



Projeto Final

Este projeto avalia a capacidade do aluno em mapear os conhecimentos adquiridos na disciplina para outro domínio. Pode ser feito de forma individual ou em duplas. Data final de entrega no dia **06/04/2024** em forma de relatório via Sigaa.

1. Motivação

O *fact-checking* é o processo de verificação de fatos, muito utilizado atualmente devido à proliferação das *fake news*. Por exemplo, ao assistir a um debate político entre dois candidatos, como neste link¹, fica impossível decidir pelo voto devido a briga de narrativas. Ou seja, não é possível saber quem está falando a verdade naquele momento. Alguns exemplos extraídos do link indicado:

- “A paraíba é o estado do Nordeste que mais ampliou sua rede obstetrícia”
- “Abrimos a maternidade Peregrino Filho em Patos”
- “Você não abriu um leito em hospital público”
- “Eu abri 1044 leitos”
- “Eu construí o Hospital de Trauma em Campina Grande”
- “Mortalidade infantil no ano de 2008 era de 21,5%, hoje está em 16%”

2. Definições

- **Claim:** uma declaração que está intencionada a ser verdade
 - ✓ “The U.S. government calculates inflation without adding in the price of food and energy”
- **Snippet:** é um trecho de alguma fonte de informação onde o *claim* é referenciado
 - ✓ “While advising his Fox News viewers to talk about inflation at their Thanksgiving dinners, Glenn Beck falsely claimed that **the government removed food and energy prices from its measure of inflation to hide rising prices**, that a survey showed economists are “worried” about inflation, and that Social Security recipients are not receiving a cost-of-living adjustment because the government “changed the calculation.”
- **Verdict:** é o veredito de um *claim* que indica o quanto o mesmo é verdadeiro

¹ <https://www.youtube.com/watch?v=HJhSei2lKuQ>

Quando utilizamos o *claim* original em uma busca (google), o mesmo não retorna nenhum link que possa ser utilizado na verificação do fato. Porém, se o mesmo texto eh modificado da forma abaixo

Original: “The U.S. government calculates inflation without adding in the price of food and energy”

Modificado: “The government removed food and energy prices from its measure of inflation to hide rising prices”

Temos como resposta o seguinte link:

<https://www.mediamatters.org/glenn-beck/beck-tells-audience-ruin-thanksgiving-dinner-misinformation-about-inflation>

Apesar das alterações léxicas e sintáticas, observe que as duas sentenças acima possuem a mesma semântica. Ou seja, estas alterações são importantes para aumentar o número de referências que podem ser verificadas de forma a se ter uma conclusão mais precisa sobre um fato qualquer.

2. Especificação do projeto

Desenvolva um sistema onde podemos incluir um *claim* e ele retorna um ou mais fontes que contenham os *snippets*. Os seguintes elementos são necessários nesta implementação:

- Um analisador léxico para a lingua portuguesa
- Um analisador sintático para organizar a estrutura original do texto
- Uma base de regras para saber como modificar o texto
- Algum recurso de acesso a sinônimos (e.g., Thesaurus)

Vamos analisar cada um destes elementos.

2.1 Análise léxica

A análise léxica é muito fácil de fazer nesse caso porque palavras são separadas por espaços. Python, por exemplo, já possui funções para fazer o tokenizer de uma sentença. Outra função é classificar os tokens nas seguintes classes gramaticais:

- ✓ Artigo
- ✓ Substantivo
- ✓ Verbo
- ✓ Adjetivo
- ✓ Advérbio
- ✓ Preposição
- ✓ Pontuação final: ! (exclamação), ? (interrogação), . (ponto final)
- ✓ Pontuação intermediária , (vírgula)

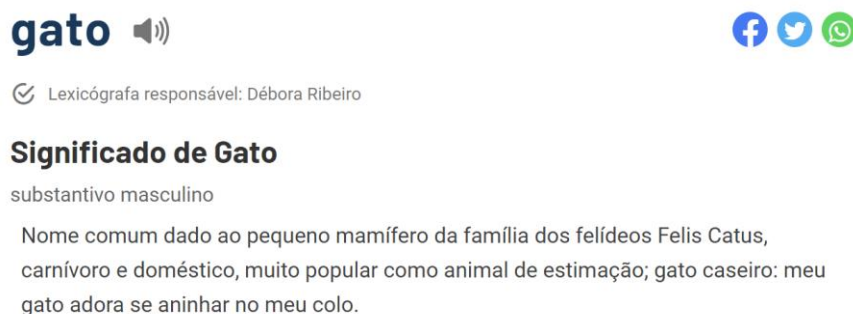
Dica: verifiquem este link para ver se ele auxilia na classificação das palavras

<http://nilc.icmc.usp.br/macmorpho/>

Outra opção é fazer consultas em tempo real a dicionários da web, uma vez que eles indicam a classe gramatical da palavra.

<https://www.dicio.com.br/>

Por exemplo, montando uma consulta para a palavra “gato” (<https://www.dicio.com.br/gato/>), o mesmo retorna:



Ou seja, devemos extrair a tag “substantivo” da resposta (web request) pra classificar o termo. Essa são apenas dicas. Outras opções podem ser utilizadas.

2.2 Análise sintática

Especificar uma gramática simples, como a definida abaixo. Tal gramática pode ser utilizada, mas vocês estão livres para estender a mesma.

- ✓ Texto → Sentença. | Sentença. Texto
- ✓ Sentença → Sintagma_nominal Sintagma_verbal
- ✓ Sintagma_nominal → Substantivo | Artigo Substantivo
- ✓ Sintagma_verbal → Verbo | Verbo Sintagma_verbal
- ✓ Artigo → ...
- ✓ Substantivo → ...

2.3 Regras e sinônimos

Sabendo quais são os elementos de um *claim* e como eles estão organizados, quais as formas que podem ser utilizadas para modificar a frase?

- ✓ Utilizar sinônimos (thesaurus), <http://www.sinonimos.com.br/>
- ✓ Utilizar reorganização dos tokens. Por isso a importância das classes gramaticais.
- ✓ A pesquisa final pode ser realizada no Google.

Ou seja, usem a imaginação!!!!

ENTREGA ATÉ O DIA 06/MAIO VIA SIGAA EM FORMA DE RELATORIO COM AS DECISOES DE PROJETOS DE CADA FASE!!!

O relatório deve conter a solução implementada para a análise léxica, análise sintática, e estratégias para modificação do *claim*. Usar exemplos de claims, mostrar o resultado das modificações, e a quantidades de links retornados pela busca (estas devem ser feitas sempre entre aspas).

Observações:

- Alguns alunos sugeriram projetos específicos. Estes podem ser desenvolvidos sem problemas.
- O projeto final original seria relacionado aos links abaixo. A ideia seria construir um otimizador de parâmetros de compilação para o GCC. Porém, ele envolve uma maior complexidade e conhecimentos em aprendizagem de máquina. As instruções estão no primeiro link abaixo. Tal implementação também serve como uma opção para terceira nota (projeto), caso algum aluno deseje se aventurar.

<https://compilergym.com/envs/gcc.html>

<https://ai.meta.com/blog/compilergym-making-compiler-optimizations-accessible-to-all/>

https://www.youtube.com/watch?v=aT2b_YAEyDw

<https://eprints.lancs.ac.uk/id/eprint/89859/1/main.pdf>