CLASH Product Description CS410 Lab 1

Blue/Purple Teams

Old Dominion University

Author: Ali Albattah

CS411W — Janet Brunelle & Hill Price

February 9, 2015

**Table of Contents**

1 Introduction .....................................................................................................2

2 Product Description …………..……….............................................................3

2.1 Key Product Features and Capabilities ..........................................................4

2.2 Major Hardware and Software Components ..................................................5

3 IDENTIFICATION OF CASE STUDY .............................................................7

4 C.L.A.S.H. Product PROTOTYPE DESCRIPTION ........................................8

4.1 Hardware and Software Prototype Architecture .............................................10

4.2 Prototype Features and Capabilities ................................................................11

4.3 Prototype Development Challenges .................................................................12

Glossary....................................................................................................................13

References...............................................................................................................14

# **1 Introduction**

CLASH, or Color Lexical Analysis algorithm and Slash Handler, is an online interface that encompasses two modules critical to its function, ‘Slash’ and ‘COLRS.’ The goal is simple, to increase reading and comprehension speeds by encouraging reading in lexical bundles. Lexical bundles are sequences of words that occur repeatedly and within the same register (Tremblay, Derwing, Libben, Westbury). The ‘Slash’ module parses the text into the aforementioned lexical bundles which would aid in retention and comprehension, as well as increasing reading speed. The ‘COLRS’ module, on the other hand, assigns distinct colors to each part of speech (P.O.S) within the text document to aid in the understanding of grammar, syntax and sentence structure. The primary focus of CLASH would be English as a second language (ESL) students. Based on data from 2004, it is estimated that 4,999,481 ESL students enrolled in public schools. Out of these approximately 5 million students, 15% had no program or a specialized tool to aid them with learning English. In 2001, out of all of ESL students residing in states that tested for reading comprehension, only 18.7% were considered at or above the norm. Also, 10% of ESL students in grades 7-12 were retained in 2001. In February of 2001, it was reported that the dropout rate for ESL students were up to four times that of their native English-speaking peers. With the aforementioned statistics in mind, CLASH aims to aid these ESL students by increasing their reading speeds and comprehension, allowing them to assimilate and rise above the poor statistics. Old Dominion University is well-known for their English Language Center (ELC), which helps international students learn to read and write English at a college level. Despite the hard work done by ELC professors, the processes of teaching the foundations of English are outdated (McKeon).

The current method for teaching English is overly simplified and extremely time consuming. In many instances, the professor would write an incorrect sentence on the whiteboard and then correct the mistakes for the students.

Currently, the process is simple, for grammar, the professor writes a sentence on the board, then circles, or marks in some way each part of speech. Not only is this time consuming, it limits the amount and size of examples that can be given. It has been proven that colors impact learning by increasing information retention and relieving eye fatigue ([HTTP://sdpl.coe.uga.edu/HTML/W305.pdf](http://sdpl.coe.uga.edu/HTML/W305.pdf)). With that in mind, ‘COLORS’ module should allow students to identify P.O.S. as well as retain fundamental grammar rules.  Extrapolating from that information the ‘COLRS.’ CLASH’s goal, however, does not center around speed, which will naturally come if students read in lexical bundles and can more rapidly identify the P.O.S. For students who desire to focus on speed, they would be directed to sites like Spreeder (HTTP://www.spreeder.com). This is only recommended as a complement to the method of reading emphasized in CLASH.

**2 Product Description**

CLASH is a web-based solution designed to aid English as a Second Language students learn English in an enjoyable, fast manner. CLASH will improve reading speed and comprehension by employing methods that teach users to become lexical bundle readers rather than word-to-word readers. CLASH will interact with the students and the database through a Graphic User Interface, which will allow instructors and administrators to observe usage data. CLASH has two modules, Slash and COLRS. Slash will divide text documents into parsed, lexical bundles. Slash will allow users to upload text documents for leisure reading. COLRS colorizes the parts of speech and thus allows student with identifying the eight parts of speech. The COLRS module assists students in identifying parts of speech through the colorizing of text. Colorization of text will be done through JavaScript Object Notation, JSON.

## 2.1 Key Product Features and Capabilities

CLASH is supposed to be a web-based application that will let the users improve their reading understanding by dividing sentences into lexical bundles and showing the parts of speech.

CLASH features can be accessible through a standard web browser and internet connection. Using the application’s user interface, users can manage the speed at which lexical bundles are showed. Parts of speech can be displayed in a separate part of the user interface.

Different types of user roles are provided according to three different controlled users. These roles are 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 have more user capabilities than students. The Instructor will be able to add/remove students and restrict the documents available.

The Instructor will be able to view activity data for the Student users. This activity data will include different kind of options such as the student’s current reading speed, and the amount of time spent on the site. CLASH will allow the instructor to upload documents to the server to be parse, edit files, and delete documents that are found on the server. The Administrator has all the privileges of the Instructor in addition to the ability to add/remove Instructors.

A feature that makes CLASH unique is that it both of the required functions such as displaying the parts of speech 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 ability found within the CLASH is the ability to pause and change the display speed. CLASH will include a text parser this enables the program of the ability to automatically identify parts of speech, and color them respectively, 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. With the required focus by ESL students, CLASH will give students and instructors enormous help to increase reading speed and comprehension.

**2.2 Major Hardware and Software Components**

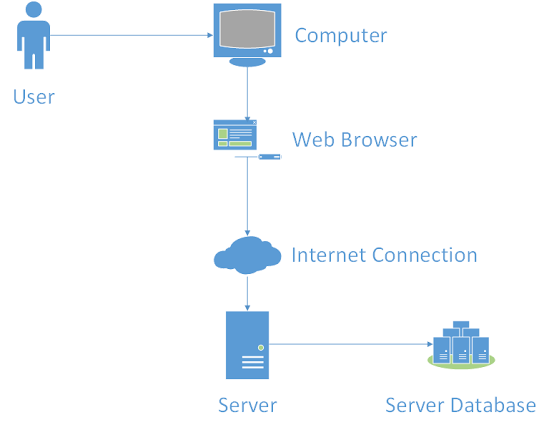
**Overview.** The CLASH will be a web-based application hosted on the server, which can be accessed through any web browser with ease. Besides an active server, a database on the server end, and a device with Internet access for the client, no other hardware or equipment is needed. CLASH will run on three major software components, including the COLRS Module, the Lexical Bundle Module and a Client-side Reader.

**COLRS module.** The COLRS Module will use an open source Natural Language Processing, or NLP, to tokenize and parse the input document in a easier-to-read manner.

**Lexical bundle module.** The Lexical Bundle module will accept the tokenized steams from the COLRS module. It will then decide to break and parse them into lexical bundles.

**Client-side reader.** The client-side reader program will receive the data from the Internet, then parse the text according to its part of speech tag. The two display modes, speed mode and color mode, will serve two different functions. Under the speed mode, the user can choose the reading speed, and under the color mode, the user can choose whether or not the Slash needs to be displayed. Also, under the color mode, the user has the ability to choose which part of speech tags need to be colorized.

**Hardware Requirements**



# **3 IDENTIFICATION OF CASE STUDY**

Old Dominion University is famous for its international students excessive presence. A lot of students from all over the world come to this university to pursue degrees in many programs ranging from finance to engineering, nursing, and computer science. However, for these students to start their academic studies, they will need to prove that their English is good enough for them to get along and get good grades just like their native English classmates. In order for them to prove that, they need to take plenty of tests such as the TOEFL test which is taken online and called IBT. If the students don't have enough knowledge or don't feel confident enough to take the test, then they apply for the English Language Center at ODU.

Greg Raver-Lampman; an instructor at the Old Dominion University English Language Center teaches students who speak English as a second or third language. These students rely on him and other ELC professors to get as much knowledge of the English language before they start their academic classes. These professors do their best to help the students grasp the language for use in future classes. They help them to understand grammar, write paragraphs that later become papers, read articles and books, listen to videos and learn how to listen to discussions, and other techniques to assist in the learning of the language. Often, however, students do not learn as quick as they want to and tend to be very slow readers. Because of this Professor Raver-Lampman requested help from the Computer Science Senior Design class, and a software solution to help students identify all the different parts and lexical bundles of a sentence was designed.

This web application was primarily created to help students at the ELC learn how to read faster. This application can be a superior solution to websites that help other people learn English. 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 listen 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 Single Page Application has an upgraded version which is the Clash prototype. A Web page changes its state, does not refresh and fits within a single page due to the highly responsive web application, which is SPA. JavaScript is the primary programming language in SPA; each single component of SPA architecture has a JavaScript built within and utilized. A good example is the Traditional SPA stack. It includes a JavaScript built user-interface, friendly application and a web server. Examples of these are Node.js, and a NoSQL database such as MongoDB. When the data is stored in JSON format, SPAs use NoSQL databases, which stands for JavaScript Object Notation and its native to the JavaScript. C.L.A.S.H differs from the traditional stack in the fact that the traditional stack uses a traditional relational database instead of using a NoSQL database; though the user-interface will be built in JavaScript and will use the Node.js as its web and application server.

**Benefits to the user.** The perfect platform for the CLASH is the SPA because it acts as a desktop application but has the accessibility of a web application. This provides the user that does not use computer savvy in two main reasons. Initially since it is a web-application all users will need to login to a web portal and there is no new software to install. Secondly, because the entire web application is focused on a single page users won’t be lost in web pages and the user interface will be clean. All actions will be available on the main page.

**Overview of prototype vs real world product.** The prototype was designed to match the real life product as much as possible. Due to the limited time we had there were certain reductions that had to be made:

1. The Database will include a traditional relational database rather than using the NoSQL.

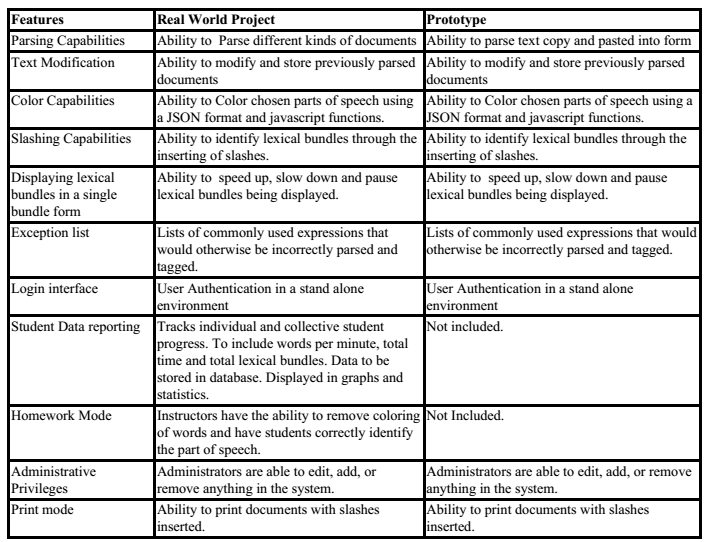
2. Homework mode for students will not have a COLRS module.

3. The COLRS module part-of-speech tagging will be basic and trusting on pre-built natural language processing tools. A custom machine learning will not exist to improve the tool’s accuracy over time.

4. The Product will not be me matched with any ODU enrollment systems. All user accounts should be set up manually.

5. The organizational hierarchy is limited to only one institution. The real world product would need to be able to expand 1to include multiple institutions.

**Overview of simulated areas and models.** The CLASH was built in a user management module since it wasn’t able to integrate with any ODU student enrollment systems. Adding and removing users will be allowed for instructors and administrators. Instructors have the accessibility to add student users on one by one basis and they will also be able to add users in bulk with an ODU enrollment file. Administrators have the ability to remove and add instructors as well as add and remove students.

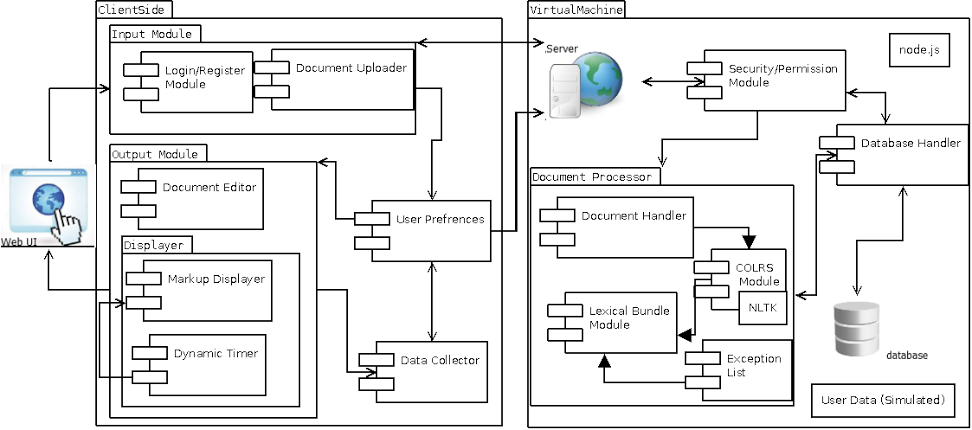


## 4.1 Hardware and Software Prototype Architecture

**Hardware.** The only hardware needed is a computer for the user.

**Software.** The software for this system will be a collection of open source or liberally licensed software to allow for the potential and possible commercialization of the software package. The prototype will encompass the packages installed under Ubuntu 14.04 LTS. The web and application server will be Node.js. The Node.js server will interact with the Natural Language Toolkit. This software will enable the HTTP server to allow for incoming network connections from users.

**Prototype MFCD**



## 4.2 Prototype Features and Capabilities

CLASH will display the ability to identify specific parts of speech by, dashing, coloring, and slashing the text in a lexical bundle stream, with the capability to speed up, slow down, or pause the word stream. In addition, the functional success of CLASH will be determined by the accuracy of the core features and the integration of team-built modules. Determining the risks of the involvement of the mentor will be mitigated by the development of the product for its ease of use, and accuracy. A well-built program with these core features and functions will result in the authentic use of CLASH against a control group. With these tasks being completed, this confirms the ability to demonstrate the feasibility of the system’s achievement; which is, the utility of CLASH in an academic environment.

## 4.3 Prototype Development Challenges

Multiple challenges may face the development of CLASH such as, computational time of the exception list, missing attributes, and the minor bugs and tweaks in the Program’s early stages. The increasing time of the Computation of the exception list will be a challenge for the simple fact that the exception list is escalating in size. Also, the missing attributes; meaning features that are not working completely or missing completely, will also create another dilemma. Although, it is impossible to write a program without minor bugs; CLASH, being unique and very useful will have its bugs and hiccups taken care of as they show up. Knowing that English is a challenging language to learn, a lot of students make a lot of mistakes while learning it. One mistake that students usually make is thinking that they instantly will be great at identifying the different types and parts of speech. As a result, CLASH aims to provide the ability to control the speed at which the lexical bundle will show with the slash feature. CLASH is meant to be simple for use, meaning having a simple interface. In doing so, the program will make the product accessible and welcoming to users with various computer skill levels. However, what we believe is simple might not be simple for certain users; which will negatively affect the appeal of CLASH in the long run. The last challenge we will face is the storage capacity, which will rise with time. Users will populate our databases with their documents. These challenges will be overcame and make CLASH one of a kind.

**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 within the same register

**MFCD –** Major Functional Component Diagram

**NLTK** – a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for the Python programming language.

**Node.js** – an open source, cross-platform runtime 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**

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.