CLASH Product Description CS410 Lab 1

Blue/Purple Teams

By: Mohammed Alabdullatif

Janet Brunelle

Hill Price

Old Dominion University

February 9, 2015

**Table of Contents**

1. Intro ..................................................................................................................2
2. Product Description ........................................................................................3
   1. Key Product Features and Capabilities ...................................................3
   2. Major Hardware and Software Components ..........................................4
3. IDENTIFICATION OF CASE STUDY ………………................................5
4. C L A S H Product PROTOTYPE DESCRIPTION .................................6
   1. Hardware and Software Prototype Architecture ……...........................8
   2. Prototype Features and Capabilities ……...............................................8
   3. Prototype Development Challenges ………………..............................9

Glossary...................................................................................................................10

References...............................................................................................................10

**1 Intro**

CLASH, Color Lexical Analysis algorithm is a website interface that has two modules; ‘COLRS’ and ‘Slash’. The ‘COLRS’ module adds colors to each part of speech (P.O.S) in a document with a particular color to help increase comprehension of a phrase structure and grammar. The ‘Slash’ module part separate a text into a bunch lexical bundles or thought groups to help students read faster and understand better. A lexical bundle is a group of words that occur repeatedly together within the same register. The program is mostly focusing on English as a second language (ESL) students.

It is estimated that 4,999,481 ESL students were enrolled in public schools. Despite those numbers however, ESL students have no other useful methods to help them learn the language.

Nowadays, the process is a little it easy, for the grammar part, any professor will write a sentence of two on the board, then he would circle each part of the sentence including the verbs and the nouns and different expressions. This is so time consuming for both the professor and the students. Psychologically, it is proven that colors have a great impact on learning.

The COLRS program should be useful for students to identify POS and improve their ways of learning the grammar. In order to improve the reading and comprehension of students, professors tend to give a lot of reading assignments, and they also advice students to use websites such as Spreeder.

# 2 Product Descriptio

Clash is a program that will help students learn English in a quicker and better way. It is going to be based on the web and will possess a graphic user interface for the interaction between students and the database. This is going to be very useful for instructors as well. It consists of two parts, the Slash part, and the COLRS part.

The Slash part will work on improving the reading quality of the students, it will be doing that by breaking up documents into lexical bundles in order for the students to better understand the passage. It will also allow the uploading of different documents for students and instructors.

The COLRS part of the program helps students identifying and understanding parts of a sentence of speech by colorizing each part of it. The program will focus on the eight traditional parts of speech which studnets have learned before. The colorizing of the text will be done by the help of the JavaScript application.

## 2.1 Key Product Features and Capabilities

Clash is an internet application that will allow students and users in general to increase their reading skills by breaking the sentences into many parts. These parts constitute of the different parts of the speech. This application can be accessed simply through any web page.

The program has three different individual password controlled user logins each has a role. There are three roles which are the administrator, instructor, and of course the student. The later will be allowed to control their reading skills and speed, as well as many documents to view. The instructor part will include a little bit more features than the students one, as well as the ability to control which documents are made available for the students.

The fact that Clash shows both the parts of the speech and its lexical bundles makes it very unique from any other web application online. The students have the ability to pause and continue the reading of many sentences, as well as the speed of reading. The program is able to identify parts of speech, and color them to make them easier to grasp. The program will be able to save the usage data and history, which will be useful for instructors in order to track their students’ improvement. This program is the first web application made ready for ESL students.

## 2.2 Major Hardware and Software Components

**Overview.**

Clash is a web application that is made available on the web server. There are no special machines that are required to use the application, all a user need is a database on the server end, and an internet enabled device for the client to access the application. There are three different parts in this program which are CLASH, COLRS, Lexical bundle module, and a client side reader.

**COLRS module.**

COLRS part of the application will use the open source natural language processing (NLP) to tokenize and parse the documents and sentences and generate them into a more easier to read way.

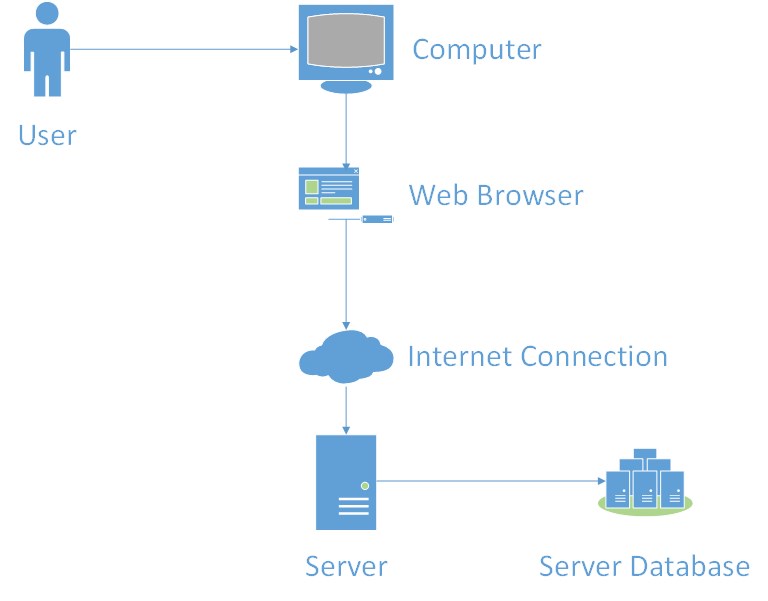
**Lexical bundle module.**

The Lexical bundle module will be allowed to accept already tokenized streams from the first module, and decide to break and parse them into lexical bundles.

**Client-side reader.**

The reader part of the program will be allowed to receive different types of data from the internet and the database, and parse the texts and sentences according to their POS tag. There will be two different display modes in the application; color mode, and speed mode. Under the color mode, any user can choose if the slash can be displayed or not. Users can also choose the reading speed under mode.

# Hardware Requirements



# 3 IDENTIFICATION OF CASE STUDY

Old Dominion University is famous for its international students excessive presence. A lot of students come from all over the world to this university to pursue degrees in many programs ranging from finance, 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 who teaches English to 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 fast 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

Clash is a program that has a prototype that will be built as a modified versions of the single page application called SPA architecture. A SPA is a very responsive application that fits in a single page and does not reload. Normally, SPA are usually made using the Java Script application and each component of the program use the same application. There is also another part called NoSQL database such as MongoDB. The reason why SPA use the NoSQL is because the data stored is stored in a JSON format.

**Benefits to the user.**

SPA is a platform for the CLASH program because it allows the users to experience the application with a look of a desktop application with the accessibility of a web application. This is really useful for users who are not very good at using a computer in a professional way. This is because of two main reasons, first because it only uses the web and not a special application or software to make it work, and second because the entire application is based in one page.

**Overview of prototype vs real world product.**

When we designed the prototype, we really wanted the application to be as close to the real product. These are some of the concessions we had to make:

1. Database will be a traditional relational database rather than NoSQL.
2. The COLRS module will not have a homework mode for students.
3. The COLORS module part of the sentence will rely on a prebuilt natural language processing tool.
4. Will not integrate with any ODU enrollment systems. All user accounts will be set up manually.
5. The organizational hierarchy will only include one institution. The real world product would need to be able to scale to include multiple institutions.

**Overview of simulated areas and models.** Because Clash users cannot integrate the application with any ODU Student enrollment systems. We decided to create a user management module system. This way instructors will be able to add and remove any users of the program. They will also be able to add or remove either students or instructors.

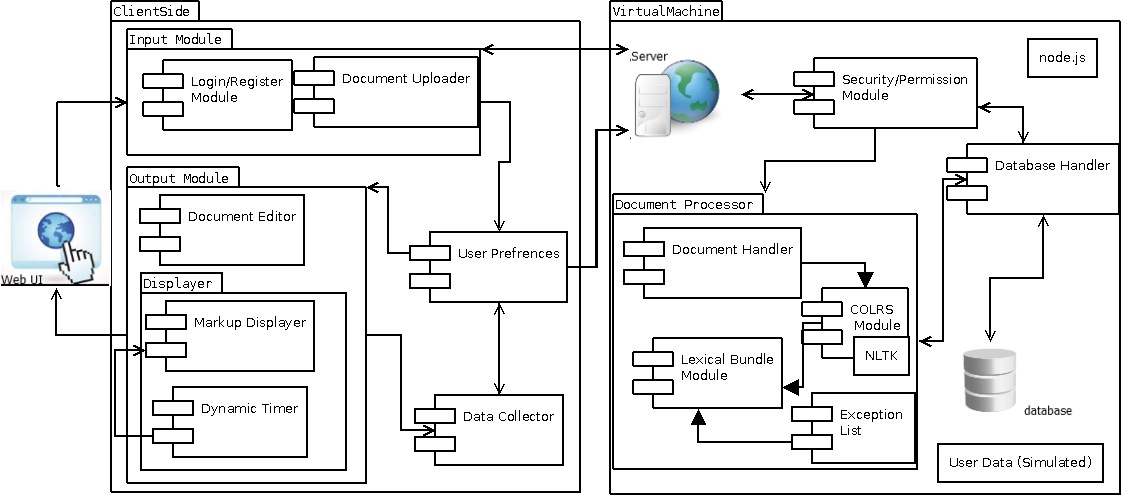
## 4.1 Hardware and Software Prototype Architecture

**Hardware**

The only hardware needed to work the application is a simple computer or laptop owned by the users.

**Software.**

When it comes to which type of software is required for the prototype, this system will need a collection of open source or licensed software to permit distribution or potential commercialization of the software package. Our prototype will consist of the following packages installed under Ubuntu 14.04 LTS. The web and application server will be Node.js. This software will provide the HTTP server to allow incoming network connections from users (the students and instructors), as well as the application handling text processing of documents.



## 4.2 Prototype Features and Capabilities

Clash is a program that will help students identify so many different parts of the speech by coloring the text and will also identify the different lexical bundles through the inserting of slashes. The prototype will also allow the users to show different bundles one at a time, and increasing and decreasing or stopping the lexical bundle stream. Success of the prototype will be chosen based on the accuracy of the tagged parts of speech and lexical bundles.

Based on the working prototype, these features will function and result in the customer and allow him to test it in the real world in any situation against a control group. Completing these proofs of concepts the program prototype will demonstrate the feasibility of such a system and together with the prototype testing group.

### 4.3 Prototype Development Challenges

Although the Clash program is well designed, there are times when the application will be facing different challenges. One of them is a computational time of the exception list. By adding more exceptions to the list, computing time increases due to the increase of size. Another challenge is the missing attributes which means that the features are not working correctly. CLASH, being unique and useful will have less hiccups that need to be taken care of when they show up.

English is not an easy language to learn that is why a lot of students make a lot mistakes usually when learning it. One main mistake that students usually make is the fact that it is impossible for students to be great at identifying the different types of speech and parts of it. CLASH aims to provide the user the ability to control the speed at which the lexical bundle will show with the slash feature.

Lab 1 CLASH Product Description

# 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

**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 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.

# 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.