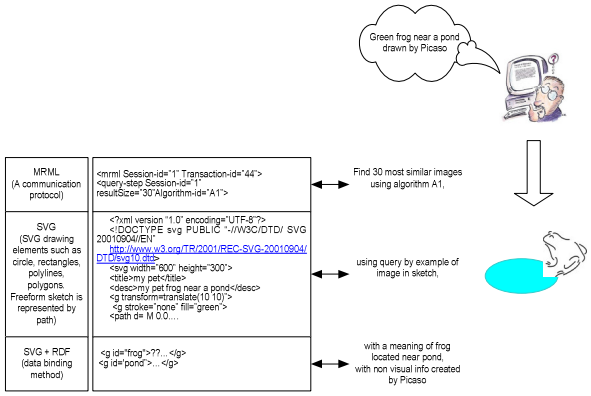

Figure 2. Proposed building blocks for the query specification framework format.

Figure illustrates how the user mental image is being described and mapped to the query specification format consisting of the MRML, SVG and RDF formats. The query will suffice on query structure such as "Find pictures using algorithm a1, returning 30 top close matches, with visual content containing shape and color as sketch1, with the meaning of frog in pond, with non-visual info created by Picasso in 1800". The corresponding encapsulation is shown in Table 1.

4. Application Design

The development of the query based specification tool encircle around the XML-based technology such as the MRML which was developed at the University of Geneva. MRML is used as a communication protocol

between a retrieval engine and a user's client software. As previously mentioned, the other main XML applications used are SVG, DOM, RDF and eXtensible Stylesheet Language Transformation (XSLT).

Table 1. Corresponding encapsulation for an example query.

Query Sentence	Enabling Format
Find pictures using algorithm a1, returning 30 top close matches	MRML
with visual content containing shape and color as sketch1	SVG
with the meaning of frog in pond	SVG
with non-visual info created by Picasso in 1800	RDF tag using Dublin Core metadata

The prototype software was written in java. From the technical aspect, we chose java version 1.4 as the minimum requirement and development was done on eclipse integrated development environment. Java was chosen as it provides many functions needed through its core Application Programming Interfaces (API) and access to third party APIs is plentiful. We largely depend on Apache (xml.apache.org) and Sourceforge (Sourceforge.net) online repositories for API libraries. Figure explains the layered approach taken to develop the prototype software.

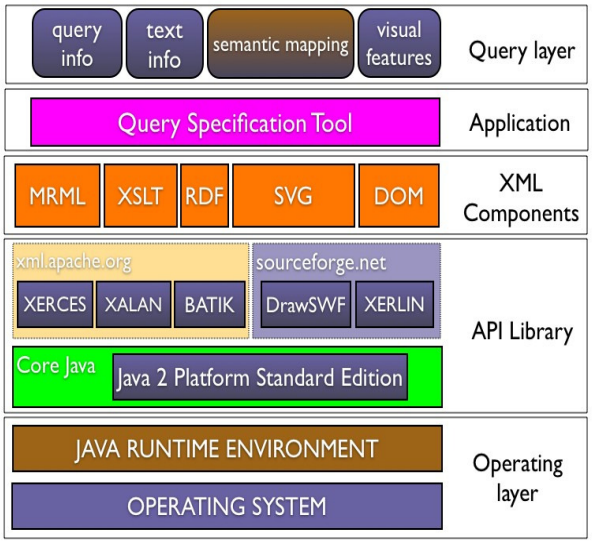

Figure 3. The software layer.

At the bottom of the architecture, a java runtime environment is required to run the developed application. The prototype software uses many APIs from sun's core java API, such as swing for GUI creation *etc*. We also used libraries from Apache, which are Xerces XML parser, Xalan as XSLT processor and batik. Batik is a toolkit for applications or applets which need to use images in the SVG format for various purposes, such as viewing, generation or manipulation. We also used Xerlin Panel, a component in Xerlin project, which can be integrated to any Java Swing application for creating and editing XML files. Last but not least, we used a strip down version of DrawSWF application, as the main component in the application layer for drawing and creating a SVG document.

The query layer encompasses all the information carried in the query message, which is the query information (provided by MRML message queue), textual and visual feature information (provided by SVG drawing and elements), and the semantic mapping between sketched object, which is provided implicitly through structured sketching and data binding strategy.

The application layer holds the prototype software developed (see figure for the inner working of the application layer). It consists of a sketching tool, SVG generator, a property editor and a query manager. Query adapter is a standalone application code that is coupled closely to the application layer. Query adapter comprises of a raster transcoder and XSLT processor tool. The query adapter enables query transformation to both textual and visual (raster) based retrieval engines.

The raster transcoder is developed using apache batik transcoder package. The package provides transcoders for transcoding a SVG document fragment to raster image. The raster image allows it to be submitted to any visual-based retrieval systems. The XSLT processor tool is based on the work of Herman and Dardailler [14]. They developed a linearizer tool, which reads SVG document, and gathers its structure and metadata or RDF association, combines each RDF resource with the textual description, and possibly extract further information from the SVG file itself. It then produces a simple HTML file as its output. According to Herman and Dardailler [14], the idea of a linearizer tool is to extract information from the SVG file using the included metadata and to produce a textual (i.e., "linear") description of the information content of the graphics. This approach was initially explored by Lovet and Dardailler [17].

Based on figure, the corresponding description on the application layer tool are as follows:

- The user projects its mental image by drawing on the canvas using the sketching toolbox provided by the Sketching tool.
- The user sketch will then be transformed to SVG format by the SVG generator. Data binding for each sketched element will be provided using a contextual menu as the user finished drawing a specific object.
- As the user sketch is drawn using a structured approach, and later transformed into SVG format, it eventually conforms to a structured template which permits a semantic-like retrieval. The template acts as a visual and semantic data placeholder and each

of its element content is editable through DOM tree node.

- To edit object property, the user will use the property editor that lists all objects defined in the sketch. Access to each property node will be handled by the DOM editor component.
- Finalize sketch will then be submitted to the query manager which encapsulates the sketch to a MRML message queue. A query document propagates the query message to the query adapter component.
- The query adapter act as a software cartridge to propagate the query output to various MRML compliant retrieval engine. It consists of a raster transcoder that provide raster image exporting feature for visual based retrieval (i.e., query by example), and a XSLT processor to process the user's query to a textual based retrieval engine.
- Currently, there are limited numbers of retrieval engines that complies the MRML proposed architecture. Since MRML is based on XML messaging approach, it is fairly simple to develop a front-end application for legacy image retrieval engines so that it can parse XML based message and process the query.

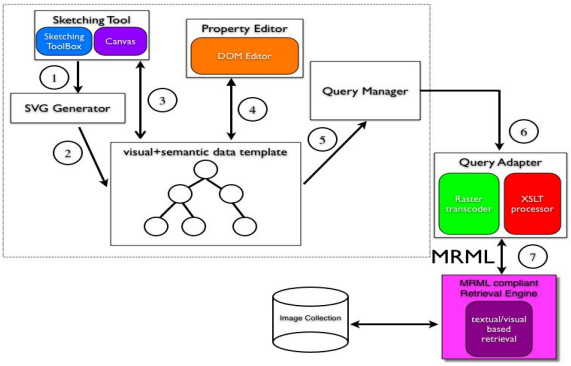

Figure 4. Application tool components.

Figures 5 and 6 illustrate an example of how a user may describe the mental image by means of sketch as well as binding the semantic elements within the query image. Appendix 1 on the other hand provides and example on the representation of sketches in terms of SVG+RDF and textual-based output.

5. Conclusion and Future Work

This research is heading towards providing image retrieval system that can digest queries from interfaces. It allows mental images to be described by means of sketches and semantic information to be bound implicitly to further described the mental images. The enabling technologies of the semantic web particularly XML and RDF formed the building blocks of our research work. Using XML in the form of SVG, provides elements of interoperability among systems and manages to bridge the semantic gap. In this sense, every stroke of sketch the user has created will be generated to SVG-like form regardless of how the original sketch is being drawn. RDF allows the provision of semantic meanings apart from the sketch images, which provides a richer semantic querying.

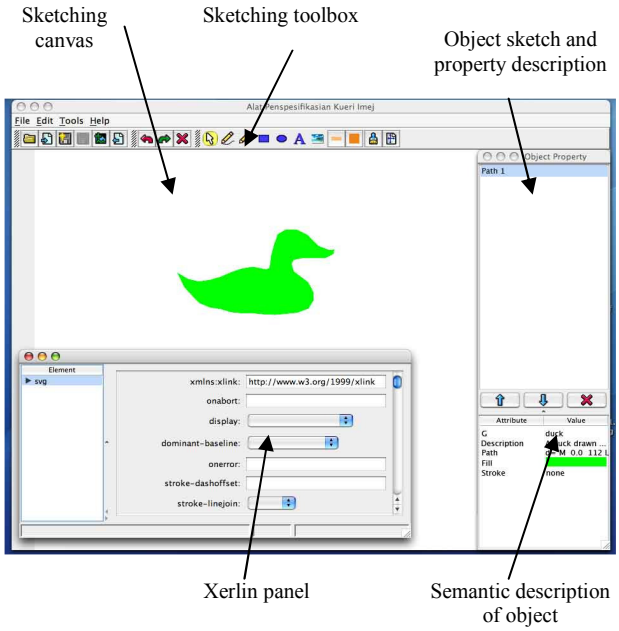

Figure 5. The query specification tool interface.

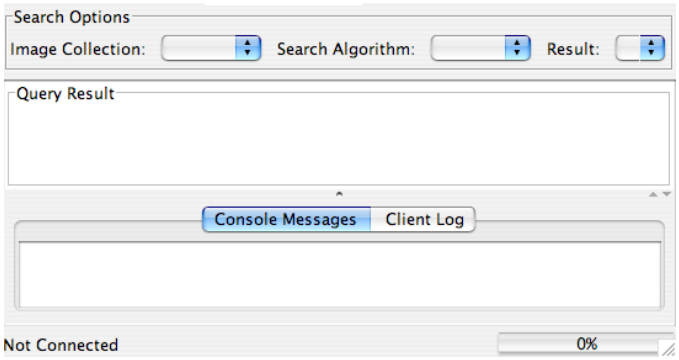

Figure 6. The query manager.

The proposed approach also addresses the implementation aspect on how to enable multiple-querying to a number of heterogeneous image retrieval systems, since the proposed query document can be extended to various image databases applications either it's a visual or textual based. From the end user perspective, this style of querying will benefit areas whereby images are associated with certain kinds of record, like in medical and geographic information systems. We believe that this simplistic approach, using and extending the SVG capability as a base format for user structured sketches with data binding value will help user projects his/her "mental image" efficiently in order to harvest good results.

In order to tackle the demand for an adept database to handle various format of data, many experimental techniques developed at research labs have crossed into publicly available databases. For example, Berkeley DB has its own native XML database engine through Berkeley XML DB (BDB XML), which has support for XQuery access. Structured data in XML documents now can be stored and queried very

effectively. Another example is in Postgresql. Postgresql have built in support for Generalized Search Tree (GiST). It is a balanced, tree-structured access method, which acts as a base template in which to implement arbitrary indexing schemes. B+-trees, R-trees and many other indexing schemes can be implemented in GiST. One advantage of GiST is that it allows the development of custom data types with the appropriate access methods, by an expert in the domain of the data type, rather than a database expert. Looking at this phenomenon, it is a wide opportunity for image database researcher to tap into the wealth of modern database offerings. Database, in image retrieval architecture, serves as a backend for storing and indexing images. In practice, many text-based image retrieval systems use Relational Database Management Systems (RDBMs) as a backend to store indexes of keywords that corresponds to an image. For an image retrieval system that uses visual features, it seldom uses a RDBMS. Many of the databases for this kind of system are built from scratch to cater the complexity of the image's visual surrogate. Since images are stored in heterogeneous databases that have different capabilities, it seems a good time to implement a simple yet effective way to query into these databases. This research has provided an initial step toward achieving this vision.

References

[1] Antani S., Kasturi R., and Jain R., "A Survey on the Use of Pattern Recognition Methods for Abstraction Indexing and Retrieval of Images and Video," *Journal of Pattern Recognition Society*, vol. 35, no. 4, pp. 945-965, 2002.

[2] Ashley J., Barber R., Flickner M., Hafner J., Lee D., Niblack W., and Petkovic D., "Automatic and Semiautomatic Methods for Image Annotation and Retrieval in Query by Image Content (QBIC)," *in Proceedings of Storage and Retrieval for Image and Video Databases III* , pp. 24-35, UAS, 1995

[3] Bradshaw B., "Semantic Based Image Retrieval: A Probabilistic Approach," *in Proceedings of the 8th ACM International Conference on Multimedia*, pp. 167-176, USA, 2000.

[4] Carson C., Belongie S., Greenspan H., and Malik J., "Blobworld: Image Segmentation Using Expectation-Maximization and its Application to Image Querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24 no. 8, pp. 1026-1038, 2002.

[5] Chang F., Thomas H., and Yong R., "Image Retrieval: Current Techniques, Promising Directions and Open Issues," *Journal of Visual Communication and Image Representation*, vol. 10, no. 4, pp. 39-62, 1999.

[6] Carneiro G., Chan B., Moreno J., and Vasconcelos N., "Supervised Learning of Semantic Classes for Image Annotation and Retrieval," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 29, no. 3, pp. 394-410, 2007

[7] Colombo C., Del A., and Pala P., "Semantics in Visual Information Retrieval," *IEEE Multimedia,* vol. 6, no. 3, pp. 38-53, 1999.

[8] Daconta C., Obrst J., and Smith T., *The Semantic Web: A Guide to Future of XML Web Services and Knowledge Management*, Wiley, Indiana, 2003.

[9] Di T., Francesconi M., Frigioni D., and Tarantino L., "Tuning CBIR System for Vector Images: The Interface Support," *in Proceedings of the Working Conference on Advanced Visual Interfaces*, Italy, pp. 425-428, 2004.

[10] Enser B., Sandom J., and Lewis H., "Automatic Annotation of Images from the Practitioner Perspective," *in Proceedings of the Image and Video Retrieval: 4th International Conference,* Singapore, pp. 497-506, 2005.

[11] Eakins P. and Graham E., "Content-Based Image Retrieval: A Report to the JISC Technology Applications Programme," *University of Northumbria at Newcastle*, Institute for Image Data Research, 2008.

[12] Hearst A., *Modern Information Retrieval*, Addison Wesley, London, 1999.

[13] Hearst A., "TileBars: Visualization of Term Distribution Information in Full Text Information Access," *in Proceedings of the Searching and Browsing Text Collections with Large Category Hierarchie, ACM SIGCHI Conference on Human Factors in Computing Systems*, Denver, USA, pp. 59-66, 1995.

[14] Herman I. and Dardailler D., "SVG Linearization and Accessibility," *Computer Graphics Forum*, vol. 21, no. 4, pp. 777, 2002.

[15] Hofstede M., Proper A., and Van P., "Query Formulation As an Information Retrieval Problem," *The Computer Journal*, vol. 39, no. 4, pp. 255-274, 1996.

[16] Jaimes A. and Chang F., *Image Databases Search and Retrieval of Digital Imagery,* John Wiley, New York, pp. 497-565, 2002.

[17] Lovet G. and Dardailler D., *SVG Linearizer Tool,* W3C Note, http://www.w3.org/WAI/ER /ASVG/ , 2000.

[18] Minsky M., "Commonsense-Based Interfaces," *Communications of the ACM*, vol. 43, no. 8, pp. 66-73, 2000.

[19] Pentland A., Picard W., and Sclaroff S., "Photobook: Tools for Content-Based Manipulation of Image Database," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233-254, 1996.

[20] Pun T. and Milanese R., "Computer Vision and Multimedia Information Systems," *in Proceedings of the International Workshop on Multimedia Information Systems and Hypermedia*, Japan, pp. 29-37, 1995.

[21] Smith R. and Chang F., "VisualSeek: A Fully Automated Content-Based Image Query System," *in Proceedings of the 4th ACM International Multimedia Conference,* India, pp. 87-98, 1996.

[22] Vailaya A., Figueiredo M., Jain K., and Zhang J., "Content-Based Hierarchical Classification of Vacation Images," *in Proceedings of IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*, vol. 1, no. 1, pp. 518-523, 1999.

[23] Van M. and De P., "The Psychology of Multimedia Databases," *in Proceedings of the 5th ACM Conference on Digital Libraries*, USA, pp. 1-9, 2000.

[24] Van J., *Information Retrieval,* 2nd Edition, Butterworth-Heinemann, London, 1979.

Shahrul Noah received his MSc and PhD in information studies from the University of Sheffiled, UK in 1994 and 1998, respectively. His research interests include information retrieval, knowledge representation, and semantic web. He currently leads the knowledge technology research group at the national university of Malaysia. He is a member of the IEEE computer society.

Saiful Sabtu received his Bsc of Information Technology degree from the same university. His research interests include information retrieval and multimedia information systems.

Appendix 1: Example of a duck drawn using the query canvas and its linearization output

Raster form

Text-based output

Title of the query: "Duck Image"
A simple duck image
Information on #duck
A duck drawn with polylines
Information on #theDuck
A drawing of the duck

Real SVG Code

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<!-- <?xml-stylesheet type="text/css" href="../../svgs/slide.css" ?> -->
<?xml-stylesheet type="text/css"
href="http://www.w3.org/Consortium/Offices/Presentations/svgs/slide.css"
?>
<svg xml:space="preserve" xmlns:xlink="http://www.w3.org/1999/xlink"
viewBox="0 0 1024 768">
  <title>Duck Image</title>
  <desc>A simple duck image</desc>
  <metadata>
    <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:axsvg="http://www.w3.org/2001/svgRdf/axsvg-schema.rdf#">
      <rdf:Description rdf:about="#duck">
        <axsvg:IsDef>true</axsvg:IsDef>
      </rdf:Description>
      <rdf:Description rdf:about="#theDuck">
        <axsvg:InstanceOf rdf:resource="#duck"/>
      </rdf:Description>
    </rdf:RDF>
  </metadata>
  <defs>
    <g id="duck" style="stroke:none; fill:lime;">
      <desc>A duck drawn with polylines</desc>
      <path d="M  0.0  112 L  20  124 L  40  129 L  60  126 L  80  120 L  100
111 L  120  104 L  140  101 L  164  106 L  170  103 L  173  80 L  178  60 L
185  39 L  200  30 L  220  30 L  240  40 L  260  61 L  280  69 L  290  68 L
288  77 L  272  85 L  250  85 L  230  85 L  215  88 L  211  95 L  215  110 L
228  120 L  241  130 L  251  149 L  252  164 L  242  181 L  221  189 L  200
191 L  180  193 L  160  192 L  140  190 L  120  190 L  100  188 L  80  182
L  61  179 L  42  171 L  30  159 L  13  140 L  00  112 Z"/>

    </g>
  </defs>
   <use xlink:href="#duck" transform="translate(700 60)" id="theDuck">
      <desc>A drawing of the duck</desc>
    </use>
</svg>
```