6ª ATIVIDADE de CES-27 / 2018
CTA - ITA - IEC
Prof Juliana e Prof Vitor

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.
```

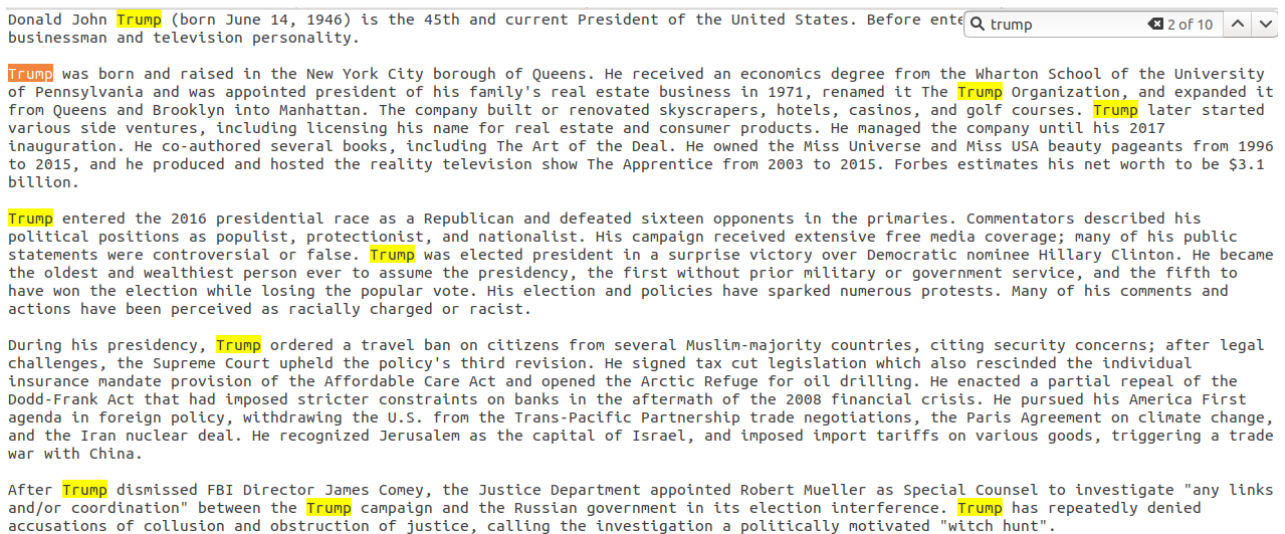**Figura 1:** "**there**" aparece 4 vezes em test.txt

{"Key":"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:

Figura 3: "**Trump**" aparece 10 vezes em test_trump_wikipedia.txt

{"Key":"interference","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).