

Alunos: Lucas da Silva Jorge  
André Marcello Figueira

## Map Reduce algorithm

### Questão 3

Rodando o programa com o arquivo **test.txt**, obteve-se as saídas **result-0** e **result-1**. As saídas estão de acordo com o esperado. Tomando, por exemplo, a palavra “**there**”, que aparece 4 vezes no arquivo teste, tem-se que a saída **result-1** conta corretamente as 4 vezes que ela aparece:

The field of distributed computing covers all aspects of computing and information access across multiple processing elements connected by any form of communication network, whether local or wide-area in the coverage. Since the advent of the Internet in the 1970s, **there** has been a steady growth of new applications requiring distributed processing. This has been enabled by advances in networking and hardware technology, the falling cost of hardware, and greater end-user awareness. These factors have contributed to making distributed computing a cost-effective, high-performance, and faulttolerant reality. Around the turn of the millenium, **there** was an explosive growth in the expansion and efficiency of the Internet, which was matched by increased access to networked resources through the World Wide Web, all across the world. Coupled with an equally dramatic growth in the wireless and mobile networking areas, and the plummeting prices of bandwidth and storage devices, we are witnessing a rapid spurt in distributed applications and an accompanying interest in the field of distributed computing in universities, governments organizations, and private institutions. Advances in hardware technology have suddenly made sensor networking a reality, and embedded and sensor networks are rapidly becoming an integral part of everyone's life from the home network with the interconnected gadgets to the automobile communicating by GPS (global positioning system), to the fully networked office with RFID monitoring. In the emerging global village, distributed computing will be the centerpiece of all computing and information access sub-disciplines within computer science. Clearly, this is a very important field. Moreover, this evolving field is characterized by a diverse range of challenges for which the solutions need to have foundations on solid principles. The field of distributed computing is very important, and **there** is a huge demand for a good comprehensive book. This book comprehensively covers all important topics in great depth, combining this with a clarity of explanation and ease of understanding. The book will be particularly valuable to the academic community and the computer industry at large. Writing such a comprehensive book has been a Herculean task and **there** is a deep sense of satisfaction in knowing that we were able complete it and perform this service to the community.

Figura 1: “there” aparece 4 vezes em test.txt

```

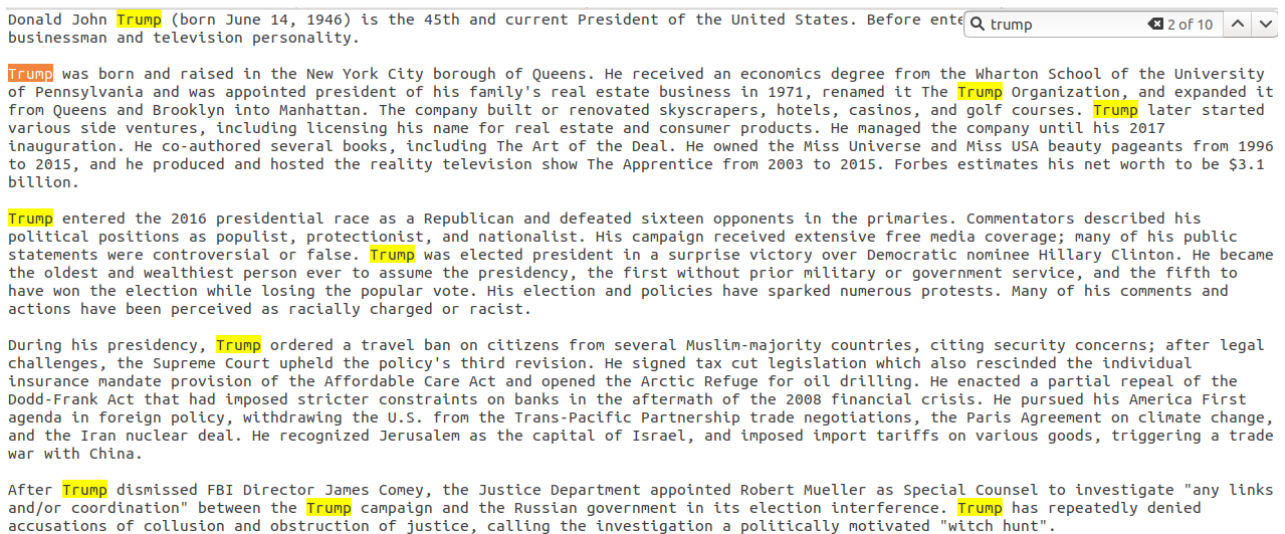
{ "key": "is", "value": "3" }
{"Key": "clarity", "Value": "1"}
{"Key": "coverage", "Value": "1"}
{"Key": "there", "Value": "4"}
{"Key": "access", "Value": "3"}
{"Key": "very", "Value": "2"}
{"Key": "solutions", "Value": "1"}
{"Key": "particularly", "Value": "1"}

```

**Figura 2:** “there” é contada 4 vezes pelo wordcount algorithm

## Questão 4

Rodando o programa com o arquivo **test\_trump\_wikipedia.txt**, obteve-se as saídas **result-0\_trump** e **result-1\_trump** (em anexo com o relatório). As saídas estão de acordo com o esperado. Tomando, por exemplo, a palavra “**Trump**”, que aparece 10 vezes no arquivo teste, tem-se que a saída **result-1\_trump** conta corretamente as 10 vezes que ela aparece:



Donald John **Trump** (born June 14, 1946) is the 45th and current President of the United States. Before entering politics, he was a businessman and television personality.

**Trump** was born and raised in the New York City borough of Queens. He received an economics degree from the Wharton School of the University of Pennsylvania and was appointed president of his family's real estate business in 1971, renamed it The **Trump** Organization, and expanded it from Queens and Brooklyn into Manhattan. The company built or renovated skyscrapers, hotels, casinos, and golf courses. **Trump** later started various side ventures, including licensing his name for real estate and consumer products. He managed the company until his 2017 inauguration. He co-authored several books, including *The Art of the Deal*. He owned the Miss Universe and Miss USA beauty pageants from 1996 to 2015, and he produced and hosted the reality television show *The Apprentice* from 2003 to 2015. *Forbes* estimates his net worth to be \$3.1 billion.

**Trump** entered the 2016 presidential race as a Republican and defeated sixteen opponents in the primaries. Commentators described his political positions as populist, protectionist, and nationalist. His campaign received extensive free media coverage; many of his public statements were controversial or false. **Trump** was elected president in a surprise victory over Democratic nominee Hillary Clinton. He became the oldest and wealthiest person ever to assume the presidency, the first without prior military or government service, and the fifth to have won the election while losing the popular vote. His election and policies have sparked numerous protests. Many of his comments and actions have been perceived as racially charged or racist.

During his presidency, **Trump** ordered a travel ban on citizens from several Muslim-majority countries, citing security concerns; after legal challenges, the Supreme Court upheld the policy's third revision. He signed tax cut legislation which also rescinded the individual insurance mandate provision of the Affordable Care Act and opened the Arctic Refuge for oil drilling. He enacted a partial repeal of the Dodd-Frank Act that had imposed stricter constraints on banks in the aftermath of the 2008 financial crisis. He pursued his America First agenda in foreign policy, withdrawing the U.S. from the Trans-Pacific Partnership trade negotiations, the Paris Agreement on climate change, and the Iran nuclear deal. He recognized Jerusalem as the capital of Israel, and imposed import tariffs on various goods, triggering a trade war with China.

After **Trump** dismissed FBI Director James Comey, the Justice Department appointed Robert Mueller as Special Counsel to investigate "any links and/or coordination" between the **Trump** campaign and the Russian government in its election interference. **Trump** has repeatedly denied accusations of collusion and obstruction of justice, calling the investigation a politically motivated "witch hunt".

**Figura 3:** “Trump” aparece 10 vezes em test\_trump\_wikipedia.txt

```

{ "key": "reference", "value": "1" }
{"Key": "family", "Value": "1"}
{"Key": "agreement", "Value": "1"}
{"Key": "jerusalem", "Value": "1"}
{"Key": "trump", "Value": "10"}
{"Key": "courses", "Value": "1"}
{"Key": "later", "Value": "1"}
{"Key": "casinos", "Value": "1"}
{"Key": "after", "Value": "2"}

```

**Figura 4:** “trump” é contado 10 vezes pelo wordcount. (Obs: o algoritmo converte as palavras contadas para lowercase com o comando **text = strings.ToLower(text)**, por isso “trump” aparece com inicial minúscula no *map* de contagem).