A SEMANTIC APPROACH TO TEXT-BASED IMAGE RETRIEVAL USING A LEXICAL ONTOLOGY

Syed Abdullah FADZLI, Rossi SETCHI

Institute of Computational Engineering, Cardiff University, Wales, UK.

ABSTRACT

This research proposed a novel semantic approach to text-based image retrieval based on a lexical ontology called OntoRo. The approach aims to narrow down the semantic gap between visual content and the richness of human semantics by using keywords-based semantic image indexing. The approach proposed involves: (a) semantic image indexing; (b) semantic search; and (c) semantic image visualization. This paper focuses on the semantic image indexing method and implements semantic search method as the evaluation process. A new concept of Semantic DNA is introduced as an indexing method which will help to retrieve semantic results. The Semantic DNA is extracted from the existing ontological structure of OntoRo. It is the key element in this approach which will be used throughout the whole research.

Keywords: *Image Retrieval, Image indexing, semantic, lexical ontology*

1. INTRODUCTION

With the widespread use of the internet, vast quantities of digital images are now available on the World Wide Web (WWW). Thus effective approaches are needed to enhance the image indexing and image retrieval methods. Famous high quality online image libraries such as ShutterStock® (http://www.shutterstock.com/) and FotoSearch® (http://www.fotosearch.com/) claim that they have over 6.5 million and 4.8 million images respectively. While with no image quality restriction, flickr® (http://www.flickr.com/) claims that they have over 3 billion and Facebook® (http://www.facebook.com/) over 4 billion digital images available online. With this current growth, the need for effective methods to index and retrieve digital images has become very crucial in order to support application in image retrieval.

So far, research in image retrieval had been divided between two approaches: (i) ones that concentrate on concept-based and (ii) ones that concentrate on content-based image retrieval [1-3]. The former focus on using text to retrieve images (i.e. title, subject, keywords, caption, etc.) while the latter focus on the visual features of the image (i.e. size, colours, textures, etc.). Although more recent research focuses on the exploitation of the content-based approach, extensive experiments on content-based image retrieval (CBIR) systems show that low-level content often fail to describe the high-level semantic concepts in a user's mind [2]. Hence the performance of content-based approach alone is still far from the users' expectation. Major image search providers such as Google Images® (http://images.google.com) and Yahoo!® (http://images.search.yahoo.com/) rely directly on textual descriptions of images found on the web pages including the image name, caption and annotations. However these search engines do not consider the semantic meanings of the descriptions, hence, cannot be used to search for high-level concepts of the images. Although most images are manually annotated by users, such annotations have a lot of disadvantages for example the annotation inaccuracy due to subjectivity of human perception or the ambiguous meaning of words used. An automated system that can extract the semantic meaning of pictures based on text analysis could assist to disambiguate the meaning of keywords and facilitate searches using high-level perception and emotion.

In this paper, the use of Semantic DNA (SDNA) based on lexical ontology in semantic image indexing and semantic search is proposed. The paper is organized as follows: Section 2 reviews the current research on automatic image annotation and indexing, which inspired this research; Section 3 presents the proposed Semantic Image Indexing method by explaining the conceptual model of the approach; Section 4 presents the Semantic Search method which will be used as the evaluation process. This paper continues with Section 5 which presents the experiment and evaluation process of the proposed method, while the whole paper is concluded in Section 6.

2. RELATED WORK

This section focuses first on current automatic image annotation and image indexing research. It then reviews several well-known methods for semantic similarity measurement.

Automatic image annotation and indexing plays an important and promising role in large-scale image organization and retrieval. Although it has been studied for several years by knowledge engineering, natural language processing, artificial intelligent, computer vision and machine learning communities, current state-of-the-art in image indexing is still far from practical. Most indexing engines need to be trained on large volume of image collections with human involvement. For example, Datta [4] and Li and Wang's [5] proposed a combination of machine learning and composite approach to map visual features with semantic concepts. The method will learn from human-labeled images to map between low-level image features with higher-level semantic labels. The machine can later use the knowledge to automatically suggest labels for other images based on visual features. But such approaches produce limited degree of accuracy since they do not incorporate domain knowledge for the disambiguation of semantic concepts. The need to measure semantic relationships between words and concepts is a traditional problem in Natural Language Processing. Measures of relatedness or distance are widely used in various applications including word-sense disambiguation, text structure analysis, text summarization, information extraction and retrieval, spelling correction, automatic annotation and automatic indexing.

Semantic similarity measurement is one of the methods used in reducing the gap between image content and the richness of human semantics. With the availability of digital taxonomies such as dictionaries, lexicons and thesauri, this human built semantic knowledge provides a formal structure of relationship links between words and concepts. WordNet [6] is one of broad-category taxonomies which are widely used by researches to study word and concept relationships. Budanitsky and Hirst [7] provide a survey of many WordNet-based measures of semantic similarity based on lexical paths of the taxonomy. For example, Jiang and Conrath [8] proposed one of the best performing measures which find the shortest path between two words in the taxonomy hierarchy, and calculate the weight of each node in the path by considering local density, node depth and link type. Many researchers used the path-length calculation approach as a way to compute semantic relationship in taxonomy. It is however proven to be highly effective only when used with domain specific taxonomy which ensures the homogeneity of the hierarchy.

Despite the popularity of WordNet, Jarmasz and Szpakowicz [9] proposed an effective lexicon-based algorithms based on *Roget's Thesaurus*. They produced a good result by treating the thesaurus as a simple hierarchy of semantic concepts. The distance measure used, which is often called edge counting, can be calculated quickly and proves to be better than WordNet in their experiments. The potential of Roget's Thesaurus has been discovered by many other researchers including Yarowsky [10] and Morris and Hirst [11]. Morris and Hirst suggested that the lexical chain of Roget's often provide enough context to resolve lexical ambiguities, an idea later used by Manabu and Takeo [12] in their approach.

In this paper, a novel approach towards semantic image indexing is proposed, which is based on Roget's hierarchical structure. This method introduced a new technique in semantic similarity measurement based on the well-constructed structure of concept classification using a digital version of Roget's, called OntoRo.

3. SEMANTIC IMAGE INDEXING

This section focuses first on the lexical ontology of OntoRo v1.7 and explains the extraction of Semantic DNA (later referred to as SDNA) from the OntoRo's hierarchical structure. It then describes the conceptual model of Semantic Image Indexing method and the SDNA weight calculation.

3.1. OntoRo v1.7 and Semantic DNA

OntoRo v1.7 is a lexical ontology based on *Roget's Thesaurus* digital version which is used by Tang [13]. It was created using the electronic version of *Roget's* built in Project Gutenberg [14]. The Gutenberg's edition was used as the main dictionary source, while the printed 2003 edition was utilized to remove outdated word entries from Gutenberg edition, and add new entries into the ontology dictionary.

Class #6: *Emotion, Religion and Morality*
└─Section #37: *Interpersonal emotion*
  └─Sub Section #83: *Social*
    └─Head #887: *Love*
      └─Part of speech: *noun*
        └─Paragraph #5: … *young woman, girlfriend, bird, squeeze…*

Figure 1: Hierarchical structure of *bird* as a symbol of love.

Different from a dictionary which explains the meaning of words, *Roget's* groups words based on ideas. A path in *Roget's* ontological structure begins with 8 highest categories called Classes. It branches to one of the 39 Sections, then to one of the 79 Sub-Sections, then to one of the 596 Head Groups and finally to one of the 990 Heads. Each Head is grouped by parts of speech (POS); nouns, adjectives, verbs and adverbs. Each group of POS is further divided into series of paragraphs, and finally, a paragraph is divided into semicolon groups of semantically closely related words. For this approach, only six levels of *Roget's* hierarchy are considered: (i) Class, (ii) Sections, (iii) Sub Sections, (iv) Heads, (v) POS, and (vi) Paragraphs. These levels are viewed as a DNA string of a word meaning, called SDNA. One unique SDNA can be shared by several words in a Paragraph in *Roget's* hierarchy.

One of the advantages of *Roget's* is the ability to identify different meaning of words according to different contexts (ambiguity). Each meaning of words is grouped in different hierarchical structure according to its contextual idea. For example, according to *Roget's*, the word *bird* can have 13 different contextual meanings, including *bird* as an animal, as a traveler, as a symbol of feminism and as a symbol of love. Based on this premise, each contextual meaning of a word can have a unique SDNA. Figure 1 shows the hierarchical structure of one of the meaning of word *bird* in *Roget's* (in context of *love*). The SDNA of the above example can be extracted as *6-37-83-887-n-5,* where *n* represents noun as the POS group (bearing in mind that this is just one of 13 SDNA strand of the word *bird*). The use of SDNA in this approach will be explained in section 3.2.

3.2. Conceptual Diagram

To perform the semantic image indexing, a two stage procedure is proposed as shown in figure 2. First the Words Analysis stage is performed based on the image annotations provided by users. To analyze the keywords, 3 separate processes are involved: (i) text parser to parse the annotations into tokens of unique words (which will afterwards be referred to as tokens), (ii) SDNA Selection to identify the OntoRo's hierarchical structure for each token, and (iii) Word-sense disambiguation process to determine the exact meaning of the token according to the image context.

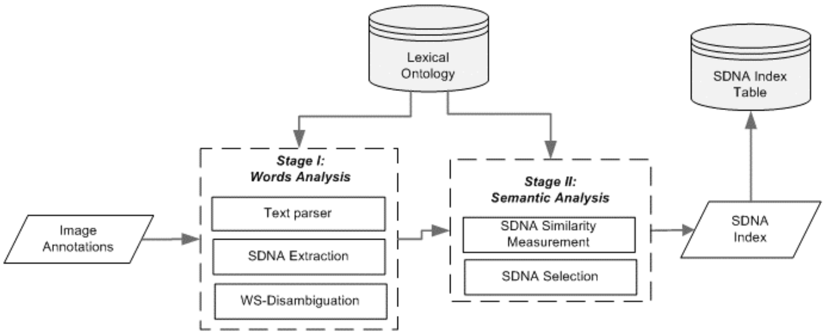


Figure 2: Conceptual diagram of Semantic Image Indexing of one image.

After the annotation has been parsed into tokens of words, the tokens are compared against OntoRo to ensure their existence in the thesaurus. Name entity such as place names, street names and people's names are generally not to be found in OntoRo, therefore these words will be ignored for this moment. Every existing token will produce several SDNA, which will then be used in the Word Sense Disambiguation process. The SDNA co-occurrence is used to determine the contextual meaning of a word. For example, the word *pen* will produce 12 unique SDNA which belongs to 12 different contextual meaning including *pen* as a stationary, a small space, a female swan and a portable enclosure for baby to play. By calculating similarity co-occurrences of *pen* SDNAs, such a measure might show that a *pen* is more likely to be a stationary rather than a place, if other words share the same ontological structure of SDNA of *pen* as a stationary. Figure 3a shows the generalized algorithm for Word Analysis stage.

```
REPEAT for every token
  CHECK OntoRo for existence
  IF token exist THEN
      EXTRACT SDNA for token
      INCLUDE SDNA into SDNA table
  END IF
END REPEAT
```

(a)

```
REPEAT for every SDNA in SDNA table
    REPEAT while n is > 1
        CHECK n-level SDNA similarity with other
        SDNA in table (see section 3.1)
        IF similarity found THEN
            COUNT similarity occurrence in unique tokens
            CALCULATE SDNA Weight (section 3.3)
            INCLUDE SDNA in SDNA Index table
        ELSE
            n := n - 1
        END IF
    END REPEAT
END REPEAT
```

(b)

Figure 3: Algorithm for (a) Word Analysis stage and (b) Semantic Analysis stage.

Semantic Analysis stage involved 2 separate processes: (i) SDNA Similarity Measurement and (ii) SDNA Selection. SDNA Similarity Measurement process will calculate the co-occurrence of similar SDNA structure of all SDNAs in the table. The SDNA with highest weight (see section 3.3) in a token will be selected in SDNA Selection process, as the most relevant OntoRo DNA of the token; hence disambiguate the meaning of the actual word. The selected SDNA will then be included in SDNA Index table as the semantic index for the image processed. Figure 3b shows the generalized algorithm for Semantic Analysis stage.

3.3. SDNA weight feature

To create the relationship between the image and SDNA, SDNA Weight (SW) is calculated to measure the importance of a particular SDNA to the image. In other words, one SDNA might be more important to an image than another SDNA if the weight is higher.

To measure the SDNA Weight, the following formula is used:

$$SW = L_i(1 + (C_i / N)) \tag{1}$$

where L_i is the hierarchical level in an SDNA i, C_i is the number of occurrences of SDNA i for the given token, which are similar up to L_i and N is the number of tokens for the image which exist in OntoRo (see section3.2). For this experiment, only 6 level of OntoRo's structure are considered in the SDNA extraction, therefore $SW = [1, 12]$.

4. SEMANTIC SEARCH

This section focuses first on the conceptual model of Semantic Search method and then describes the SDNA Relationship Weight calculation.

4.1. Conceptual Diagram

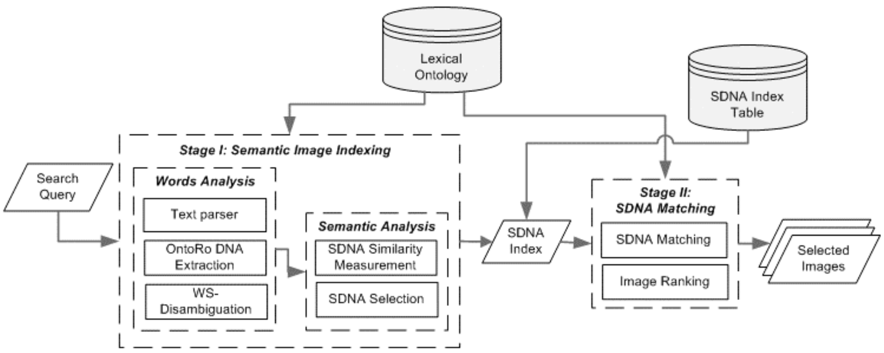


Figure 4: Conceptual diagram of Semantic Search

Figure 4 shows two stage procedures for Semantic Search method. This methods starts with the search query going through the Semantic Image Indexing (see section 3.1) method introduced above. The first stage will produce SDNA Index selected for each token which will be used in the next stage. SDNA Matching stage is divided into two separate processes: (i) SDNA Matching which will compare the query SDNA with SDNA Index table and produce SDNA Relationship Weight (SRW), and (ii) Image Ranking which will ranks the images found based on SRW produced by previous process. Figure 5 shows the generalize algorithm for SDNA Matching stage in Semantic Search method.

```
REPEAT for every images in SDNA Index table
    REPEAT for every SDNA of query
        COMPARE every SDNA of query with every SDNA of image
        IF nodes of SDNA of query is equal to nodes of SDNA of image THEN
            CALCULATE SDNA Relationship Weight
        END IF
    END REPEAT
    SUM the SDNA Relationship Weight of all matches
END REPEAT
```

Figure 5: Algorithm for DNA Matching stage in Semantic Search method.

4.2. SDNA Relationship Distance Feature

SDNA Relationship Distance (SRD) of a particular SDNA of a token is determined by calculating the mean of the SDNA weight for a token and the matching SDNA index weight, using the following formula:

$$SRD_{(\alpha,\beta)} = \sum_{i}^{|\alpha|} \sum_{j}^{|\beta|} ((SW_{(\alpha,i)} + SW_{(\beta,j)})/2) \tag{2}$$

Where α represents the set of search query SDNA, β represents the set of SDNA of an image in SDNA Index table, i is the number of SDNA in α and j is the number of SDNA in β. $SRD_{(\alpha,\beta)}$ will determine the similarity measure between a search query α and an image in the SDNA Index table β. Images with high $SRD_{(\alpha,\beta)}$ will be considered as highly related to the search query α.

5. CASE STUDY

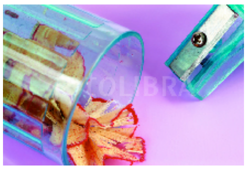
A research collaboration with an online image library website, fotoLibra® [15], had provided this case study with 160,000 digital images. fotoLibra® was selected because of their large collection (335,060 images by the time this paper was written) of high quality images covering a broad range of themes which already being manually annotated by image owners. The annotations are used to help the search process using their current image retrieval method.

5.1. Experiment Results

Using the semantic image indexing algorithm which had been introduced above, all 160,000 digital images from fotoLibra® had been indexed producing 710,750 SDNA with their respective weight. The average annotation for each image is 20.8 words with 59.8% of the words used in the annotations exist in OntoRo. The result shown 99.2% of all images had at least 1 SDNA selected as their index (0.8% images cannot be indexed due to bad annotation which produces no SDNA similarity).

Table 1 shows two of the Semantic Image Indexing example results that were obtained during the experiment. The selected SDNA shown that the first image is highly related to *nursery, children, tiny tot* and *school,* and second image related to *background, material, education, pencil sharpener* and *object.* These SDNA is included in the SDNA Index table as the SDNA of respective images, and later be used in the Semantic Search which will be explain in the next section.

Table 1: Two of the Semantic Image Indexing example results using fotoLibra® image library.

Image	**Tokens (extract from annotations)**	**Selected SDNA for image**	**Tokens related to SDNA**
Image ID: 58776 	a day in the nursery, 1 aaefl, em15, t3b, nursery, children playing, school, pre school, teachers, teaching, tiny tots, prestatyn, north wales	1-6-22-130-1-2	nursery, children, tiny tot
		1-8-28-171-1-3	nursery, children, school
		4-24-57-539-1-3	nursery, school
		1-6-22-132-1-1	children, tiny tot
		1-8-28-170-1-1	children, nursery, school
Image ID: 137702 	razit ror, background, pencil sharpener parings office object, education elements, school, material	3-15-48-445-1-2	background, material
		4-20-53-490-1-1	background, education
		4-24-57-524-1-1	background, material
		4-25-58-594-1-4	background, pencil sharpener, object
		4-25-58-586-1-4	pencil sharpener, background, pencil sharpener, object

5.2. Evaluation

Based on the results from the above experiment, Semantic Search method has been implemented to evaluate the image search result using random search query. For the purpose of this paper, five keywords have been chosen as the search query to explain a school environment for children: *child*, *school*, *education*, *learning* and *teacher*. These five words have been used to search for images using Google Images® and fotoLibra® search query to be compared with the proposed Semantic Search method. The first 20 search results from Google Images® (figure 6a) reveal several unrelated images that were retrieved based on the image captions. Only 6 out of 20 images considered as related to the search query. Meanwhile, fotoLibra® search (figure 6b) only returns 3 images where all of them focus only on classroom environment.

The proposed Semantic Search based on the Semantic Image Indexing method is proven to produce better results using the proposed search query. Figure 7 shows the first 20 results for the experimental query that are sorted according to SRD. These results provide variety kinds of images explaining the school environment of children including classroom environment, pupils, class decoration, classroom board, 2 school buildings, pencil sharpener and school library. Only 4 out of 20 results considered as unrelated to the search query including an image of a school prospectus (image ID 52133) and 3 images of a music school competition (image 120728, 120727 and120729). The results show that, given a descriptive search query, Semantic Image Indexing approach using OntoRo, manages to retrieve good image results from different perspectives compared to Google Images® which only produces less than half related images, and fotoLibra® search with limited results from one perspective.

(a) (b)

Figure 6: Search result using (a) Google Image® and (b) fotoLibra® search query.

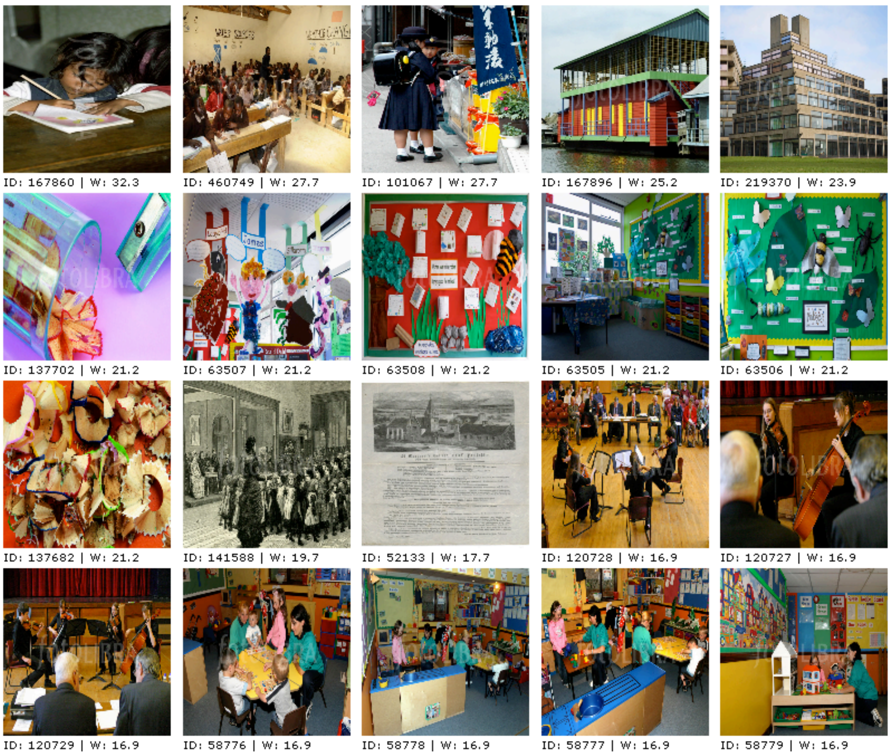


Figure 7: Search result using Semantic Search

6. CONCLUSION

The Semantic Image Indexing and Semantic Search methods are designed to help industrial designers retrieve images that reflect the semantic elements of a desired design, using intelligent mood boards. In the automotive industry, mood boards are used by industrial designers to express emotions and inspirations. A mood board is a collection of images, text, objects and textures, compiled with the intention of communicating or provoking emotion and creativity during the product design process. In modern days, digital mood boards utilize the effectiveness of digital images, image effects, animations and even audios and videos. The huge and fast growing number of digital images can be a perfect source for digital mood board materials. According to Edwards [16] the main challenge in developing a mood board is to collect and identify all relevant materials which describe the concept or theme of the desired design. This paper proposes and evaluates both methods as one important step toward meeting this need. The variety of search results produced during the experiment is crucial for the development of mood boards, where designers need images from different sources that could semantically describe a design concept, as their source of inspiration. The experimental results also show that OntoRo as the lexical ontology source for SDNA extraction is very effective in determining the semantic relationships between annotation words. The proposed methods can be applied using any existing lexical ontology or taxonomies with a consistent and well-organized hierarchical structure. Further work includes semantic indexing on other multimedia contents including audios and videos as potential materials in developing intelligent mood boards.

ACKNOWLEDGEMENT

We thank all members of Knowledge Engineering Systems (KES) group in Institute of Computational Engineering (ICE) for their help with this research; the Ministry of Higher Education of Malaysia for their sponsorship; and VisconPro Ltd (fotoLibra®) for images provided for the experiment.

REFERENCES

1. Rui, Y., T.S. Huang, and S.F. Chang, Image retrieval: Current techniques, promising directions, and open issues. Journal of Visual Communication and Image Representation, 1999. 10(1): p. 39-62.
2. Liu, Y., et al., A survey of content-based image retrieval with high-level semantics. Pattern Recognition, 2007. 40(1): p. 262-282.
3. Datta, R., et al., Image retrieval: Ideas, influences, and trends of the new age. 2008.
4. Datta, R., et al. Tagging over time: real-world image annotation by lightweight meta-learning. 2007: ACM Press New York, NY, USA.
5. Li, J. and J.Z. Wang, Real-Time Computerized Annotation of Pictures. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, 2008: p. 985-1002.
6. Fellbaum, C., WordNet: An Electronic Lexical Database. 1998, Cambridge, MA: The MIT Press.

7. Budanitsky, A. and G. Hirst, Evaluating WordNet-based measures of lexical semantic relatedness. Computational Linguistics, 2006. 32(1): p. 13-47.
8. Jiang, J.J. and D.W. Conrath, Semantic similarity based on corpus statistics and lexical taxonomy. Arxiv preprint cmp-lg/9709008, 1997.
9. Jarmasz, M. and S. Szpakowicz. Roget's Thesuarus and Semantic Similarity. in Recent advances in natural language processing III. 2003.
10. Yarowsky, D. Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. 1992: Association for Computational Linguistics Morristown, NJ, USA.
11. Morris, J. and G. Hirst, Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computational Linguistics, 1991. 17(1): p. 21-48.
12. Manabu, O. and H. Takeo. Word sense disambiguation and text segmentation based on lexical cohesion. 1994: Association for Computational Linguistics Morristown, NJ, USA.
13. Tang, Q., Knowledge Management Using Machine Learning, Natural Language Processing and Ontology, in Cardiff School of Engineering. 2006, Cardiff University: Cardiff. p. 243.
14. Hart, M. and G. Newby. Project Gutenberg. 2003 [cited 2009 Jan]; Available from: http://www.gutenberg.org/wiki/Main_Page?sess=25607a1aa5.
15. VisConPro. fotoLibra, The Image Warehouse. 2009 [cited 2009 01/08/2009]; An online picture library consist of more than 330,000 images.]. Available from: http://www.fotolibra.com/.
16. Edwards, A., S.A. Fadzli, and R. Setchi, A Comparative Study of Developing Physical and Digital Mood Boards, in IPROMS Conference 2009. 2009: http://conference.iproms.org.