CLASH Product Description CS410 Lab 1

Blue/Purple Teams

Old Dominion University

Author: Ali Albattah

CS411W

Janet Brunelle & Hill Price

Version 3

April 3, 2015

Table of Contents

[**1 Introduction** 2](#_Toc413191930)

[**2 Product Description** 3](#_Toc413191931)

[2.1 Key Product Features and Capabilities 3](#_Toc413191932)

[2.2 Major Hardware and Software Components 5](#_Toc413191933)

[**3 IDENTIFICATION OF CASE STUDY** 7](#_Toc413191934)

[**4 CLASH Product PROTOTYPE DESCRIPTION** 8](#_Toc413191935)

[4.1 Hardware and Software Prototype Architecture 9](#_Toc413191936)

[4.2 Prototype Features and Capabilities 9](#_Toc413191937)

[4.3 Prototype Development Challenges 10](#_Toc413191938)

[**Glossary** 11](#_Toc413191939)

[**References** 12](#_Toc413191940)

**List of Figures**

[Figure 1 - Major Functional Component Diagram 5](#_Toc413188314)

[Figure 2 - CLASH Process Flow 6](#_Toc413188315)

Figure 3 - Prototype Hardware/Software Architecture…...………………………………………………...7

**List of Tables**

[Table 1 - Real World Product vs Prototype 10](#_Toc413185566)

# **1 Introduction**

COLRS and Slash are the modules, which make up CLASH. COLRS aids users in better understanding grammar and sentence structure by highlighting particular components of the text, while the Slash module helps users comprehend and read more fluently by grouping texts into lexical bundles, to aid English as a Second Language (ESL) students read faster and be able to comprehend what they read faster as well.

The processes used to teach reading and grammar to ESL students have been outdated for over a decade (McKeon). In 2001 only 18.7 percent of ESL students were successful in accomplishing normal or above normal comprehension level. In 2004 an estimate of 15 percent out of 4,999,481 ESL students had no resources to help aid them in achieving a greater comprehension level of the English language. Furthermore, 10 percent of ESL students between 7-12 grades were held back (McKeon). It is obvious that learning a second language is a difficult process. Many international college students in the United States find that especially true. They are required to pass one of two tests with the required respective scores to be admitted into their colleges: the Test of English as a Foreign Language (TOFEL) and the International Benchmark Test (IBT). Here at Old Dominion University, if a student does not receive the minimum required score, he will be place in a bridge program designed to fine tune the student’s English. However, even then many students find it extremely difficult to reach these respective benchmarks in the allotted time, usually 18 months.

The current methods of teaching ESL students English, such as writing on chalk boards and using reading websites, are not only outdated, but time consuming and not efficient. Breaking up words into lexical bundles using slashes has been shown to increase the likelihood of a student grasping the English language quicker and more efficiently, which is why the SLASH Handler module was created. The other module, known as the COLRS module, aids the student by breaking the sentences up by their Parts of Speech (POS). Each respective POS has a unique color, which would allow students to understand grammatical patterns and sentence syntax. This would make reading in English much more predictable.

# **2 Product Description**

CLASH is designed to help ESL students learn English more efficiently and quicker than customary means. It will have a Graphic User Interface to connect the student and the database; it will also permit teachers to monitor the usage of the program by every student. The CLASH application comprises of two modules SLASH Handler and COLRS.

SLASH Handler’s straightforward program interface provides ESL students an easy to follow and understand program. SLASH will enhance the reading speeds of ESL students by separating texts into lexical bundles to allow for faster comprehension by using the Handler’s playback environment. It will likewise significantly enhance reading perception and convert readers from word to word readers into lexical bundle readers. It will also allow for the conversion of physical text into electronic text for recreational reading.

The COLRS module helps students recognize different parts of speech by colorizing the eight essential POS within the document. The eight essential POS (verbs, nouns, adverbs, adjectives, conjunctions, prepositions and pronouns) will each have a respective color. The user will be allowed to select which POS is to be colorized. The goal is to aid the student in learning the uses of different words and to make the language more structured and thus predictable.

## 2.1 Key Product Features and Capabilities

CLASH’s aim is to improve reading speed and comprehension of students by using its two modules to encourage users to read in lexical bundles and identify POS. A lexical bundle is a group of words that often occur together within the same sentence, such as “day-to-day.” Reading in lexical bundles allow for students to read and comprehend much quicker than students who would read individual words of the sentence in a sequence.

Students are able to control reading speeds and document viewing options by using SLASH PLAYBACK. Instructors are able to add/remove student accounts, add/edit documents and monitor student activity. The administrators have the same abilities as the instructors, but can also add and remove instructor accounts.

CLASH allows the user to view both parts of speech and lexical bundles. CLASH also has a feature that highlights certain parts of the text to allow for easier identification of specific words or parts of speech. Instructors can also ensure that the text is logically and accurately broken up into POS and lexical bundles using the COLRS and Slash Handler modules respectively. This would allow the instructor to tailor his lectures based on specific student needs. CLASH is designed with the needs of ESL students in mind.

A feature that makes CLASH unique is that it encompasses both of the required modules to display the POS and lexical bundles. This provides the students with the ability to improve their reading comprehension by having a speed reading application that does not break up the individual lexical bundle. Another feature of CLASH allows users to pause and change the display speed. CLASH will include a text parser which enables the program of the ability to automatically identify parts of speech and colorize them, allowing for easy identification. It will also allow instructor to review the parsed text to ensure accuracy. Another feature that CLASH will have is the ability to save usage data, allowing instructors to display the student progress, and fine tune their lectures accordingly. CLASH is the first web-based speed-reader designed specifically for use in ESL instruction. Instructors will be able to have documents available to the students at their specified level. CLASH will give students and instructors enormous assistance in increasing reading speed and comprehension.

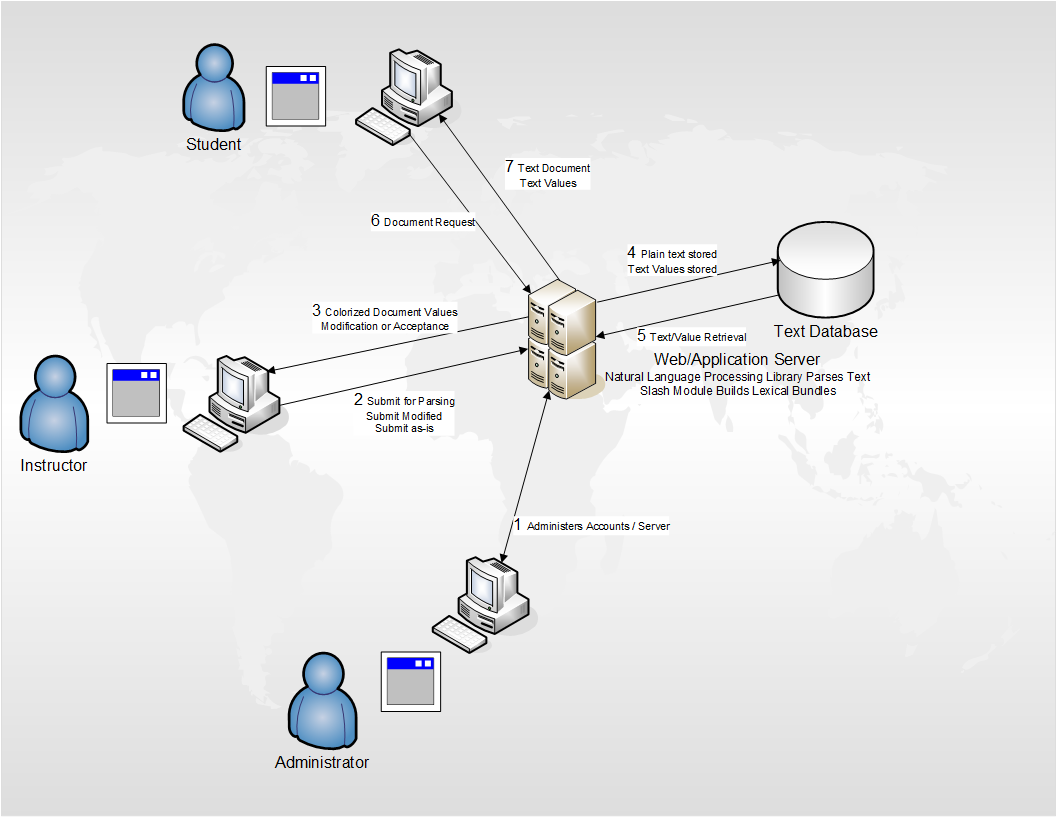
## 2.2 Major Hardware and Software Components

#### **Figure 1 - Major Functional Component Diagram**

**Overview**: CLASH can be easily accessed with no extra equipment except for internet and an internet enabled device. With CLASH, users will have access to 2 major area: Server side and Client side, as shown in Figure 1.

**Server-side:** COLRS Module use open source Natural Language Processing (NLP), and Natural Language Tool Kit (NLTK) to tokenize and parse the input document, and it will generalize the Parts of Speech (POS) tags in more readable format. The Document Processer (DP) takes the text submitted by the instructor on the Client Side and parses it into discrete words called tokens. The output will be a tokenized stream with a token and tag pair.

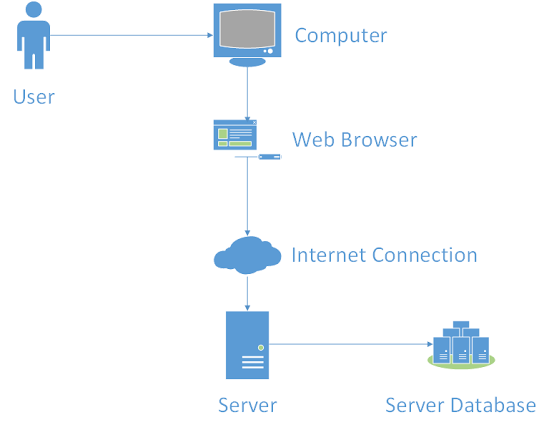
The SLASH Handler Module will accept tokenized stream from the COLRS Module and decide how to break sentences into digestible lexical bundles. There will be a sub-component to deal with the Exception List that improves the accuracy of the final output. The Exception List lists commonly used expressions that would otherwise be incorrectly handled by the SLASH Algorithm.



#### Figure 2 - CLASH Process Flow

**Client-side:** The reader program will receive data from the server and parse the text according to their POS tag. There will be two display modes, color mode and speed mode. Under color mode, user can choose what POS tag to be colorized, and choose whether the Slash needs to be displayed. The user can adjust the reading speed under the speed mode, and the reader program will display each lexical bundles based on the reading speed.

(This space intentionally left blank.)



#### Figure 3 - Prototype Hardware/Software Architecture

# **3 IDENTIFICATION OF CASE STUDY**

Old Dominion University boasts a large and diverse international student population. Many students from all over the world come to this university to pursue degrees in many programs ranging from nursing to computer science. However, for these students to start their academic studies, they will need to meet a language benchmark to ensure that their English can allow them to succeed in a university setting. These language benchmarks come in the form of internationally acclaimed tests, either the TOFEL or IBT. If the student does not meet these requirements or does not feel confident enough to take the test, then he can apply for the English Language Center at ODU.

Greg Raver-Lampman, an instructor at the Old Dominion University’s English Language Center, teaches students who speak English as a second language. These students rely on the ELC professors to gain as much knowledge of the English language before they start their academic classes. The teachers, in turn, help the students understand grammar, write sentences that later become paragraphs, listen to the language using various tools, and provide the students with an environment that encourages learning English. Often, however, ESL students do not learn as quickly as they need to and tend to be very slow readers. Keeping in mind, the allotted maximum time given by universities and immigration restrictions puts pressure on professors and students to meet the aforementioned benchmarks in order to progress to the next phase of their academic studies. Professor Raver-Lampman knows this all too well and has thus requested help from the Computer Science Senior Design class to develop a software solution to help students identify the POS and read in lexical bundles.

This web application was primarily created to help students at the ELC learn how to read faster. This website can be used by not only students, but also professionals that travel to English speaking countries frequently and want to learn how to read and understand correctly. This is a solution that can be used by anyone that simply wants to learn the English Language and read it efficiently.

# **4 CLASH Product PROTOTYPE DESCRIPTION**

The CLASH prototype will be a modified version of the single page application. The difference between this and CLASH is that instead of using a NoSQL database it will use a traditional relational database, however the interface will be built in JavaScript and Node.js will be used as the web and application server. Table 1 highlights the differences in features between the real world product and the prototype.

(This space intentionally left blank.)

|  |  |  |
| --- | --- | --- |
| **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. |

Table 1 - Real World Product vs Prototype

## 4.1 Hardware and Software Prototype Architecture

**Hardware:** Initially the program will require a laptop from the developers, but will eventually be migrated with a virtual machine (VM) once received by the ODU Computer Science Department.

**Software:** The CLASH prototype will be hosted on a Virtual Machine (VM) provided by the Old Dominion University Department of Computer Science. The CLASH prototype will be designed as a modified version of the SPA architecture. SPA, or Single Page Application, loads all of the resources that are pivotal to navigating throughout the webpage on the first page load. This aforementioned modified version will suffice for CLASH’s prototype. The SPA built into CLASH's Node.js server will create the web pages for the users by integrating custom programming which will parse out the input text via NLTK. The UI will be built in HTML, CSS and JavaScript. This would allow the prototype to function with relative ease and simplicity. Furthermore, the natural language processing software that will be utilized by the prototype is the Natural Language Toolkit (NLTK). A python script will parse the text once the text is uploaded to the application server.

## 4.2 Prototype Features and Capabilities

The CLASH prototype will be able to colorize texts and identify lexical bundles by the use of slashes and allow students to change the speed and display of the bundles. The lexical bundles are controlled by the SLASH Handler module, which inserts slashes that break up words into common groups of words known as lexical bundles. The COLRS module will colorize words based on one of the 8 essential parts of speech. The success of the prototype will be determined by the feedback provided by its users. By having a controlled group test this prototype and its core features we can identify how well the system works in an academic environment and also further develop the application.

## 4.3 Prototype Development Challenges

One particular error that may occur is the identification parts of text within the lexical bundles using the slash feature. CLASH is meant to be an easy to use, user-friendly application for all levels of computer users, although we realize that not everyone may find it to be that way. Another problem that may occur is the overload of documents surpassing the amount of storage space available (10GB) for long-term use but even with the minor faults in the prototype, CLASH will be the only application of its kind.

# **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 at Old Dominion University.

**ESL**: English as second language.

**ELL**: English Language Learner.

**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.

**NLTK**: A suite of libraries and programs for symbolic and statistical natural language

processing (NLP).

**Node.js**: Open source, cross-platform run-time environment for server-side and networking

applications.

**POS**: Part-of-Speech such as noun, adjective, verb, etc….

**Server Side**: The back-end of the CLASH system responsible text processing, the database, user-authentication, and web-hosting.

**SPA**: Single page application. A highly responsive web application that fits on a single page

and does not reload as the web page changes states.

**TOEFL**: English language proficiency test required by universities for enrollment for internationally based students.

**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

**VM**: Virtual Machine.

# **References**

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.

Engelbrecht, K. (2003, June 18). The Impact of Color on Learning. Retrieved February 25, 2015, from [http://sdpl.coe.uga.edu/HTML/W305.pdf](http://sdpl.coe.uga.edu/HTML/W305.pdf" \t "_blank)

Hoffman, D. (n.d.). Academictips.org - Reading and Highlighting Tips. Retrieved February 25, 2015, from [http://www.academictips.org/acad/literature/readingandhighlighting.html](http://www.academictips.org/acad/literature/readingandhighlighting.html" \t "_blank)

Lab 1 CS 410