



# Assignment 1

## 3-gram Language models

Mourad Akandouch  
INFOY004  
Academic year : 2017-2018  
ULB - VUB

### TABLE OF CONTENTS

<i>Introduction</i> .....	2
---------------------------	---

## *INTRODUCTION*

The aim of our first assignment was to write language models based on a training set. After that, regarding the generated language models, we had to detect in which varieties of English a document is written. That classification was based on *perplexity*.

The language models we had to build are based on trigrams letter and the training set consists of a corpus of three varieties of English: British, Australian and American.

Also, we had to generate a random output based on the generated language models for each varieties of English. Those generated strings have a fixed length  $k$ . In this report, I will put a string of length  $k = 200$  as mentioned in the assignment.

Finally, we had to score each sentence in the test set with each of our generated language models.