Grupo 19

Gonçalo Bustorff Silva 82050

Guilherme Gomes 89378

**MP2 – Segundo Miniprojecto de Língua Natural**

**Descrição do problema**

É necessário desenvolver uma métrica de similaridade que permite identificar o tipo de uma questão de cinema. Dado o ficheiro **QuestoesConhecidas.txt**, onde se encontram questões pré-anotadas, e outros recursos como lista de filmes ou pessoas consideradas na solução, deverá ser possível classificar as novas perguntas com um dos 16 tipos possíveis:

*actor\_name, budget, character\_name, genre, keyword, original\_language, original\_title, overview, person\_name, production\_company, production\_country, release\_date, revenue, runtime, spoken\_language, vote\_avg*

Exemplificando, diga-se que nas **QuestoesConhecidas.txt** encontram-se as perguntas:

*budget How much budget did Valkyrie have?*

*budget What was the budget of Furious 7?*

Dada a nova questão:

*What was American Gangster's budget?*

A métrica implementada deverá classificar a pergunta do tipo:

*budget*

**Proposta de solução**