

Relatório de IC

Projeto 3

Universidade de Aveiro



DETI

Tiago Pedrosa

93389

João Gonçalves

98287

20 de dezembro de 2022

Conteúdo

1	Fcm	2
2	Lang	3
3	Findlang	4
4	Contribuições dos Autores	5

Capítulo 1

Fcm

O objetivo do programa é fornecer ferramentas que permitam recolher informações estatísticas sobre os dados inseridos. Ao serem fornecidos os parâmetros do ficheiro de entrada, a ordem do modelo e o "*smoothing parameter*" este consegue estimar a entropia do texto com base no modelo de contextos finitos anteriormente calculados.

A função utiliza cadeias de Markov, sendo que nestes modelos a próxima letra na palavra é inspirado pelas letras anteriores a essa. Ao ser usado esse modelo a maneira de armazenamento da estrutura de dados é feito com o registo de todas as palavras do tamanho que foi fornecido, na ordem do modelo, e também a probabilidade de cada letra que poderá vir a seguir a essa palavra.

Capítulo 2

Lang

Na implementação do código de **lang.cpp**, recebe-se dois ficheiros, e com estes ficheiros, é contruído um modelo através do método **loadModel** presente na classe **fcm**.

Na classe **fcm** têm-se disponível o método que permite a estimativa do número de bits necessários para comprimir o segundo ficheiro, isto é feito através do método **estimate**, que ao ser chamado imprime os valores calculados. Os valores que são gerados serão maiores o quão mais diferente a linguagem do segundo ficheiro for da do modelo criado, ou seja caso as linguagens sejam semelhantes os valores apresentados serão menores.

Capítulo 3

Findlang

Para este programa é pretendido este fazer reconhecimento de linguagem consoante a estimativa do número de bits necessários para a compressão no modelo, ou a sua distância. Para isso são inseridos diversos modelos e comparados a um só texto de modo a descobrir a qual linguagem este texto se assemelha mais.

Neste programa utiliza-se modelos já disponíveis, e faz-se a comparação com o ficheiro inserido, o modelo que tiver a menor distância irá ser registada ao longo do processamento do programa, tanto o seu nome como a distância, ficando assim a saber qual a linguagem que o ficheiro utiliza.

Capítulo 4

Contribuições dos Autores

O Trabalho foi analisado, discutido e realizado de modo homogêneo por:

- Tiago Pedrosa 93389
- João Gonçalves 98287