Lab II - Product Specification Outline

CS 411W Lab II

Prototype Product Specification
For
CLASH

Professor Janet Brunelle
Professor Hill Price

Prepared by: Artem Fisan

Date: 04/20/2015

Table of Contents

1 Introduction

In Old Dominion University (ODU), Greg Raver-Lampman, an instructor at the English Language Center (ELC) has done a case study on English as a Second Language (ESL) students and finds that they tend to shows fixation and regression problems as well as being word-by-word readers (site his pp). Lampman's case study shows that out of fifteen Advance I students at the ELC, 87% of the students read between 69-137 Words per Minute (WPM) (site his pp). In 2011-2012 school year, there were an estimated total of 4.4 million ESL students in public schools (English Language Learners). Students have different learning styles, different abilities and learning pace. This causes difficulties for professors teaching English as a foreign language because they are unable to identify the correct level of teaching for all the students (A Critical Study). Old Dominion University has an entire department focused on addressing this problem, though have an outdated method of teaching reading and grammar.

1.1 Purpose

CLASH is a Single Page Application (SPA) that contains two modules COLRS and SLASH. The COLRS module colors seven parts of speech (POS) of a provided text using NLTK. The POS that will be used are adjective, adverb, conjunction, noun, preposition, pronoun, and verb. The SLASH module inserts a "/" in between lexical bundles. Lexical bundles are 3 to 5 word group of words, representing one thought. Slash Handle displays lexical bundles one at the time with the ability to change display speed via a menu. Targeted community of our product are English as Second Language Students (ESL).

1.2 Scope

The objectives are to increase English comprehension and reading speed. A SPA is an ideal platform for CLASH because it offers the abilities of a desktop application via web browser. Our goal is to provide most accurate slashing and coloring to increase English comprehension and reading speed. Prototype will contain multiple features: parsing, text modification, color capabilities, slashing capabilities, displaying lexical bundles in a single bundle form, exception list, login interface, administrative privileges, and print mode.

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

Lab 1_v3

A Critical Study on the problems of ESL Students. (2013). *(IJITR) INTERNATIONAL JOURNAL OF INNOVATIVE TECHNOLOGY AND RESEARCH, Volume No. 1*(Issue No. 1), 028-031. Retrieved February 27, 2015, from http://www.ijitr.com/index.php/ojs/article/viewFile/12/10

Academics. (2015, February 9). Retrieved March 4, 2015, from http://www.odu.edu/academics

English Language Learners. (2014, May 1). Retrieved March 3, 2015, from http://nces.ed.gov/programs/coe/indicator_cgf.asp

English Proficiency. (2015, February 4). Retrieved March 4, 2015, from https://www.odu.edu/content/odu/admission/proficiency.html

McKeon, D. (n.d.). Research Talking Points on English Language Learners. Retrieved December 11, 2014.

Mikowski, M., & Powell, J. Single Page Applications. Manning Publications 2014.

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

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 software configuration, capabilities and features of the Product CLASH prototype. The information provided in the remaining sections of this document includes a detailed description of the software, and external interface architecture of the CLASH prototype, as well as the key features of the prototype and the parameters that will be used to control or establish that feature.

2 General Description

CLASH is a web application, specifically a Single Page Application (SPA), that has three deliverables COLRS, SLASH, and the Slash Player. The COLRS module colors the parts of speech of a given text using NLTK to parse the text into parts of speech. Each POS will have a specific unique color. The SLASH module inserts a forward slash (/) in between lexical bundles. Lexical bundles are a three to five word group of words, representing one thought. Per the parsed document the lexical bundles will be identified. The Slash Player will display the lexical bundles at a reading speed, one lexical bundle at a time. Speed control will be available to users to moderate the speed as preferred.

2.1 Prototype Architecture Description

The CLASH prototype will be a Single Page Application that does not reload. CLASH will use a traditional database, and Node.js as the web and application server. The hardware required for CLASH is an active server and database on the server end.

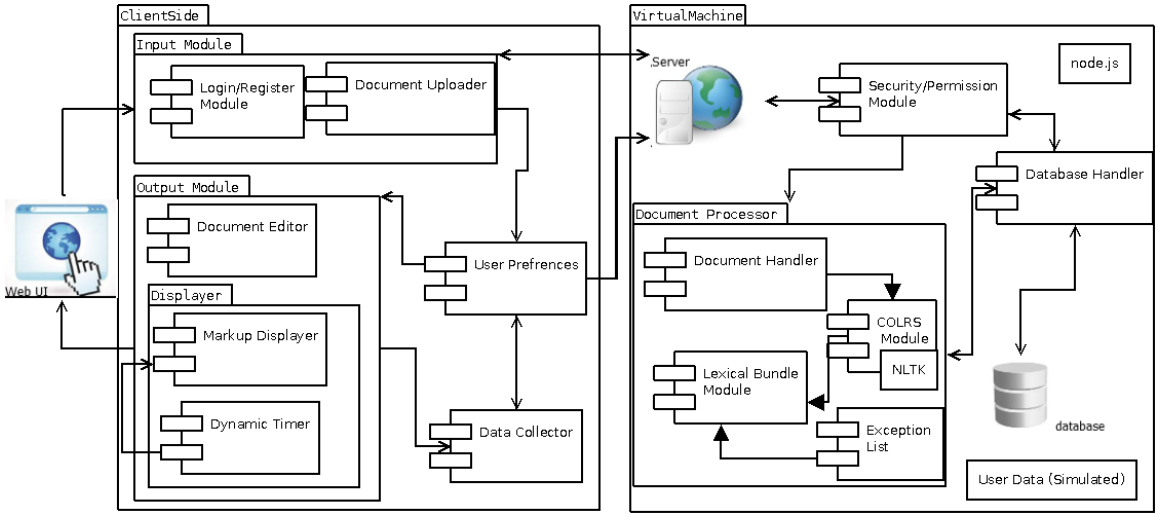
The server will store the website and database as well as documents uploaded by users. The database will contain the document handler, SLASH Module, COLRS Module, Exception list, tracking data, user account information, and user credentials. Figure 1, the Major Functional Component Diagram (MFCD) shows the key features of CLASH.

The users interacts with the website, uploading text file and selecting module. The COLRS Module uses Natural Language Tool Kit (NLTK). The NLTK highlights the POS in a text document, creating a POS tag. The different POS tags will be displayed in the specifically unique color. The user will be able to select single or multiple POS to display. The SLASH Module will use the POS tagged document, and break the sentences into lexical bundles, of multiple words. An exception list was created to improve the accuracy of the product. The exception list will contain lexical bundles that are manually added by the Instructor because they do not follow the normal slashing rules. For example, banana pudding is one thought, but is made up of two nouns. The Slash Player will receive input from the user and display the lexical bundle created one at a time. Users will be able to choose the reading speed the lexical bundles will be displayed.

CLASH prototype contains the following major features:

- **Parsing:** Ability to parse text.
- **Text modification:** Ability to modify and store previously parsed text.
- **Color:** Ability to color chosen parts of speech using a JSON format and JavaScript functions.
- **Slashing:** Ability to identify lexical bundles through the insertion of slashes.
- **Slash Handler:** Displaying lexical bundles in a single bundle form with ability to speed up, slow down and pause lexical bundles being displayed.
- **Exception list:** List of commonly used expressions that would otherwise be incorrectly parsed and slashed
- **Login interface:** User authentication in a standalone environment.
- **Administrative privileges:** Administrators has full control of the system (edit/add/remove/etc.)

Figure 1. CLASH Prototype Major Function Component Diagram.



2.2 Prototype Functional Description

The CLASH prototype will have the ability to identify the following parts of speech: adjective, adverb, conjunction, noun, preposition, pronoun, and verb. Once POS is identified it will be colorized with a specifically unique color that is designated for its POS tag. The lexical bundles will be correctly placed according to the tags as well as the exception list. If an incorrect slash is inserted, the Instructor user will be able to edit text as necessary. The Slash Player will display lexical bundles one at a time with a speed feature. The speed feature will allow user to increase, decrease, or pause the lexical bundle stream. User will be able to print parsed text with slashes and colored POS.

Appendix

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