

**TMNT**

**and Janie**

SOUTHERN METHODIST UNIVERSITY

MustangWiki Search Engine

Administrative User Manual

Mustangwiki

Administrative User Manual

Janie Pascoe Alex Saladna Ziga Cerkovnik

Table of Contents

How to Use 1

[1] Interactive Mode 1

[2] Maintence Mode 1

[3] Hash Table 2

[4] AVL Tree 2

Index 3

Chapter

1

What is MustangWiki?

M

ustangWiki is a collaborative, open place where students, faculty, staff, and the community can create online, freely-accessible content. From historical events to detailed chemistry experiments, MustangWiki has it all, but it’s broken! Containing over 170,000 entries, MustangWiki is a search engine designed to find relevant information from a Wiki Books dump and display it to the user but it has lost its search. We have been asked to implement a faster engine and our new one is exactly that.

# How to Use

MustangWiki is designed to be incredibly easy to use. With your Wiki Books dump [in XML format] in your source folder, you can begin running MustangWiki, which will take some time to complete based on your own computer's speed. During this time, the program is also removing all "stop words" as well as stemming all of the words it is parsing in. Once the data has been loaded the terminal will initially ask you which mode you would like to be in. Here are the differences between your two options and what they allow you to do:

## [1] Interactive Mode

* Choose a data structure
* Enter a query
* Return to mode selection

## [2] Maintenance Mode

* Add Documents
* Clear the index
* Return to mode selection

If the user chooses interactive mode, you then have the ability to choose between two different data structures and these are the differences between your options:

## [3] Hash Table

* Uses an associative array
* Generates an index of buckets using a hash function
* O (1)

## [4] AVL Tree

* Binary search tree
* Balances itself
* O (log n)

Once you, the use have chosen a data structure by either entering ‘1’ for AVL tree or ‘2’ for Hash Table, the data will be stored in the correct structure and you will be asked if you would like to search through the document. Upon entering ‘Y’ for yes or ‘N’ for no the program will either terminate because of entering ‘N’ or you will be prompted to enter a search because of entering ‘Y’. The search terminal allows you to enter words or nested searches, most of which prefixed with ‘AND’ or ‘OR’ if multiple words are being searched. Upon pressing enter, the search engine will return the top 15 documents that this word appears most on in order of relevancy. It will display the title, author, date, document number, and relevancy number for you to view. You are then able to select whether or not you would like to view the text of a document by entering ‘Y’ or ‘N’ and which document you would like by inputting the respective number associated with that document’s ordering in the output. (i.e. 1-15 and not the actual document number). Then MustangWiki will display the entire document for your viewing pleasure.

Once you are finished with your search you have a few options. You will be prompted to answer ‘Y’ for yes or ‘N’ for no if you would like to search for another term or phrase. If you answer no, then the program will terminate, but if you answer yes, you may enter a new search and the program will provide the top 15 output as before based on your input query.

# Index

Interactive Mode ……………………………..2

Maintence Mode ……………………………..2

AVL 3

Hash Table 3

SMU 2

Table of Contents 1

XML 2