



**ITMO UNIVERSITY**

Saint Petersburg, Russia

# Development of an automatic spelling correction tool for analyzing clinical text in Russian

Student: Dmitry Pogrebnoy, J4132c

Head of Practice from the trainee's host organization: Sergey Kovalchuk, PhD

Scientific assistant: Anastasia Funkner, engineer

# Electronic medical records

- In healthcare, there are various predictive and decision-making models based on information from patients' medical records
- The quality of such models strongly depend on the quality of the source texts
- Electronic patient data is usually presented in plain text and contains a lot of spelling errors
- Spelling errors in the source texts greatly reduce the quality of the final models and therefore require correction

# Goal and tasks

**Goal:** Develop an automatic spelling correction tool for analyzing clinical texts in Russian. The tool should accept raw medical text as input and return corrected text with a minimum number of incorrect words.

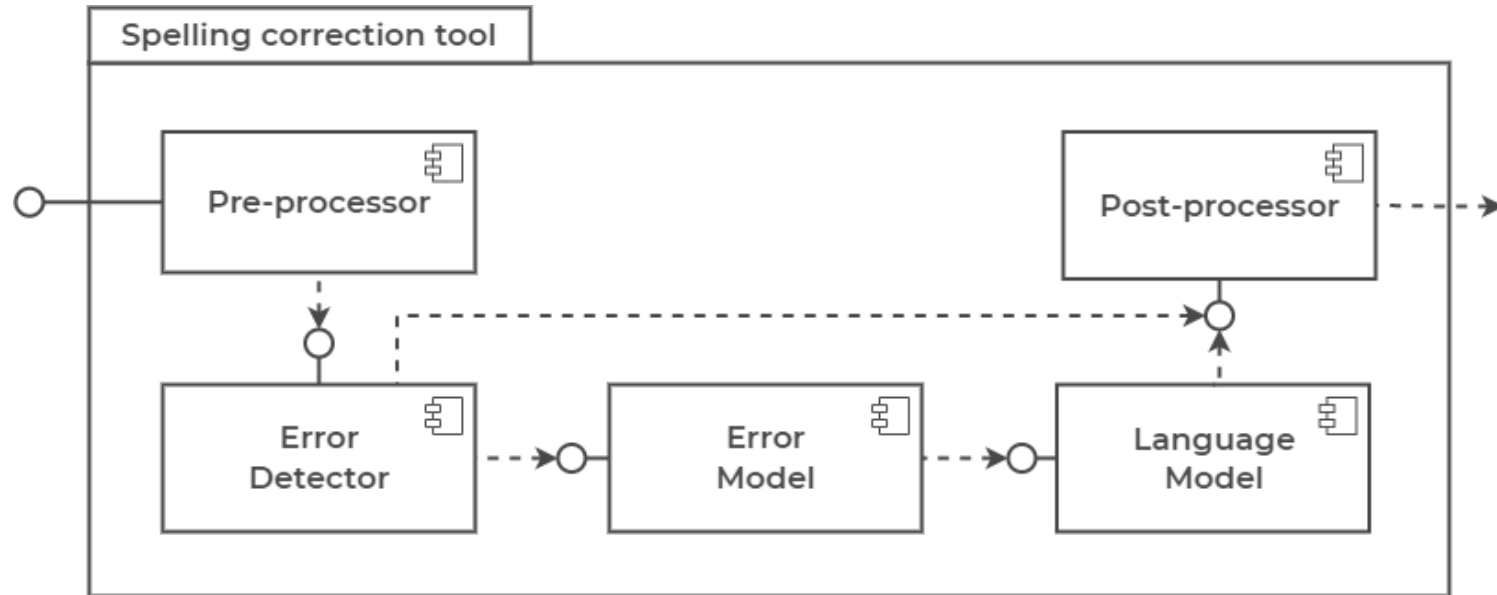
## **Tasks for first semester:**

- Make an overview of the Russian medical texts correction
- Analyze existing solutions in this area
- Develop the first draft of the architecture of the tool

# Existing tools

Tool	Error precision	Lexical precision	Overall precision	Average words per second
Aspell-python	0.65	0.775	0.7125	353
PyHunspell	0.59	0.49	0.54	11.5
PyEnchant	0.6	0.455	0.5275	26.4
LanguageTool-python	0.64	0.845	0.7425	19.1
PySpellChecker	0.335	0.765	0.55	4.3
SymSpellPy	0.42	0.78	0.6	15892.1
Jumspell	0.395	0.925	0.66	2043.2
Spellchecker prototype	0.41	0.83	0.62	0.07

# Principle architecture



# Conclusion and further plans

- The overview of the Russian medical texts correction are made
- The analysis of existing spelling correctors is carried out
- The principle architecture of the tool is developed

## Further plans:

- Finish the development of the architecture of the tool
- Implement a tool for correcting spelling errors
- Conduct testing of the developed tool

# Thank you for your attention!

[www.ifmo.ru](http://www.ifmo.ru)

IT'sMO *re than a*  
UNIVERSITY

# Spelling errors in question

Type of mistake	Incorrect text	Correct text
Missing characters	туб□ркулез	туб <u>е</u> ркулез
Extra characters	туберк <u>п</u> улез	туберкулез
Shuffled characters	туб <u>ре</u> кулез	туб <u>е</u> <u>р</u> кулез
Missing word separator	острый туберкулез	острый_туберкулез
Extra word separator	туб_еркулез	туберкулез



# Testing an existing tools

- **Error precision** - the ratio of the number of correctly corrected words to the total number of incorrect words
- 200 medical words with spelling errors were taken to compute error precision
- **Lexical precision** - the ratio of the number of unchanged modified words to the total number of correct words
- 200 correct medical words were taken to compute lexical precision
- **The performance test** was done on a laptop on Ubuntu 20.04 with 16 GB RAM and Intel Core i7-9750H CPU @ 2.60GHz \* 12