Lab I - CLASH Product Description CS411

Artem Fisan

Professor Janet Brunelle

Professor Hill Price

Old Dominion University

**Table of Contents**

1 Introduction 3

2 Product Description

2.1 Key Product Features and Capabilities 3

2.2 Major Components (Hardware/Software) 3

3 Identification of Case Study 3

4 CLASH Product Prototype Description 5

4.1 Prototype Architecture (Hardware/Software) 6

4.2 Prototype Features and Capabilities 4

4.3 Prototype Development Challenges 5

Glossary 11

References 12

List of Figures

Figure 1. Major Functional Component Diagram (MFCD) 5

Figure 2. Major Functional Component Diagram (MFCD) 8

List of Tables

Table 1. Prototype vs. Real World Product 7

**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 Old Dominion University has an entire department focused on addressing this problem, though have an outdated method of teaching reading and grammar.

CLASH is a single-page web application that is consists of 2 modules; COLRS and SLASH. The SLASH module includes a UI Slash Player. The COLRS module will color each part of speech (POS) in a text document with a particular color code (red for noun, pink for verb etc.). The POS that will be used are adjective, adverb, conjunction, noun, preposition, pronoun, and verb. The SLASH module parses text into a group of words that occur repeatedly together, or present one single thought, lexical bundles.

**2.0 Product 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 Key Product Features and Capabilities**

CLASH has a login system that contains three types of users: Administrator, Instructor, and Student. Key features available to certain roles are as followed: ability to create and remove accounts, generate passwords, upload, edit and removing of text documents, ability to view text slashed and colored in a text file or via Slash Player. Slash Player will provide users control of the speed a lexical bundles will be displayed. The Slash Player module will display a lexical bundle one at a time with the ability to pause and change the display speed. The COLORS module gives the user the ability to choose which POS will be displayed. The eight POS are adjective, adverb, conjunction, noun, preposition, pronoun, and verb. All users will access CLASH through a supported web browser via Internet connection.

The Student user will be able to log into CLASH after account is created, select which module they would like to view, which document to view, control POS colorized, and control reading speed. The Instructor user will be able to add and remove a Student users, upload documents, edit documents, and select the documents available to the Student user. Instructor user will also have access to view activity data for the Student users. Activity data will include student user current reading speed, and the amount of time spent on the site. Administrator will be able to add and remove Instructor and Student users, as well as having their privileges.

**2.2 Major Hardware and Software Components**

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.

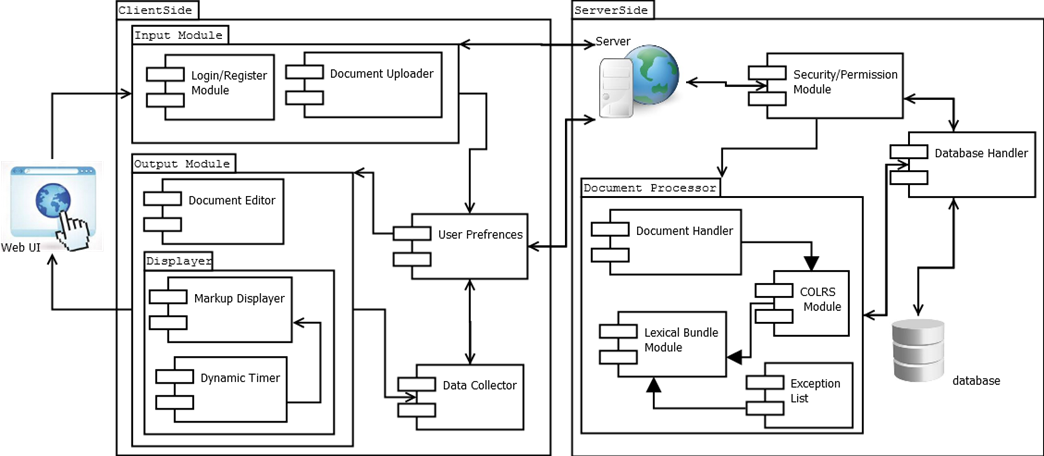


Figure 1. Major Functional Component Diagram (MFCD)

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.

**3. IDENTIFICATION OF CASE STUDY**

Multiple international students attending Old Dominion University to pursue degree in 124 Undergraduate programs, 54 Master’s programs, and 42 Doctoral programs (Academics). International students are required to show a passing score in the TOEFL or iBT. The TOEFL requires a score between 500 and 550, and the iBT required a score between 61 and 79 (English Proficiency). For students that are unable to pass the TOEFL or iBT are directed to the English Language Center at Old Dominion University. Greg Raver-Lampman, an instructor at the ELC, is attempting to improve educational techniques with modern technology. The current teaching process in the classroom is to read books and write paragraphs that later become papers. CLASH is the perfect tool for teachers to track the improvement of students learning English. It also provides an easier way of presenting material.

**4. CLASH Product Prototype Description**

The CLASH prototype will be a Single Page Application (SPA). A SPA is a highly responsive web application that does not reload. CLASH will use a traditional database, and Node.js as the web and application server. A SPA is an ideal platform for CLASH because it offers the abilities of a desktop application, but the representation via web browser. All functionality will be accessible through menus on the web page. CLASH will have a login system that will contain three users with different roles: Administrator, Instructor, and Student. Table 1 represents the difference between the Prototype and Real World Product.

|  |  |  |
| --- | --- | --- |
| **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 standalone environment | User Authentication in a standalone 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. | Ability to print documents with slashes inserted. |

Table 1 – Prototype vs. Real World Product

**4.1 Hardware and Software Prototype Architecture**

The hardware required for the prototype system is a computer that can run a supported browser to access CLASH through the Internet. Software required for the prototype is a collection of open source code. On the Client Side, user will paste text into CLASH text field to be sent to the server for processing. The process of COLRS and SLASH can be seen in Figure 2 – MFCD. The process will use the NLTK to tokenize and tag POS to color or slash the text.

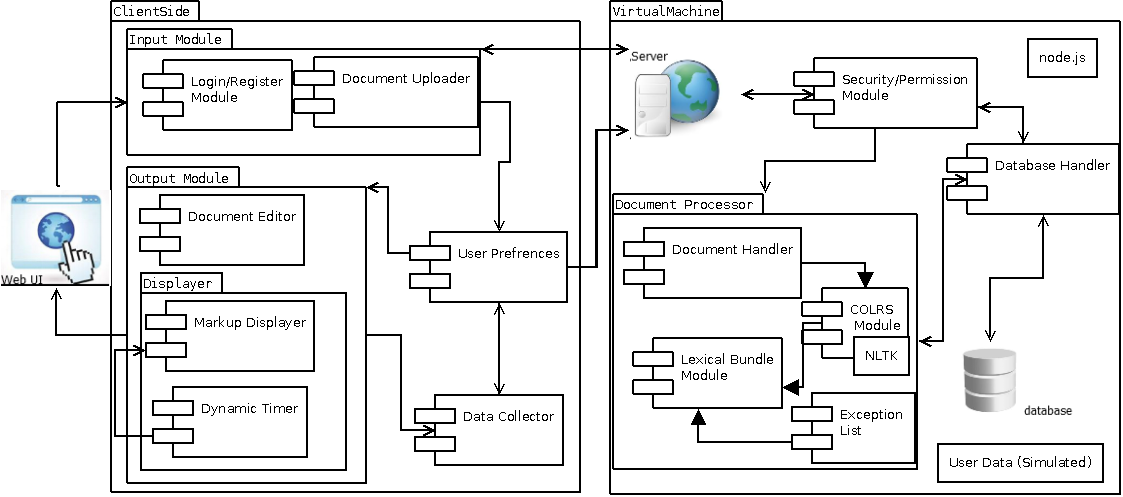


Figure 2 – MFCD

**4.2 Prototype Features and Capabilities**

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.

**4.3 Prototype Development Challenges**

There are multiple challenges that CLASH may face on its developmental journey. One of them 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. CLASH, being unique and very useful, will have minor hiccups that have to be taken care of when they show up. It is impossible to write a program without minor bugs, but those issues will be discovered by actual testing or as they occur.

English is a challenging language and with that comes errors. One particular error that will possibly arise is the incorrect identification of parts of speech and causing incorrect placement of slashes. 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 will 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 long run 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 10GB will not be enough for long term use. Overall, our prototype is going to have a few challenges but CLASH will be one of a kind (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**

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.