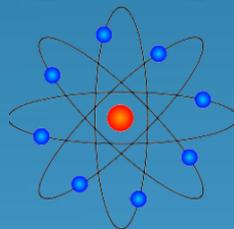




*Russia*  
*Voronezh*



**Voronezh-2**



**IYNT 2015**

Work performed pupil 8B class:

*Danil Kozlov*

Supervisor:

*Lepeshkina Natalia Valeryevna*

## 8. Library

- One person has decided to download all of the fiction existing in the English language and store it on a single USB stick. He expects to find or generate the respective text files, compress them, and then index them conveniently. Is this ambition realistic? Suggest a plan to approach this goal and solve a partial problem of this plan.

# The goals of the task

- To determine the estimated amount of literature for compression.
- To determine the amount of memory after compression of the books.
- To consider options for the desired media memory.
- To define realistic intentions.

# The general volume of material.

- To determine the exact amount of all the existing pieces of the English literature is not possible.
- The largest library in the world is the Library of Congress. It has more than 32 million books.
- The biggest e-library is the Google books library. It has nearly 130 million books (129 864 880).

# The Calculation of data for archiving.

The average number of pages - 250

One page on average - 1986 characters

1 character = 1 byte

$250 * 1986 * 1 = 496,500$  bytes (memory for 1 book)

$496\,500 * 130\,000\,000 = 64\,545\,000\,000\,000$  bytes is the approximate amount of memory required for all English-language literature.

$1\text{TB} = 2^{40}$

bytes

Thus, you need backup TB 58.70333552593

# Compressing text using various archivers

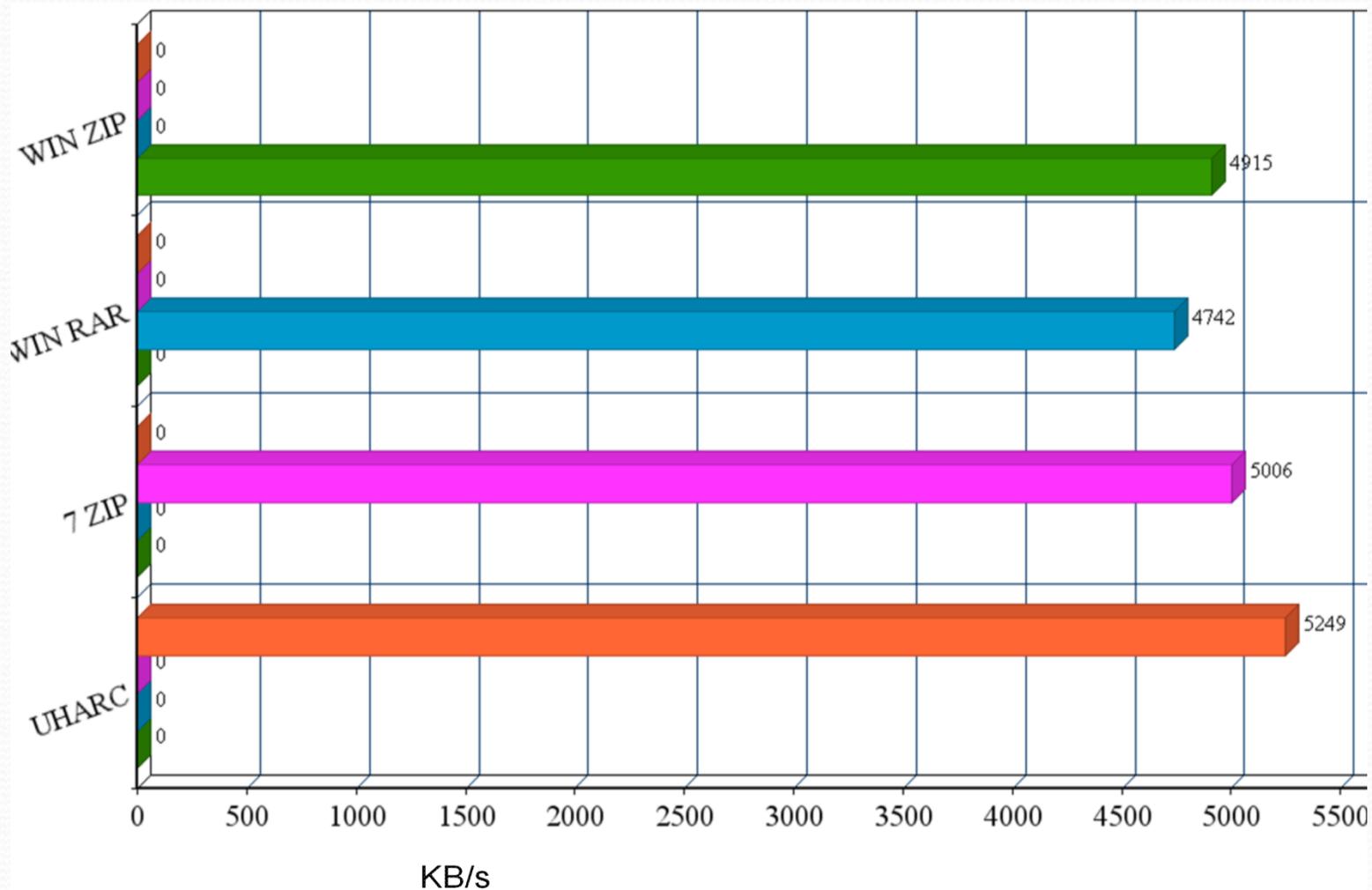
The program	Reducing the amount of post-compression (times)
UHARC 	24
WIN RAR 	17
7 ZIP 	15
WIN ZIP 	10

**Of all of these archives the most capacious is UHARC**  
**Consequently,  $58.70333552593 / 23 = 2.44547231358$  TB**  
**That is, the material can be compressed up to  $\approx 2,4$  TB**

# Types of flash drives

Type of Device		The amount of memory	How many are needed
Device HyperStor-6200		100 Tb	1
Drive WD		1-3Tb	1-3
Flash-drive		64 GB до 128 GB	1008-504
HD DVD		4.7 Гб	523
FDD		1.44 Мб	1 832 520

# The rate of contraction



# The time of compression

- The highest rate of compression in archives: UHARC (5249 kb / s) and 7 ZIP (5006 kb / s)
- To archive 64545000000 kb requires 12,296,627 sec., Or 142.32 days (for UHARC)
- For the 7 ZIP 12,893,527 sec., Or 149.23 days
- The archiving process of the intended scope of the literature takes a lot of time in continuous operation of the archiver.

# Conclusion

- The problem statement says only about the fiction in the English language, but the exact information about the number of books in the English language is not available.
- Compression process takes a considerable period of time - by 142 days or longer (depending on the choice of compression method).
- Data can take more than 3 terabytes - that is required by more than one carrier.
- During the whole time compression computer must operate constantly.
- To carry out the job currently with the existing technical means is real, but it is very difficult.

# The sources

- <http://fulltienich.com/samaya-bolshaya-fleshka-po-obemu/>
- [https://ru.wikipedia.org/wiki/Поисковый\\_индекс](https://ru.wikipedia.org/wiki/Поисковый_индекс)
- <http://products.wdc.com/largecapacitydrives>
- <http://timfan.info/forum/viewtopic.php?pid=37#p37>
- <http://www.books.ru/news/id.php?id=2485>
- <http://www.uznayvse.ru/interesting-facts/samyie-bolshie-biblioteki-v-mire.html>