Yuchen Tian, Bailey Kingsley and Michael Tran
11/13/2017
Assignment Number: 4
CMSC 409 Artificial Intelligence

1 Term Document Matrix

1.1 Answer

Table

2 Mining Steps

2.1 A

2.1.1 Use

Divide sentences into a list of substrings or words

2.1.2 Data Lost

Lost Spaces

2.2 B

2.2.1 Use

Removing any anomalies that might affect reading the word

2.2.2 Data Lost

Punctuation and special characters

2.3 C

2.3.1 Use

Removing any anomalies that might affect reading the word

2.3.2 Data Lost

Numbers

2.4 D

2.4.1 Use

Removing any anomalies that might affect reading the word

2.4.2 Data Lost

Upper cased letters

2.5 E

2.5.1 Use

Removing overly used words that are very common in the English language

2.5.2 Data Lost

Removed stopped words

2.6 F

2.6.1 Use

Process for removing the commoner morphological and inflexional endings from words in English

2.6.2 Data Lost

Suffix, prefix and plural words

2.7 G

2.7.1 Use

To reform a sentence

2.7.2 Data Lost

N/A

Appendix A   Workload

A.1   Michael Tran

Worked on documentation

A.2   Bailey Kingsley

Worked on code

A.3   Yuchen Tian

Worked on code