**RII - Tarefa 2**

**Extração de Informação Textual**

**Descrição geral:**

Trabalho em duplas

**Tarefa:** realizar Extração de Informação a partir dos documentos indicados pela professora. Dois tipos de sistema de extração deverão ser construídos:

1. Um Wrapper para extrair informação fixa, baseada em tags
   * Construído manualmente por vocês (podem usar aprendizagem de máquina ou regras).
   * Corpus de 20 documentos fornecidos pela professora
2. Um sistema para EI aberta, para extrair relações
   * Usar ferramenta disponível = OpenIE
     + http://openie.allenai.org/
   * 10 documentos

**Dados para extração com wrapper**

Exemplo: Ver sites de classificados

* <http://www.jornalclassificados.com.br/anuncios.php?cat=2>

### Anunciante

* Telefone: (19) 4119-0459 (19) 9 9269-9019
* Email:
* Endereço:

### Produto/ Serviço

* **ESCORT GL 1.6 84**
* alcool segundo dono impecável para colecionador equipado sem detalhes
* R$ 17.000,00

### Anunciante

* Telefone: (19) 9 7407-5369
* Email: paiva.celo@ig.com.br
* Endereço:

### Produto/ Serviço

* **CLASSIC 13**
* flex completo (-) ar, com manual e chave reserva, alarme, som, ac. troca menor
* R$ 21.500,00

**Outro site:** <https://www.classificadosodia.com.br/anuncio/veiculos/2872494916/chevrolet-astra78465>

#### valor:

## R$ 13.500,00

**ASTRA HATCH 1.8 - GASOLINA E GNV - CARRO DE GARAGEM COM MANUTENÇÃO EM DIA. PNEUS NOVOS. SEM BATIDAS. LINDO E INTEIRÃO.**

#### Tipo:

**Automoveis**

#### Marca:

**Chevrolet**

#### Modelo:

**Astra**

#### Ano:

**2001**

#### Quilometragem:

**176000**

#### Portas:

**3**

#### Cor:

**Cinza**

#### Combustivel:

**Gasolina e Gás Natural**

#### Câmbio:

**Manual**

#### Novo/Usado:

**Novo**

#### Motor:

**1.8**

**Características e Acessórios:**

**Pneus Novos, Alarme, Ar Quente, Ar-Condicionado, Desembaçador Traseiro, Direção Hidráulica, Limpador Traseiro, Rádio e CD-Player, Retrovisores Elétricos, Travas Elétricas, Vidros Elétricos, Insulfilm,**

**Projeto & Relatório** (tudo misturado):

Escrevam um relatório simples, com as informações solicitadas abaixo:

# Nome dos membros da equipe:

# Descrição dos documentos (corpus) de entrada

Os documentos serão oriundos do repositório ‘arxiv.org’, que armazena artigos de várias áreas de estudo.

* Exemplo: <https://arxiv.org/list/quant-ph/new>

### New submissions for Mon, 8 May 17

[1]  **[arXiv:1705.01942](https://arxiv.org/abs/1705.01942" \o "Abstract) [[pdf](https://arxiv.org/pdf/1705.01942" \o "Download PDF),**[**other**](https://arxiv.org/format/1705.01942)**]**

**Improving the Accuracy of an Adiabatic Quantum Computer**

[John E. Dorband](https://arxiv.org/find/quant-ph/1/au:+Dorband_J/0/1/0/all/0/1)

Subjects: **Quantum Physics (quant-ph)**; Emerging Technologies (cs.ET)

The purpose of the D-Wave adiabatic quantum computer is to find a set of qubit values that minimize its objective function. For various reasons, the set of qubit values returned by the D-Wave has errors. This paper presents a method of improving the results returned by the D-Wave. The method individually modifies the qubit values returned by the D-Wave to find a set of values which is a minimum of the objective function. That set however is not necessarily guaranteed to be a global minimum. The method is simple and easily incorporated into any algorithm that has direct access to the sets of values returned by the D-Wave. Examples are also presented that demonstrate the merit of using such a sample improvement method.

# Template de saída

* Authors
* Title
* Subjects
* Comments
* Link
* Relations

Exemplos:

[1]

Authors: John E. Dorband

Title: Improving the Accuracy of an Adiabatic Quantum Computer

Subjects: Quantum Physics (quant-ph); Emerging Technologies (cs.ET)

Comments:  NULL

Link: arXiv:1705.01942

Relation1: find (D-Wave adiabatic quantum computer, set of qubit values)

Relation2: presents (paper, a method of improving the results)

Relation3: etc

**Obs.: essas relações são ilustrativas. O formato da relação extraída depende da ferramenta usada.**

# Arquitetura do sistema

Apresentar uma figura simples com os módulos do sistema. Prover uma descrição breve de cada módulo, e informar qual a técnica usada na construção do wrapper.

# Testes e avaliação dos resultados

A avaliação dos resultados terá que ser feita de forma manual. Neste caso, não temos matriz de relevância montada a priori ☺

* **Precisão:**
* **Cobertura:**

Incluir no relatório uma matriz de resultados para CADA módulo do sistema.

**Matriz de resultados para o Wrapper usando 20 documentos de entrada**

|  |  |  |
| --- | --- | --- |
|  | Precisão | Cobertura |
| Campo 1 | 18/18 | 18/20 |
| Campo 2 | 15/20 | 15/20 |
| Campo 3 | ... | ... |
| Campo 4 |  |  |
| Campo 5 |  |  |
| Média |  |  |

**Matriz de resultados para a EI aberta = relações em 5 documentos**

|  |  |
| --- | --- |
|  | Extraída corretamente  (positivo verdadeiro) |
| Relação 1 | sim |
| Relação 2 | não |
| Relação 3 | sim |
| Relação 4 | ? – não extraída (ausente) |
| Etc... |  |
| Precisão | 2/3 |
| Cobertura | 2/4 |

# Conclusão

Um texto curto para explicar o que vocês concluem a partir do resultado dos experimentos.