**Mineração da Web - Tarefa 2**

**Construção de um sistema de Extração de Informação textual**

**Descrição geral:**

Trabalho em equipes de, no máximo, 4 pessoas => nem um integrante a mais!

**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 = CoreNLP
	* 3 documentos escolhidos dentre os 20 iniciais

Quero ver o sistema em execução.

**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%22%20%5Co%20%22Abstract) [[pdf](https://arxiv.org/pdf/1705.01942%22%20%5Co%20%22Download%20PDF),**[**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%3A%2BDorband_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.

[3]  **[arXiv:1705.01962](https://arxiv.org/abs/1705.01962%22%20%5Co%20%22Abstract) [[pdf](https://arxiv.org/pdf/1705.01962%22%20%5Co%20%22Download%20PDF),**[**other**](https://arxiv.org/format/1705.01962)**]**

**Experimental verification of entanglement generated in a plasmonic system**

[F. Dieleman](https://arxiv.org/find/quant-ph/1/au%3A%2BDieleman_F/0/1/0/all/0/1), [M. S. Tame](https://arxiv.org/find/quant-ph/1/au%3A%2BTame_M/0/1/0/all/0/1), [Y. Sonnefraud](https://arxiv.org/find/quant-ph/1/au%3A%2BSonnefraud_Y/0/1/0/all/0/1), [M. S. Kim](https://arxiv.org/find/quant-ph/1/au%3A%2BKim_M/0/1/0/all/0/1), [S. A. Maier](https://arxiv.org/find/quant-ph/1/au%3A%2BMaier_S/0/1/0/all/0/1)

Comments: 7 pages, 4 figures

Subjects: **Quantum Physics (quant-ph)**; Mesoscale and Nanoscale Physics (cond-mat.mes-hall); Optics (physics.optics)

A core process in many quantum tasks is the generation of entanglement. It is being actively studied in a variety of physical settings - from simple bipartite systems to complex multipartite systems. In this work we experimentally study the generation of bipartite entanglement in a nanophotonic system. Entanglement is generated via the quantum interference of two surface plasmon polaritons in a beamsplitter structure, i.e. utilising the Hong-Ou-Mandel (HOM) effect, and its presence is verified using quantum state tomography. The amount of entanglement is quantified by the concurrence and we find values of up to 0.77 +/- 0.04. Verifying entanglement in the output state from HOM interference is a nontrivial task and cannot be inferred from the visibility alone. The techniques we use to verify entanglement could be applied to other types of photonic system and therefore may be useful for the characterisation of a range of different nanophotonic quantum devices.

# 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.**

[3]

Authors: [F. Dieleman](https://arxiv.org/find/quant-ph/1/au%3A%2BDieleman_F/0/1/0/all/0/1), [M. S. Tame](https://arxiv.org/find/quant-ph/1/au%3A%2BTame_M/0/1/0/all/0/1), [Y. Sonnefraud](https://arxiv.org/find/quant-ph/1/au%3A%2BSonnefraud_Y/0/1/0/all/0/1), [M. S. Kim](https://arxiv.org/find/quant-ph/1/au%3A%2BKim_M/0/1/0/all/0/1), [S. A. Maier](https://arxiv.org/find/quant-ph/1/au%3A%2BMaier_S/0/1/0/all/0/1)

Title: Experimental verification of entanglement generated in a plasmonic system

Subjects: Quantum Physics (quant-ph); Mesoscale and Nanoscale Physics (cond-mat.mes-hall); Optics (physics.optics)

Comments: 7 pages, 4 figures

Link: [arXiv:1705.01962](https://arxiv.org/abs/1705.01962)

Relation1: is (the generation of entanglement, a core process)

Relation2: is generated (entanglement, via the quantum interference)

Relation3: etc

# 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 3 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.