An Automatic and Innovative Approach for Converting Pedagogical Text Documents to Visual Learning Object

Ali Shariq Imran[1], Atif Mansoor[2], and ABM Tariqul Islam[3]

[1] Faculty of Comp. Science and Media Technology, Gjøvik University College, Norway
{ali.imran}@hig.no

[2] Institute of Avionics and Aeronautics, Air University, Islamabad, Pakistan
{atif.mansoor}@gmail.com

[3] Visual Computing Lab, Dept. of Comp. Science, University of Rostock, Germany
{tariqul.islam}@uni-rostock.de

Abstract. In this paper, we present a novel idea of converting pedagogical text documents to visual learning objects by automatically extracting nouns and semantic keywords from the text documents, and representing these keywords as a word cloud. A word cloud contains words that are weighted based on frequency, time, appearance, etc., depending on the concept they are used for. Each word in the word cloud would correspond to a visual representation of that word. A visual representation may contain drawings, figures, images, etc. that explains the given concept. The extracted keywords are used to query the Internet to find the corresponding visual representation of a given word. The idea is to bring text documents to life by creating a visual representation of the important concepts from the text documents. This paper is a work in progress.

Keywords: pedagogic, text documents, visual learning object, word cloud.

1 Introduction

Burmark, in his book [1] stated that the visual aids such as images, drawings, graphics, videos, etc. can improve learning by up to 400%, as they actively incorporate all the human senses into learning experience. It is also a well-established fact that about approximately 65% of the world population are visual learners and the rest of 35% of the world population consist of textual, auditory and kinesthetic learners [2]. Furthermore, 90 percent of information that comes to the brain is visual [3]. Thereby, turning the important concepts in text documents to their visual representation can improve the learning experience of the users, and to some extent the learning outcome.

The complex ideas are better retainable and implementable as depicted in graphical manner. In computer science, flow chart gives more understandable depiction of complex relations and dependencies. The spatial representation of

C. Stephanidis (Ed.): HCII 2014 Posters, Part II, CCIS 435, pp. 84–88, 2014.

information coupled with images carry much more meaning and also helps the learner to identify and associate similar ideas. Along with improvement in retention ability, it also aids in critical thinking by developing the associability aspect.

The human brain consists of different areas that work together to translate the viewed images into retained information. In human brain, the visual information is processed in the visual cortex. It is positioned in occipital cortex, situated at the rear side of human brain. As an image is formed at the retina, the different information like shape, orientation, color etc. are transported by different neurons to the occipital cortex. The human brain is divided into two hemispheres, each containing one visual cortex. The left hemisphere collects the view from right eye and the right hemisphere receives the view from left eye through respective visual cortex. The occipital cortex consists of subparts that play an important role in detection, classification and learning. The object location and identification is performed here. The signal from occipital cortex is transmitted to temporal cortex that performs recognition, forming memory and understanding different reactions. The signals are also transferred to parietal cortex that cohesively forms the view from various signals received from senses. It performs spatial attention and spatial mapping. The cerebellum located in the rear lower part of the brain is responsible for cognitive activity like learning. Cerebellum though having a weight just 10% of total weight of the brain contains higher number of neurons compared to total sum of neurons in other part of brain. The whole biological action in the brain with respect to visual learning demonstrates a lot of activity that in turn result in overall better retention and memory.

2 Methodology

In this work, we aim to automatically turn the educational text document into a visual learning object (VLO) (an image) by extracting and finding important words from the document using natural language processing (NLP) techniques, including statistical measures such as frequency count or term frequency inverse document frequency (TF-IDF) [4]. The proposed algorithm reads the pedagogical documents consisting of text files and performs the semantic analysis on them as shown in Figure 1. In the semantic analysis a text document is analyzed automatically to extract important words, known as keywords. A preprocessing step is required before performing the semantic analysis on the potential keywords. The preprocessing steps required to extract potential keywords from the pedagogical text documents are described as follows:

1. In the first step, the text document file is read and all punctuation and capitalization is removed.
2. A tokenizer is then used to separate the text into individual words.
3. Next, all the stop words are removed. Stop words consists of those words in English language which do not convey any significant meaning on their own, such as ‘and’, ‘is’, etc [4].

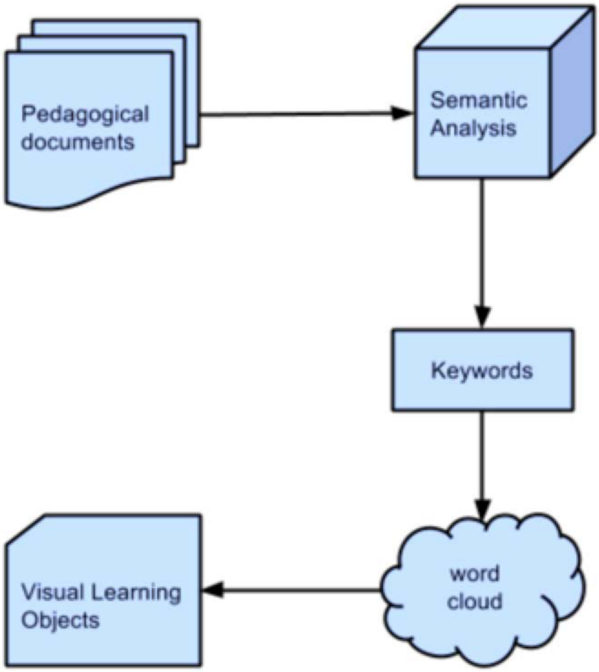

Fig. 1. A block process diagram representing how the pedagogical documents are converted into visual learning objects

4. The duplicate words are removed afterwards.
5. The last step is to apply stemming [5]. Stemming is the process of bringing words to its root form. For example, the word 'stemming' can be converted to its root form as 'stem'.

At this stage, statistical measure such as frequency count or TF-IDF can be employed on the candidate keywords to further short-list the potential keywords. The potential keywords obtained as a result can either be a noun, a verb, an adverb or an adjective. These are different parts-of-speech (POS) of a language. Generally, the nouns POS are used to represent real world entities and concepts, so the nouns are further processed in the semantic analysis. At the end of pre-processing steps, the algorithm analyzes all the nouns POS candidate keywords semantically.

The semantic analysis consist of two steps. In the first step the importance of short-listed candidate keywords is calculated based on the visual similarity concept [6]. Visual similarity measures the visualness of a given word, which in fact defines how important a word is in the given context. It is measured by computing a similarity of a given candidate word and visual seed words. The seed words are the ground truth against which the visualness of a candidate word is determined. The seed words can be defined manually or they can be extracted automatically from the document's corpus, metadata, and/or title. Visualness is calculated using semantic similarity metric between words. There are many different metrics to compute the visualness [7]. For any given word w, It is defined as:

$$vis(w) = \sum_{i=1}^{n} vis(s_i) \frac{sim(w, s_i)}{\sum sim(w, s_i)}, \qquad (1)$$

where $i = 1, 2, ...n$, for the whole set of n seed words $s_1, s_2, ...s_n$, and $sim(w, s_i)$ is the semantic similarity between the word w and a seed word s_i. The visualness of a given seed word s_i is denoted by $vis(s_i)$. The resulting visualness score $vis(w)$ is in the range of 0 to 1.

Calculating visualness will give us a refined set of potential candidate keywords. We may however still need to refine the words further to have better semantic meaning. For example, the refined keywords might contain few words with dual meanings. For instance, if 'cancer' is one of the short-listed keyword, then it can either refer to a zodiac sign or a disease. The algorithm then needs to disambiguate the correct sense of a word in order to fetch the right image from the Internet. This can be addressed using a word sense disambiguation (WSD) process [8], given as:

$$Disambiguation(s_i) = |context(s) \cap gloss(s_i)|, \qquad (2)$$

Where s is a seed word, $context(s)$ represents the whole set of all seed words, and $gloss$ defines all possible sense i of s. The assigned sense of s is the i^{th} where s_i maximizes the $Disambiguation$ value.

In short, the score is calculated by counting the number of overlapping words between the gloss of each sense s_i and the word's context. Visualness and WSD will give us the refined set of semantically accurate keywords.

The keywords are then be depicted as a word cloud. The word cloud is pasted onto various objects like a globe or other objects, chosen on the basis of interest it infuses in the intended learner group. Further, using 'Microsoft Kinect gestures' the users are able to interact with the word cloud. For example, they can rotate the globe or the object and see different words appear on it from the extracted document with variable font size, color and appearance etc. The choice of font size, color, appearance etc. depends upon the number of word occurrence in the text or importance ascertained by semantic analysis, along with the target group like children, students, adults from specific profession etc.

By interacting with the word cloud, the users can see corresponding visual objects associated with the words. Users may further be directed to relevant information at Wikipedia, PowerPoint presentations, and videos over the Internet corresponding to chosen keywords. This will allow an interactive learning experience, and it could be a fun and interesting tool to teach young children.

3 Conclusion

In this paper, we propose an innovative and novel approach to convert text documents to visual learning objects by automatically extracting semantic nouns from the document. The proposed algorithm reads the text document files, do the tokenization of words, shortlist nouns, remove duplicates, and extract semantically accurate and meaningful keywords. Semantically accurate keywords are extracted using visualness and word sense disambiguation process. The keywords are represented as a word cloud. The algorithm then query the search engines with extracted keywords to find correspondence images from the Internet. It then associates the appropriate image to the extracted keyword. This paper is a work in progress and we are evaluating the outcome and the results. We believe that creating a visual representation of the important words in a text documents will improve the learning outcome, especially for teaching young kids.

References

1. Burmark, L.: Visual Literacy: Learn to See, See to Learn. Association for Supervision and Curriculum Development, p. 115 (2002) ISBN 0871206404, ISSN 978087120640
2. David, H.J., Carr, C., Yueh, H.-P.: Computers as Mindtools for Engaging Learners in Critical Thinking. TechTrends 43(2), 24–32 (1998)
3. Hyerle, D.: Visual Tools for Transforming Information into Knowledge, 2nd edn., Corwin (2009)
4. Manning, C.D., Raghavan, P., Schutze, H.: Introduction to Information Retrieval. Cambridge University Press, Cambridge (2008)
5. Sharma, D.: Stemming Algorithms: A Comparative Study and Their Analysis. International Journal of Applied Information Systems 4(3), 7–12 (2012)
6. Qiu, Y., Guan, G., Zhiyong, W., Feng, D.A.: Improving News Video Annotation with Semantic Context. In: Proceedings of the 2010 International Conference on Digital Image Computing: Techniques and Applications, DICTA 2010, pp. 214–219. IEEE Computer Society, Washington, DC (2010)
7. Li, H., Tian, Y., Ye, B., Cai, Q.: Comparison of Current Semantic Similarity Methods in WordNet. In: International Conference on Computer Application and System Modeling (ICCASM), October 22-24, vol. 4, pp. V4-408–V4-411 (2010)
8. Agirre, E., Edmonds, P.: Word Sense Disambiguation: Algorithms and Applications. Text, Speech and Language Technology, vol. 33, XXII (2006) ISBN 978-1-4020-4809-8