**Lab 1 – CLASH Product Description**

Team Blue/Purple

Fredrick Murry Jr.

CS411W

Janet Brunelle and Hill Price

Version 2

Table of Contents

[**1** **INTRODUCTION** 3](#_Toc413235530)

[**2** **CLASH PRODUCT DESCRIPTION** 4](#_Toc413235531)

[**2.1** **Key Product Features and Capabilities** 4](#_Toc413235532)

[**2.2** **Major Components (Hardware/Software)** 6](#_Toc413235533)

[**3** **IDENTIFICATION OF CASE STUDY** 7](#_Toc413235534)

[**4** **CLASH PRODUCT PROTOTYPE DESCRIPTION** 7](#_Toc413235535)

[**4.1** **Prototype Architecture (Hardware/Software)** 9](#_Toc413235536)

[**4.2** **Prototype Features and Capabilities** 9](#_Toc413235537)

[**4.3** **Prototype Development Challenges** 10](#_Toc413235538)

[**Glossary** 11](#_Toc413235539)

[**References** 12](#_Toc413235540)

[Figure 1. CLASH Major Functional Components 6](#_Toc413234846)

[Figure 2. CLASH Prototype Major Functional Components 9](#_Toc413234847)

[Table 1. Real World Product Vs. CLASH Prototype 8](#_Toc413235495)

Lab 1 –CLASH Product Description (Murry)

1. **INTRODUCTION**

During the year 2004, it was estimated more than 4.9 million English as a second language(ESL) students were enrolled in public schools in the United States. Of those 4.9 million ESL students, 15% lacked resources or programs to aid in their learning of the English language. In a study it was found that only approximately 18.7% of ESL students tested for reading comprehension were at or above the standard during 2001. That same year it was reported that the dropout rate of ESL students was almost four times the rate of native English-speakers. (McKeon)

At the present time, Old Dominion University has a department that specializes in instructing ELC students. Currently, ODU’s process for teaching ESL involves an instructor manually writing a sentence on the board, then marking the parts of speech (POS) and lexical bundles (a group of words that occur repeatedly together, or represent a single thought). This can be tedious. Also this help is unavailable to the student when he or she is studying away from class. There are reading applications available to aid in reading comprehension but none specifically designed for ESL students. The ones currently available do not identify both POS and lexical bundles.

A proposed solution is CLASH, or Color Lexical Analysis algorithm and Slash Handler. CLASH is a web application that consists of two modules, COLRS and Slash. The COLRS module identifies the POS in a sentence using different colors. This could help ESL students increase their comprehension of sentence structure when using the application on their own. The Slash module separates the sentences into lexical bundles to be displayed at various user-controlled speeds. This could help ESL students increase their reading speed and comprehension.

1. **CLASH PRODUCT DESCRIPTION**

The goal of CLASH is to help ESL students increase their reading speed and comprehension at a rate higher than traditional methods. CLASH will be a web-based application assessable through a standard web-browser. CLASH will consist of a server on the backend that will handle the COLRS and Slash modules. The COLRS module will identify POS using colorization. The Slash module separates sentences into lexical bundles and displays the results in the web-browser at a speed controlled by the student for reading. CLASH will allow the instructor to provide reading assignments to the students. Student usage data can also be monitored by the instructor.

* 1. **Key Product Features and Capabilities**

CLASH will be a web-based application that will allow users to improve their reading comprehension. CLASH does this by separating sentences into lexical bundles to be viewed by the user. CLASH will also highlight the POS. The features of CLASH can be accessed using a standard web-browser and Internet connection. Through the application’s user interface, a user can control the speed at which lexical bundles are displayed for reading. POS can be viewed in a separate part of the user interface. The product features individual password-controlled user logins with three different types of user roles. These roles include Administrator, Instructor, and Student. The Student will be able to control their reading speed, type of view, and which available document to view. The Instructor will include the Student user capabilities and more. The Instructor will be able to add/remove Student users and select documents available to be viewed by the Student users. The application allows for the Instructor to view the activity data of Student users. This activity data will include the Student user’s current reading speed, and the amount of time spent reading. The Instructor will be able to upload documents to the server, edit files, and deletes documents currently on the server. The Administrator has all the capabilities of the Instructor plus the ability to add/remove Instructors as users.

One of the features that make CLASH unique is that it displays both POS and lexical bundles. This allows students to improve their reading comprehension by having a speed reading application that does not break up the individual lexical bundle. There is the ability to pause and change the display speed. CLASH will include a text parser that is able to automatically identify and color POS, allowing for easy identification. It will also allow the Instructor user to review the parsed text for accuracy and editing purposes. CLASH will be able to save usage data to allowing instructors to review student progress, so they can alter their lecture lectures accordingly. CLASH will be the first web-based application specifically designed for use in ESL instruction. Instructors can have documents available to the students at their appropriate reading level. CLASH has the capability of giving students and instructors a powerful tool to increase reading speed and comprehension.

* 1. **Major Components (Hardware/Software)**

CLASH will feature three major components software components. They will include the COLRS module, SLASH module, and the text reader viewed through the web browser. Being that CLASH is a web-based application, the only hardware requirements are an active server on the back-end and an Internet-enabled device for the user. The major hardware components can be seen in Figure1.

The COLRS module will use Natural Language Processing (NLP) to tag the POS in the input document. An open source tokenizer will be used for the process. The Slash module will use the same tokenizer from the COLRS module in its algorithm to identify lexical bundles. The text reader will use the tokenized output from the server to be viewed by the user. It will have a User Interface for control of user viewing speed and POS identification using color-coded highlights.

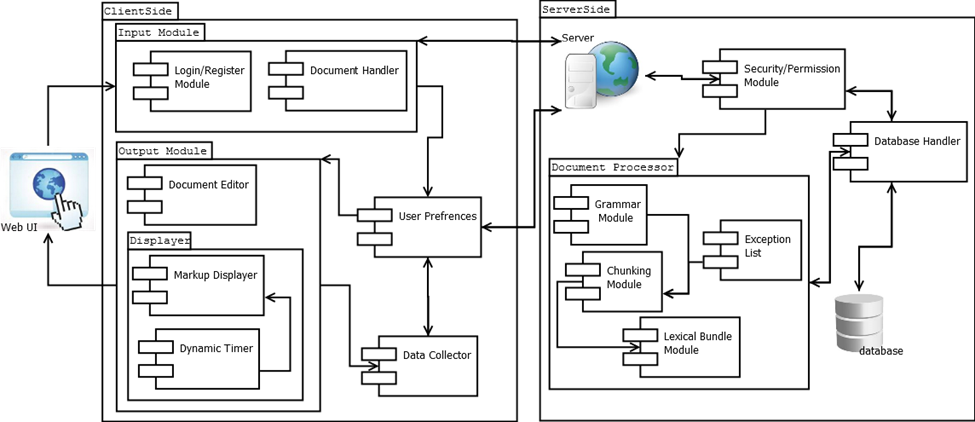


Figure 1. CLASH Major Functional Components

1. **IDENTIFICATION OF CASE STUDY**

Before international students come to Old Dominion University, they must take an International Benchmark Test (IBT) for English proficiency. This is the TOEFL test. When students do not meet the required proficiency level, they must sign up for the English Language Center at ODU.

During 2014, Greg Raver-Lampman, an instructor at the English Learning Center, was searching for software to assist his students in their ESL studies. He found that there was not any software that used lexical bundles in its reader. Use of lexical bundles was Mr. Raver-Lampman’s preferred method for improving the reading speed and comprehension of his students. Mr. Raver-Lampman proposed his idea for software designed with ESL students in mind to the Computer Science Senior Design class. This software solution would identify POS and display lexical bundles in a web-based speed reader. Though designed for ESL students, it could be used by anyone who wanted to improve their reading speed and comprehension. Mr. Raver-Lampman would also like to share the software with his professional colleagues.

1. **CLASH PRODUCT PROTOTYPE DESCRIPTION**

The CLASH prototype will be a Single Page Application (SPA) that does not reload when changes are made to the web page. The SPA will be built using JavaScript and will utilize a JavaScript-built Graphical User Interface, application and web server, and a database. Ideally a database with JavaScript object Notation (JSON) would be used, such as NoSQL. The prototype will use a relational database instead of NoSQL. The SPA will combine the user the accessibility of a web application with the look and feel of a desktop application. There will be no software to download and install by the users.

The prototype will be a proof of concept for the real world product(see Table 1. below). The POS tagging will use open source NLP tools and will very basic in its operation. The accounts for the various users will manually set up instead of the real world product’s integration with ODU’s enrollment systems.

Table 1. Real World Product Vs. CLASH Prototype

|  |  |  |
| --- | --- | --- |
| **Features** | **Real World Product** | **Prototype** |
| Parsing Capabilities | Ability to Parse different kinds of documents | Ability to parse text copy and pasted in text block |
| 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 handled by the SLASH Algorithm. | Lists of commonly used expressions that would otherwise be incorrectly handled by the SLASH Algorithm. |
| 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. | Limited basic student metrics will be available such as Lexical Bundles per Minute. |
| 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 users and saved documents in the system. | Administrators are able to edit, add, or remove users and saved documents in the system. |
| SLASH Document Viewing Mode | Ability to view documents with slashes inserted and SLASH Reader. | Ability to view documents with slashes inserted and SLASH Reader. |

* 1. **Prototype Architecture (Hardware/Software)**

The main hardware for the prototype will be a Virtual Machine (VM) hosted at ODU. The user will be able to use their own internet-enabled device for access. The software for the prototype will include a collection of open source software and custom algorithm programs. The VM will run Ubuntu 14.04 LTS. Nod.js will be used for the application and web server, so that there will be interaction with the Natural Language Toolkit. (NLTK), and the database server. The Node.js server will provide the SPA that will allow user access to the CLASH prototype. The Major Functional Component Diagram (MFCD) is shown in Figure 2.

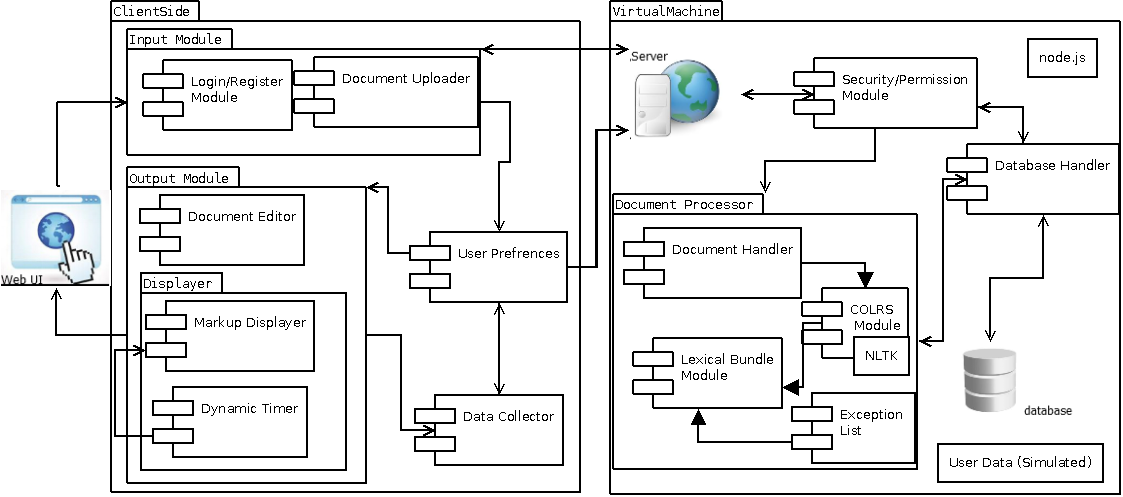


Figure 2. CLASH Prototype Major Functional Components

* 1. **Prototype Features and Capabilities**

The CLASH prototype will offer a proof of concept for the real world product by demonstrating a speed reader, with lexical bundles and use of colorized POS, is feasible. CLASH’s success will be determined by feedback from Mr. Raver-Lampman on the data collected from student usage and his testing of his students. Success would be the adoption and further development of the CLASH product. Risks are to be mitigated by the oversight of Mr. Raver-Lampman during the development of the prototype. It involves the testing of ease of use, accuracy of CLASH functions and the integration of the ‘COLRS’ and ‘Slash’ modules.

* 1. **Prototype Development Challenges**

There are many challenges expected in the development of the CLASH prototype. One is the integration of the ‘COLRS’ module and ‘Slash’ module into one application. There is also the may be missing knowledge by some of members of the team on certain aspects of the development of the prototype. The challenge of accurately creating an algorithm to identify lexical bundles is expected when developing the prototype. The main challenge is the creating a simple user interface that is intuitive and facilitates the adoption of the product for further development.

# **Glossary**

**CLASH-** **-** Color Lexical Analysis algorithm and Slash Handler

**COLRS**- Colored Organized Lexical Recognition Software

**ELC**- English Learning Center

**ESL**- English as second language

**IBT**- International benchmark test

**JSON**- JavaScript Object Notation

**Lexical Bundle-** a group of words that occur repeatedly together, or represent a single thought group.

**MFCD**- Major Functional Component Diagram

**NLTK**- a suite of libraries and programs for symbolic and statistical natural language processing for the Python programming language.

**Node.js**- an open source, cross-platform run-time environment for server-side and networking applications.

**POS**- Parts of Speech

**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.

**TOEFL**- Test of English as a Foreign Language

**Ubuntu**- a Debian-based Linux operating system

**VM**-Virtual Machine

# **References**

Blue/Purple teams (2014, December). Clash Product Description CS410 Lab 1.

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

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.

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