CS 411 - Lab 1 CLASH Product Description (Blue/Purple Team)

Francia Molina

Old Dominion University

Professor Janet Brunelle

Professor Hill Price

3/4/2015

Table of Contents

1 Introduction 3

2 Product Description 6

2.1 Key Product Features and Capabilities 4

2.2 Major Hardware and Software Components 6

3 Identification of Case Study 7

4 CLASH Product Prototype Description 8

4.1 Hardware and Software Prototype Architecture 9

4.2 Prototype Features and Capabilities 9

4.3 Prototype Development Challenges 10

Glossary 11

References 12

List of Figures

Figure 1 - Prototype Major Function Component Diagram 7

Figure 2 - Prototype Major Function Component Diagram 10

List of Tables

Table 1 - Table 1 - Prototype versus Real World Product 9

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. An ESL is a person who is learning the English language in addition to his or her native language. The percentage of public school students in the United States who were ESL learners during school year 2011–12 were 9.1 percent, or an estimated 4.4 million students, compared to school year 2002–03 8.7 percent, or an estimated 4.1 million students (The Condition of Education). In 2011–12, ESL students made up an average of 14.2 percent of total public school enrollment, ranging from 10.9 percent in small cities to 16.7 percent in large cities (The Condition of Education). ODU has an entire department, the ELC, focused on aiding ESL students with their personal, career, and academic goals involving the English language. The ELC uses an outdated method of teaching for reading and grammar.

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.). Sites like Spreeder have originated in attempts to aid with reading comprehension, but in truth don’t help comprehension due to their focus on speed (Improve Reading). 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.

Color Lexical Analysis and Slash Handler (CLASH) is a web interface that includes two modules, COLRS and SLASH. 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 SLASH module parses text into thought groups to help increase reading speed and comprehension. Thought groups are also used inversely with lexical bundles (LB), which are groups of words that occur repeatedly together, representing one thought. CLASH’s primarily focus is to be used by English as a Second Language students or English Language Learners. CLASH, using the principal Engelbrecht, will implement the COLRS module to use color correlation with POS to increase grammar recognition and retention. CLASH is the application that will be aiding teachers’ methods of teaching reading and comprehension by providing a two-in-one product that focuses on colorizing grammar and reading lexical bundles.

2 Product Description

CLASH is a Single Page Application (SPA) that has two modules, COLRS and SLASH. The COLRS module takes the text provided and parse it according to the parts of speech found in the sentence. The module will color only the following eight POS: noun, pronoun, adverb, verb, preposition, adjective, interjection, and conjunction. The purpose of COLRS is to help the user identify and understand English grammar. 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 purpose of SLASH is to make reading and comprehension of the English language easier and faster.

2.1 Key Product Features and Capabilities

CLASH is a web-based application possessing a Graphic User Interface (GUI), database, slash and coloring algorithms. The GUI will be an aesthetically pleasing interface that is simple to navigate. The database will hold the text files, parsing algorithm, coloring algorithm, and program code. The SLASH module will take a text file and break up the document into lexical bundles with a Slash algorithm inspired by the Japanese technique of teaching English (Nishida, H). SLASH also has a feature called Slash Player that will be a text stream showing one lexical bundles, of three to five words, at a time with the feature of speed control for display time. The COLRS module will take a text file previously parsed and color the POS. The parts of speech that will be colored will be noun, pronoun, adverb, verb, preposition, adjective, and conjunction.

CLASH can be used on any supported web browser and Internet connection. All user accounts will be created by the Administrator. Through the user interface, users can copy and paste text, select between the two modules SLASH and COLRS, and control the speed at which lexical bundles are displayed in the Slash Player. The COLRS module will display a separate menu with the eight variances of POS being colorized. The product features three different types of user roles, these roles include Administrator, Instructor, and Student. The Administrator has all the abilities of the Instructor plus the ability to add and or remove Instructors as users. The Instructor will be able to add and remove Students as users, edit slashed documents, upload documents, and view documents in COLRS and SLASH. The Student will be able to select type of view, SLASH or COLRS, POS colorized, and control reading speed in Slash Player. The application allows for the Instructor to view activity data for the Student users. This activity data will include the student’s current reading speed, and the amount of time spent on the site. This is for instructors to be able to track improvement of individual students as desired. CLASH will provide two separate modules that will allow all users to improve reading comprehension and grammar.

2.2 Major Hardware and Software Components

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. Figure 1 will show the overall hardware and software components being used.

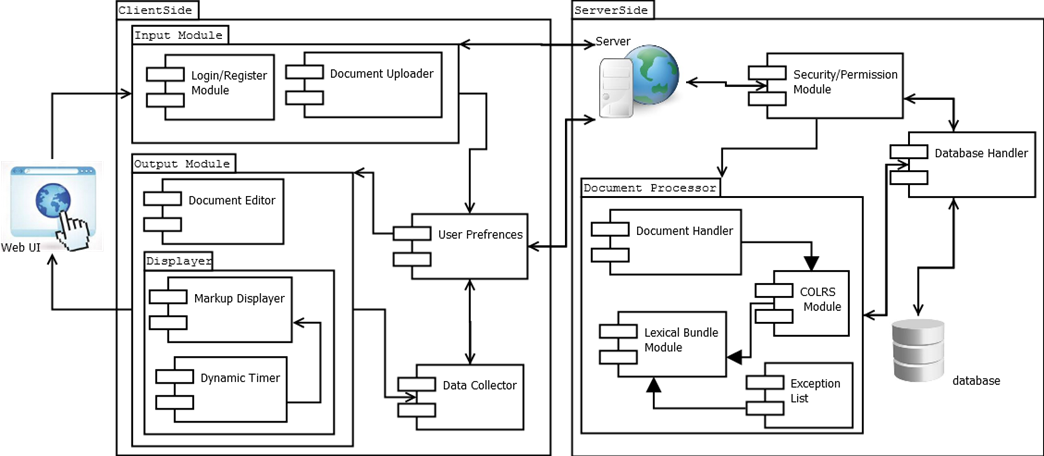


Figure 1 - Prototype Major Functional Component diagram

3

Identification of Case Study

ODU is home to many international students. For the 2012-2013 academic school year there were a total of 1,086 international students (Monarch Diversity). International students must show proficiency in English to apply for admission. That is, if the native language is not English, evidence of English proficiency must be demonstrated to enroll in credited courses. If applicants do not have the required test scores, they are allowed admittance through Conditional Admission. The Bridge Program is designed for conditionally admitted students with TOEFL iBT scores from 61 to 79 (PBT 500-550) or IELTS scores from 5.5 to 6.5 (English Proficiency). ELC provides the Bridge Program which runs over two semesters. Once completed, the Undergraduate or Graduate Bridge Program satisfies the university's English proficiency requirement. ELC professors teach the students English grammar, reading skills, and conversing skills. ESL students attempt to grasp as much knowledge of the English language before they start their credited classes. The range of teaching techniques ranges from writing paragraphs that later become papers, reading articles and books, listening to videos and learning how to listen to discussions, and more. Often times, however, students’ experience that they do not learn as quickly as they want to and as a result are slow readers. Greg Raver-Lampman is an instructor at ODUs English Language Center that is looking for a product that will help him and other English teachers’ aid ESL students in their studies. CLASH focuses on what Professor Raver-Lampman considers to be the two key methods in improving ESL students’ recognition and comprehension of grammar and reading.

4 CLASH Product Prototype Description

The CLASH prototype will be a modified version of a Single Page Application (SPA) architecture. A SPA is a web application that is on a single web page with the goal of providing a similar experience to that of a desktop application that does not reload as the web page changes states. The SPA structure of CLASH will be built completely in JavaScript (JS), meaning each component and operation will use JavaScript. CLASH will have a traditional html webpage that is called forth by JS, a JavaScript friendly web server, Node,js, and a traditional relational database. A SPA is an ideal platform for CLASH because it offers the look and feel of a desktop application, but the accessibility of a web application. The platform was chosen with users who are not computer savvy in mind. CLASH will have no software to install by the user. A user management module will be implemented to administrators to add and remove users as needed. Users will login to their accounts after being created by Administrator. All functions will be accessible through menus on the website. In Table 1 – Prototype versus Real Life Product (RLP) shows were the prototype differs from the RLP with a short description to each feature.

|  |  |  |
| --- | --- | --- |
| **Features** | **Real World Project** | **Prototype** |
| Parsing Capabilities | Ability to Parse different kinds of documents | Ability to parse text copy and pasted into form |
| Text Modification | Ability to modify and store previously parsed documents | Ability to modify and store previously parsed documents |
| Color Capabilities | Ability to Color chosen parts of speech using a JSON format and javascript functions. | Ability to Color chosen parts of speech using a JSON format and javascript functions. |
| Slashing Capabilities | Ability to identify lexical bundles through the inserting of slashes. | Ability to identify lexical bundles through the inserting of slashes. |
| Displaying lexical bundles in a single bundle form | Ability to speed up, slow down and pause lexical bundles being displayed. | Ability to speed up, slow down and pause lexical bundles being displayed. |
| Exception list | Lists of commonly used expressions that would otherwise be incorrectly parsed and tagged. | Lists of commonly used expressions that would otherwise be incorrectly parsed and tagged. |
| Login interface | User Authentication in a stand alone environment | User Authentication in a stand alone environment |
| Student Data reporting | Tracks individual and collective student progress. To include words per minute, total time and total lexical bundles. Data to be stored in database. Displayed in graphs and statistics. | Not included. |
| Homework Mode | Instructors have the ability to remove coloring of words and have students correctly identify the part of speech. | Not Included. |
| Administrative Privileges | Administrators are able to edit, add, or remove anything in the system. | Administrators are able to edit, add, or remove anything in the system. |
| Print mode | Ability to print documents with slashes inserted.  Table 1 - Prototype versus Real World Product | Ability to print documents with slashes inserted. |

4.1 Hardware and Software Prototype Architecture

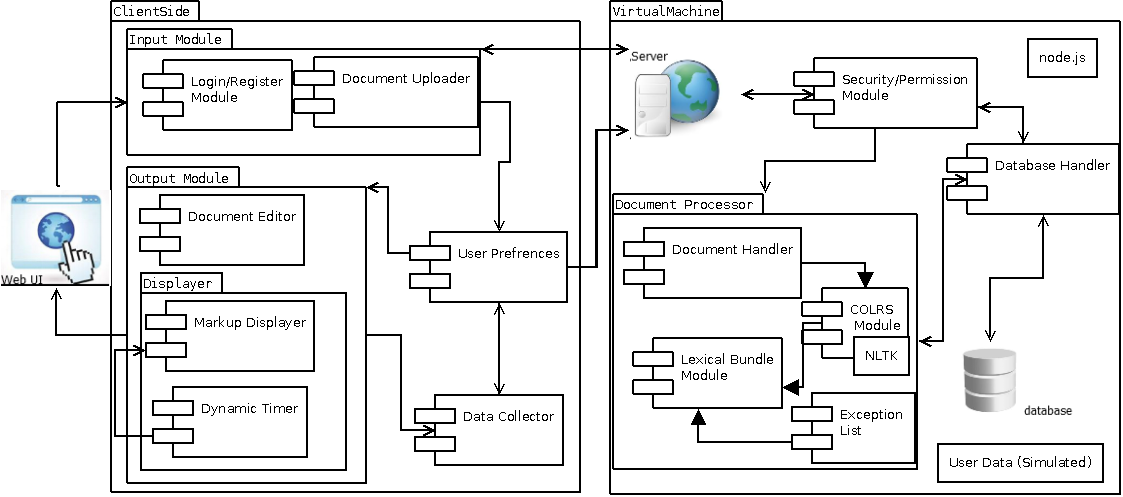
 The hardware required for the prototype is a device with Internet connection and a supported web browser. The software required for the prototype will be a collection of open source or liberally licensed software to permit distribution or potential commercialization of the software package. The web and application server will be Node.js. This will allow the HTTP server to establish an incoming network connections from users (the students and instructors), as well as the application handling text processing of documents. Figure 2 – MFCD shows the updated process that CLASH will use for the COLRS module and SLASH module. The SLASH module includes the Slash Player feature.

Figure 2 – Prototype Major Functional Component Diagram

4.2 Prototype Features and Capabilities

The CLASH prototype will have three text processing modules as follows: COLRS, SLASH, and Slash Player. The COLRS module will demonstrate the ability to identify parts of speech, distinguishing amongst the various POS with color. The SLASH module will identify lexical bundles throughout a paragraph and display them by the insertion of slashes. The Slash Player will display lexical bundles one at a time, and allow the user the ability to increasing and decreasing the lexical bundle stream. CLASH will be able to take text in two forms, by file upload or by copy and paste. The Administrator will be able to add and remove Instructors and Student roles. Instructor will be able to add and remove Student roles as well as manually edit slashed and colored documents. CLASH will be able to create and update an exception list in the SLASH module. The exception list, is a list that will hold lexical bundles that are common and reoccur multiple times.

4.3 Prototype Development Challenges

There are several challenges that may arise in the developmental stage. An example is, computational time of the exception list. By adding more exceptions to the list, computational time increases due to the increase of size of the list itself. Another challenge is missing attributes; meaning features are not working properly, not working at all, or missing completely. An example of this would be the failure of the NLP system that is being used to create the tags needed for POS recognition and slash insertion. English is a challenging language and with that comes errors. For example, a possible error that will arise is the incorrect identification of parts of speech and incorrect placement of slashes. This can occur because English is a vast language and has many exceptions that may lead to incorrect slashing and identification of POS.

CLASH aims to provide the user the ability to control the speed at which the lexical bundle will show with the SLASH feature. There is a possibility of having difficulty getting the algorithm that will control the speed to do the proper task. CLASH is meant to have a simple interface. This is to make the product accessible to users with various computer skills levels. There is that possibility that CLASHs simple is not simple which will in the affect the appeal it has for certain users. Furthermore, another challenge that will arise with time is storage capacity. Users will populate our database with their documents and the designated storage capacity will not be enough for long term use. Building a working prototype with these features will be tested in a real world situation against a control group (Lab 1 CS 410 Team Blue).

Glossary

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](http://www.spreeder.com/)

TOEFL – Test of English as a Foreign Language

Ubuntu – a Debian-based Linux operating system

VM – Virtual Machine

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/

Lab 1 CS 410 Team Blue

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.