Visualization of Subjective Extracted Text Using the Parse Tree

Erfan Mirzabeiki*
Department of Computer Vision and Multimedia
Kualalumpur , Malaysia
Email: erfan_66m {at} yahoo.com

Lili A.Nurliyana
Department of Computer Vision and Multimedia
Kualalumpur , Malaysia

***Abstract*—Text visualization is a well-known technique used in note taking and obviating the difficulties of the learners in focusing on important portions of the text materials to be learnt. On the other hand applying the text visualization methods on the extracted sections required by the learners will amplify the efficiency of perceiving from the text. This paper intends to propose an automatic approach for subjective extracted text visualization to improve leaning from digital text. The main function of this approach divided into two parts, firstly: sentence extraction based on the inclusion of a significant part of speech and role of the words in sentence (i.e. Noun/Names/Places/dates Numbers/ Events) by using the parse tree, secondly: Visualization of the extracted information for maintaining the user's attention on the content were developed on the basis of psychological, multimedia, and typographic cueing foundations and techniques (visual effects). This experiment conducted on different sample texts in the field of "History" in order to visualize the extracted part.**

Keywords----Text visualization, Information Extraction, Typographic Effects, Color Coding Effect, Parse Tree

I. INTRODUCTION

With today massive growth of digital textual materials due to the rapid and often exponential scientific advancements, more value has been accrued to and more consideration has been given to the process of acquiring subject matter conceptual matrix through text. On the other hand, students' learning process is greatly challenged when reading, remembering and perceiving the information from the digital text contents. So in order to surmount this challenge, text visualization can be used as a recognized approach to accelerate and optimize learning practices. [1] The main purpose of this study is to develop the visualization of extracted text approach on the texts with the subject of "History" and in order to increase the efficiency of learner's comprehension and perceiving of the important portions of the digital text. In order to evaluate the proposed method of visualization an, a pre-project survey and post project experiment will be conducted to measure the influences of visual factors on learning achievement

II. THEORETICAL BACKGROUND FOR VISUALIZATION OF SUBJECTIVE EXTRACTED TEXT

Through the extensive analysis of the relevant text visualization researches, it has been learnt that the theoretical background for the techniques offered by different papers are mostly related to the major branches of psychological, text structure , human computer interaction (HCI) and text structure concepts;

According to the psychological basis, the gestalt laws of perception are highly applicable for visual designing because they are explaining the inclination of human to" group "and "unified" things, and technically defining the human perception based on the scientific rules.[8] the basic notion of the Gestalt Theory is arguing about the structure, arrangement or an overview of some individual things that have greater specific characteristics in a unified form rather than sum of its single elements and discuss about the profound influences of structural changes like spacing and timing on the meaning of an informative content.

As the basic and fundamental objective of text visualization is to enhance the knowledge acquisition, so the cognitive theory of multimedia learning and dual coding theory[7] can buttress this goal in the way that, visual and verbal information are going to be processed in visual and verbal channels separately, while verbal information which is represented in a visual format will be processed in both channels, on the other hand the capacity of the both channels are limited, so the information to be processed should be in a format that the learner has to use both channels for information processing to avert the cognitive overload in the pictorial and verbal channels. [10], [6], [7]

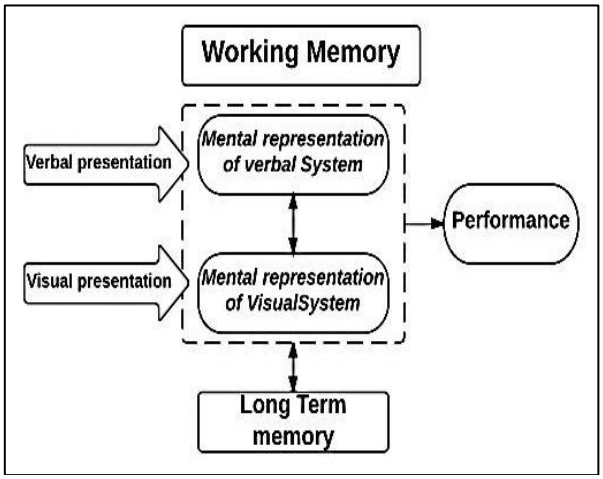

Figure 1: Processing of verbal and visual information in visual and verbal channels[25], [24]

In addition to the above-mentioned psychological influential theories the characteristics of the human brain (left/right) can also be effective in realm of text visualization. according to [13] analysis of the human brain functional structure, left brain is responsible for analytical tasks, investigating for detailed data, doing the logical and definitional jobs, and it is reliant on literal meaning; in contrast, the right side more dependent on images, colours and vision than words, moreover, right brain's functions are more literal rather than being logical, and innately holistic.

The instructional foundation used to visualize the extracted text in order to intensify the comprehension level of the learners can be supported by the typographic effect, text segmentation and spacing, that describe the effective designing and arranging the type in the page. according to [15] the typographic effects have immensely influence on how readers get inspired by reading a text, the main purpose the typographic effects is to render a text into a pleasant, easy to read and attractive format, some of these effects are size, font, colour, or position. In addition based on the main objectives of Human Computer Interaction (HCI) for optimized visualization concepts, some of the useful prospect of visualization and text visualization are derived from the advantages of human computer interaction (HCI) aspects that the most important of them is to provide an effective tool that the user can perceive the needed knowledge with minimal effort (perception characteristic of HCI). As another auxiliary approach that helps users to concentrate of the content is the approach of colour coding; this approach is a graphical method that use the colour to for analysing and visualizing the text according to [11], colours can be effective in this approach for some of their specific characteristics as: first, being visually separate and this attribute make them capable of encoding the different word classes in a sentence or a text content additionally the colours are innately evocative, hence the colour coding approach force the usage of colour meaning and expression based on a criteria(cultural meaning of the colours, psychological expression and …) for the meaning of the colours. [12]

Besides to the above foundations and concepts regarding the structure of text, parse tree generator approach can be applied for the purpose of extracting the portion of interest from text, based on the definition of the parse tree, it is "A concrete syntax tree /parse tree / parsing tree is an ordered, rooted tree that represents the syntactic structure of a string according to some formal grammar". [16] Applying the parse tree algorithm help us to find the role of each words and grammatical dependencies of them in the sentences

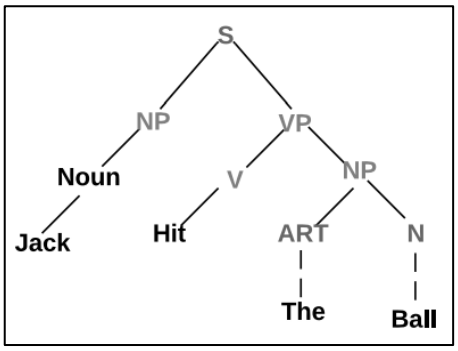

Figure 2: sample of tagging used parse tree[29]

Regarding to the definition of "text visualization" by [1], it is "a tool or method for interpreting visual data fed into a computer and for generating images from complex Multi-dimensional data sets", hence as texts are containing information, therefore visualization of textual material, is a sort of information visualization, that search for the possible approaches for visual representation of a large amount of information.

With consideration to the integration of the text visualization with the empirical researches related to the subjects as: human computer interaction (HCI), perceptual psychology, and cognitive elements, social sciences approaches will be applied as an influential issue for evaluating the different methods of text visualization.

III. RELATED WORK

There are several approaches of text visualization provided based on the psychological, text structural, multimedia and HCI; the method of "Visualization in interactive 3D virtual environment (I3DVE)" proposed by [18] which uses the 3d visualization collaborated with text description to visualize the text material related to science.

Similarly [19] performed an experiment convert the text content into a pictorial single level and multilevel Mind map, The mind map produced by this application has two main attributes: Visualizing the meaning with the suitable image (by using the Google image search engine) ; Secondly,

providing the layout for the input text by using the Microsoft Automatic Graph Layout.

As another analogous example of related work, [21] measures the impact of using comic style comic style reading material on the motivation and attitudes of the Japanese readers called Extensive Reading (ER).A more relevant application is a result of a study conducted by [14] that provide of a proposed visual design guidelines that are mainly emphasis on the enhancement of the text structure and attracting the reader's attention, on the quality of readers learning.

Additionally [12] uses the two major characteristics of the coolers to measure the attractiveness of a texts based on its writing style and to buttress this hypothesis that the different texts genres can be colour coded differently cause of the different writing style.

According to the outcomes of the above studies, the general effectiveness of text visualization can be clustered into: Reader's comprehension [7],[3], reader's concentration on the content, understanding of text structure, attractive learning experience and improvement of learner's attitude and capabilities in using technologies in the field of learning. [4], [5], [25]

IV. PROPOSED METHOD

Visualization of Subjective Extracted Text implemented in order to augment the comprehension level of the users in studding and learning from a text; additionally this method is going to help them to focus on important content of the text. This approach is going to visualize the historical text based on the focus-context techniques like colour coding, and typographic effects on the text. Applying these techniques allow the viewer to inspect the specific part in detail without losing global context Different colours assigned to any the significant extracted part help the user to understand and memorize what content is important, besides the visualizing effects of size and colour attract considerably the attention of the viewer through the presentation of the text.

A. Extraction using parse tree:

Information extraction is typically used as a process for the extraction of a particular type of relationship of interest in a document.

A parse tree is a graphic representation of the a derivation that shows the hierarchical structure of the language based on the standard and formal grammar, parse tree includes the interior nodes labelled by non- terminals of the Grammar and the leaf nodes are labelled by terminals of the grammar,[27] parse tree is conducting the process of parsing in order to determine whether a sentence or a string is grammatically in a suitable formation and to identify the function of each component in a sentence.

As the input text for extraction and visualization is chosen in the title of "history", some text components will get more values in this realm, such as: names/nouns, Numbers, Dates, and Events, the main functionalities of the parse tree generator is helping us to extract the primary information that are highly valuable in a historical text content. The below diagram illustrates the process of syntax analysis and parse tree generation:

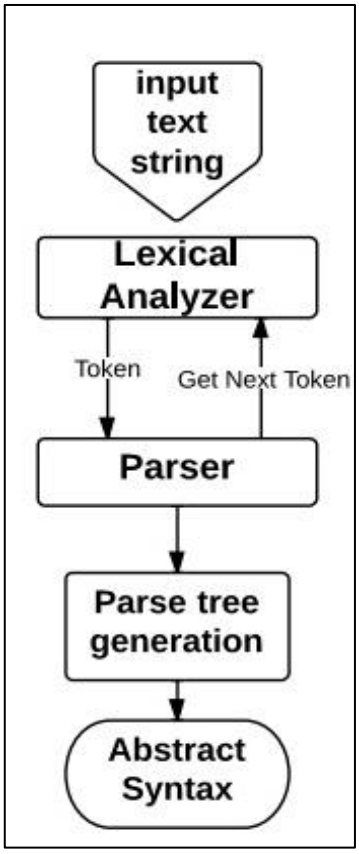

Figure 3: Parse Tree Generation [29]

B. lexical analysis

Lexical analysis in a text, is the process of correlating each word with its corresponding label, for this reason with the help of delimiters (punctuation and blanks), morphological analyser will identify the role and the tag of each word in a text alone and mark each of them with a token symbol; proceeding the recognized tokens will be categorized based on their grammatical class. [26].

The below table illustrates the specifications of each study /researches based on their theory basis and result briefly:

Table 1: specification of the researches based on their basis and the results

Author's Name	*The year of publication*	*The Research Title*	*Use of Established Theories*	*Summary of the Model/research results*	*National Context of the study*
Sung-Hee Jin	2012	Visual design guidelines for improving learning from dynamic and Interactive digital text	▪ Gestalt theory of visual perception. ▪ Cognitive load theory (CLT) [9], [22] ▪ Human Computer Interaction (HCI) concepts for optimized visualization ▪ Dual coding theory	Visualizing the text contents by applying the visualization guidelines (structure design and selective attention guidelines) for content structure comprehension, and assisting the reader to concentrate on the text content.	South Korea
George Korakakis	2012	The impact of 3D visualization types in instructional multimedia Applications for teaching science	▪ Human Computer Interaction (HCI) concepts for optimized visualization ▪ Multimedia modeling and the cognitive theory of multimedia learning[25]	Visualization and learning by a model called "Atomic Orbitals" (the learning virtual environment) which applies interactive 3D virtual environment (I3DVE) for visualization.	Greece
Mohamed Elhoseiny	2012	An Automated System for Mind Map Generation from English Text	▪ Human brain (left/right) functionality ▪ Cognitive load theory (CLT) ▪ Human Computer Interaction (HCI) ▪ Dual coding theory[7]	Mind Maps Automation System which generate and represent a pictorial single level and multilevel Mind map out of a text content	USA
Evan Jones	2010	The Use of Comic Book Style Reading Material in an EFL Extensive Reading Program	▪ Cognitive load theory (CLT) ▪ Dual coding theory ▪ Human brain (left/right) functionality	The result of this experiment illustrated that the visualized text help the reader to have a better comprehension through reading the content, however it doesn't provide any evidence for increasing the reader's motivation toward reading in English through the extensive reading (ER) program.	Japan
Wibke Weber	2007	What Colors Tell About a Text	▪ Color coding approach ▪ Human brain (left/right) functionality	The result of this study indicates that the color-coded text representation will illustrate that the level of understanding can be gained by reading the text based on the darkness or brightness of colors applied on the context.	Germany

C. Part of speech Dependencies and parsing:

In order to identify the dependencies and the parsing task which generates the final parse-tree, the Stanford typed dependencies [28] tokenization processing used to function the lexical analysis; by applying this method the representation of the part of speech will be classified into 55 grammatical relations categories; the final product of this generator is a parse tree that labels each word, the following schema illustrates the result of tokenization for the sample sentence of:

"*While on a visit to the united states in March 2010, Ericdecided not to go back to India*"

The syntax analysis for the syntax for the sample sentence will be generated as below:

```
(ROOT
  (SBAR (`` ``)
    (S
      (SBAR (IN While)
        (FRAG
          (PP (IN on)
            (NP
              (NP (DT a) (NN visit))
              (PP (TO to)
                (NP
                  (NP (DT the) (VBN united) (NNS
states))
                  (PP (IN in)
                    (NP (NNP March) (CD
2010)))))))))
      (NP (NNP Eric))
      (VP (VBD decided)
        (S (RB not)
          (VP (TO to)
            (VP (VB go)
              (PRT (RP back))
              (PP (TO to)
                (NP (NNP India) ('' '')))))))))
```

Figure 4: Example of parsed sentence by Stanford dependencies [29], [28]

As the subject of the input sample text is related to the "Science of History" besides to all parts of speech for extraction, the most important parts of speech and required context regions to be extracted from the text are:

- Nouns, Noun phrase, Noun phrase as adverbial Modifier and Nominal Subject
- Phrasal verb particle
- Numeric Modifier, Numbers
- Adverbial Modifier
- Object of a preposition in order indicate an incident

A. Visualization Approach

Visualization of knowledge are extensively applied in realms of learning and knowledge management in order to assist the learners to process, having access and coping with learning of the complex knowledge and large amount of information. Hence to increase the level of learner's text comprehension the psychological and instructional foundations of visual designing should be reflected effectively in the text.

Visual approaches proposed are mainly aimed at visualization of perceptual organization by uniformly visualize the subjects (part of speech/ role of the words) required by the user , on the other hand application of directive attraction and focal factors can be used to help the learners to focus and concentrate on the content to be learnt.

B. Visualization using colour -coding to help memorizing the extracted parts

with consideration to "visual distinction" as the main characteristics of the colour [11], visualizing the different part of speech can be feasible by the usage of the colour ,for this reason each word class and part of speech will be visualized by an individual colour , based on the grammar used in the extraction method all the word in the text can be grouped into 36 category that the 7 of these main part of speech are coloured by distinct colour and the rest of part of speech that are in minor prior are coloured by black in order to prevent the user from the occurrence of the split attention [21] This occurs due to split when two visual information existing simultaneously to achieve the meaning. Usage of dictionary or glossary to look up the further information increases the probability of split attention occurrence for the learners.

The table below illustrate the colour code schema:

Noun :singular, plural, proper	
Numbers: Cardinal, Date,	
Adverb: WH adverb, Adverb of Time, Places	
Particles	
Adjectives	
Foreign word	
Conjunction: Coordinating Conjunction	

Table 2: Colour coding scheme

C. Typographic effects(spacing, highlight/bold effects, font size/type) to draw the user's attention on the extracted part

according to the instructional basis of text designing, applying the typographic effects [15] is an influential approach to intensify the learner's comprehension from learning material; additionally word spacing will provide an approach for presentation of the text in an internally consistent idea sections [25], [24] on the other hand typographic effects can be beneficial to attract the learner's attention to the specific portions of the text and effectively improve all the cued information. [14]

Besides to the above-mentioned attributes of text designing, it can be mentioned that recruitment of the digital text and its technological characteristics, can be effective in interactive text designing approaches and visualization; which represent the visual form of the required text for the learners and directing their attention. [14]

Below figure illustrates the typographic effects on the text:

Figure 5: Extraction and visualization of Noun (singular, plural, proper) by the proposed method

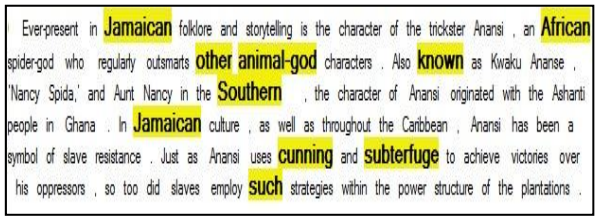

Figure 6: Extraction and visualization of Adjectives) by the proposed method

V. EVALUATION METHOD

The evaluation method proposed for this approach includes a pre-project and post -project surveys and will use the 50 student of University Putra Malaysia as a sample class.

A pre-project questionnaire (in English) will be administrated to the participants, firstly: in order to measure the importance of each "part of speech" in the Historical text and to obtain the most priorities of textual information through reading and learning the historical text. For this reason number of 10 student of "English Literature" assigned to indicate what are the most priorities of textual information through reading and learning the Historical text secondly: participants will be asked about the level of each techniques of visualization (Spacing, effects of colour, typographic cuing) applied on the extracted parts of text based on the subject indicated by the user (Adjectives, adverbs, numbers dates).

a post project experiment will be conducted in order to gauge the participant' s learning efficiency by using the proposed visualization method, in addition to measure the effects of the focal and directive attention design factors used in the text visualization approach; at this level the participants will be asked to read a text (about "History of Malaysia") two times with and without visual effects , proceeding the learning factors like the time of answering , accuracy and consistency of the participant's answers , and their concentration level on the text will be measured by analysing each question sheet after reading the texts (non-visualized and visualized volumes of the texts) .

VI. DISCUSSION

The usage of the visual aids [2], [17] enhances the learner's capability to make connections with the textual word and its meaning. By doing so, this greatly reduces the cognitive load on the working memory (short term memory). Besides applying the approaches of visualization on the digital text attract the learner's attention to fixate on the learning contexts. As well visualization provides an amusing experience of learning conditions for the learners.

Theoretical contribution:

regarding the related theories in the literature review section, the proposed method empirically buttresses the Dual coding theory and Multimedia modelling and the cognitive theory of multimedia learning of [25], [7]; according to this theory The cognitive theory of multimedia learning (CTML) discusses on the idea that learners attempt to build meaningful connections between words and visual aid and that they learn more deeply than they could have with words or visual alone. [23]

Practical contribution:

Visualization of subjective extracted text providing the effective methods of visualizing and designing text based on the psychological foundation; in addition the lexical analysis of the words in the text and put forward potential solutions to obviate the difficulties in learning from complex texts. The holistic attributes of this approach can be in parallel to the principals of the effective visualization and colour designing, whereas this approach is presenting the organized information

and it categorizes pertinent information and assigning a specific colour to each category. Refer to the effective design principals mentioned these factors can assist learners to better understand the relationships between items and their roles in the text. [20], application of colour in information visualization can be influential in drawing the learners' attention in proportion to the significance and importance of information.

VII. CONCLUSION

this study developed the "visualization of subjective extracted text using parse-tree that the main structures of this approach are mainly rely on the psychological foundations and the role of the words and lexical category in the text . the main objectives of the proposed method are1) extract the learner's-required part of speech and portion of the text 2) visualize the extracted part to attract the learners' attention and assist them to focus on learning materials; this is conducted by applying the effect of: typographic cueing, font size modification, colour coding and spacing, on the digital text to be learnt. The overall finding of this study is shown that, firstly using the visual assistance collaborated with the digital text; assist the learners to effectively engage with the mental activities and learning concepts. On the other hand providing the visual aid is going to help learners understand the complex problem efficiently and reduces the cognitive effort required to overcome such problems.

REFERENCES

[1] Sili c, 2010 ,Visualization of Text Streams: A Survey, Department of Electronics, Microelectronics, Computer and Intelligent Systems, University of Zagreb, Faculty of Electrical Engineering and Computing

[2] Huimin Lu, Gang Liu, (1122) Text Visualization and Visual Analytics Based on Multi-layer Topic Maps, Journal of Information & Computational Science 8: 12 (2011) 2459–2464 Available at http://www.joics.com

[3] S.Maravic, Cisar, D. Radosav, R. Pinter , P. Cisar ,(2011),Effectiveness of Program Visualization in Learning in Learning Java: a Case Study with Jeliot 3.Int.J.of Computers, Communications and Control, ISSN 1841-9836,E-ISSN 1841-9844 Vol.VI (2011),No.4 (December),pp.669-682

[4] Hu, Paul J. Hwa, (2012), "Examining the role of learning engagement in technology-mediated learning and its effects on learning effectiveness and satisfaction." Decision Support Systems 53.4 782-792.

[5] Falakmasir, M. Hassan, Mazzola, Nancy Grant. (2012). the Impact of Social Performance Visualization on Students, 12th IEEE International Conference on Advanced Learning Technologies

[6] Richard E. Mayer, 2005, Cognitive Theory of Multimedia Learning, the Cambridge handbook of multimedia learning, Cambridge university press

[7] A. Paivio, 2006, dual coding theory and education, conference on "Pathways to Literacy Achievement for High Poverty Children," The University of Michigan School of Education, September 29-October 1, 2006

[8] L. Graham, 2008, Gestalt Theory in Interactive Media Design, journal of humanities and social sciences, ISSN 1934-7227, volume 2 , issue 1

[9] F.pass, J.E Tuovinen, 2004, Cognitive Load Theory: Instructional Implications of the Interaction between Information Structures and Cognitive Architecture, Instructional Science 32: 1–8, © 2004 Kluwer Academic Publishers. Printed in the Netherlands

[10] S.Olaf; T.Tanja Keller (Eds), 2005, [book]: knowledge and information visualization: searching for synergies, ISBN: 3-540-26921-5 Springer berlin Heidelburg, New York

[11] Hugo Liu, Henry Lieberman, Ted Selker. (2003). A Model of Textual Affect Sensing using Real-World Knowledge. Proceedings of the 2003 International Conference on Intelligent User Interfaces, IUI 20January 12-15, 2003, Miami, FL, USA. ACM 2003, ISBN 1-58113-586-6, pp. 125-132. Miami, Florida.

[12] W. Weber, 2007, Text Visualization – What Colours Tell About a Text, 11th International Conference Information Visualization (IV'07) 0-7695-2900-3/07 © 2007 IEEE

[13] E. Frangenheim, 2005, [book]: reflection on classroom thinking strategies: practical strategies to encourage thinking in your classroom, Sage publication Inc., 6th edition, ISBN 1-4129-1829-4 (Pub)

[14] Sung-Hee Jin, 2012, Visual design guidelines for improving learning from dynamic and interactive digital text, 0360-1315/$ – see front matter 2012 Elsevier Ltd. All rights reserved. http://dx.doi.org/10.1016/j.compedu.2012.12.010

[15] Swann, C. 1991, Language andTypography, London, Lund Humphries

[16] [Book]Modern Programming Languages: A Practical Introduction 2nd Edition (2010) Details: ...Programming, Prentice Hall, 2nd edition, 2010, ISBN-10

[17] B. Bogue& R. Marra (Eds.), 2005, Overview: Visual Spatial Skills Visual Spatial Skills AWE Copyright © 2005, A Product of AWE-Assessing Women in Engineering NSF Grant # 012064 available through: www.aweonline.org

[18] G. Korakakis, E.A. Pavlatou, 2008, and Computers & Education: 3D visualization types in multimedia applications for science learning: A case study for 8th grade students in Greece, 0360-1315/$ - see front matter _ 2008 Elsevier Ltd. All rights reserved. doi:10.1016/j.compedu.2008.09.011

[19] M. Elhoseiny, A. Elgammal, 2012, English2MindMap: an Automated System for MindMap Generation from English Text, 2012 IEEE International Symposium on Multimedia, 978-0-7695-4875-3/12 © 2012 IEEE DOI 10.1109/ISM.2012.103
[20] Maureen C. Stone, 2006, [book]: Field Guide to Digital Colour: Choosing colours for Data Visualization and Copyright 2003 by A K. Peters Ltd
[21] E. Jones, 2010, The Use of Comic Book Style Reading Material in an EFL Extensive Reading Program: A Look at the Changes in Attitude and Motivation to Read in English in a Japanese University, Language Education in Asia, 2010, 1(1), 228-241.http://dx.doi.org/10.5746/LEiA/10/V1/A19/Jones
[22] P. A. Kirschner1, P Ayres2, 2005, Contemporary Cognitive Load Theory Research: The Good, the Bad and the Ugly
[23] S.D Sorden,(2005) Cognitive Theory of Multimedia Learning , Mohave Community College/Northern Arizona University, online accessavailable from:http://sorden.com/portfolio/sorden_draft_multimedia2012.pdf
[24] Baddeley, A.D. (1997) Human memory: Theory and Practice (Revised Edition). Hove: Psychology Press. Baddeley, A.D. (1996). Your Memory: A User's Guide, 3rd Edition. Prion Books, London
[25] R.E Mayer,2005 , Multimedia Learning, Cambridge university press
[26] S. Prado Daud,2010, NLP – LEXICAL ANALYSIS APPLIED TO REQUIREMENTS, Proceedings of the 9th Brazilian Conference on Dynamics Control and their Applications Serra Negra, SP - ISSN 2178-3667
[27] L. Tari et al,, 2009,Parse Tree Database for Information Extraction, IEEE TRANSACTIONS ON KNOWLEDGE & DATA ENGINEERING
[28] M.Catherine de Marne C. D. Manning, 2008,Stanford typed dependencies manual, Revised for the Stanford Parser v. 3.3 in November 2013
[29] S. Abney, 1996, Part-of-Speech Tagging and Partial Parsing