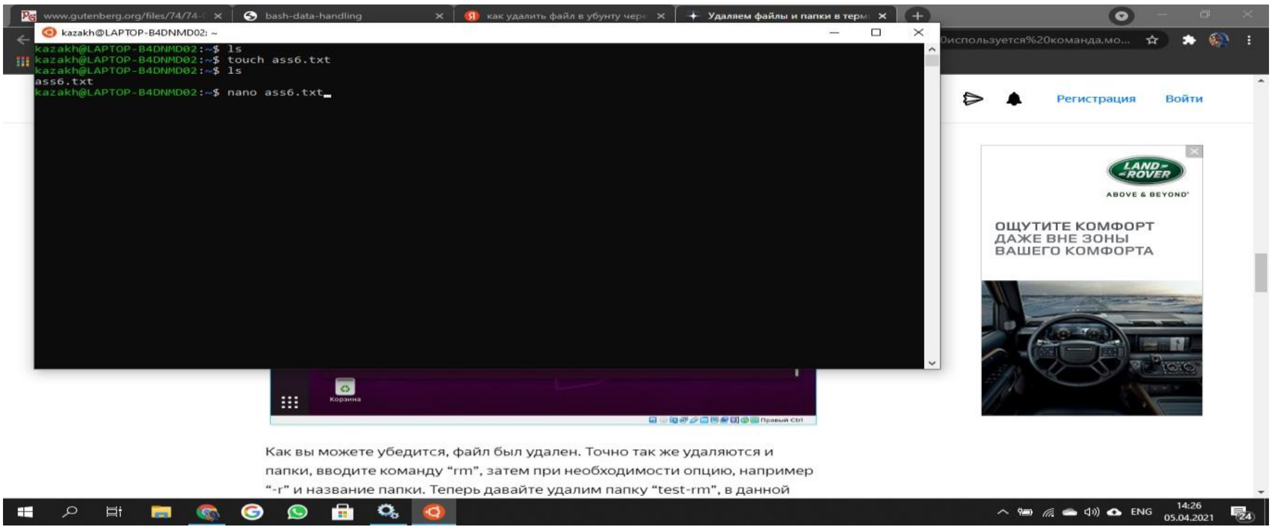
Assignment 6

Exercise I

Download the book „The Adventures of Tom Sawyer“ from http://www.gutenberg.org/ebooks/74 (utf-8 format) and store it locally.

SOLVE

Firstly, I checked that there is no files in ls. Then I opened a new txt file with touch. I copied the given text into this file.

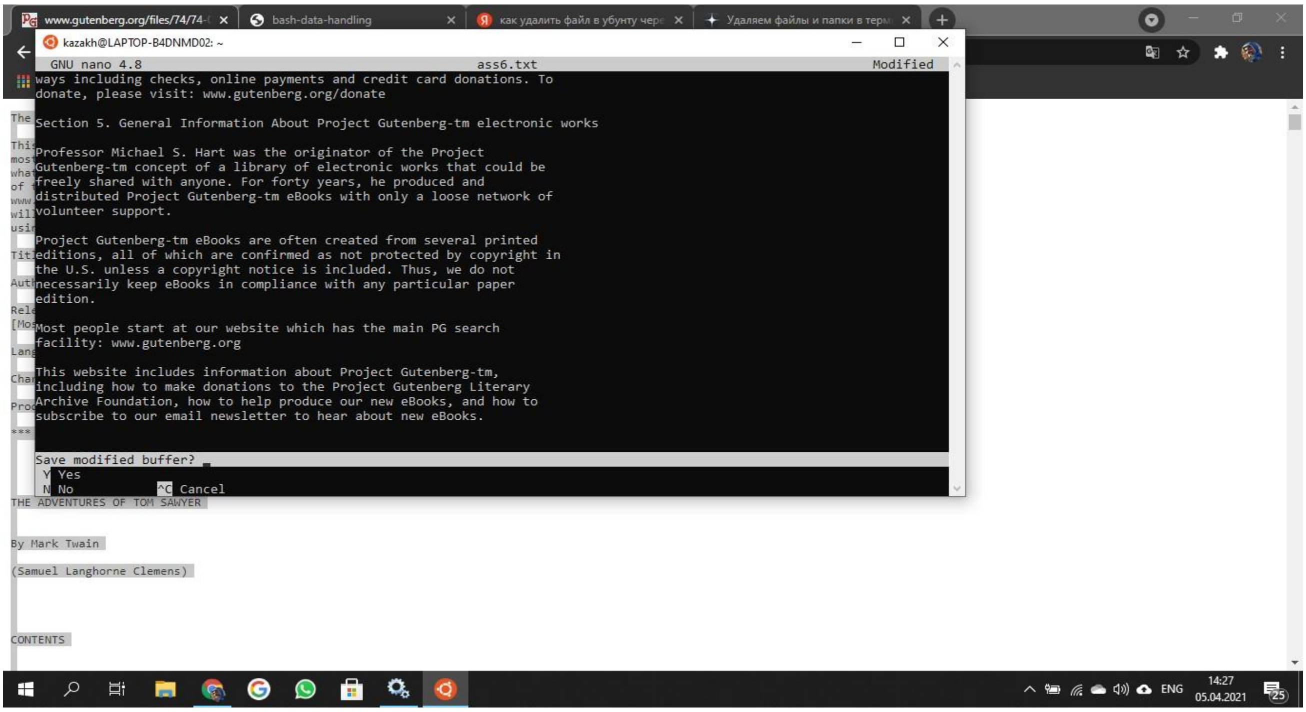


```
kazakh@LAPTOP-B4DNMD02:~$ ls
kazakh@LAPTOP-B4DNMD02:~$ touch ass6.txt
kazakh@LAPTOP-B4DNMD02:~$ ls
ass6.txt
kazakh@LAPTOP-B4DNMD02:~$ nano ass6.txt
```

Как вы можете убедится, файл был удален. Точно так же удаляются и папки, вводите команду “rm”, затем при необходимости опцию, например “-r” и название папки. Теперь давайте удалим папку “test-rm”, в данной

www.gutenberg.org/files/74/74-
bash-data-handling
как удалить файл в убунту чере
Удаляем файлы и папки в терм
kazakh@LAPTOP-B4DNMD02: ~
Регистрация
Войти
LAND ROVER
ABOVE & BEYOND
ОЩУТИТЕ КОМФОРТ ДАЖЕ ВНЕ ЗОНЫ ВАШЕГО КОМФОРТА
ENG
14:26
05.04.2021

And I saved the given text to the file.



```
  GNU nano 4.8                         ass6.txt                          Modified
ways including checks, online payments and credit card donations. To
donate, please visit: www.gutenberg.org/donate

Section 5. General Information About Project Gutenberg-tm electronic works

Professor Michael S. Hart was the originator of the Project
Gutenberg-tm concept of a library of electronic works that could be
freely shared with anyone. For forty years, he produced and
distributed Project Gutenberg-tm eBooks with only a loose network of
volunteer support.

Project Gutenberg-tm eBooks are often created from several printed
editions, all of which are confirmed as not protected by copyright in
the U.S. unless a copyright notice is included. Thus, we do not
necessarily keep eBooks in compliance with any particular paper
edition.

Most people start at our website which has the main PG search
facility: www.gutenberg.org

This website includes information about Project Gutenberg-tm,
including how to make donations to the Project Gutenberg Literary
Archive Foundation, how to help produce our new eBooks, and how to
subscribe to our email newsletter to hear about new eBooks.


Save modified buffer?
 Y Yes
 N No           ^C Cancel
```

```
THE ADVENTURES OF TOM SAWYER

By Mark Twain

(Samuel Langhorne Clemens)

CONTENTS
```

www.gutenberg.org/files/74/74-
bash-data-handling
как удалить файл в убунту чере
Удаляем файлы и папки в терм
kazakh@LAPTOP-B4DNMD02: ~
ENG
14:27
05.04.2021

2. For cygwin users only: Convert file to Unix Format with command: dos2unix.exe

SOLVE: I'm done

3. Browse through the first pages of the book using less (for help type man less)

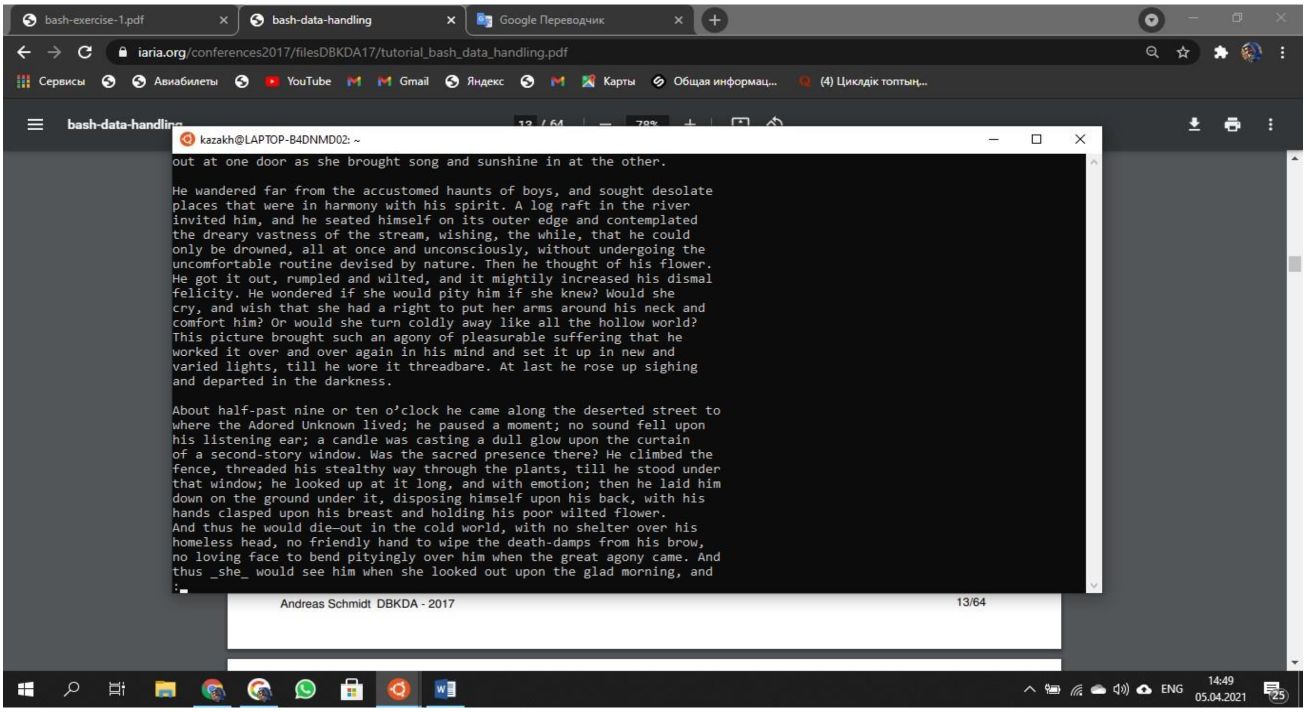
a) Go to line 1234 of the file. What ist the third word?

--

SOLVE

Firstly, I entered command **less ass6.txt,** then entered **1234g** and went to a given line.

I can read the third word without any code.) Third word is one.

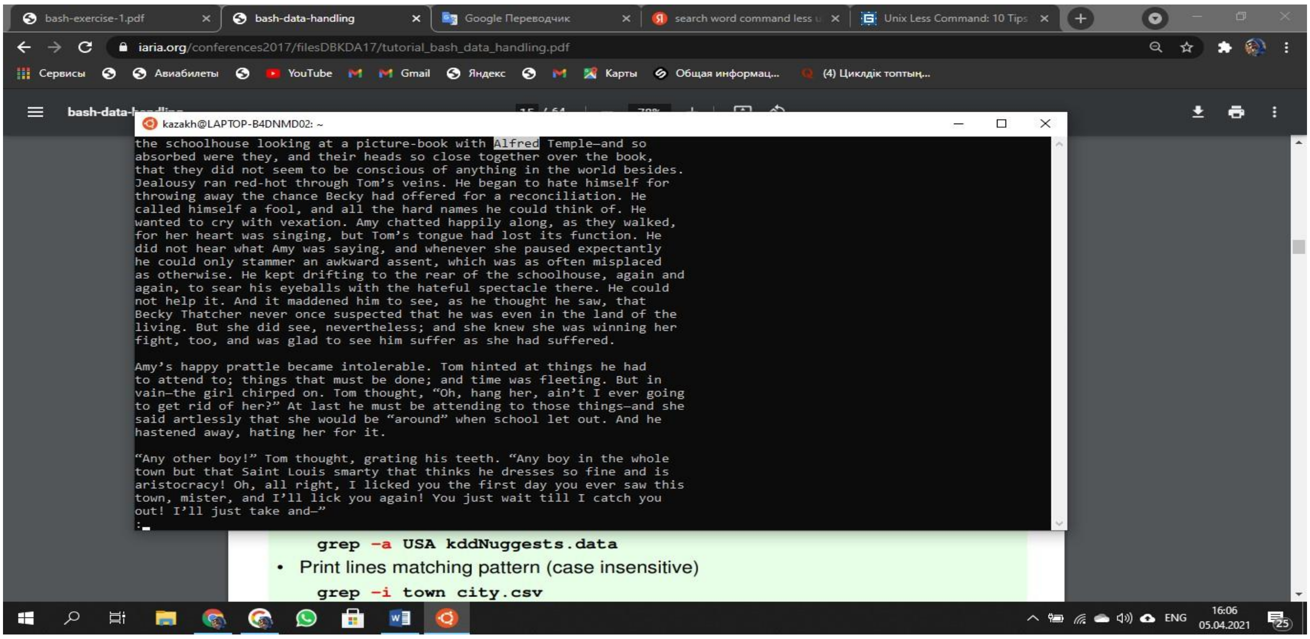


```
out at one door as she brought song and sunshine in at the other.

He wandered far from the accustomed haunts of boys, and sought desolate
places that were in harmony with his spirit. A log raft in the river
invited him, and he seated himself on its outer edge and contemplated
the dreary vastness of the stream, wishing, the while, that he could
only be drowned, all at once and unconsciously, without undergoing the
uncomfortable routine devised by nature. Then he thought of his flower.
He got it out, rumpled and wilted, and it mightily increased his dismal
felicity. He wondered if she would pity him if she knew? Would she
cry, and wish that she had a right to put her arms around his neck and
comfort him? Or would she turn coldly away like all the hollow world?
This picture brought such an agony of pleasurable suffering that he
worked it over and over again in his mind and set it up in new and
varied lights, till he wore it threadbare. At last he rose up sighing
and departed in the darkness.

About half-past nine or ten o'clock he came along the deserted street to
where the Adored Unknown lived; he paused a moment; no sound fell upon
his listening ear; a candle was casting a dull glow upon the curtain
of a second-story window. Was the sacred presence there? He climbed the
fence, threaded his stealthy way through the plants, till he stood under
that window; he looked up at it long, and with emotion; then he laid him
down on the ground under it, disposing himself upon his back, with his
hands clasped upon his breast and holding his poor wilted flower.
And thus he would die—out in the cold world, with no shelter over his
homeless head, no friendly hand to wipe the death-damps from his brow,
no loving face to bend pityingly over him when the great agony came. And
thus _she_ would see him when she looked out upon the glad morning, and
:
```

b) Search for the first apperance of Alfred (what is his last name?)

SOLVE

I used the command “less” and entered “/Alfred”. And found first line and his last name. His last name is Temple.



```
the schoolhouse looking at a picture-book with Alfred Temple—and so
absorbed were they, and their heads so close together over the book,
that they did not seem to be conscious of anything in the world besides.
Jealousy ran red-hot through Tom’s veins. He began to hate himself for
throwing away the chance Becky had offered for a reconciliation. He
called himself a fool, and all the hard names he could think of. He
wanted to cry with vexation. Amy chatted happily along, as they walked,
for her heart was singing, but Tom’s tongue had lost its function. He
did not hear what Amy was saying, and whenever she paused expectantly
he could only stammer an awkward assent, which was as often misplaced
as otherwise. He kept drifting to the rear of the schoolhouse, again and
again, to sear his eyeballs with the hateful spectacle there. He could
not help it. And it maddened him to see, as he thought he saw, that
Becky Thatcher never once suspected that he was even in the land of the
living. But she did see, nevertheless; and she knew she was winning her
fight, too, and was glad to see him suffer as she had suffered.

Amy’s happy prattle became intolerable. Tom hinted at things he had
to attend to; things that must be done; and time was fleeting. But in
vain—the girl chirped on. Tom thought, “Oh, hang her, ain’t I ever going
to get rid of her?” At last he must be attending to those things—and she
said artlessly that she would be “around” when school let out. And he
hastened away, hating her for it.

“Any other boy!” Tom thought, grating his teeth. “Any boy in the whole
town but that Saint Louis smarty that thinks he dresses so fine and is
aristocracy! Oh, all right, I licked you the first day you ever saw this
town, mister, and I’ll lick you again! You just wait till I catch you
out! I’ll just take and—”
:
```

```
grep -a USA kddNuggests.data
```

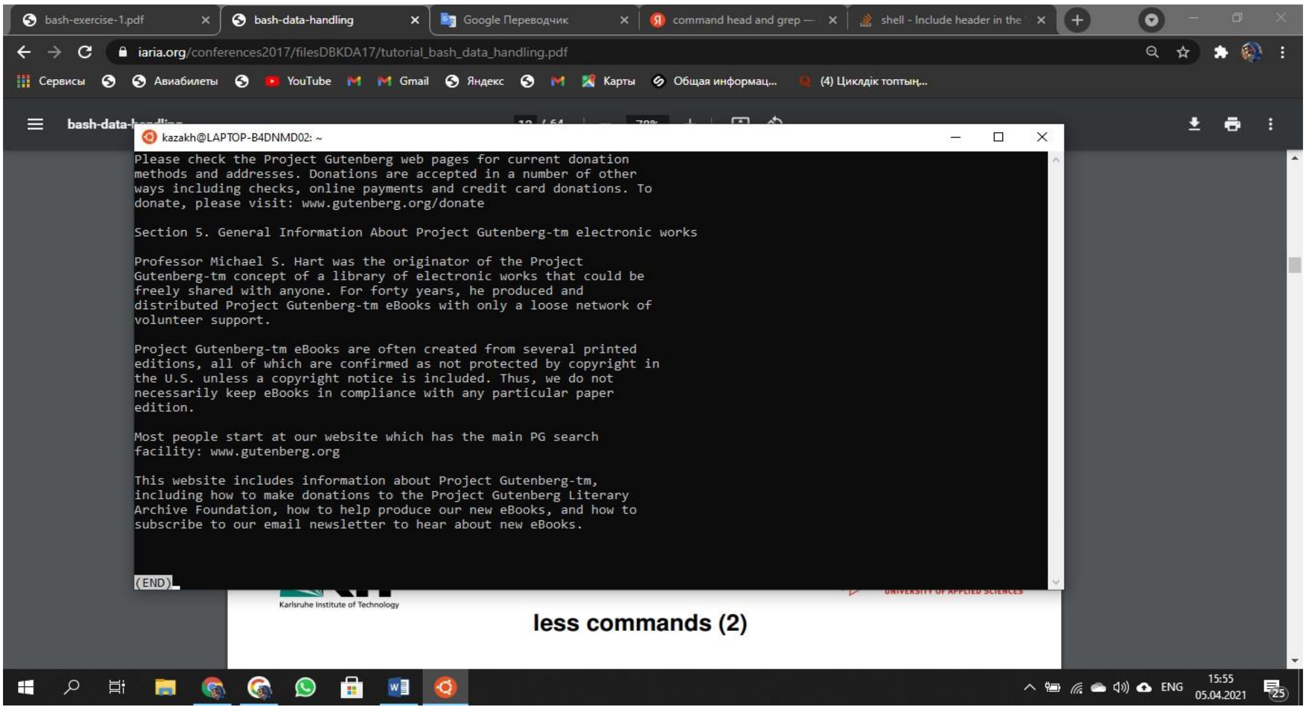
- Print lines matching pattern (case insensitive)

```
grep -i town city.csv
```

c) Go to the end of the document.

SOLVE

I used the “less” command again and entered “>”. This command takes us to the end of the text.



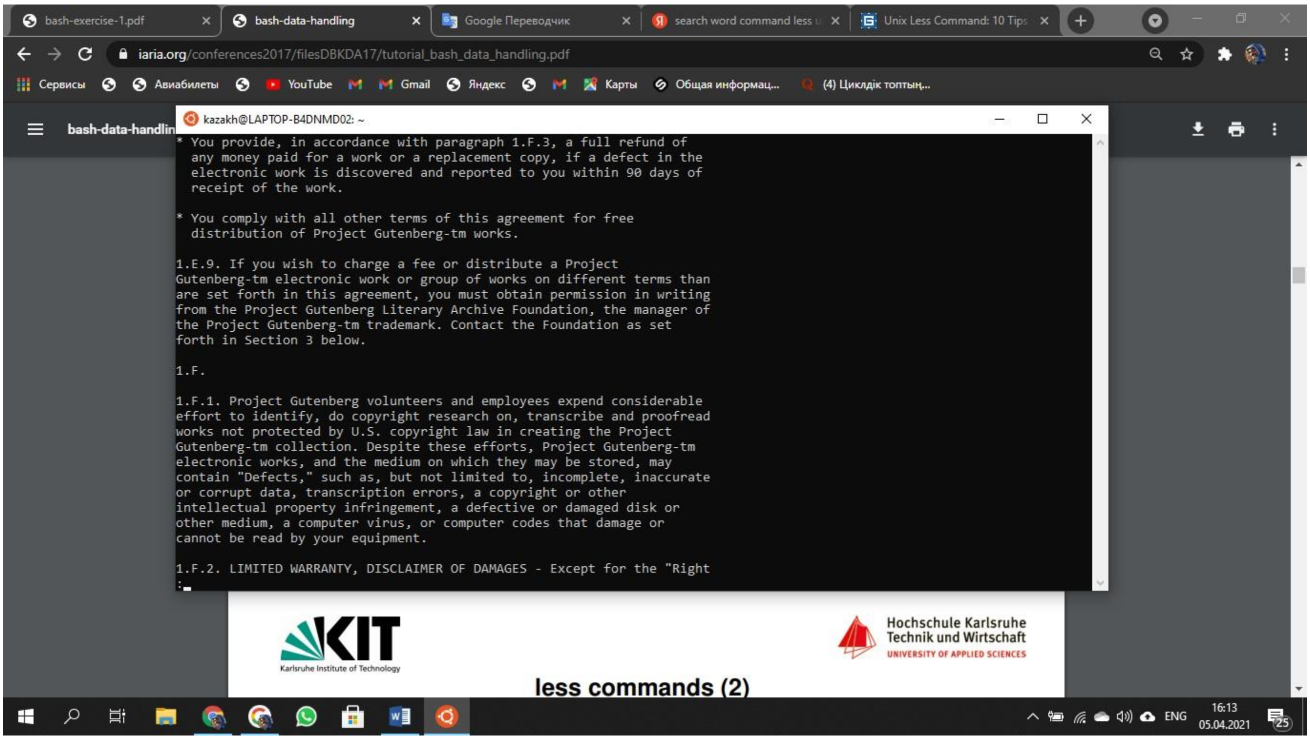
```
Please check the Project Gutenberg web pages for current donation
methods and addresses. Donations are accepted in a number of other
ways including checks, online payments and credit card donations. To
donate, please visit: www.gutenberg.org/donate

Section 5. General Information About Project Gutenberg-tm electronic works

Professor Michael S. Hart was the originator of the Project
Gutenberg-tm concept of a library of electronic works that could be
freely shared with anyone. For forty years, he produced and
distributed Project Gutenberg-tm eBooks with only a loose network of
volunteer support.

Project Gutenberg-tm eBooks are often created from several printed
editions, all of which are confirmed as not protected by copyright in
the U.S. unless a copyright notice is included. Thus, we do not
necessarily keep eBooks in compliance with any particular paper
edition.

Most people start at our website which has the main PG search
facility: www.gutenberg.org

This website includes information about Project Gutenberg-tm,
including how to make donations to the Project Gutenberg Literary
Archive Foundation, how to help produce our new eBooks, and how to
subscribe to our email newsletter to hear about new eBooks.



(END)
```

less commands (2)

d) Scroll backward 5 pages.

SOLVE

Going to the end of the text, I gave the "b" command 5 times.



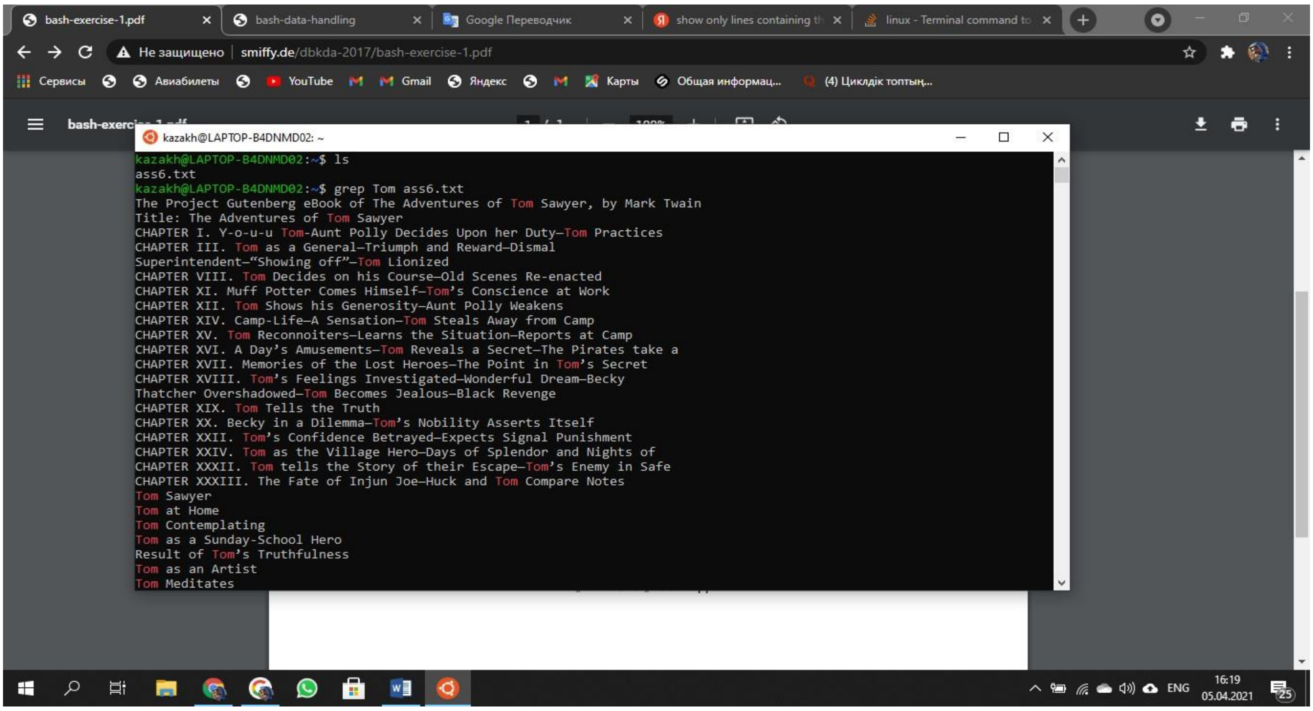
bash-exercise-1.pdf
bash-data-handling
Google Переводчик
search word command less
Unix Less Command: 10 Tips
iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
kazakh@LAPTOP-B4DNMD02: ~
KIT
Karlsruhe Institute of Technology
Hochschule Karlsruhe Technik und Wirtschaft
UNIVERSITY OF APPLIED SCIENCES
less commands (2)
ENG
16:13
05.04.2021

```
* You provide, in accordance with paragraph 1.F.3, a full refund of
  any money paid for a work or a replacement copy, if a defect in the
  electronic work is discovered and reported to you within 90 days of
  receipt of the work.

* You comply with all other terms of this agreement for free
  distribution of Project Gutenberg-tm works.

1.E.9. If you wish to charge a fee or distribute a Project
Gutenberg-tm electronic work or group of works on different terms than
are set forth in this agreement, you must obtain permission in writing
from the Project Gutenberg Literary Archive Foundation, the manager of
the Project Gutenberg-tm trademark. Contact the Foundation as set
forth in Section 3 below.

1.F.

1.F.1. Project Gutenberg volunteers and employees expend considerable
effort to identify, do copyright research on, transcribe and proofread
works not protected by U.S. copyright law in creating the Project
Gutenberg-tm collection. Despite these efforts, Project Gutenberg-tm
electronic works, and the medium on which they may be stored, may
contain "Defects," such as, but not limited to, incomplete, inaccurate
or corrupt data, transcription errors, a copyright or other
intellectual property infringement, a defective or damaged disk or
other medium, a computer virus, or computer codes that damage or
cannot be read by your equipment.

1.F.2. LIMITED WARRANTY, DISCLAIMER OF DAMAGES - Except for the "Right
:
```

e) Show only lines containing the word „Tom“.

SOLVE

I used “grep” for search all lines. I entered “grep Tom ass6.txt”.



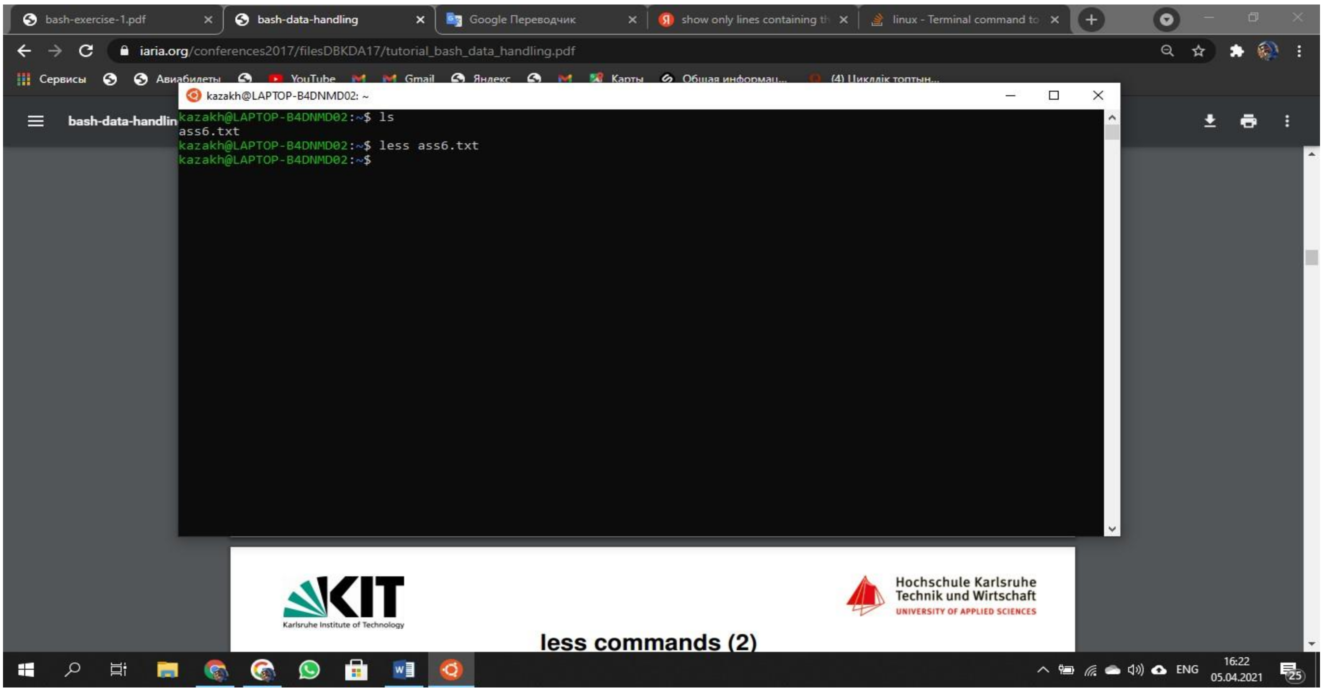
bash-exercise-1.pdf
bash-data-handling
Google Переводчик
show only lines containing th
linux - Terminal command to
Не защищено | smiffy.de/dbkda-2017/bash-exercise-1.pdf
kazakh@LAPTOP-B4DNMD02: ~
ENG
16:19
05.04.2021

```
kazakh@LAPTOP-B4DNMD02:~$ ls
ass6.txt
kazakh@LAPTOP-B4DNMD02:~$ grep Tom ass6.txt
The Project Gutenberg eBook of The Adventures of Tom Sawyer, by Mark Twain
Title: The Adventures of Tom Sawyer
CHAPTER I. Y-o-u-u Tom-Aunt Polly Decides Upon her Duty—Tom Practices
CHAPTER III. Tom as a General—Triumph and Reward—Dismal
Superintendent—“Showing off”—Tom Lionized
CHAPTER VIII. Tom Decides on his Course—Old Scenes Re-enacted
CHAPTER XI. Muff Potter Comes Himself—Tom’s Conscience at Work
CHAPTER XII. Tom Shows his Generosity—Aunt Polly Weakens
CHAPTER XIV. Camp-Life—A Sensation—Tom Steals Away from Camp
CHAPTER XV. Tom Reconnoiters—Learns the Situation—Reports at Camp
CHAPTER XVI. A Day’s Amusements—Tom Reveals a Secret—The Pirates take a
CHAPTER XVII. Memories of the Lost Heroes—The Point in Tom’s Secret
CHAPTER XVIII. Tom’s Feelings Investigated—Wonderful Dream—Becky
Thatcher Overshadowed—Tom Becomes Jealous—Black Revenge
CHAPTER XIX. Tom Tells the Truth
CHAPTER XX. Becky in a Dilemma—Tom’s Nobility Asserts Itself
CHAPTER XXII. Tom’s Confidence Betrayed—Expects Signal Punishment
CHAPTER XXIV. Tom as the Village Hero—Days of Splendor and Nights of
CHAPTER XXXII. Tom tells the Story of their Escape—Tom’s Enemy in Safe
CHAPTER XXXIII. The Fate of Injun Joe—Huck and Tom Compare Notes
Tom Sawyer
Tom at Home
Tom Contemplating
Tom as a Sunday-School Hero
Result of Tom’s Truthfulness
Tom as an Artist
Tom Meditates
```

f) Quit less

SOLVE

It’s easy. I entered only “q” and I got out of there.

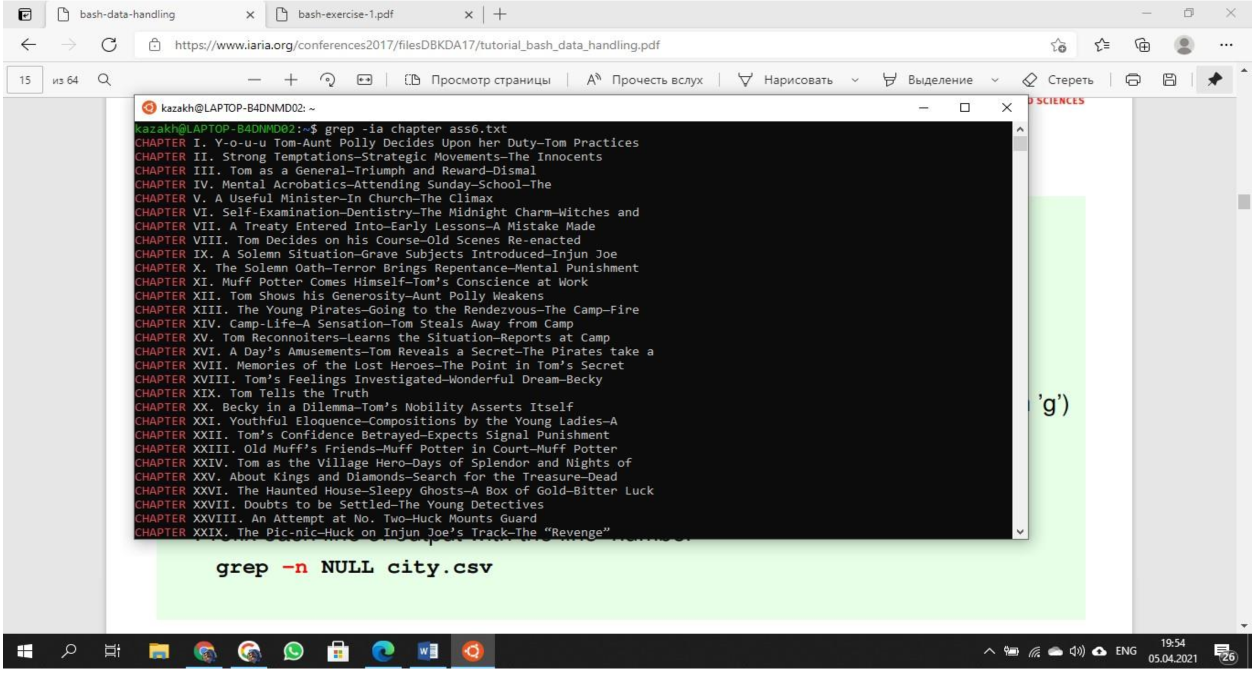


```
kazakh@LAPTOP-B4DNMD02:~$ ls
ass6.txt
kazakh@LAPTOP-B4DNMD02:~$ less ass6.txt
kazakh@LAPTOP-B4DNMD02:~$
```

4. How many chapters has the book (hint use grep with –a option)?

SOLVE

I use “grep” command and entered “grep –ia ass6.txt”. ~i~ is ignore case and a – matching lines a binary file. Here are all the chapters and their number is 35



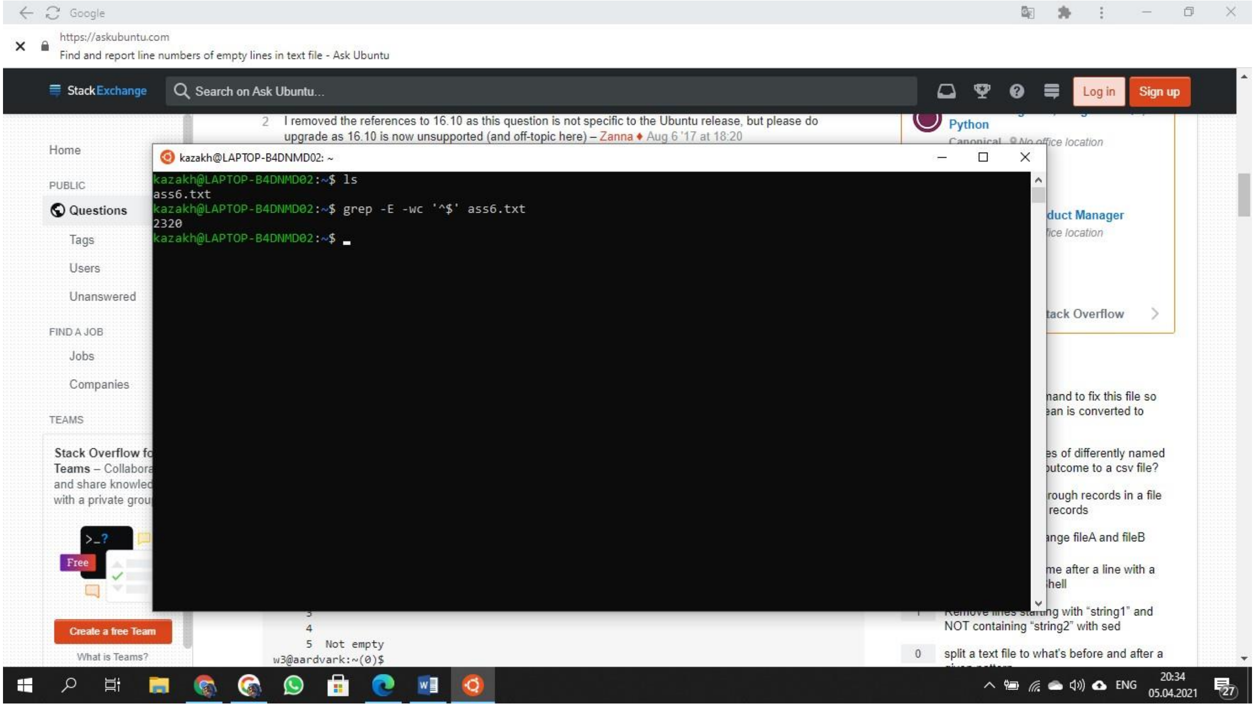
```
kazakh@LAPTOP-B4DNMD02:~$ grep -ia chapter ass6.txt
CHAPTER I. Y-o-u-u Tom-Aunt Polly Decides Upon her Duty–Tom Practices
CHAPTER II. Strong Temptations–Strategic Movements–The Innocents
CHAPTER III. Tom as a General–Triumph and Reward–Dismal
CHAPTER IV. Mental Acrobatics–Attending Sunday–School–The
CHAPTER V. A Useful Minister–In Church–The Climax
CHAPTER VI. Self-Examination–Dentistry–The Midnight Charm–Witches and
CHAPTER VII. A Treaty Entered Into–Early Lessons–A Mistake Made
CHAPTER VIII. Tom Decides on his Course–Old Scenes Re-enacted
CHAPTER IX. A Solemn Situation–Grave Subjects Introduced–Injun Joe
CHAPTER X. The Solemn Oath–Terror Brings Repentance–Mental Punishment
CHAPTER XI. Muff Potter Comes Himself–Tom’s Conscience at Work
CHAPTER XII. Tom Shows his Generosity–Aunt Polly Weakens
CHAPTER XIII. The Young Pirates–Going to the Rendezvous–The Camp–Fire
CHAPTER XIV. Camp-Life–A Sensation–Tom Steals Away from Camp
CHAPTER XV. Tom Reconnoiters–Learns the Situation–Reports at Camp
CHAPTER XVI. A Day’s Amusements–Tom Reveals a Secret–The Pirates take a
CHAPTER XVII. Memories of the Lost Heroes–The Point in Tom’s Secret
CHAPTER XVIII. Tom’s Feelings Investigated–Wonderful Dream–Becky
CHAPTER XIX. Tom Tells the Truth
CHAPTER XX. Becky in a Dilemma–Tom’s Nobility Asserts Itself
CHAPTER XXI. Youthful Eloquence–Compositions by the Young Ladies–A
CHAPTER XXII. Tom’s Confidence Betrayed–Expects Signal Punishment
CHAPTER XXIII. Old Muff’s Friends–Muff Potter in Court–Muff Potter
CHAPTER XXIV. Tom as the Village Hero–Days of Splendor and Nights of
CHAPTER XXV. About Kings and Diamonds–Search for the Treasure–Dead
CHAPTER XXVI. The Haunted House–Sleepy Ghosts–A Box of Gold–Bitter Luck
CHAPTER XXVII. Doubts to be Settled–The Young Detectives
CHAPTER XXVIII. An Attempt at No. Two–Huck Mounts Guard
CHAPTER XXIX. The Pic-nic–Huck on Injun Joe’s Track–The “Revenge”
```

bash-data-handling
bash-exercise-1.pdf
https://www.iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
kazakh@LAPTOP-B4DNMD02: ~
grep -n NULL city.csv

5. Count the number of empty lines in the file. --

SOLVE

I use grep and entered “grep -E -wc '^$' ass6.txt”. Here -E receives only the marked text. And -wc gives their number. So, empty lines are 2320.



I removed the references to 16.10 as this question is not specific to the Ubuntu release, but please do upgrade as 16.10 is now unsupported (and off-topic here) – Zanna ♦ Aug 6 '17 at 18:20

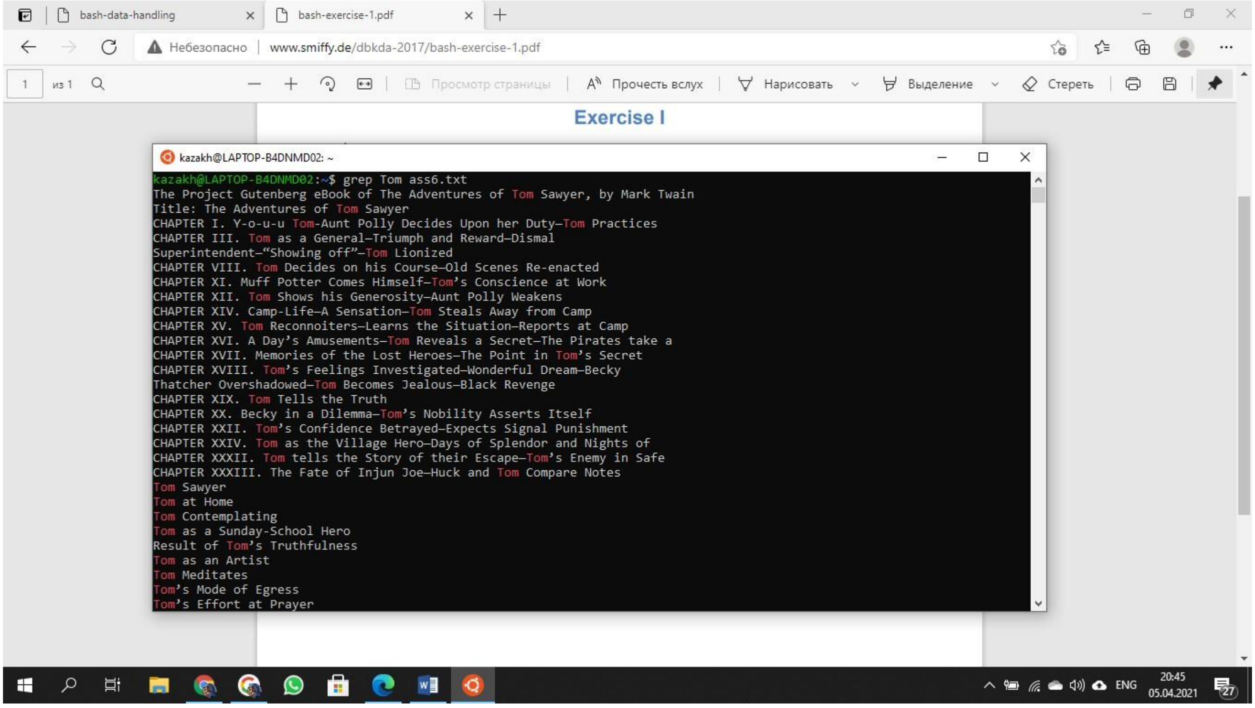
```
kazakh@LAPTOP-B4DNMD02:~$ ls
ass6.txt
kazakh@LAPTOP-B4DNMD02:~$ grep -E -wc '^$' ass6.txt
2320
kazakh@LAPTOP-B4DNMD02:~$ _
```

6. Execute grep Tom and grep –o Tom . What is the difference?

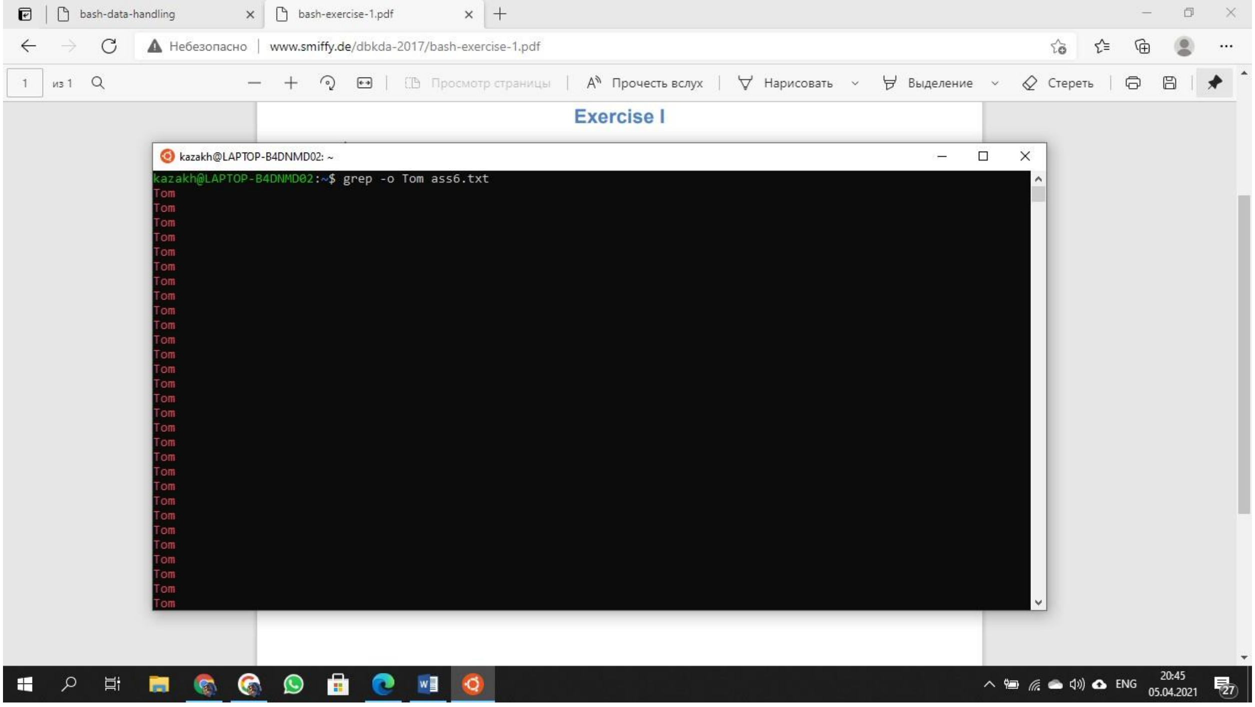
SOLVE

I search for sentences with the word "Tom" in the "grep Tom file" command.

In the "grep -o Tom file" command, words other than "Tom" were cut out.



```
kazakh@LAPTOP-B4DNMD02:~$ grep Tom ass6.txt
The Project Gutenberg eBook of The Adventures of Tom Sawyer, by Mark Twain
Title: The Adventures of Tom Sawyer
CHAPTER I. Y-o-u-u Tom-Aunt Polly Decides Upon her Duty—Tom Practices
CHAPTER III. Tom as a General—Triumph and Reward—Dismal
Superintendent—“Showing off”—Tom Lionized
CHAPTER VIII. Tom Decides on his Course—Old Scenes Re-enacted
CHAPTER XI. Muff Potter Comes Himself—Tom’s Conscience at Work
CHAPTER XII. Tom Shows his Generosity—Aunt Polly Weakens
CHAPTER XIV. Camp-Life—A Sensation—Tom Steals Away from Camp
CHAPTER XV. Tom Reconnoiters—Learns the Situation—Reports at Camp
CHAPTER XVI. A Day’s Amusements—Tom Reveals a Secret—The Pirates take a
CHAPTER XVII. Memories of the Lost Heroes—The Point in Tom’s Secret
CHAPTER XVIII. Tom’s Feelings Investigated—Wonderful Dream—Becky
Thatcher Overshadowed—Tom Becomes Jealous—Black Revenge
CHAPTER XIX. Tom Tells the Truth
CHAPTER XX. Becky in a Dilemma—Tom’s Nobility Asserts Itself
CHAPTER XXII. Tom’s Confidence Betrayed—Expects Signal Punishment
CHAPTER XXIV. Tom as the Village Hero—Days of Splendor and Nights of
CHAPTER XXXII. Tom tells the Story of their Escape—Tom’s Enemy in Safe
CHAPTER XXXIII. The Fate of Injun Joe—Huck and Tom Compare Notes
Tom Sawyer
Tom at Home
Tom Contemplating
Tom as a Sunday-School Hero
Result of Tom’s Truthfulness
Tom as an Artist
Tom Meditates
Tom’s Mode of Egress
Tom’s Effort at Prayer
```



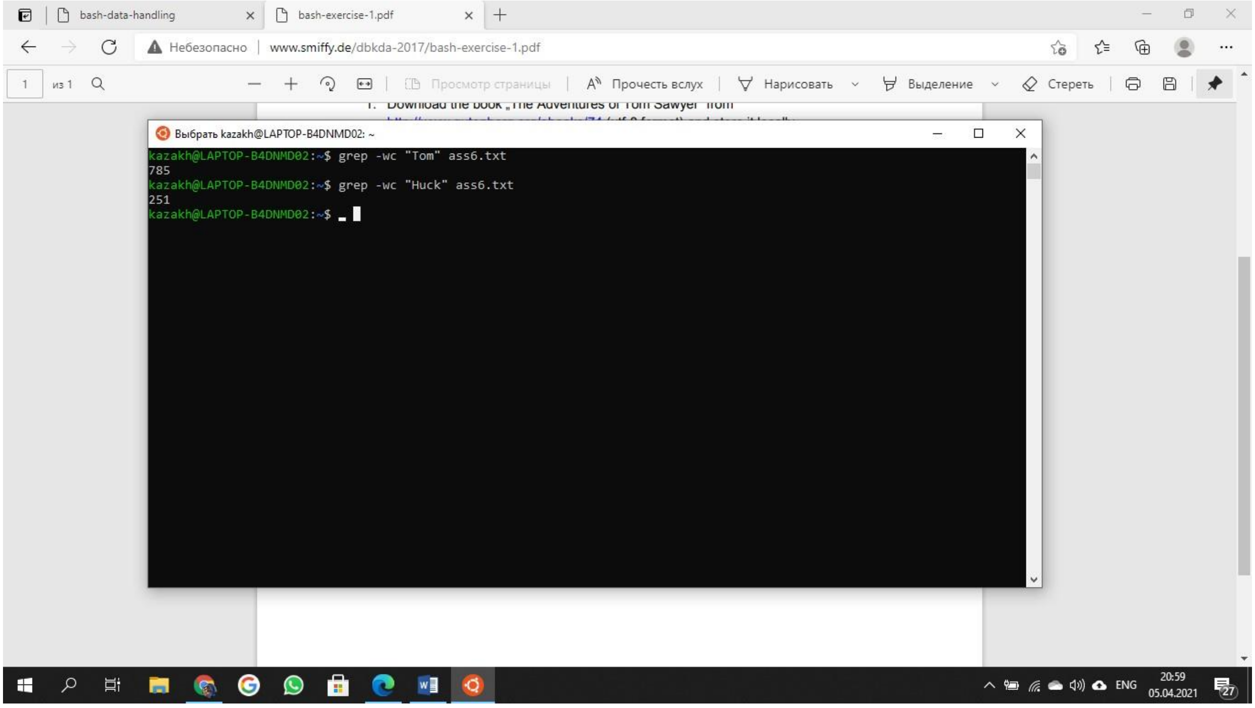
```
kazakh@LAPTOP-B4DNMD02:~$ grep -o Tom ass6.txt
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
Tom
```

7. How often does the names „Tom“ and „Huck“ appears in the book?

SOLVE

I use grep and entered “grep –wc name file”. –wc counts the given word.

So, Tom-785 and Huck-251.



1. Download the book „The Adventures of Tom Sawyer" from

```
kazakh@LAPTOP-B4DNMD02:~$ grep -wc "Tom" ass6.txt
785
kazakh@LAPTOP-B4DNMD02:~$ grep -wc "Huck" ass6.txt
251
kazakh@LAPTOP-B4DNMD02:~$ _
```

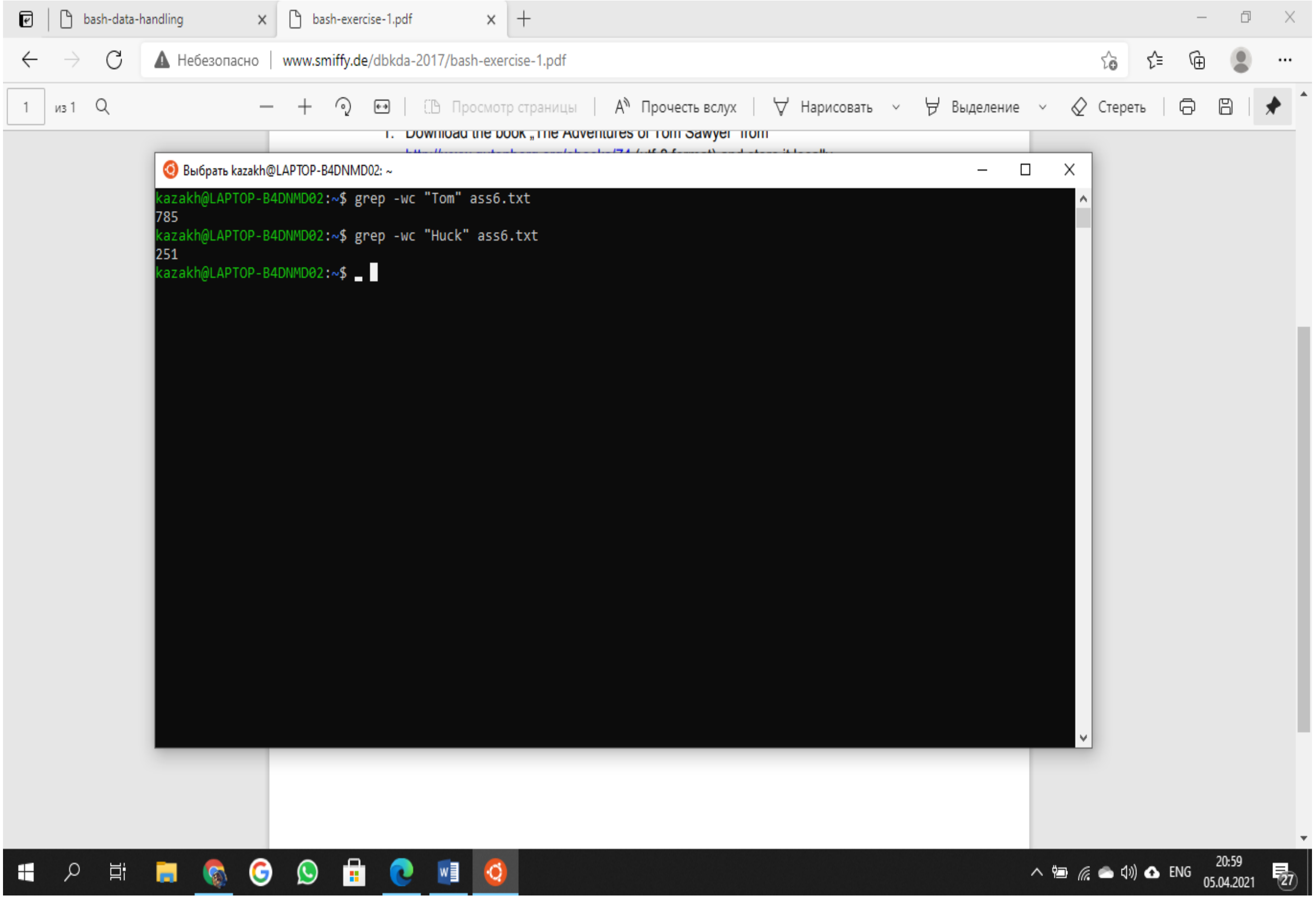
Assignment 6

Exercise 2

1. Download the book ’The Adventures of Tom Sawyer‘ from http://www.gutenberg.org/files/74/74-0.txt and store it locally in a file called ’The-Adventures-of-Tom-Sawyer.txt.’

SOLVE

I opened a new txt file with touch. I copied the given text into this file. And I named it “The-Adventures-of-Tom-Sawyer.txt”.


bash-data-handling
bash-exercise-1.pdf
Небезопасно
www.smiffy.de/dbkda-2017/bash-exercise-1.pdf
1
из 1
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
1. Download the book „The Adventures of Tom Sawyer“ from
Выбрать kazakh@LAPTOP-B4DNMD02: ~
ENG
20:59
05.04.2021
27

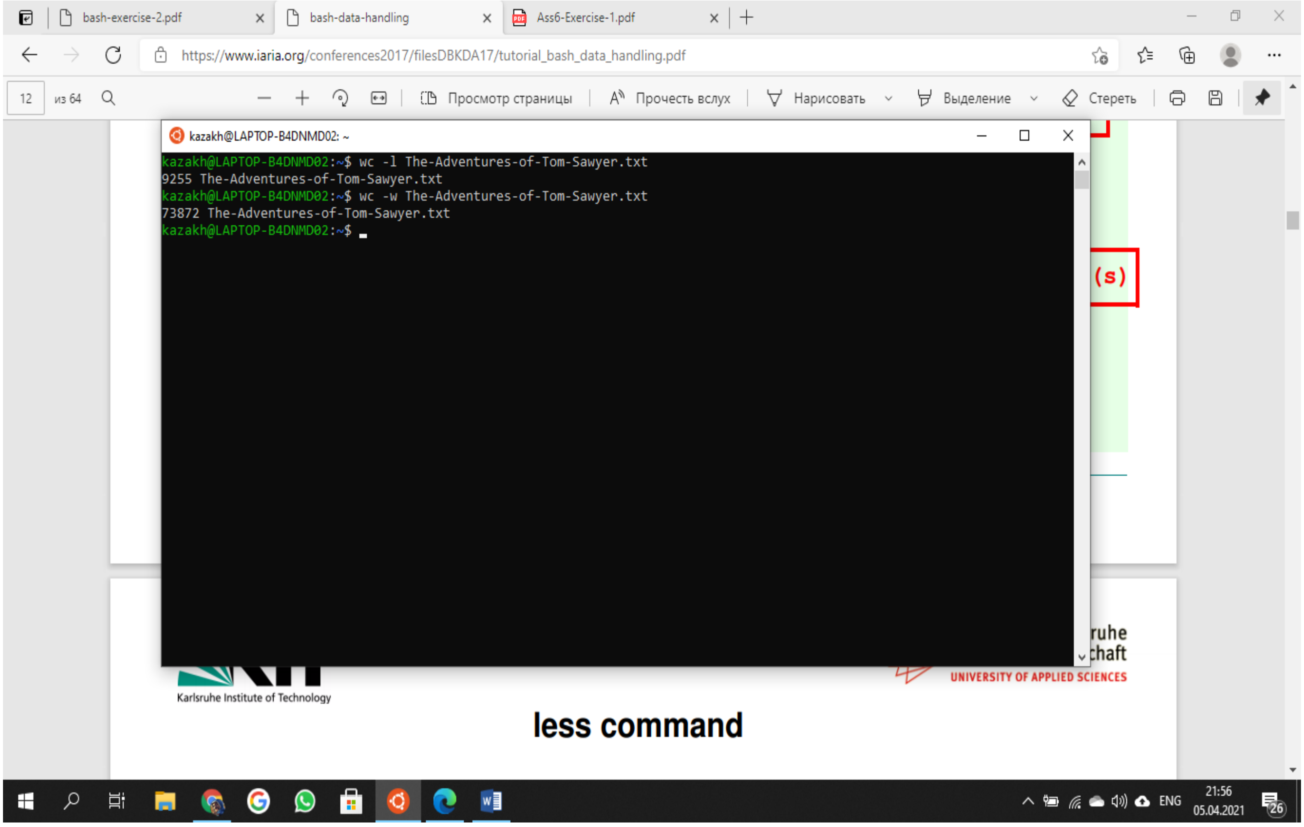
```
kazakh@LAPTOP-B4DNMD02:~$ grep -wc "Tom" ass6.txt
785
kazakh@LAPTOP-B4DNMD02:~$ grep -wc "Huck" ass6.txt
251
kazakh@LAPTOP-B4DNMD02:~$ _
```

2. **Count the words and lines in the book ’The-Adventures-of-Tom-Sawyer.txt’.**

SOLVE

I entered “wc –l file.txt” for lines. There are 9255 lines.

I entered “wc –w file.txt” for words. There are 73872 words.



```
kazakh@LAPTOP-B4DNMD02:~$ wc -l The-Adventures-of-Tom-Sawyer.txt
9255 The-Adventures-of-Tom-Sawyer.txt
kazakh@LAPTOP-B4DNMD02:~$ wc -w The-Adventures-of-Tom-Sawyer.txt
73872 The-Adventures-of-Tom-Sawyer.txt
kazakh@LAPTOP-B4DNMD02:~$ 
```

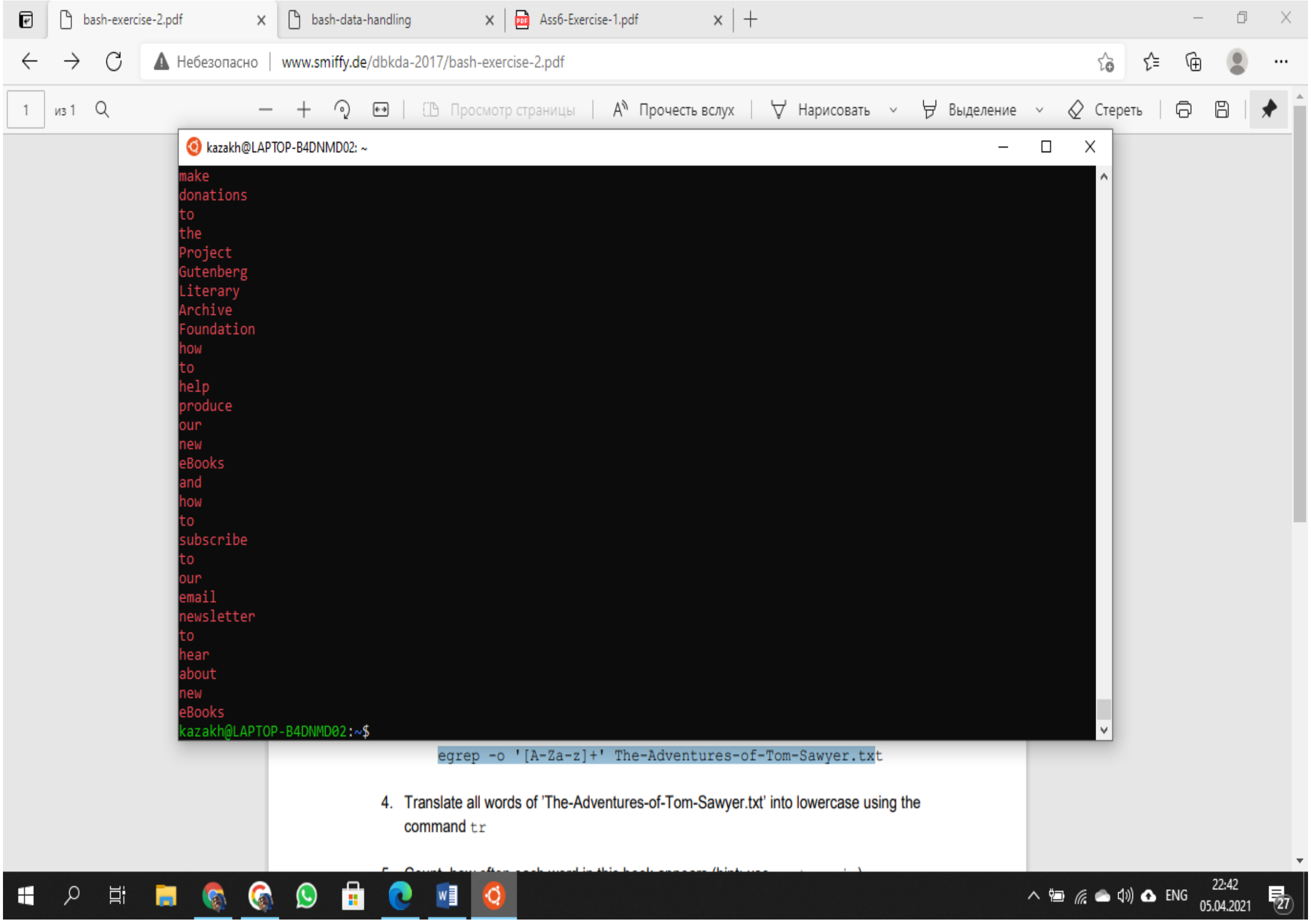
bash-exercise-2.pdf
bash-data-handling
Ass6-Exercise-1.pdf
https://www.iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
12
из 64
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
kazakh@LAPTOP-B4DNMD02: ~
(s)
Karlsruhe Institute of Technology
UNIVERSITY OF APPLIED SCIENCES
less command
ENG
21:56
05.04.2021

3. What does the following command perform?

egrep -o '[A-Za-z]+' The-Adventures-of-Tom-Sawyer.txt

SOLVE

This code divides words in the text with the symbols [A-Z, a-z] into each line. And -o removes the rest.



```
make
donations
to
the
Project
Gutenberg
Literary
Archive
Foundation
how
to
help
produce
our
new
eBooks
and
how
to
subscribe
to
our
email
newsletter
to
hear
about
new
eBooks
kazakh@LAPTOP-B4DNMD02:~$
```

```
egrep -o '[A-Za-z]+' The-Adventures-of-Tom-Sawyer.txt
```

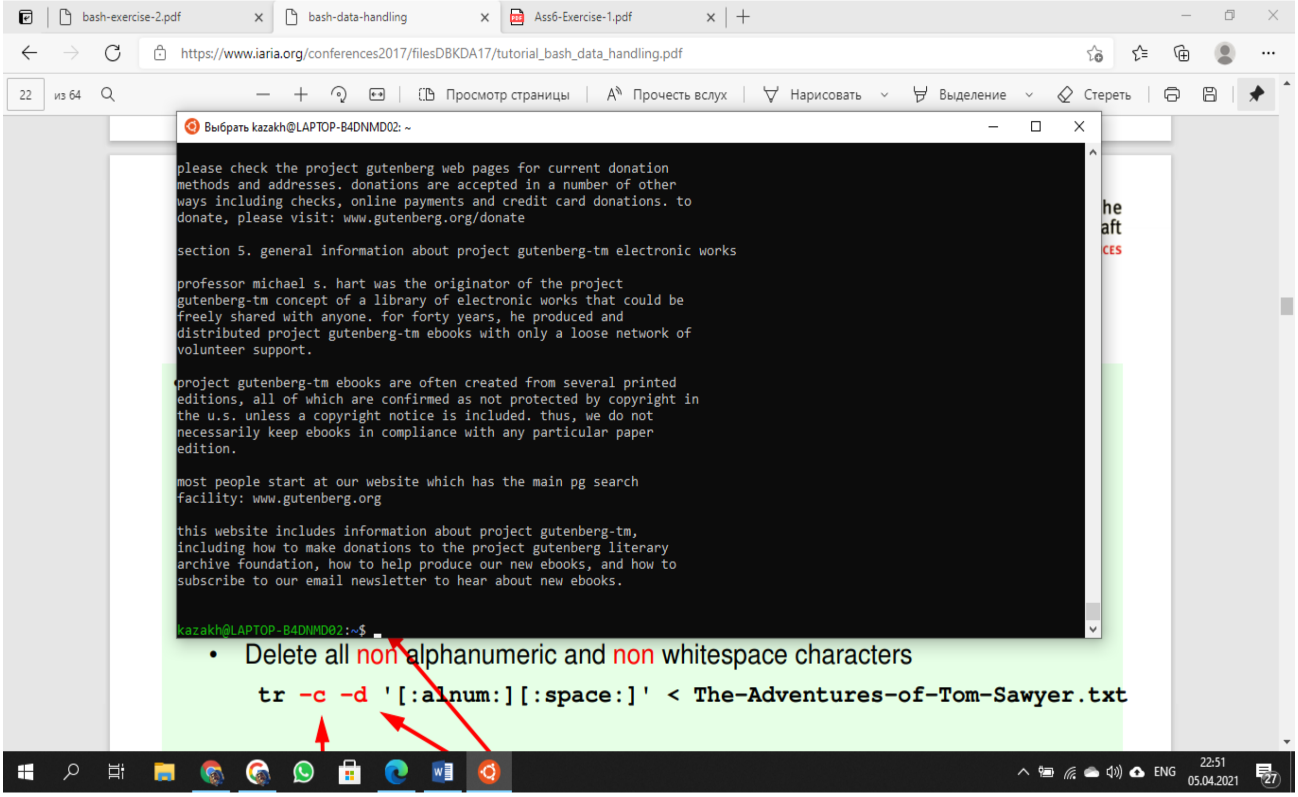
4. Translate all words of 'The-Adventures-of-Tom-Sawyer.txt' into lowercase using the command `tr`

4. Translate all words of ’The-Adventures-of-Tom-Sawyer.txt’ into lowercase using the command tr.

SOLVE

I used tr and I wrote all the words in lower case using this code.

tr 'A-Z' 'a-z' < The-Adventures-of-Tom-Sawyer.txt



```
please check the project gutenberg web pages for current donation
methods and addresses. donations are accepted in a number of other
ways including checks, online payments and credit card donations. to
donate, please visit: www.gutenberg.org/donate

section 5. general information about project gutenberg-tm electronic works

professor michael s. hart was the originator of the project
gutenberg-tm concept of a library of electronic works that could be
freely shared with anyone. for forty years, he produced and
distributed project gutenberg-tm ebooks with only a loose network of
volunteer support.

project gutenberg-tm ebooks are often created from several printed
editions, all of which are confirmed as not protected by copyright in
the u.s. unless a copyright notice is included. thus, we do not
necessarily keep ebooks in compliance with any particular paper
edition.

most people start at our website which has the main pg search
facility: www.gutenberg.org

this website includes information about project gutenberg-tm,
including how to make donations to the project gutenberg literary
archive foundation, how to help produce our new ebooks, and how to
subscribe to our email newsletter to hear about new ebooks.


kazakh@LAPTOP-B4DNMD02:~$
```

- Delete all non alphanumeric and non whitespace characters

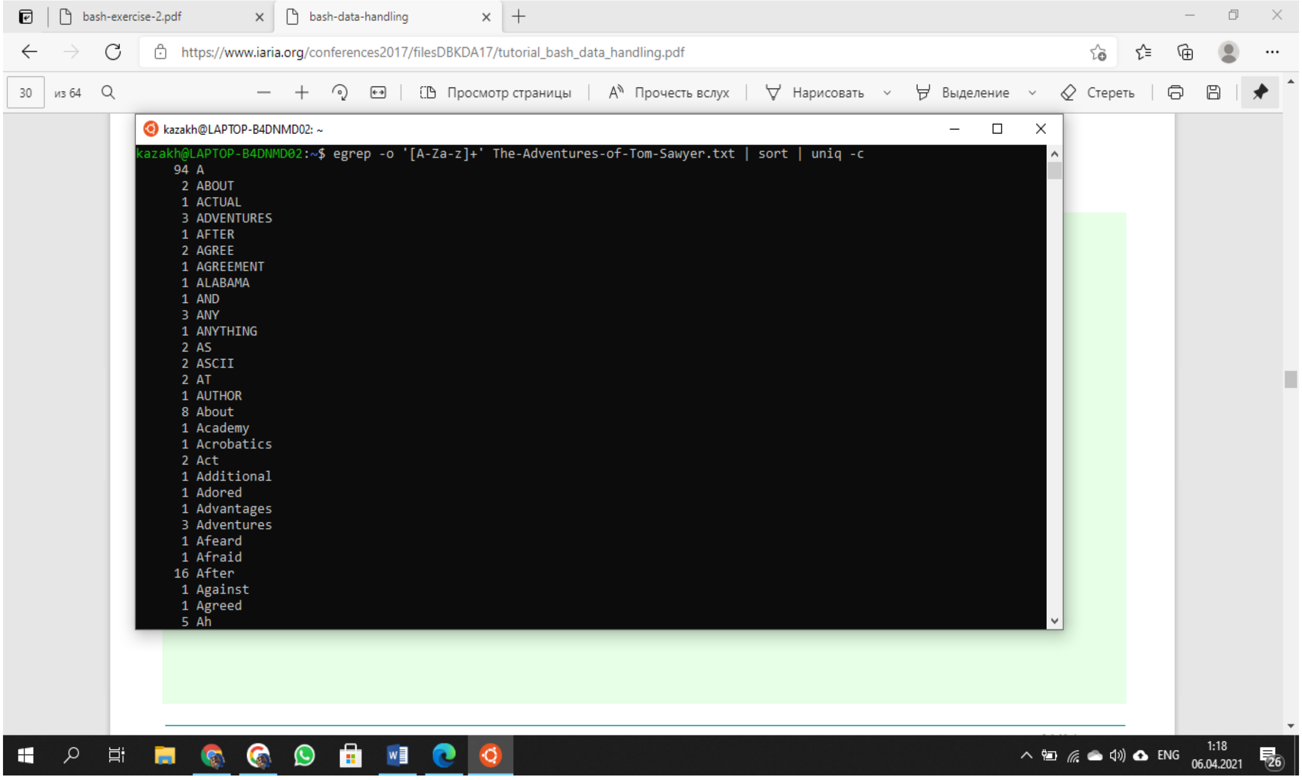
```
tr -c -d '[:alnum:][:space:]' < The-Adventures-of-Tom-Sawyer.txt
```

5. Count, how often each word in this book appears (hint: use sort, uniq).

SOLVE

I tried to do it based on the solution of the previous task.

I sorted the words, then counted them.

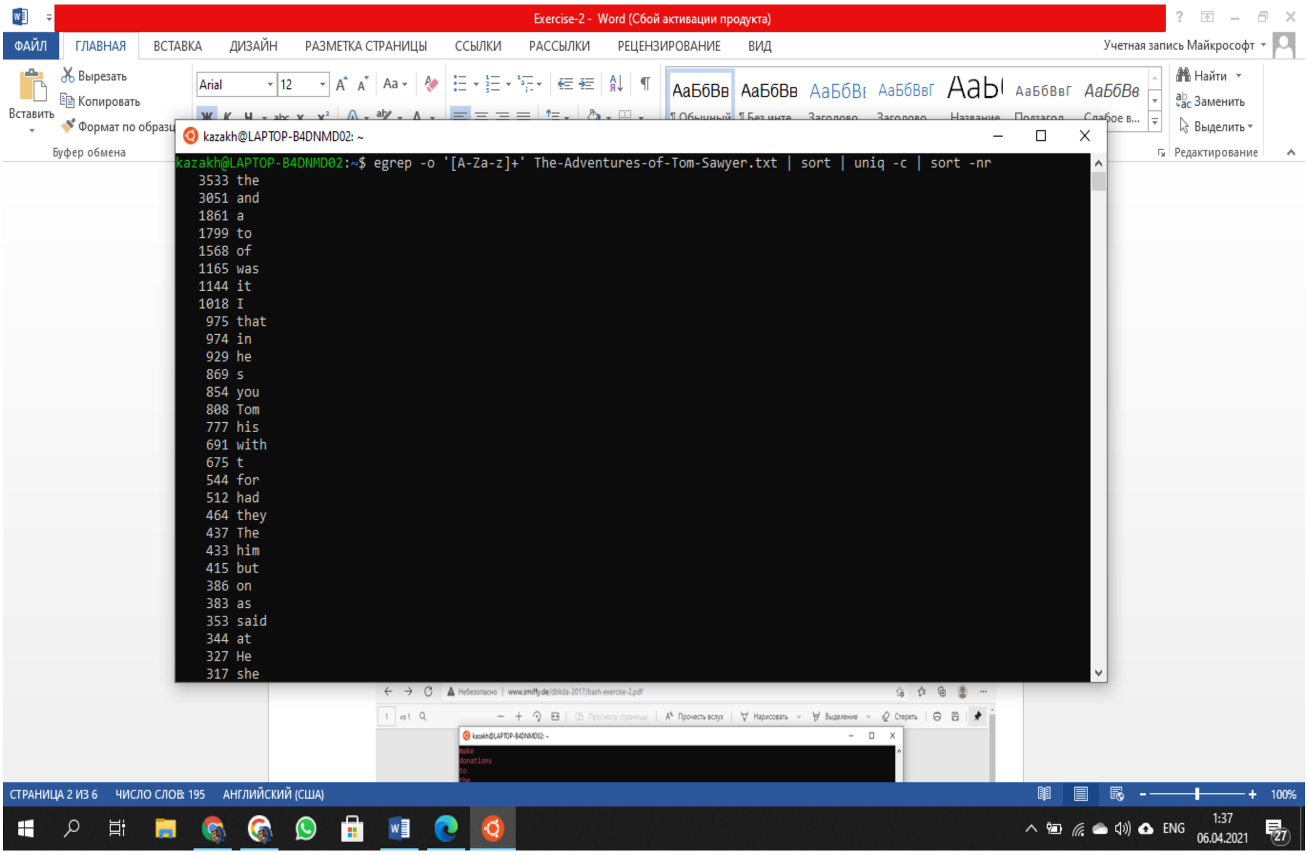


```
kazakh@LAPTOP-B4DNMD02:~$ egrep -o '[A-Za-z]+' The-Adventures-of-Tom-Sawyer.txt | sort | uniq -c
     94 A
      2 ABOUT
      1 ACTUAL
      3 ADVENTURES
      1 AFTER
      2 AGREE
      1 AGREEMENT
      1 ALABAMA
      1 AND
      3 ANY
      1 ANYTHING
      2 AS
      2 ASCII
      2 AT
      1 AUTHOR
      8 About
      1 Academy
      1 Acrobatics
      2 Act
      1 Additional
      1 Adored
      1 Advantages
      3 Adventures
      1 Afeard
      1 Afraid
     16 After
      1 Against
      1 Agreed
      5 Ah
```

6. Order the result, starting with the word with the highest frequency. Which word is it?

SOLVE

I added sort -nr to my code. This word is “the”.



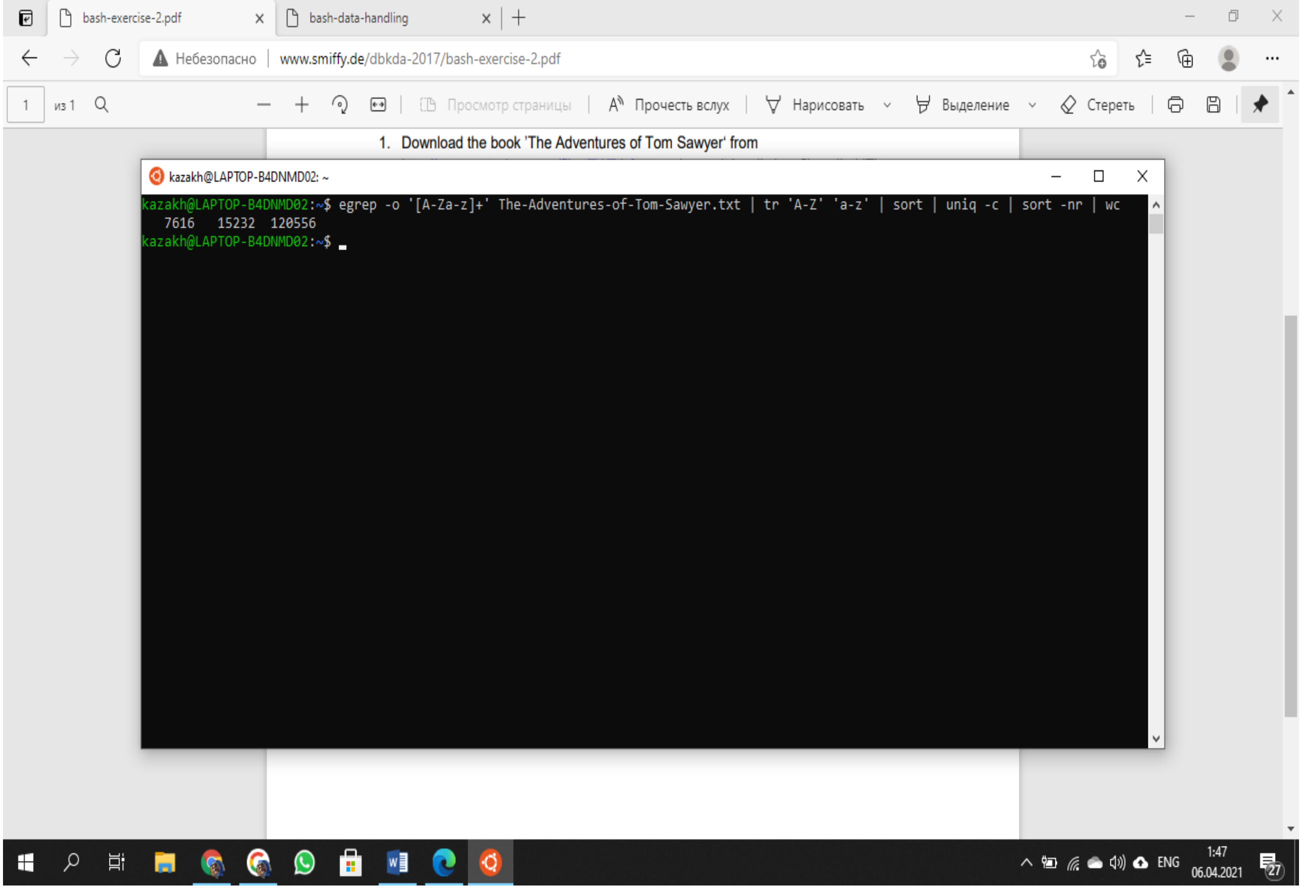
```
kazakh@LAPTOP-B4DNMD02:~$ egrep -o '[A-Za-z]+' The-Adventures-of-Tom-Sawyer.txt | sort | uniq -c | sort -nr
   3533 the
   3051 and
   1861 a
   1799 to
   1568 of
   1165 was
   1144 it
   1018 I
    975 that
    974 in
    929 he
    869 s
    854 you
    808 Tom
    777 his
    691 with
    675 t
    544 for
    512 had
    464 they
    437 The
    433 him
    415 but
    386 on
    383 as
    353 said
    344 at
    327 He
    317 she
```

7. Write all the above steps in one statement (using pipes)

SOLVE

I wrote one code using all the commands. Lines – 7616 and words – 15232.

I used pipes | |.



1. Download the book 'The Adventures of Tom Sawyer' from

```
kazakh@LAPTOP-B4DNMD02:~$ egrep -o '[A-Za-z]+' The-Adventures-of-Tom-Sawyer.txt | tr 'A-Z' 'a-z' | sort | uniq -c | sort -nr | wc
   7616   15232  120556
kazakh@LAPTOP-B4DNMD02:~$ 
```

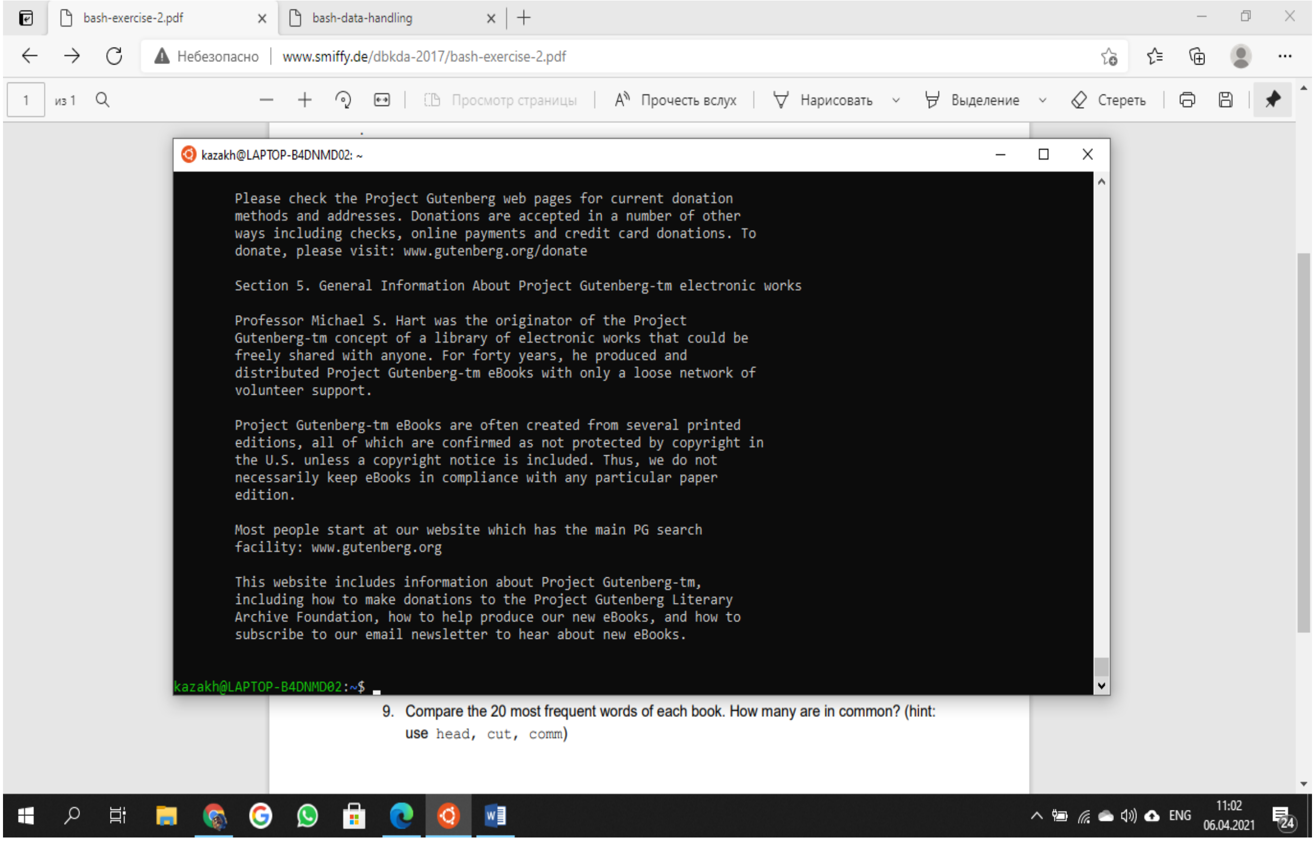
8. Compare the result with the result from the following book: http://www.gutenberg.org/files/2701/2701-0.txt. At which positions (rank) appear the first book specific words?

SOLVE

I use “comm”. **comm** is compare two sorted files line by line.

My code: **comm file1 file2**

First book specific words in first column.



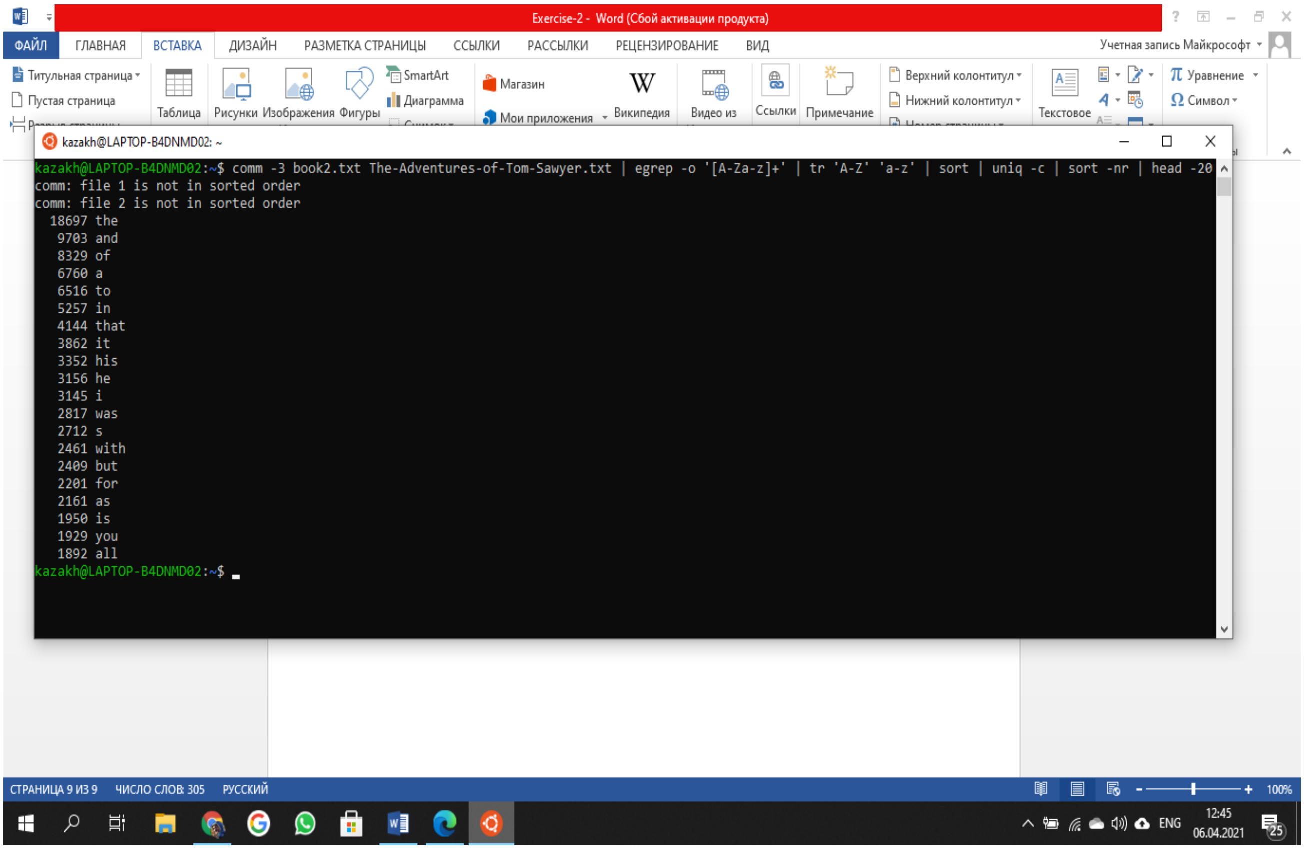
```
        Please check the Project Gutenberg web pages for current donation
        methods and addresses. Donations are accepted in a number of other
        ways including checks, online payments and credit card donations. To
        donate, please visit: www.gutenberg.org/donate

        Section 5. General Information About Project Gutenberg-tm electronic works

        Professor Michael S. Hart was the originator of the Project
        Gutenberg-tm concept of a library of electronic works that could be
        freely shared with anyone. For forty years, he produced and
        distributed Project Gutenberg-tm eBooks with only a loose network of
        volunteer support.

        Project Gutenberg-tm eBooks are often created from several printed
        editions, all of which are confirmed as not protected by copyright in
        the U.S. unless a copyright notice is included. Thus, we do not
        necessarily keep eBooks in compliance with any particular paper
        edition.

        Most people start at our website which has the main PG search
        facility: www.gutenberg.org

        This website includes information about Project Gutenberg-tm,
        including how to make donations to the Project Gutenberg Literary
        Archive Foundation, how to help produce our new eBooks, and how to
        subscribe to our email newsletter to hear about new eBooks.

kazakh@LAPTOP-B4DNMD02:~$ _
```

9. Compare the 20 most frequent words of each book. How many are in common? (hint: use `head, cut, comm`)

9. Compare the 20 most frequent words of each book. How many are in common? (hint: use head, cut, comm).

SOLVE

I chose the lines in both books with comm. I put each word on a line with egrep. I ignored the indexes with tr. Sorted, counted, sorted by frequency and print the 20 most frequent words.



```
kazakh@LAPTOP-B4DNMD02:~$ comm -3 book2.txt The-Adventures-of-Tom-Sawyer.txt | egrep -o '[A-Za-z]+' | tr 'A-Z' 'a-z' | sort | uniq -c | sort -nr | head -20
comm: file 1 is not in sorted order
comm: file 2 is not in sorted order
  18697 the
   9703 and
   8329 of
   6760 a
   6516 to
   5257 in
   4144 that
   3862 it
   3352 his
   3156 he
   3145 i
   2817 was
   2712 s
   2461 with
   2409 but
   2201 for
   2161 as
   1950 is
   1929 you
   1892 all
kazakh@LAPTOP-B4DNMD02:~$
```

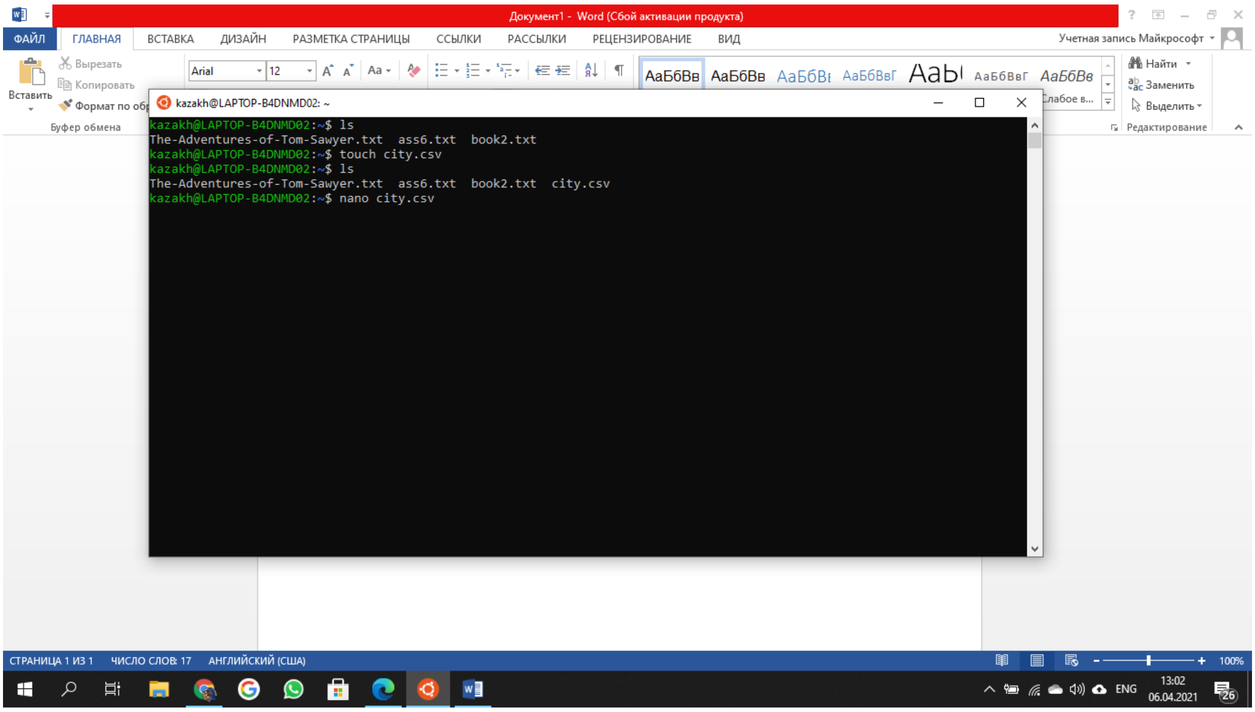
Assignment 6

Exercise 3

1. Create a working copy of your file city.csv (for security reasons).

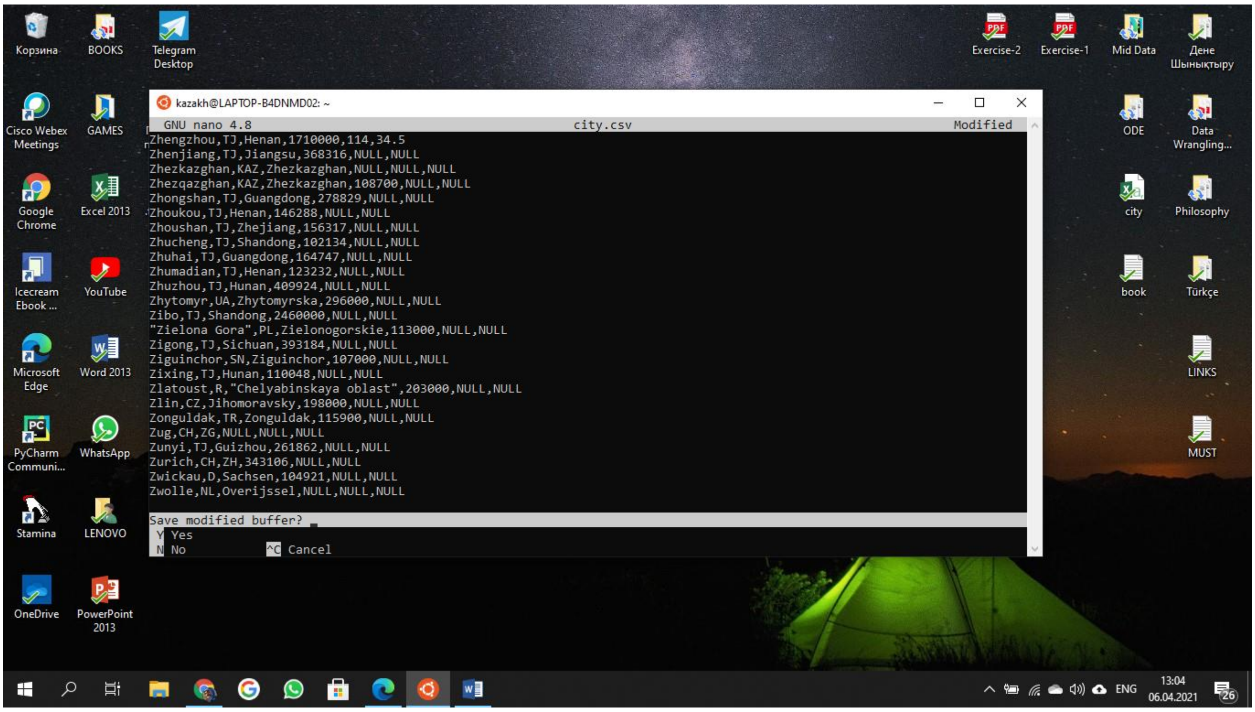
SOLVE

I create new csv file and named it, “city.csv”.



```
kazakh@LAPTOP-B4DNMD02:~$ ls
The-Adventures-of-Tom-Sawyer.txt  ass6.txt  book2.txt
kazakh@LAPTOP-B4DNMD02:~$ touch city.csv
kazakh@LAPTOP-B4DNMD02:~$ ls
The-Adventures-of-Tom-Sawyer.txt  ass6.txt  book2.txt  city.csv
kazakh@LAPTOP-B4DNMD02:~$ nano city.csv
```

And I copied all text and pasted to the csv file.



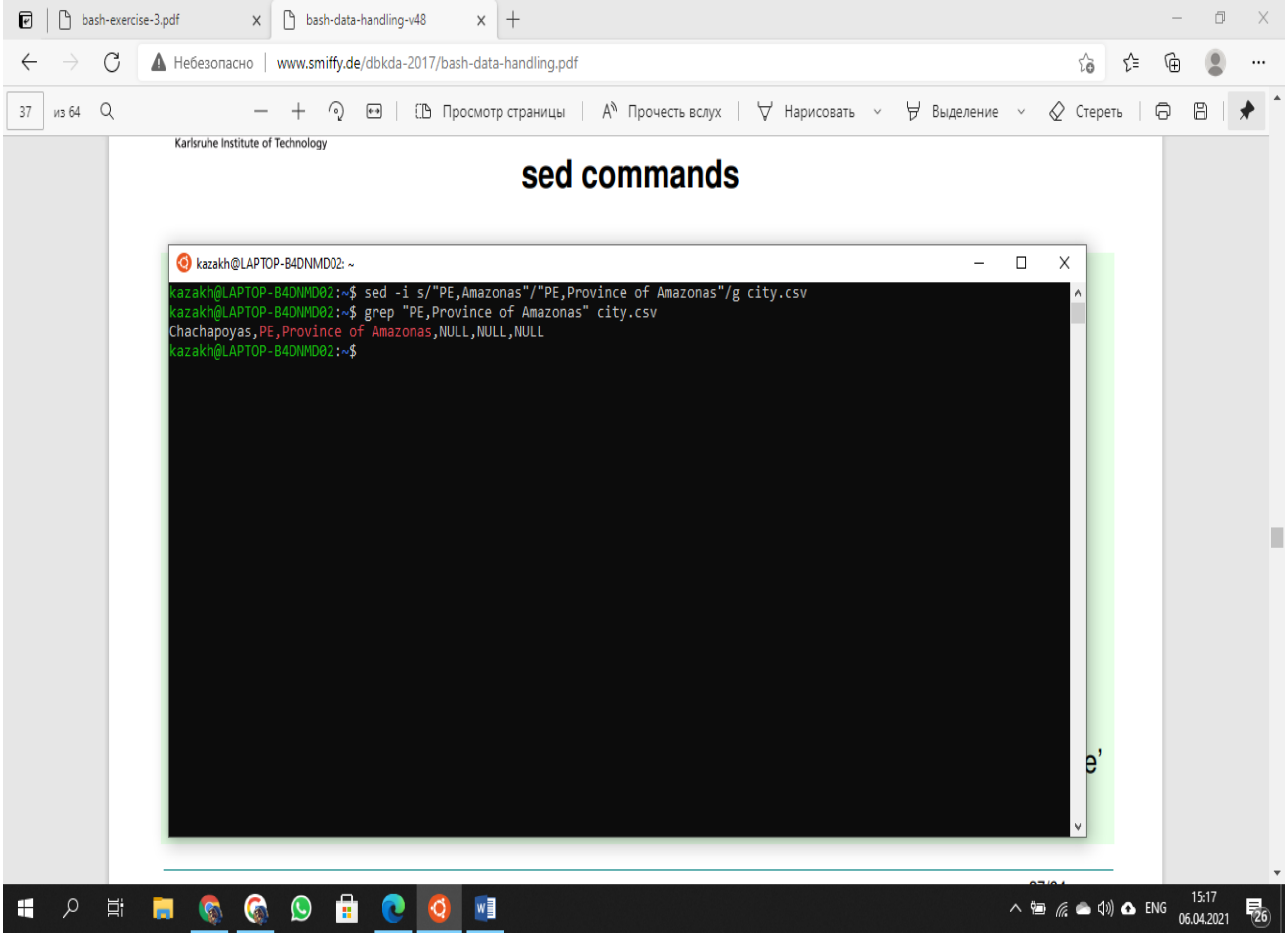
```
  GNU nano 4.8                         city.csv                         Modified
Zhengzhou,TJ,Henan,1710000,114,34.5
Zhenjiang,TJ,Jiangsu,368316,NULL,NULL
Zhezkazghan,KAZ,Zhezkazghan,NULL,NULL,NULL
Zhezqazghan,KAZ,Zhezkazghan,108700,NULL,NULL
Zhongshan,TJ,Guangdong,278829,NULL,NULL
Zhoukou,TJ,Henan,146288,NULL,NULL
Zhoushan,TJ,Zhejiang,156317,NULL,NULL
Zhucheng,TJ,Shandong,102134,NULL,NULL
Zhuhai,TJ,Guangdong,164747,NULL,NULL
Zhumadian,TJ,Henan,123232,NULL,NULL
Zhuzhou,TJ,Hunan,409924,NULL,NULL
Zhytomyr,UA,Zhytomyrska,296000,NULL,NULL
Zibo,TJ,Shandong,2460000,NULL,NULL
"Zielona Gora",PL,Zielonogorskie,113000,NULL,NULL
Zigong,TJ,Sichuan,393184,NULL,NULL
Ziguinchor,SN,Ziguinchor,107000,NULL,NULL
Zixing,TJ,Hunan,110048,NULL,NULL
Zlatoust,R,"Chelyabinskaya oblast",203000,NULL,NULL
Zlin,CZ,Jihomoravsky,198000,NULL,NULL
Zonguldak,TR,Zonguldak,115900,NULL,NULL
Zug,CH,ZG,NULL,NULL,NULL
Zunyi,TJ,Guizhou,261862,NULL,NULL
Zurich,CH,ZH,343106,NULL,NULL
Zwickau,D,Sachsen,104921,NULL,NULL
Zwolle,NL,Overijssel,NULL,NULL,NULL

Save modified buffer?
 Y Yes
 N No           ^C Cancel
```

2. Exchange in the file all occurences of the Province "Amazonas" in Peru (Code PE) with "Province of Amazonas" using sed (inplace).

SOLVE

I used “sed” and –i for inplace. I replaced "Amazonas" in Peru (Code PE) with Province of Amazonas. And I checked it with grep.

Code: *sed –i s/”PE,Amazonas”/”PE,Province of Amazonas”/g city.csv*



Karlsruhe Institute of Technology

sed commands

kazakh@LAPTOP-B4DNMD02: ~

```
kazakh@LAPTOP-B4DNMD02:~$ sed -i s/"PE,Amazonas"/"PE,Province of Amazonas"/g city.csv
kazakh@LAPTOP-B4DNMD02:~$ grep "PE,Province of Amazonas" city.csv
Chachapoyas,PE,Province of Amazonas,NULL,NULL,NULL
kazakh@LAPTOP-B4DNMD02:~$
```

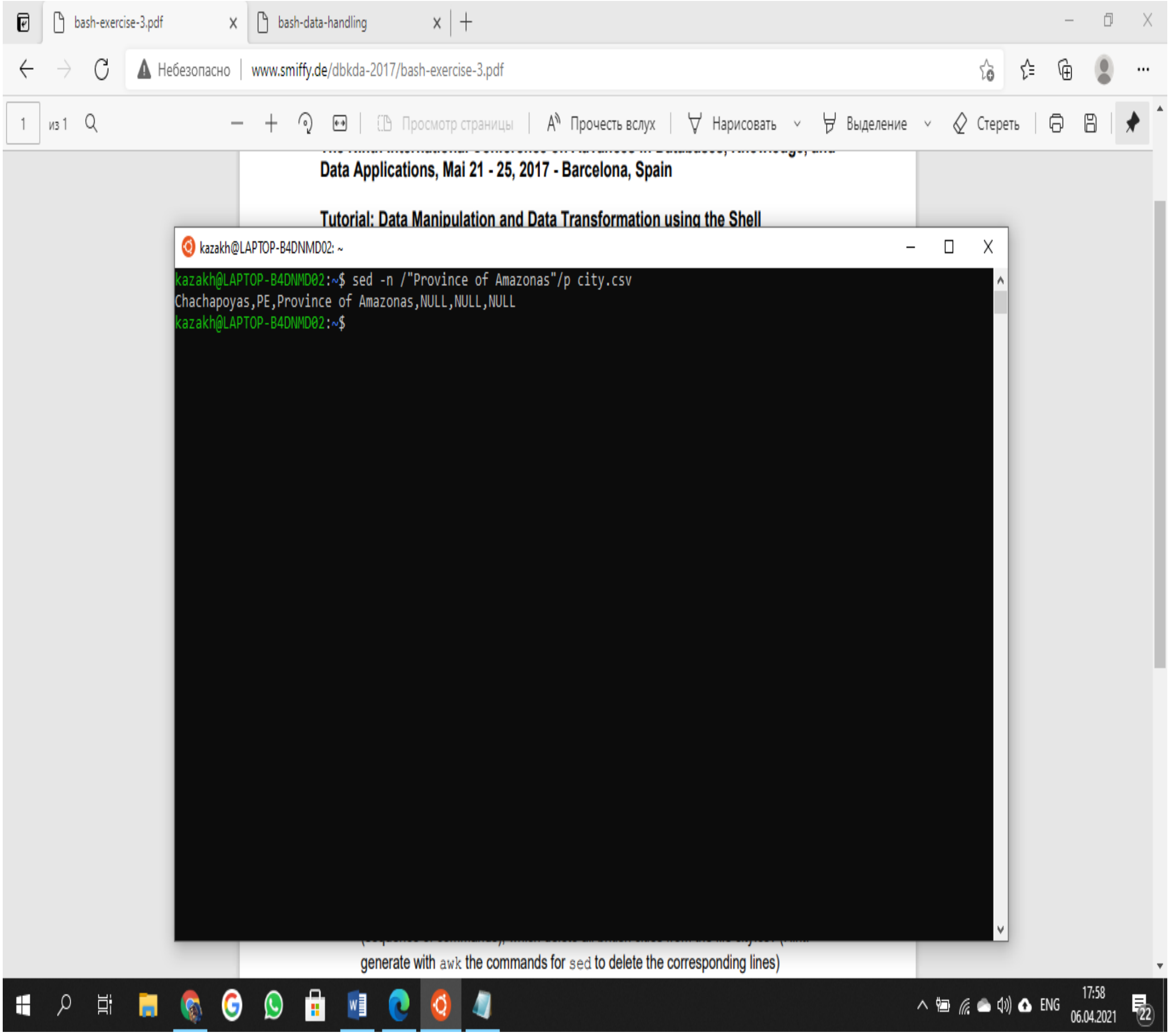
3. Look for entries with the String "ce of Amazonas" - it should be only 1 !

SOLVE

I used sed and entered *sed –n /"Province of Amazonas"/p city.csv*

It works like a grep.

Code: *sed –n /"Province of Amazonas"/p city.csv*



Data Applications, Mai 21 - 25, 2017 - Barcelona, Spain

Tutorial: Data Manipulation and Data Transformation using the Shell

```
kazakh@LAPTOP-B4DNMD02:~$ sed -n /"Province of Amazonas"/p city.csv
Chachapoyas,PE,Province of Amazonas,NULL,NULL,NULL
kazakh@LAPTOP-B4DNMD02:~$
```

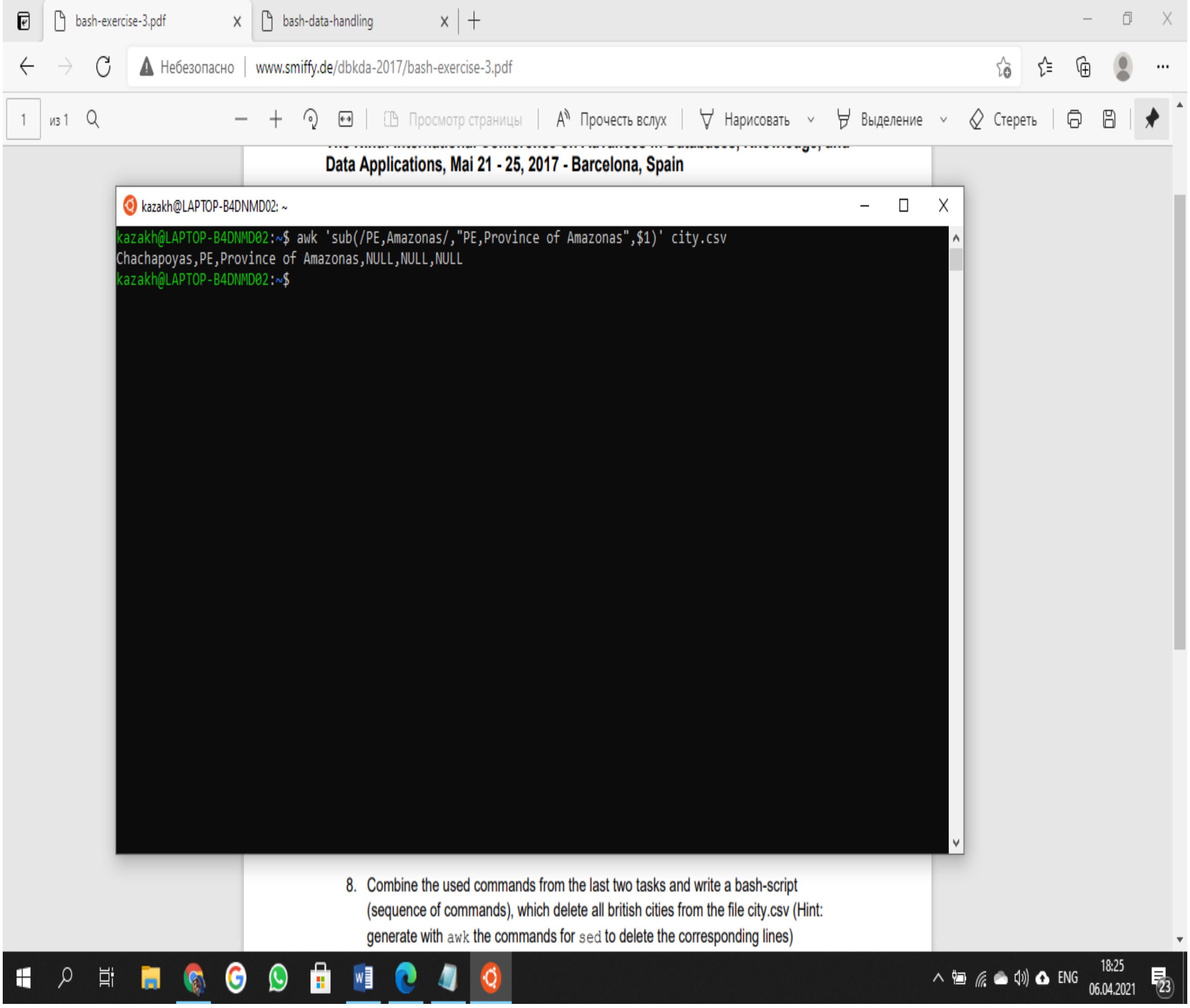
generate with awk the commands for sed to delete the corresponding lines)

4. Make the same operations using awk.

SOLVE

I did it with “awk”

Code: *awk ‘sub(PE,Amazonas/,”PE,Province of Amazonas”, $1)’ city.csv*



Data Applications, Mai 21 - 25, 2017 - Barcelona, Spain

```
kazakh@LAPTOP-B4DNMD02:~$ awk 'sub(/PE,Amazonas/,"PE,Province of Amazonas",$1)' city.csv
Chachapoyas,PE,Province of Amazonas,NULL,NULL,NULL
kazakh@LAPTOP-B4DNMD02:~$
```

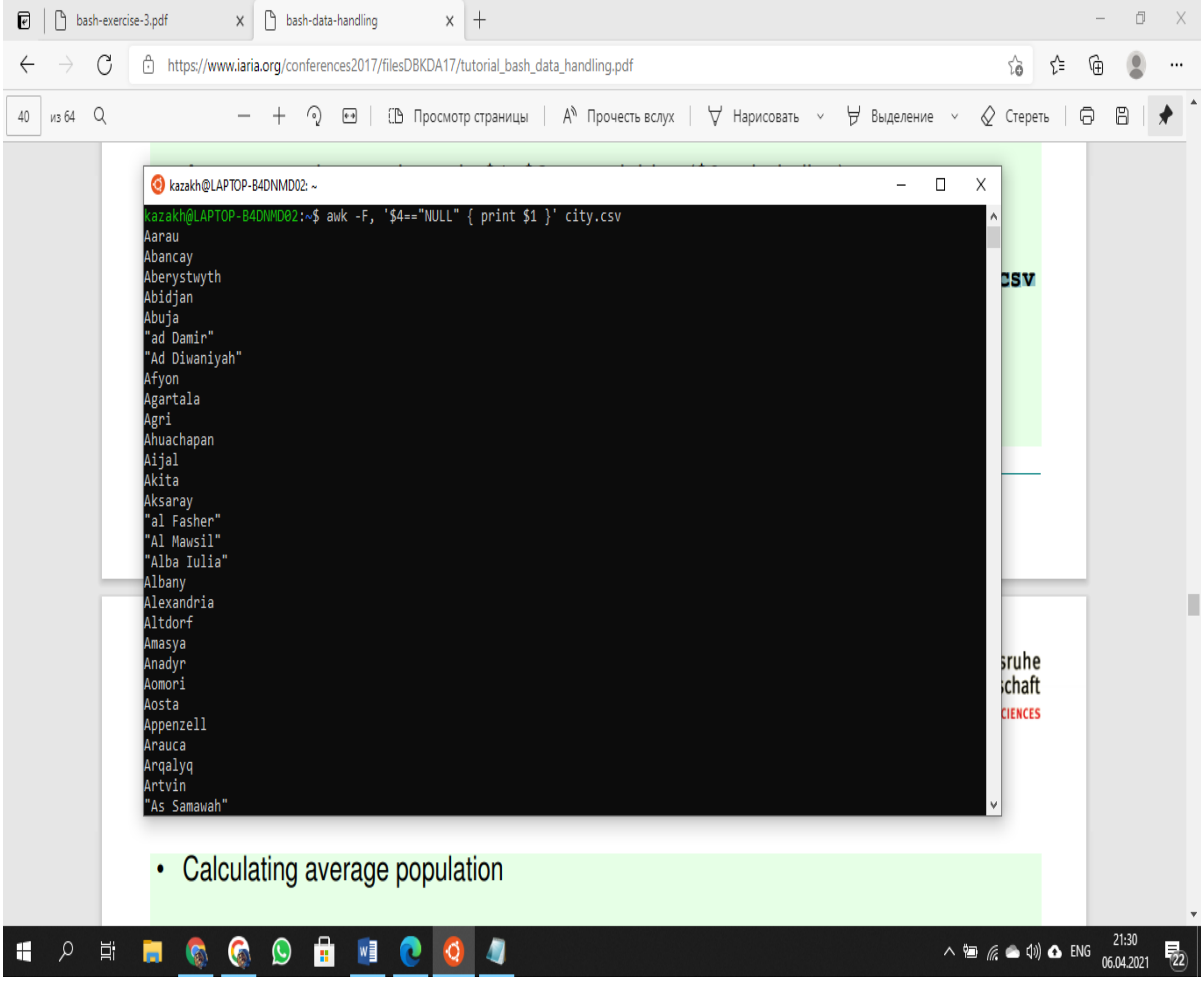
8. Combine the used commands from the last two tasks and write a bash-script (sequence of commands), which delete all british cities from the file city.csv (Hint: generate with `awk` the commands for `sed` to delete the corresponding lines)

5. Print all cities which have no population given.

SOLVE

I use awk. If fourth column equals to zero, then it print name of city.

Code: *awk –F, ‘$4==”NULL” { print $1 }’ city.csv*



```
kazakh@LAPTOP-B4DNMD02:~$ awk -F, '$4=="NULL" { print $1 }' city.csv
Aarau
Abancay
Aberystwyth
Abidjan
Abuja
"ad Damir"
"Ad Diwaniyah"
Afyon
Agartala
Agri
Ahuachapan
Aijal
Akita
Aksaray
"al Fasher"
"Al Mawsil"
"Alba Iulia"
Albany
Alexandria
Altdorf
Amasya
Anadyr
Aomori
Aosta
Appenzell
Arauca
Arqalyq
Artvin
"As Samawah"
```

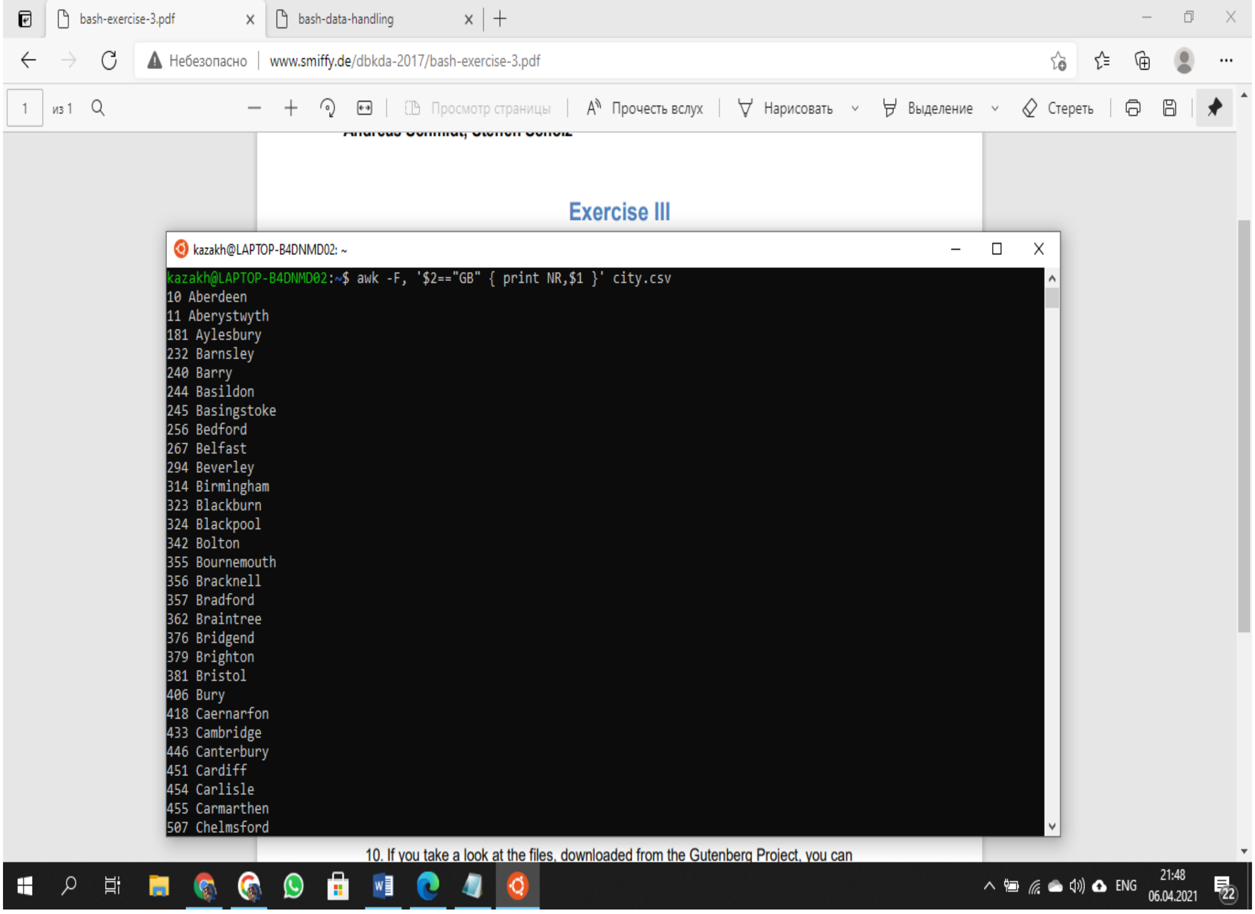
bash-exercise-3.pdf
bash-data-handling
https://www.iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
40
из 64
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
kazakh@LAPTOP-B4DNMD02: ~
Calculating average population
ENG
21:30
06.04.2021

6. Print the line numbers of the cities in Great Britain (Code: GB) using awk.

SOLVE

I use awk. If second column is GB, then it print number of line and city.

Code: *awk -F, '$2=="GB" { print NR,$1 }' city.csv*



bash-exercise-3.pdf
bash-data-handling
Небезопасно
www.smiffy.de/dbkda-2017/bash-exercise-3.pdf
из 1
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
Exercise III
kazakh@LAPTOP-B4DNMD02: ~
10. If you take a look at the files, downloaded from the Gutenberg Project, you can
ENG
21:48
06.04.2021

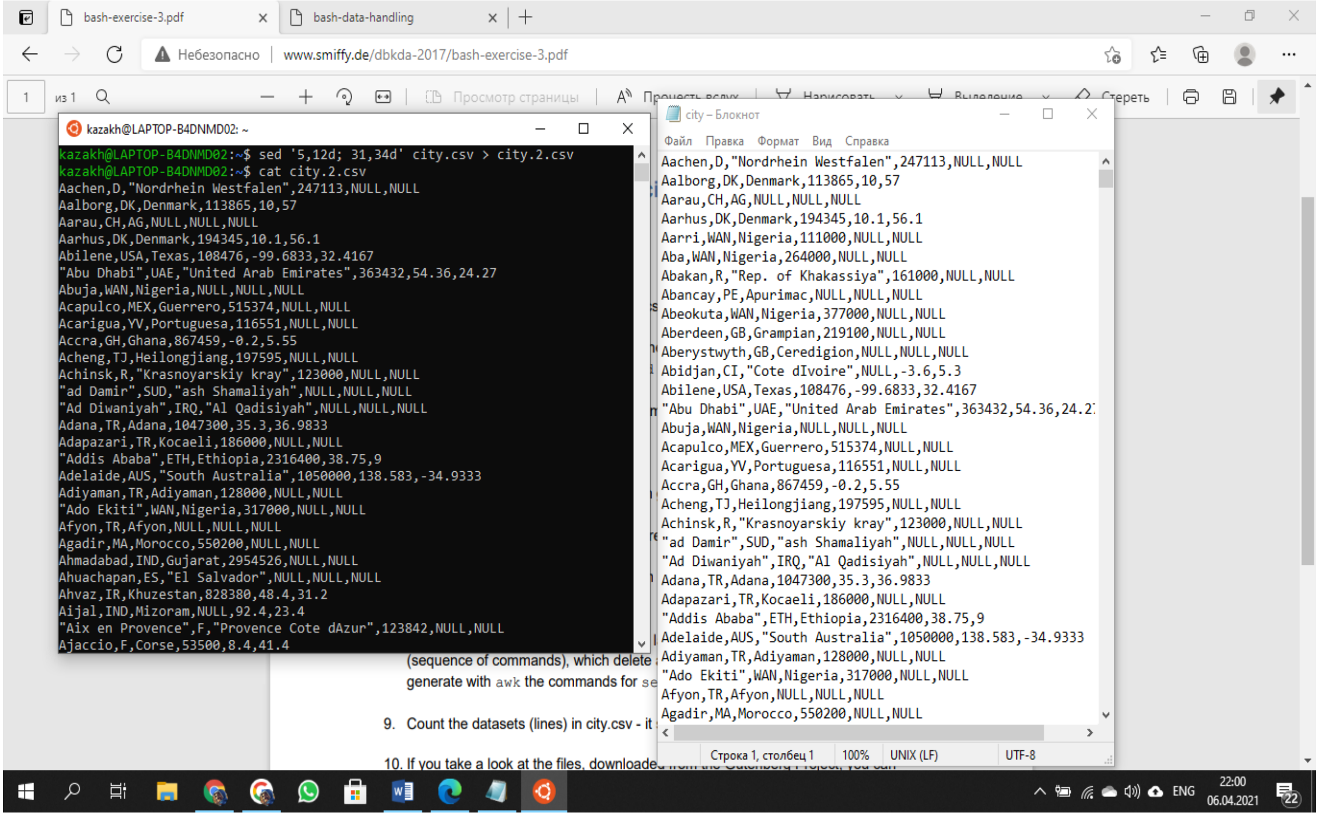
```
kazakh@LAPTOP-B4DNMD02:~$ awk -F, '$2=="GB" { print NR,$1 }' city.csv
10 Aberdeen
11 Aberystwyth
181 Aylesbury
232 Barnsley
240 Barry
244 Basildon
245 Basingstoke
256 Bedford
267 Belfast
294 Beverley
314 Birmingham
323 Blackburn
324 Blackpool
342 Bolton
355 Bournemouth
356 Bracknell
357 Bradford
362 Braintree
376 Bridgend
379 Brighton
381 Bristol
406 Bury
418 Caernarfon
433 Cambridge
446 Canterbury
451 Cardiff
454 Carlisle
455 Carmarthen
507 Chelmsford
```

7. Delete the records 5-12 and 31-34 from city.csv and store the result in city.2.csv using sed.

SOLVE

I use sed. I deleted 5-12 lines and 31-34 lines. And saved result to city.2.csv.

Code: *sed '5,12d; 31,34d' city.csv > city.2.csv*



```
kazakh@LAPTOP-B4DNMD02:~$ sed '5,12d; 31,34d' city.csv > city.2.csv
kazakh@LAPTOP-B4DNMD02:~$ cat city.2.csv
Aachen,D,"Nordrhein Westfalen",247113,NULL,NULL
Aalborg,DK,Denmark,113865,10,57
Aarau,CH,AG,NULL,NULL,NULL
Aarhus,DK,Denmark,194345,10.1,56.1
Abilene,USA,Texas,108476,-99.6833,32.4167
"Abu Dhabi",UAE,"United Arab Emirates",363432,54.36,24.27
Abuja,WAN,Nigeria,NULL,NULL,NULL
Acapulco,MEX,Guerrero,515374,NULL,NULL
Acarigua,YV,Portuguesa,116551,NULL,NULL
Accra,GH,Ghana,867459,-0.2,5.55
Acheng,TJ,Heilongjiang,197595,NULL,NULL
Achinsk,R,"Krasnoyarskiy kray",123000,NULL,NULL
"ad Damir",SUD,"ash Shamaliyah",NULL,NULL,NULL
"Ad Diwaniyah",IRQ,"Al Qadisiyah",NULL,NULL,NULL
Adana,TR,Adana,1047300,35.3,36.9833
Adapazari,TR,Kocaeli,186000,NULL,NULL
"Addis Ababa",ETH,Ethiopia,2316400,38.75,9
Adelaide,AUS,"South Australia",1050000,138.583,-34.9333
Adiyaman,TR,Adiyaman,128000,NULL,NULL
"Ado Ekiti",WAN,Nigeria,317000,NULL,NULL
Afyon,TR,Afyon,NULL,NULL,NULL
Agadir,MA,Morocco,550200,NULL,NULL
Ahmadabad,IND,Gujarat,2954526,NULL,NULL
Ahuachapan,ES,"El Salvador",NULL,NULL,NULL
Ahvaz,IR,Khuzestan,828380,48.4,31.2
Aijal,IND,Mizoram,NULL,92.4,23.4
"Aix en Provence",F,"Provence Cote dAzur",123842,NULL,NULL
Ajaccio,F,Corse,53500,8.4,41.4
```

```
Aachen,D,"Nordrhein Westfalen",247113,NULL,NULL
Aalborg,DK,Denmark,113865,10,57
Aarau,CH,AG,NULL,NULL,NULL
Aarhus,DK,Denmark,194345,10.1,56.1
Aarri,WAN,Nigeria,111000,NULL,NULL
Aba,WAN,Nigeria,264000,NULL,NULL
Abakan,R,"Rep. of Khakassiya",161000,NULL,NULL
Abancay,PE,Apurimac,NULL,NULL,NULL
Abeokuta,WAN,Nigeria,377000,NULL,NULL
Aberdeen,GB,Grampian,219100,NULL,NULL
Aberystwyth,GB,Ceredigion,NULL,NULL,NULL
Abidjan,CI,"Cote dIvoire",NULL,-3.6,5.3
Abilene,USA,Texas,108476,-99.6833,32.4167
"Abu Dhabi",UAE,"United Arab Emirates",363432,54.36,24.2
Abuja,WAN,Nigeria,NULL,NULL,NULL
Acapulco,MEX,Guerrero,515374,NULL,NULL
Acarigua,YV,Portuguesa,116551,NULL,NULL
Accra,GH,Ghana,867459,-0.2,5.55
Acheng,TJ,Heilongjiang,197595,NULL,NULL
Achinsk,R,"Krasnoyarskiy kray",123000,NULL,NULL
"ad Damir",SUD,"ash Shamaliyah",NULL,NULL,NULL
"Ad Diwaniyah",IRQ,"Al Qadisiyah",NULL,NULL,NULL
Adana,TR,Adana,1047300,35.3,36.9833
Adapazari,TR,Kocaeli,186000,NULL,NULL
"Addis Ababa",ETH,Ethiopia,2316400,38.75,9
Adelaide,AUS,"South Australia",1050000,138.583,-34.9333
Adiyaman,TR,Adiyaman,128000,NULL,NULL
"Ado Ekiti",WAN,Nigeria,317000,NULL,NULL
Afyon,TR,Afyon,NULL,NULL,NULL
Agadir,MA,Morocco,550200,NULL,NULL
```

(sequence of commands), which delete
generate with awk the commands for se

9. Count the datasets (lines) in city.csv - it

10. If you take a look at the files, downloaded

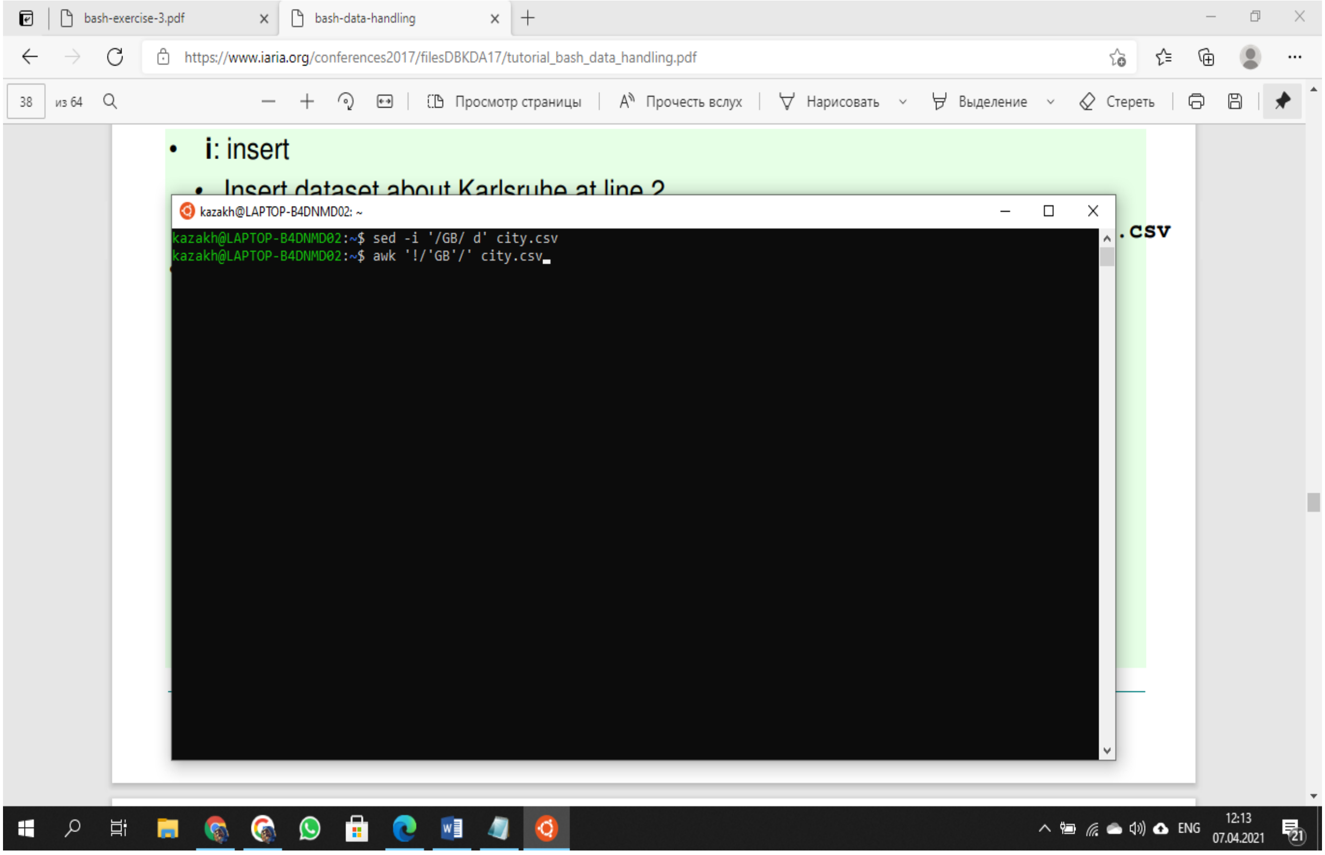
8. Combine the used commands from the last two tasks and write a bash-script (sequence of commands), which delete all british cities from the file city.csv (Hint: generate with awk the commands for sed to delete the corresponding lines)

SOLVE

I tried a lot, but could not find a solution. But instead of using sed and awk equally, I wrote code other ways. There are 2 ways. With awk and with sed.

Code 1: sed –i '/GB/ d' city.csv

Code 2: awk '!/'GB'/' city.csv


bash-exercise-3.pdf
bash-data-handling
https://www.iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
38
из 64
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
i: insert
Insert dataset about Karlsruhe at line 2
.csv
kazakh@LAPTOP-B4DNMD02: ~
ENG
12:13
07.04.2021

```
kazakh@LAPTOP-B4DNMD02:~$ sed -i '/GB/ d' city.csv
kazakh@LAPTOP-B4DNMD02:~$ awk '!/'GB'/' city.csv
```

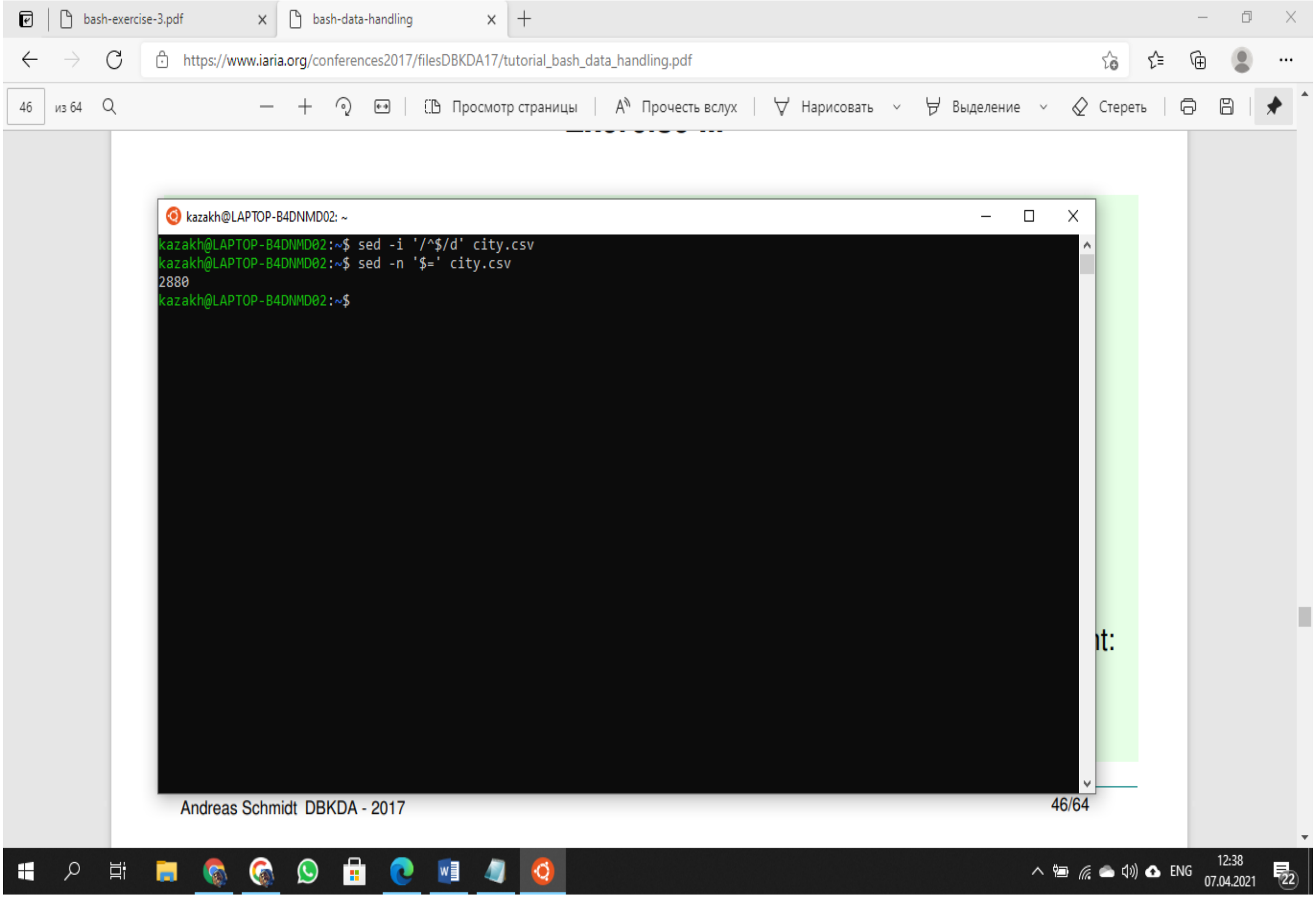
9. Count the datasets (lines) in city.csv - it should be 2880.

SOLVE

I use sed for counting. Firstly I removed all empty lines. And counted the dataset.

Code1: *sed –i '/^$/d' city.csv*

Code2: *sed –n '$=' city.csv*



```
kazakh@LAPTOP-B4DNMD02:~$ sed -i '/^$/d' city.csv
kazakh@LAPTOP-B4DNMD02:~$ sed -n '$=' city.csv
2880
kazakh@LAPTOP-B4DNMD02:~$
```

bash-exercise-3.pdf
bash-data-handling
https://www.iaria.org/conferences2017/filesDBKDA17/tutorial_bash_data_handling.pdf
46
из 64
Просмотр страницы
Прочесть вслух
Нарисовать
Выделение
Стереть
kazakh@LAPTOP-B4DNMD02: ~
Andreas Schmidt DBKDA - 2017
46/64
ENG
12:38
07.04.2021

10. If you take a look at the files, downloaded from the Gutenberg Project, you can identify some boilerplate text at the begin and the end of the book. Which are the lines, who separate the literary text from the boilerplate text?

SOLVE

9. Count the datasets (lines) in city.csv - it should be 2880

SOLVE