Lab II - Product Specification Outline

CS 411W Lab II

Prototype Product Specification
For
CLASH

Professor Janet Brunelle
Professor Hill Price

Prepared by: Francia Molina

Date: 4/20/2015

Table of Contents

List of Figures

1 Introduction

There is a lack of software designed to improve the reading speed and comprehension specifically for English as a Second Language (ESL) students. According to Greg Raver-Lampman, an instructor at Old Dominion University (ODU) English Language Center (ELC), ESL readers tend to be word-by-word readers (Lampman). This issue effects ESL readers across the country not just at ODU. The current teaching style followed by professors is to write a sentence on the board, then highlight each part of speech. It has been proven that color stimulates learning, color relieves eye fatigue, increase information retention, increase productivity and accuracy (Engelbrecht, K). Reading assignments are provided to students to increase reading comprehension and reading speed. It is believed that the more exposure a student has to language through reading, the greater the possibilities that overall language proficiency will increase (Anderson, N.). The issue correlated with this is that in efforts to assist students to increase their reading, teachers overemphasize accuracy. When this occurs, reading fluency is impeded (Anderson, N.). It has been proven in sentence recall experiments that lexical bundles were remembered and read more efficiently than word by word reading. The experiments suggest that regular multi-word sequences, lexical bundles, leave memory traces in the brain allowing for better comprehension (Tremblay, Derwing, Libben, Westbury). Lexical bundles are a group of words that occur repeatedly together, representing one thought.

1.1 Purpose

Color Lexical Analysis and Slash Handler (CLASH) is a web interface program. CLASH is a Single Page Application (SPA) that has two modules, COLRS and SLASH. CLASH's primarily focus is to be used by English as a Second Language students or English Language Learners. The COLRS module colorizes each part of speech (POS) in a text document with a designated color to increase understanding of sentence structure and grammar. This is done by an algorithm developed to recognize the following parts on speech: noun, pronoun, adverb, verb, preposition, adjective, and conjunction. The COLRS module takes the provided text and parses it according to the parts of speech found in the sentences. The user will be able to select which of the POS to be colorized. All non-selected POS will remain black.

The SLASH module will parse given text into lexical bundles to help increase reading speed and comprehension The SLASH module takes the text provided and breaks it up into groups of three to five words known, as lexical bundles. A lexical bundle is a group of words that occur repeatedly together, representing one thought. The module will have a feature known as, Slash Handler. Slash Handler will stream one lexical bundle at the time with ability to control speed of display time. The speed controls are as follows: Play, Pause, Speed Display, and Speed Setting.

1.2 Scope

CLASH is a combination of two teaching methods. One is the use of color to stimulate learning, by relieving eye fatigue resulting in an increase of information retention (Engelbrecht, K). The latter is reading by lexical bundles, inspired by the Japanese technique of teaching English (Nishida, H). CLASH is intended to aid instructors at Old Dominion University (ODU) English Language Center (ELC) and English as a Second Language (ESL) students. The program will have an esthetically pleasing interface that is simple to navigate. The COLRS module will focus on the following POS: noun, pronoun, adverb, verb, preposition, adjective, and conjunction. The CLASH prototype will have the ability to parse text copied from a source that will be pasted into text form, modify and store previously parsed documents, display chosen POS in a unique color, identify lexical bundles and insert slashes. CLASH will also have an Exception List. The Exception List is a list of commonly used expressions that would otherwise be incorrectly parsed and tagged. There will be three type of user roles: Students, Instructor, and Administrator. Users will have ability to print documents with slashes inserted.

1.3 Definitions, Acronyms, and Abbreviations

CLASH - Color Lexical Analysis algorithm and Slash Handler
Client Side – The user-interface of CLASH
COLRS – Colored Organized Lexical Recognition Software
Document Processor – A Server Side component responsible for processing the text entered by an Instructor user type.
ELC – English Learning Center
ELL - English Language Learners
ESL – English as second language
GUI - Graphic User Interface
HTML - HyperText Markup Language
IBT – International Benchmark Test
Intensive English Program – A short and intensive English language training program offered by US colleges and universities to improve the English language skills of international students who did meet the minimum TOEFL scores for typical enrollment.
JS – JavaScript
JSON – JavaScript Object Notation. A nested data structure commonly used to pass data between a server and a client.
Lexical Bundle – a group of words that occur repeatedly together within the same register
MFCD – Major Functional Component Diagram.
NLP – Natural Language Processing
NLTK – A suite of libraries and programs for symbolic and statistical natural language processing (NLP).
Node.js – an open source, cross-platform run-time environment for server-side and networking application.
POS – Parts of Speech
Server Side – The back-end of the CLASH system responsible text processing, the database, user-authentication, and web-hosting.
SLASH – Aspect of CLASH that displays slashed text
Slash Player – Aspect of CLASH that displays a text stream showing one lexical bundle, of three to five words, at a time with the feature of speed control for display time.
Software as a Service (SaaS) – Software distribution model in which applications are hosted by a vendor or service provider and made available to customers over a network, typically the Internet.
Token: Text that has been processed into individual words by the Document Processor
SPA – single page application, is a highly responsive web application that fits on a single page and does not reload as the web page changes states.
Spreeder – Speed reading tool; www.spreeder.com
TOEFL – Test of English as a Foreign Language
Ubuntu – a Debian-based Linux operating system
VM – Virtual Machine

1.4 References

Anderson, N. (1999, April 1). Improving Reading Speed - Activities for the Classroom. Retrieved February 1, 2015, from http://dosfan.lib.uic.edu/usia/E-USIA/forum/vols/vol37/no2/p2.htm

Engelbrecht, K. (2003). The Impact of Color on Learning (p. 5). Chicago, Illinois: Perkins & Will.

English Proficiency. (2015, February 2). Retrieved February 6, 2015, from https://www.odu.edu/content/odu/admission/proficiency.html

Improve Reading Speed and Comprehension. (2006, January 1). Retrieved February 25, 2015, from http://spreeder.com/

Francia Molina Lab 1 CS 411 (Francia's Lab- Personal)

Monarch Diversity. (n.d.). Retrieved February 6, 2015, from https://www.odu.edu/admission/international/global

Nishida, H. (2013). The Influence of Chunking on Reading Comprehension: Investigating the Acquisition of Chunking Skill. *THE JOURNAL OF ASIA TEFL, Vol.10*(No. 4), Pp. 163-183.

Raver-Lampman, Greg. (2014. August). Personal Interview.

The Condition of Education 2014. (2014, January 1). Retrieved February 6, 2015, from http://nces.ed.gov/fastfacts/display.asp?id=96

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011, January 15). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. Retrieved December 10, 2014.

1.5 Overview

This product specification provides the capabilities and features of the CLASH prototype. The information provided in the remaining sections of this document includes a detailed description of the hardware and software of the CLASH prototype; the key features of the prototype; the parameters that will be used to control, manage, or establish that feature; and the performance characteristics of that feature in terms of outputs, displays, and user interaction.

2 General Description

CLASH is a modified version of a Single Page Application. A SPA is a web application that is on a single web page that provides a desktop application appearance that does not reload as the web page changes states. CLASH will be built completely in JavaScript (JS), meaning each component and operation will use JavaScript. CLASH will have no software to install by the user. All functions will be accessible through menus on the website. The CLASH prototype will have the ability to parse text copy and pasted into form, modify and store previously parsed documents, display chosen POS in a unique color, identify lexical bundles and insert slashes. CLASH will also have an Exception List. The Exception List is a list of commonly used expressions that would otherwise be incorrectly parsed and tagged. There will be three types of users: Students, Instructor, and Administrator. The Administrator role will have the ability to edit, add, or remove anything in the system.

Users will have ability to print documents with slashes inserted. This paragraph orients the reader to what the product (prototype) looks like.

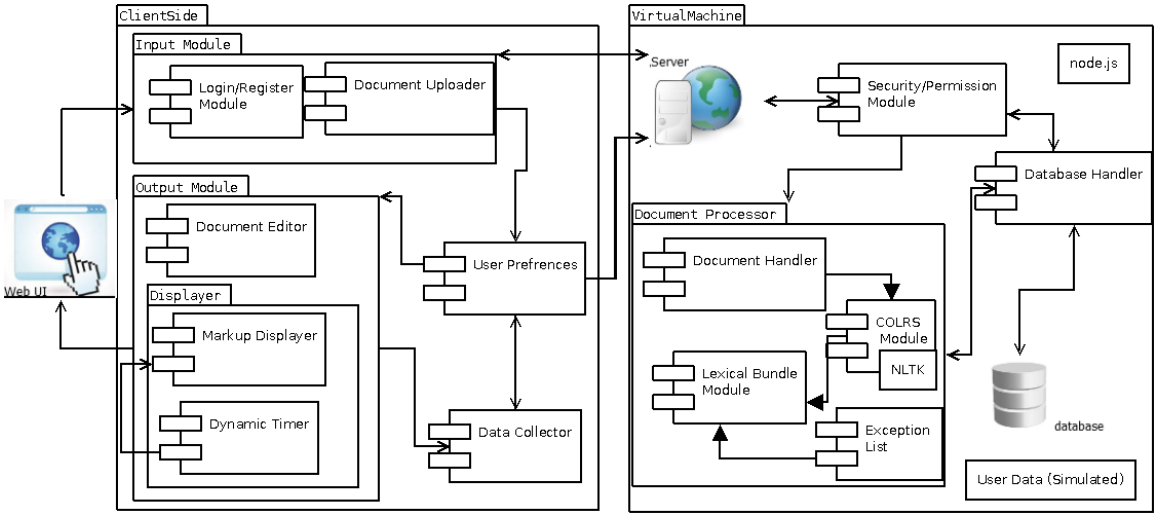
2.1 Prototype Architecture Description

CLASH will be a web-based application hosted on a server and will be accessed through any supported web browser. Hardware components of CLASH are a server, and a database on the server end. Users do not have any special hardware requirements. The user will need a computer with Internet or any Internet enabled device to access the application. There are three major software components that make up CLASH; the COLRS module, SLASH Module, and Slash Player. The COLRS Module uses the Natural Language Tool Kit (NLTK). NLTK is a collection of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python programming language. The NLP software will parse and tokenize the inputted document that will generate the POS tags. Once the POS tags are generated, CLASH will be able to colorize the tags created and insert slashes according to the module selected. The SLASH Module will take the tokenized stream from the COLRS Module and using the slash algorithm break the sentences into 3 to 5 word lexical bundles. There will be a feature known as, the exception list that will improve the accuracy of the final output. The exception list is a populated list of commonly known lexical bundles, such as banana pudding, that can be incorrectly slashed by CLASH. Teachers have the ability to update the exception list

CLASH is comprised of the following major components:

- COLRS Module: Uses open source Natural Language Processing (NLP) to tokenize and parse the input document. The results will be the POS tags that will later be attached to an independent color for POS colorization. The output will be the stream of text with a token and corresponding tag pair.
- SLASH Module: Will use the tokenized tags to break sentence into smaller comprehensible lexical bundles. The Slash Module will have a feature known as Slash Handler.
- Slash Handler: Will display the lexical bundles in a stream with control speed features. The Slash Handler will display one lexical bundle at a time.
- Exception List: The Exception List is a populated list of commonly known lexical bundles that can be incorrectly slashed by CLASH.

Figure 1. CLASH Prototype Major Function Component Diagram.



2.2 Prototype Functional Description

The CLASH prototype has two modules, the COLRS module and SLASH module. The COLRS module will demonstrate the ability to identify parts of speech, distinguishing amongst the various POS with color. The SLASH module will have a feature named Slash Handler. The Slash Handler will display lexical bundles one at a time, and allow the user the ability to increasing and decreasing the lexical bundle stream. The SLASH module will identify lexical bundles through a paragraph and display them by the insertion of slashes. CLASH will be able to create and update the exception list in the SLASH module. The exception list is a list that will hold lexical bundles that are common and reoccur multiple times. CLASH will use a traditional database and Node.js as the web and application server. There will be three user roles: Student, Instructor, and Administrator. The Student role will be able to select type of view, COLRS or SLASH, POS colorized, and control reading speed in Slash Handler. The Instructor role will be able to add and remove Students as users, edit slashed documents, upload documents, and view documents in COLRS and SLASH. The Instructor has all the abilities of the Student roles. The Administrator has all the abilities of the Instructor and Student roles, with the exception of the ability to add and or remove Instructors as users.

Appendix

- No additional equipment, software, and other materials required for the prototype to be functional.
