UNIVERSITI PUTRA MALAYSIA

VISUALIZATION OF EXTRACTED GRAMMATICAL ROLE OF WORDS USING PARSE TREE CONVERSION TO IMPROVE UNDERSTANDING OF ENGLISH TEXTS

ERFAN MIRZABEIKI

FSKTM 2014 28

VISUALIZATION OF EXTRACTED GRAMMATICAL ROLE OF WORDS USING PARSE TREE CONVERSION TO IMPROVE UNDERSTANDING OF ENGLISH TEXTS

By

ERFAN MIRZABEIKI

Thesis Submitted to the School of Graduate Studies,Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Master of Science

December 2014

COPYRIGHT

DEDICATION

To My Parents

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfillment of the requirement of the degree of Master Science

VISUALIZATION OF EXTRACTED GRAMMATICAL ROLE OF WORDS USING PARSE TREE CONVERSION TO IMPROVE UNDERSTANDING OF ENGLISH TEXTS

By

ERFAN MIRZABEIKI

December 2014

Chairman : Assoc. Prof. Lili Nurliyana Abdullah, PhD

Faculty : Computer Science and Information Technology

The plethora of English text materials and advancement of digital texts are the factors that have sustained interest in automatic methods for text visualization as an auxiliary tool to improve reading abilities. To better understanding of the English text, firstly, we must be familiar with the English language structure in the text and grammatical role of each word in the sentences and after that, we should be capable of learning, memorizing and understanding the main core of the text. In this study, to identify the grammatical role of the words in text, and improvement of understanding (memorization, recalling, attention extraction, and overall comprehension), an automatic method for grammatical role of the words extraction and visualization is proposed. Our approach identifies the role of each word in the sentences based on generating the dependencies tree and then will extract and visualize each grammatical role of the word in the text based on the userôs requirement. The proposed method for text visualization has the ability of extracting more than 30 grammatical ñrole of the wordsò in the text and capable of visualizing the extracted parts using text colour-coding and applying typographic effects (highlighting, bold/italic effects, and font size and font type) on the text. This method aimed at visualization of perceptual organization by uniformly visualize the subjects (role of the words/ role of the words) selected by the users, on the other hand, model of directive attraction and focal factors going to help users to focus and concentrate on the content to be learnt.

To validate the effectiveness of extraction and visualization of our work, 40 academic members of Universiti Putra Malaysia (UPM) were asked to give their feedbacks after reading a visualized piece of English text using our proposed method. The result illustrated that compared to other subscales of text visualization; the item which was in higher level of efficiency is the combination of selected visualization items including ñfont size/colour/boldò, ñhighlight font size/colourò,

and ñfont style/bold/ colourò format in the text. Therefore, according to the analysis of the expertsô feedback, three groups of visualization: group1 (highlight/size/bold); group 2 (style/colour/size); group 3 (size/colour/bold) performed to design the experiment for this research work to determine the effect of text visualization on user reading abilities. 100 students were the participants of the experiment and they have been assigned into three visualization groups and one control group. The result of Multivariate Analysis Of Variance (MANOVA) revealed that although the text visualization approach for the three groups were more effective on learner's attention extraction, time of skimming, memorization, recalling, and text structure comprehension, the group with combination of (font size/colour/bold style) showed more efficiency compared to other groups. In addition, there is a significant pair wise difference observed between three groups based on reading ability to improve understanding of English text. Moreover, control group scored much lower ratings than the three groups based on understanding abilities (attention extraction, time of skimming, memorization, recalling, and text structure comprehension). Finally, to determine the analysis result of usability factors, all the visualization groups claimed the satisfactory level for each of the three usability scales, which are ease of use, awareness of the structure, and satisfaction of use.

Abstrak tesis yang dibentangkan kepada Senat Universiti Putra Malaysia untuk memenuhi keperluan ijazah untuk Master Sains

VISUALISASI PERANAN NAHU PERKATAAN YANG DIEKSTRAKKAN MENGGUNAKAN TERJEMAHAN PEPOHON HURAIAN UNTUK MENINGKATKAN KEFAHAMAN TEKS BAHASA INGGERIS

Oleh

ERFAN MIRZABEKI

Disember 2014

Pengerusi : Prof. Madya Lili Nurliyana Abdullah, PhD

Fakulti : Sains Komputer Dan Teknologi Maklumat

Kepelbagaian bahan teks Bahasa Inggeris dan kemajuan teks berbentuk digital adalah faktor yang mengekalkan minat kepada kaedah-kaedah automatik visualisasi teks sebagai satu alat penting dalam memperbaiki kebolehan pembelajaran. Untuk lebih memahami teks Bahasa Inggeris, pertama sekali kita perlu biasa dengan struktur Bahasa Inggeris dalam teks dan peranan nahu setiap perkataan dalam ayat dan selepas itu kita perlu mampu mempelajari, mengingati dan memahami intipati teks tersebut. Dalam kajian ini, untuk menangani isu yang disebutkan dalam pengenalpastian peranan nahu dalam perkataan-perkataan dalam teks, dan penambahbaikan kebolehan belajar (daya hafalan, mengingat kembali, pengestrakkan perhatian dan keseluruhan kefahaman), satu kaedah automatik peranan nahu pengestrakkan nahu dan visualisasi perkataan telah dicadangkan. Pendekatan kami untuk mengenalpasti peranan setiap perkataan dalam ayat adalah berdasarkan penjanaan pokok ketergantungan (*dependencies tree*) dan kemudian mengestrak dan menvisualisasi setiap bahagian nahu berdasarkan keperluan pengguna. Kaedah yang disarankan untuk visualisasi teks mempunyai kebolehan mengestrak lebih dari 30 peranan nahu perkataan dalam teks dan juga ia mampu menvisualisasi bahagian-bahagian yang diestrak menggunakan pengkodan warna teks dan mengaplikasikan kesan tipografi (*highlighting*, kesan *bold/italic*, saiz dan jenis fon) pada teks berkenaan. Kaedah ini bertujuan untuk visualisasi secara organisasi dengan melihat keseragaman subjek (bahagian ungkapan/peranan perkataan) yang dipilih oleh pengguna, dan sebaliknya pengaplikasian perhatian secara langsung dan faktor penting yang akan membantu pengguna memberi tumpuan dan perhatian kepada isi kandungan yang akan dipelajari.

Untuk mengesahkan keberkesanan pengestrakkan dan visualisi komponen-komponen ini, 40 ahli akademik Universiti Putra Malaysia (UPM) telah diminta memberi maklumbalas selepas membaca satu teks dalam Bahasa Inggeris mengunakan kaedah

cadangan kami. Keputusan menunjukkan bahawa berbanding dengan sub-skala visualisasi teks; item yang lebih tinggi aras keberkesanannya adalah satu kombinasi item-item visualisasi terpilih termasuklah: ñsaiz fon/warna/*bold*ò, ñsaiz dan warna fon dicerahkan (highlight)ò, dan format ñgaya fon/*bold*/warnaò dalam teks tersebut. Oleh itu, menurut analisis pakar, 3 kumpulan visualisasi: kumpulan 1 (*highlight/size/bold*); kumpulan 2 (*style/colour/size*); kumpulan 3 (*size/colour/bold*) direka bentuk untuk eksperimen ini untuk menentukan kesan visualisasi teks ke atas kebolehfahaman pengguna. 100 orang pelajar menjadi peserta eksperimen dan mereka diagihkan kepada tiga kumpulan visualisasi dan satu kumpulan kawalan. Keputusan *Multivariate Analysis Of Variance* (MANOVA) menunjukkan bahawa walaupun pendekatan visualisasi teks untuk tiga kumpulan adalah lebih efektif dari aspek pengestrakkan perhatian dan masa penyiring pelajar, hafalan, mengingat semula, dan kefahaman struktur teks, kumpulan dengan kombinasi (saiz fon /warna / gaya *bold*) menunjukkan lebih efisien berbanding dengan kumpulan-kumpulan lain. Tambahan pula, terdapat satu perbezaan *pair-wise* yang diperhatikan di antara tiga kumpulan itu berdasarkan kebolehan membaca untuk memperbaiki kefahaman teks Bahasa Inggeris. Seterusnya, kumpulan kawalan mempunyai perkadaran yang lebih rendah dari ketiga-tiga kumpulan berdasarkan kebolehan membaca mereka. Akhir sekali, untuk menentukan keputusan faktor-faktor kebolehgunaan, kesemua kumpulan visualisasi mendakwa bahawa setiap satu dari ketiga-tiga skala kebolehgunaan mempunyai aras yang dikehendaki dan memuaskan iaitu penggunaan yang mudah, kesedaran kepada struktur teks, dan juga kepuasan menggunakannya.

ACKNOWLEDGEMENT

In the name of God, the compassionate, and the most merciful, who endowed me strength and helped me to complete this work.

First and foremost, I offer my sincerest gratitude to my supervisor, Assoc. Prof. Dr. Lili Nurliyana Abdullah, for her excellent guidance, encouragement, continues support, and providing her technical expertise during the period of my study at the University Putra Malaysia. Only the successful completion of this study has become possible due to her supervision. Moreover, I want to appreciate her invaluable advice, and professional comments on this research.

I would like to express my appreciation to my advisor Assoc. Prof. Dr. Mohammad Taufik Abdullah for his generous assistance and his supportive comments for my research that are extremely valuable for the improvement of my progress.

My deepest gratitude to my father, Dr. Mohammad Ali Mirzabeiki and my great mother Pooran Mirzaie for their love, support and warm encouragement.
I also thank my dear brother and sister, Vahid and Mahsa who always give me so much joy and happiness.

Finally, I would like to address a special word of thanks to Ms. Baharak Talebloo for her generous assistance and sparing her precious time for discussions that helped me a lot in arrangement of this work.

Erfan Mirzabeiki
August 2014

I certify that a Thesis Examination Committee has met on 1 December 2014 to conduct the final examination of Erfan Mirzabeiki on his thesis entitled "Visualization of Extracted Grammatical Role of Words using Parse Tree Conversion to Improve Understanding of English Texts" in accordance with the Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Master of Science.

Members of the Thesis Examination Committee were as follows:

Shyamala a/p C. Doraisamy, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Fatimah binti Khalid, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Internal Examiner)

Zulikha Jamaludin, PhD
Professor
School of Computing
Universiti Utara Malaysia
(External Examiner)

ZULKARNAIN ZAINAL, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 19 March 2015

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Master of Science**.** The members of the Supervisory Committee were as follows:

Lili Nurliyana Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Chairman)

Muhamad Taufik Abdullah, PhD
Associate Professor
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
(Member)

BUJANG BIN KIM HUAT, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by Graduate Student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: _____ Date: _____

Name and Matric No.: Erfan Mirzabeiki GS34881

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: _____	Signature: _____
Name of Chairman of Supervisory Committee: _____	Name of Member of Supervisory Committee: _____

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

IE	Information Extraction
IR	Information Retrieval
NLP	Natural Language Processing
MANOVA	Multivariate Analysis Of Variance
DEP	Dependent
AUX	Auxiliary
AUXPASs	Passive Auxiliary
COP	Copula
ARG	Argument
AGENT	Agent
COMP	Complement
ACOMP	Adjectival Complement
CCOMP	Clausal Complement With Internal Subject
XCOMP	Clausal Complement With External Subject
OBJ	Object
DOBJ	Direct Object
IOBJ	Indirect Object
POBJ	Object of Preposition
SUBJ	Subject
NSUBJ	Nominal Subject
NSUBJPASS	Passive Nominal Subject
CSUBJPASS	Passive Clausal Subject
CC	Coordination
CONJ	Conjunct
EXPL	Expletive (Expletive ñThereò)
MOD	Modifier
PREP	Prepositional Modifier
POSS	Possession Modifier
PREDET	Predeterminer
PRECONJ	Preconjunct
VMOD	Reduced Non-Finite Verbal Modifier
MWE	Multiword Expression Modifier
MARK	Marker
ADVMOD	Adverbial Modifier
POSSESSIVE	Possessive Modifier (óS)
PRT	Phrasal Verb Particle
PUNCT	Punctuation
REF	Referent
SDEP	Semantic Dependent
XSUBJ	Controlling Subject
AMOD	Adjectival Modifier
APPOS	Appositional Modifier
ADVCL	Adverbial Clause Modifier
DET	Determiner
NEG	Negation Modifier
RCMOD	Relative Clause Modifier

QUANTMOD	Quantifier Modifier
NN NOUN	Compound Modifier
NPADVMOD	Noun Phrase Adverbial Modifier
TMOD	Temporal Modifier
NUM	Numeric Modifier
NUMBER	Element Of Compound Number
RBS	Adverb, Superlative
RP	Particle
SYM	Symbol
RBR	Adverb, Comparative
UH	Interjection
VB	Verb, Base Form
VBD	Verb, Past Tense
VBG	Verb, Gerund, Present Particle
VBN	Verb Past Participle
VBZ	Verb 3rd Person, Singular Presenter
VBP	Verb Non-3rd Person, Singular Presenter,
WDT	Wh -Determiner
WP	Wh Pronoun
WP$	Possessive Wh -Pronoun
WRB	Wh Adverbs
RB	Adverb
PRP$	Possessive Pronoun
CD	Cardinal Number
CC	Coordination Conjunction
DT	Determiner
EX	Extential There
FW	Foreign Word
IN	Preposition
JJ	Adjectives
JJR	Adjectives, Comparative
JJS	Adjectives, Superlatives
LS	List Item Maker
MD	Modal
NN	Noun, Singular Or Mass
NNS	Noun, Plural
NNP	Proper Noun, Singular
NNPS	Proper Noun, Plural
PDT	Predetermine
POS	Possessive Ending
PRP	Personal Pronoun

CHAPTER 1

INTRODUCTION

1.1.1 Background of the study

With today massive growth of digital textual materials due to rapid scientific advancements, more value has been accrued to the process of acquiring and learning form text. On the other hand, the development of new technologies is a great potential for the usage of textual and visual information in understanding from text contents. When a reader studying a text, the understanding process is highly challenged when memorizing, recalling and perceiving the information and focusing on the contents from the digital text. To understand the English text better, the chief requirement is being familiar with the English language structure in the text and grammatical role of each word in the sentences and the next step is the capability of learning and memorizing and understanding the main core of the text.

The plethora of textual materials is one of the main issues that have sustained interests in automatic methods for extracting and visualizing information from text.

Presentation of text with visual (text visualization) can simplify the information processing by the learners in the way that, verbal information (text) and visual information are going to be processed in two separate verbal and visual system cognitive subsystems. Hence, the memory-enhancement effect of using visuals in the texts is the plus point of using visual and text compare to the single text. (Mayer, 2006)

1.1.1 Text visualization

Text visualization is a branch of information visualization, and it means the usage of computerïsupported, interactive, visual representation of abstract data to enhance the cognitive processing. In another comprehensive definition, text visualization is the visual presentation of natural language text and its relationship. According to definition of text visualization, recruitment of interactive and dynamic efficient characteristics of digital text is effective in order to perform the unified processing of verbal and visual information that is most likely to occur in a circumstance that verbal and visual information are concurrently available in working memory (Baddeley, 1992; Chandler & Sweller, 1991). In addition to the processing of verbal and visual information, visual comprehension is happening, based on a precise interaction between visual perception and higher-order cognitive processing (Hee Jin, 2012; Mayer, 2006).

One of the critical and important prerequisites of implementing an appropriate text visualization method is to firstly decide what kind of information to be

visualized. Information Extraction (IE) and information retrievalsô (IR) fundamentals used to extract information, but IE functionality differs from IR in the way that, the focus of IE is to find the useful information and portions in the document rather than finding the whole documents. IE is one of the models of Natural Language Processing (NLP), as the term implies, the goal is to extract information from text without requiring the end user to read the complete text.

Grammar is the way of formatting the sentences and it plays as the joint of holding the language together and with incorrect usage of grammar, sentences might become meaningless and their message might become unclear. Thus for learners of English as a foreign language, understanding grammar and role of the words within the English text is quite necessary. On the other hand, during understanding of a text, learners might have problem with focusing on important portions and they might have difficulties in text structure comprehension, recalling and memorization of the content where in the text (Dillon & Jobst, 2005; Sung & Mayer, 2012). According to the technological attributes of digital text, it can become interactively and dynamically efficient based upon the requirement of the users, so it can be applied in order to surmount the challenges of reading and learning of a text and to accelerate and optimize understanding practices (Silic, 2010).

1.1.2 Information Extraction

Extraction of information is an model of natural language processing (NLP), as the term indicates; the main objective is to extract information from text content, without requiring the end user to read the whole text. Simply the IEôs main characters are as follow:

- ¾ Take an input text from a document source and extract the fundamental facts the text related to one or more predefined fact types.
- ¾ Indicate each element as a template whose spaces are filled based on what is found in the text source.

According to provided definition of IE, the main goal of this work is to extract the most essential information without requiring the user to read the whole text nonetheless having used of IR system that now tends to be called ñsearch enginesò requires the user to read each document to know the facts reported in it. (Yakushiji *et al*, 2001) However, the objective is to tabulate the facts reported in large amount of text sources. Therefore based on the differences in the goals of IR and IE systems, IE can be used as a fact retrieval service.

The other specific capability of the IE model is that the fact extraction aspects of IE are highly dependent on the userôs specification of what facts they wish to extract.

Synthetic analysis and presentation of language mostly are on the basis of context-free grammar (CFG) formalism, based on this method of analysis; a given

sentenceôs syntactic structure will be presented in a form of tree structure, known as parse tree (Goncalves, 2008).

A parse tree generally includes the constituentôs structure of a block of text and wordôs grammatical class in literature text. Based on the requirement of the grammar learning, some textôs components will get more values in this realm. For the above reason the different grammatical roles of the words and grammar subjects are extractable by using the parse tree, the output of the parse tree is syntactic analysis of every phrase words including the role of the words tagging some of the possible tags are: adjective, adverb, article, determiner, conjunction, pronoun, proper noun, and verb. So the main functionalities of the parse tree generator are helping us to extract the primary information that is highly valuable in the text.

In order to identify the dependencies and the parsing task which generates the final parse-tree, the Stanford typed dependencies tokenization processing partly implemented to perform the lexical analysis and generating the trees for each sentence in the text; by applying this method the representation of the role of the words is categorized.

In order to intensify the understanding abilities, comprehension and usersô attention extraction on the specific portion of the text, some basis and foundations are mostly in common among the visualization techniques that are related to technological aspects of digital text, multimedia, typographic cueing (visual effects). Visual approaches and techniques used for visualization of role of the words in the text are divided into two parts. First part is the visualization of extracted roles using colour-coding to help memorizing, recalling the extracted parts, and increasing the attractiveness of the text. Second part is visualization of text by typographic effects (spacing, highlight, bold effects, font size, and Font type) to draw the user's attention on the extracted part. The proposed method is mainly aimed at visualization of perceptual organization by uniformly visualize the subjects (grammatical role of the words) which are selected by the user. On the other hand, model of attention extraction and focal factors can be used to help the learners to focus and concentrate on the content to be learnt. (Chuan *et al.*, 2012)

1.2 Problem statement

Based on the significant technological characteristics of the digital text, one of its main potential is the capability of getting the visual and interactive design in itself. Increasingly accretion of reading and understanding from digital text, has led this issue become more significant to determine how digital text visualization can improve reading and understanding abilities.

The main challenges of text visualization are visual text design and text extraction. Text design transforms the text contentôs information into a new visual representation for enhancement of text analysis and browsing (Chuan *et al*., 2012; Wise et al., 1995). To deal with large amount of information specifically a text content, readers are expected to explore the important parts of the text. Thus the identification and extraction of the important part of the text, and subjectively extraction of the important elements of a sentence, that the user must understand is another challenge. During reading of an English text, Extraction of userôs attention on the essential contents, understanding the text structure and recognizing where readers are in the text are challenging factors. Additionally memorization, recalling, difficulties in focusing on important parts and overall text comprehension other problematic issues while reading a text. (Hee Jin, 2012). Plus to the above matters usability and user-directed structure of the text visualizationôs model must be considered during the implementation of the model to extract the required role of the word in the text and represent the required visual form of digital text based on the userôs choice in a suitable manner. According to the result of studies done by Dillon & Jobst (2005); Sung & Mayer, (2012); Loman & May Dyson & Gregory, (2002); Hee Jin, (2012) several critical design principles based on, multimedia, text designing, instructional, and psychological foundations have been suggested. Regarding the current researches, text visualization principles provide useful recommendation but have limitation in:

i. Visualizing and extracting the specific and important required portions of the text upon userôs selection
ii. Implementation of syntax analysis, and grammatical role of the words extraction from natural language text
iii. Function the visualization and IE task without losing or eliminating any part of the text
iv. Interactive performing of the IE and text visualization on digital text

Based on challenges of reading and understanding of an English text, it is highly required to figure out the best way to help the userôs to better deal with the reading content. To this point, the current research work attempts to discover, ñHow text visualization and subjective text extraction can help the readers to improve their reading and understanding abilities. This research is going to develop a model for ñvisualizing role of the words using parse-tree extraction to improve understanding of English text (recall, memorization, time of skimming, attention extraction) as well as validating components implemented, the functioning and overall effectiveness of the model.

1.3 Project objectives

The main objective of this study is to propose a method for text visualization that automatically extract and visualize the grammatical role of the words in the text, to help readers to effectively improve their understanding abilities specifically, their comprehension level of English text structure, focus on the content, perceiving the important extracted portions of the digital text in an efficient manner.

The main purpose of this research is to achieve the following objectives:

1) To implement an appropriate syntax analysis and dependencies checking approach for extraction of grammatical role of the words in the text
2) To implement the visualization approach to visualize the extracted grammatical role of the words in the English text for improvement of English text understanding
3) To increase the usability level of text visualization model

1.4.1 Research questions

1. What is the efficient approaches for dependencies checking and word tagging through the text sô elements
2. What is an appropriate method for text mining and parts of speech extraction
3. What are the most important parts of speech to be visualized?
4. What are the most efficient visualization effects should be implemented on the text?
5. Which of the proposed text visualization effects, influence on the userôs reading abilities
   a) Which of the visual effects (font size, colour, bold/highlight style and size/highlight) influence the userôs focus /attention extraction and time of skimming?
   b) Which of the visual effects (font size, colour, bold/highlight style and size/highlight) influence the userôs memorization and recalling?
   c) Which of the visual effects (font size, colour, bold/highlight style and size/highlight) influence the userôs visual/structure understanding and comprehension?
6. What is the overall usability level of the proposed approach?

1.5 Scope of research

The scope of this research is to extract and visualize the grammatical role of the words in the English text. According to the importance of understanding from text and perceiving English text structure for learners of English as foreign language, this study will hereafter focuses on text extraction, visualization and automatic manipulation of text elements because of their effective characteristics in learning, text comprehension, recalling, memorization and directive attention.

<table>
<tr><td rowspan="2">Extraction</td><td>Role of the words extraction</td><td>Related researches</td><td>Purpose</td></tr>
<tr><td>1. Generation of the parse tree and lexical analyzer for each block of the text
2. Extraction
ƒ Nouns
ƒ Verbs(Past, Present, Past participle)
ƒ Adjectives
ƒ Adverbs
ƒ WH determiner
ƒ Pronouns and 25 more grammatical roles of the words in the text</td><td>ElHosseiny,(2012)
Chuan et al.,(2012)
Tohidi, (2011)
Weber, (2007)</td><td>Overall text structure comprehension</td></tr>
<tr><td rowspan="4">Visualization</td><td>Colour effect</td><td>Related researches</td><td>Purpose</td></tr>
<tr><td>Applying the two major colour-based effects on the extracted part
ƒ Extracted text colouring,
ƒ Extracted highlighting effect</td><td>Hee Jin, (2012)
Weber, (2007)
Chuan et al.,(2012)</td><td>ƒ Attention extraction
ƒ Recalling
ƒ Memorization</td></tr>
<tr><td>Typographic effects</td><td>Related researches</td><td>Purpose</td></tr>
<tr><td>Applying the typographic effects related to the figure of text:
ƒ Bold/Italic
ƒ Effects,
ƒ Font size
ƒ Font style
ƒ Underline</td><td>Hee Jin, (2012)
Weber, (2007)</td><td>ƒ Attention extraction
ƒ Recalling
ƒ Memorization</td></tr>
</table>

1.6 Thesis organization

The rest of this dissertation organized as follows. Chapter 2 presents the background and fundamental concepts of the text visualization and subjective text extraction. Further, the theoretical contributions of text visualization and research trends for different text visualization methods are described. Moreover, in the critical review section the different aspects of text visualization models are explained.

Chapter 3 presents the methodology of this research work. The main framework of the research for visualization of extracted role of the words from English text is provided in this chapter. Additionally this chapter explains the research design and identifies the procedure of collection, measurement and analysis of data for this project.

Chapter 4 presents and clarifies the phases of the proposed approach for visualization of extracted role of the words from English text. Our proposed approach has the main following phases; (i) parse tree generation, (ii) role of the words extraction and (iii) visualization.

Chapter 5 presents the analysis and finding of the research. Modelôs components data analysis and examinations are provided in this chapter. The overall evaluation and effectiveness of the proposed approach which is ñvisualizing role of the words using parse tree extraction to improve understanding of English textò explained in this chapter

Chapter 6 presents the conclusion and further suggestion for the project.

1.7 Keywords definition

Stanford Type dependency: The Stanford typed dependencies representation is designed by Marnee & Manning, (2008) to provide a simplified explanation of the grammatical relationships in a text blocks that can be perceived and effectively used by user without linguistic knowledge who want to extract textual relationships.

Text visualization: In this research text, visualization means model of typographic effects on the text, which are text size, font style, and font size, colour effect (text colouring and highlighting).

Subjective Text Extraction: is to extract the at least 30 grammatical role of the words (verb, adjective, adverb, noun, and 30 grammatical role of the words Table (4.1) in English texts based on the user selection.

REFERENCES

Abney, S. (1997). Part-of-speech tagging and partial parsing. In Corpus-based methods in language and speech processing (pp. 118-136). Springer Netherlands.

Ary, D., Jacobs, L., Sorensen, C., & Walker, D. (2013). Introduction to research in education. Cengage Learning.

Baddeley, A. D. (1998) Human memory: Theory and Practice (Revised Edition). A User's Guide, 3rd Edition. London, Prion Books, Hove: Psychology Press.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature reviews neuroscience*, 4(10), 829-839.

Beck, C. R. (1991). Strategies for Cueing Visual Information: Research Findings and Instructional Design Implications. *Educational Technology*, 31(3), 16-20.Chicago

Ben-Dov, M., & Feldman, R. (2005). Text mining and information extraction. In*Data Mining and Knowledge Discovery Handbook* (pp. 801-831). Springer US.

Bresciani, S., Blackwell, A. F., & Eppler, M. (2008, January). A Collaborative Dimensions Framework: Understanding the mediating role of conceptual visualizations in collaborative knowledge work. *In Hawaii International Conference on System Sciences, Proceedings of the 41st Annual (pp. 364-364).* IEEE.

Chang, D., & Nesbitt, K. V. (2006, April). Developing Gestalt-based design guidelines for multi-sensory displays. *In Proceedings of the 2005 NICTA-HCSNet Multimodal User Interaction Workshop*-Volume 57 (pp. 9-16). Australian Computer Society, Inc..

Chang, D., Dooley, L., & Tuovinen, J. E. (2002, July). Gestalt theory in visual screen design: a new look at an old subject. *In Proceedings of the Seventh world conference on computers in education conference on Computers in education: Australian topics-Volume 8 (pp. 5-12).* Australian Computer Society, Inc.

Chuang, J., Manning, C. D., & Heer, J. (2012). ñWithout the clutter of unimportant wordsò: *Descriptive key phrases for text visualization.* ACM Transactions on Computer-Human Interaction (TOCHI), 19(3), 19.

Cimiano, P., Ladwig, G., & Staab, S. (2005, May). Gimme'the context: context-driven automatic semantic annotation with C-PANKOW. *In Proceedings of the 14th international conference on World Wide Web* (pp. 332-341). ACM.

Cisar, S. M., Radosav, D., Pinter, R., Cisar, P., Radosav, D., & Cisar, P. (2011). Effectiveness of program visualization in learning java: a case study with jeliot 3. *International Journal of Computers Communications & Control*,6(4), 669-682.

Cohen, D. K. (1996). Standards-based school reform: Policy, practice, and performance. In H. F. Ladd (Ed.), *Holding Schools Accountable, Washington, D.C: The Brookings Institution*.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research.Educational *Evaluation and Policy Analysis,* 25(2), 119-142.

Daud, S. P., & Ribeiro, C. H. C. (2010). NLPïLEXICAL ANALYSIS APPLIED TO REQUIREMENTS. *In Proceedings of the 9th Brazilian Conference on Dynamics Control and their Models Serra Negra,* SP-ISSN (pp. 2178-3667).

De Marneffe, M. C., MacCartney, B., & Manning, C. D. (2006, May). Generating typed dependency parses from phrase structure parses. InProceedings of LREC *(Vol. 6, pp. 449-454).*

Dillon, A., & Jobst, J. (2005). Multimedia learning with hypermedia. *The Cambridge handbook of multimedia learning*, 569-588. (pp. 569ï588). Cambridge, New York

Dyson, M. C., & Gregory, J. (2002). Typographic cueing on screen. Visible Language, 36(3), 326.

Frangenheim, E. (2005). Reflections on classroom thinking strategies: Practical strategies to encourage thinking in your classroom. Paul Chapman Educational Publishing.

Elhoseiny, M., & Elgammal, A. (2012, December). Comput. Sci. Dept., Rutgers, State Univ. of New Jersey, Piscataway, NJ, USA. *In Multimedia (ISM), 2012 IEEE International Symposium on (pp. 326-331*). IEEE.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. Instructional science, 32(1), 1-8.

Falakmasir, M. H., Hsiao, I. H., Mazzola, L., Grant, N., & Brusilovsky, P. (2012, July). The Impact of Social Performance Visualization on Students. In Advanced Learning Technologies (ICALT), 2012 *IEEE 12th International Conference on (pp. 565-569).* IEEE.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). How to design and evaluate research in education (Vol. 8). New York: McGraw-Hill.

Gonçalves, T., & Quaresma, P. (2008, October). Using linguistic information to classify portuguese text documents. In Artificial Intelligence, 2008. MICAI'08. Seventh Mexican *International Conference on (pp. 94-100).* IEEE.

Grishman, R. (1997). Information extraction: Techniques and challenges. In Information Extraction: *A Multidisciplinary Approach to an Emerging Information Technology* (pp. 10-27*).* Springer Berlin Heidelberg.

Hu, P. J. H., & Hui, W. (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.

Wise, J. A., Thomas, J. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., & Crow, V. (1995, October). Visualizing the non-visual: spatial analysis and interaction with information from text documents. *In Information Visualization, 1995. Proceedings. (pp. 51-58).* IEEE.

Jacko, J. A. (Ed.). (2009). Human-Computer Interaction. Interacting in Various Model Domains: 13th *International Conference, HCI International 2009, San Diego, CA, USA, July 19-24, 2009, Proceedings. Springer.*

Jin, S. H. (2013). Visual design guidelines for improving learning from dynamic and interactive digital text. *Computers & Education*, 63, 248-258.

Jones, E. (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*, 1(1), 228-241.

Kirschner, P. A., Ayres, P., & Chandler, P. (2011). Contemporary cognitive load theory research: The good, the bad and the ugly. *Computers in Human Behavior*, 27(1), 99-105.

Korakakis, G., Pavlatou, E. A., Palyvos, J. A., & Spyrellis, N. (2009). 3D visualization types in multimedia models for science learning: A case study for 8th grade students in Greece. *Computers & Education*, 52(2), 390-401.

Graham, L. (2008). Gestalt theory in interactive media design. *Journal of Humanities & Social Sciences, 2(s1).*

LaViola, J. J. (2007). Advances in mathematical sketching: Moving toward the paradigm's full potential. *Computer Graphics and Models*, IEEE, 27(1), 38-48.

Liu, H., Lieberman, H., & Selker, T. (2003, January). A model of textual affect sensing using real-world knowledge. *In Proceedings of the 8th international conference on Intelligent user interfaces* (pp. 125-132). ACM.

Lu, H., & Liu, G. (2011). Text visualization and visual analytics based on multi-layer topic maps. *J. Information and Computational Science, 8(12), 2459-2464.*

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. Journal of Marketing, 71(1), 160-177.

Cer, D. M., De Marneffe, M. C., Jurafsky, D., & Manning, C. D. (2010, May). Parsing to Stanford Dependencies: Trade-offs between Speed and Accuracy. In LREC.

De Marneffe, M. C., & Manning, C. D. (2010). Stanford typed dependencies manual, 2008. URL: http://nlp. stanford. edu/software/dependencies_manual. pdf [access date: 30h].

De Marneffe, M. C., & Manning, C. D. (2008). Stanford typed dependencies manual. URL http://nlp. stanford. edu/software/dependencies manual. pdf.

De Marneffe, M. C., & Manning, C. D. (2008, August). The Stanford typed dependencies representation. In Coling 2008: *Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation (pp. 1-8). Association for Computational Linguistics.*

Marra, R. M., Rodgers, K. A., Shen, D., & Bogue, B. (2009). Women Engineering Students and Self Efficacy: A Multi Year, Multi Institution Study of Women Engineering Student Self Efficacy. *Journal of Engineering Education*, 98(1), 27-38.

Stone, M. C. (2006). book]: Field Guide to Digital Colour: Choosing colours for Data Visualization and Copyright 2003 by A K. Peters Ltd.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. The Cambridge handbook of multimedia learning, 31-48.

Moens, M. F. (2007). Summarizing court decisions. *Information processing & management*, 43(6), 1748-1764.

Chatterjee, P., & Design, H. I. (2010). Entertainment, engagement and education in e-learning. Training & Management Development Methods, 24(2), 601-21.

Paivio, A. (2006). Dual coding theory and education. Ponencia presentada en: Pathways to Literacy Achievement for High Poverty Children, University of

Michigan. Recuperado de: http://www. umich. edu/~ rdytolrn/pathwaysconference/pathways. html.

Palmer, S. E., Brooks, J. L., & Nelson, R. (2003). When does grouping happen?. Acta psychologica, 114(3), 311-330.

Reis, A., Faísca, L., Ingvar, M., & Petersson, K. M. (2006). Colour makes a difference: Two-dimensional object naming in literate and illiterate subjects. Brain and Cognition, 60(1), 49-54.

Sagiv, N., Heer, J., & Robertson, L. (2006). Does binding of synesthetic colour to the evoking grapheme require attention?. Cortex, 42(2), 232-242.

Siddharthan, A. (2006). Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1), 77-109.

ĠiliÍ, A., & Baġiİ, B. D. (2010). Visualization of text streams: A survey. In Knowledge-Based and Intelligent Information and Engineering Systems (pp. 31-43*). Springer Berlin Heidelberg.*

Sorden, S. D. (2005). A cognitive approach to instructional design for multimedia learning. Informing Science: *International Journal of an Emerging Transdiscipline*, 8, 263-279.

Sung, E., & Mayer, R. E. (2012). Affective impact of navigational and signaling aids to e-learning. *Computers in human behavior*, 28(2), 473-483.

Swann, C. (1991). [Book:] Language and typography. John Wiley & Sons Incorporated.

Tari, L., Tu, P. H., Hakenberg, J., Chen, Y., Son, T. C., Gonzalez, G., & Baral, C. (2012). Incremental information extraction using relational databases. Knowledge and Data Engineering, IEEE Transactions on, 24(1), 86-99.

Tergan, S. O., & Keller, T. (Eds.). (2005). *Knowledge and information visualization: Searching for synergies* (Vol. 3426). Springer.

Tohidi, H., Ibrahim, H., & Azmi, M. A. (2012). Statistical character-based syntax similarity measurement for detecting biomedical syntax variations through named entity recognition. *In Networked Digital Technologies (pp. 164-178).* Springer Berlin Heidelberg.

Tufte, E. R., & Graves-Morris, P. R. (1983). *The visual display of quantitative information* (Vol. 2). Cheshire, CT: Graphics press.

Turmo, J., Ageno, A., & Català, N. (2006). Adaptive information extraction.ACM *Computing Surveys (CSUR)*, 38(2), 4.

Webber.A.D , *Modern Programming Languages: A Practical Introduction* 2nd Edition (2010) Details: Programming, Prentice Hall, 2nd edition, 2010, ISBN-10

Weber, W. (2007, July). Text Visualization-What colours tell about a text. In Information Visualization, 2007. IV'07. *11th International Conference (pp. 354-362).* IEEE.

Yakushiji, A., Tateisi, Y., Miyao, Y., & Jun'ichi Tsujii. (2001, January). Event extraction from biomedical papers using a full parser. *In Pacific Symposium on Biocomputing* (Vol. 6, pp. 408-419).