Saint Petersburg State University

Leonid P. Churilov, Vladimir A. Starostenko,
Yuri I. Stroev, Natalia N. Vlasova, Sergei I. Zykov

Distant Learning and Handwritten Medical Records Recognition in Clinical Pathophysiology

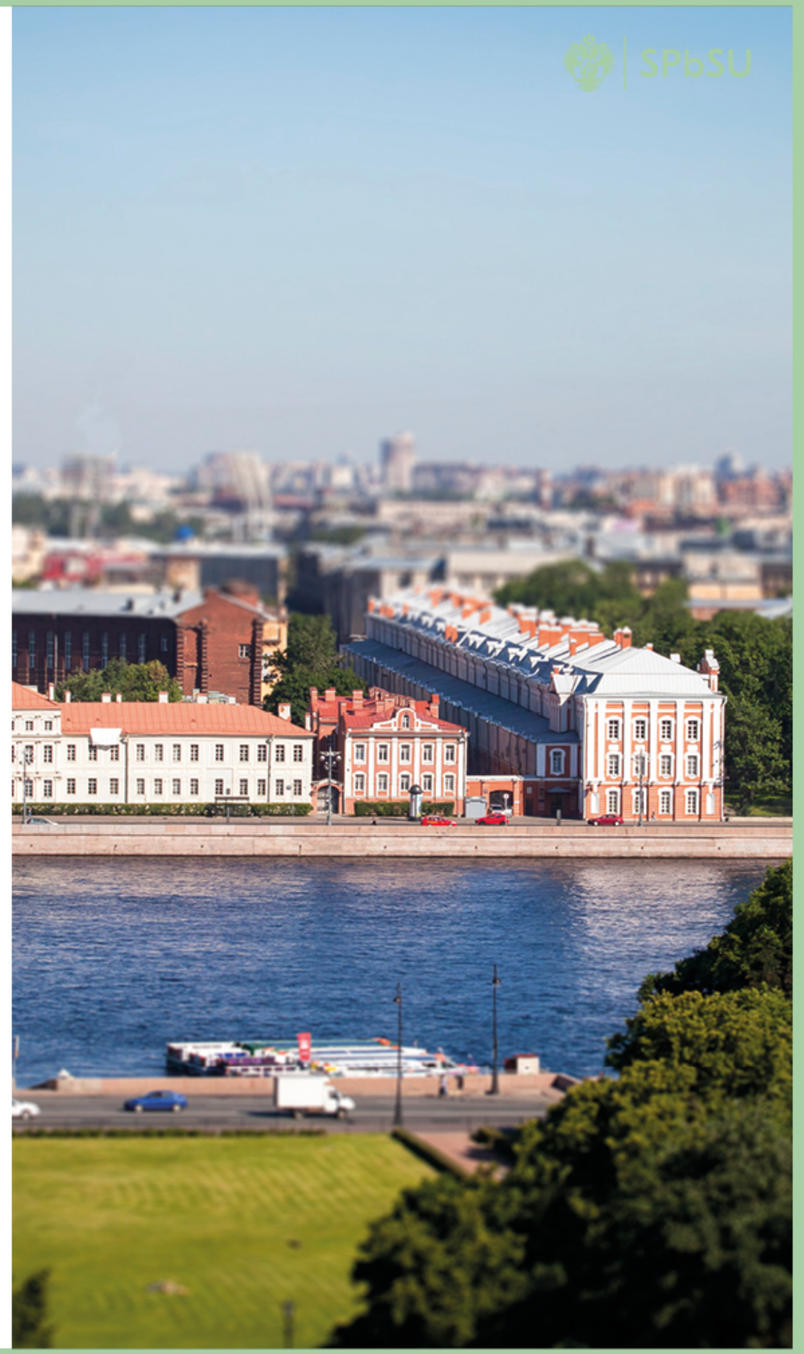
St. Petersburg State University is the world's largest scientific and educational center

It offers almost four hundred basic educational programs and more than a thousand additional educational programs, from African studies to econophysics.

There are more than **40,000** students enrolled, including international students.
More than **6,000** teachers, including Nobel and Fields Prize winners, work at the University.

Among **24** Faculties and Institutes of University Faculty of Medicine and Faculty of Dental Medicine & Medical Technologies.

Through its history University produced and\or harbored **8** Nobel Prize winners, **2** of them – in Physiology & Medicine



Science & Research at SPbU

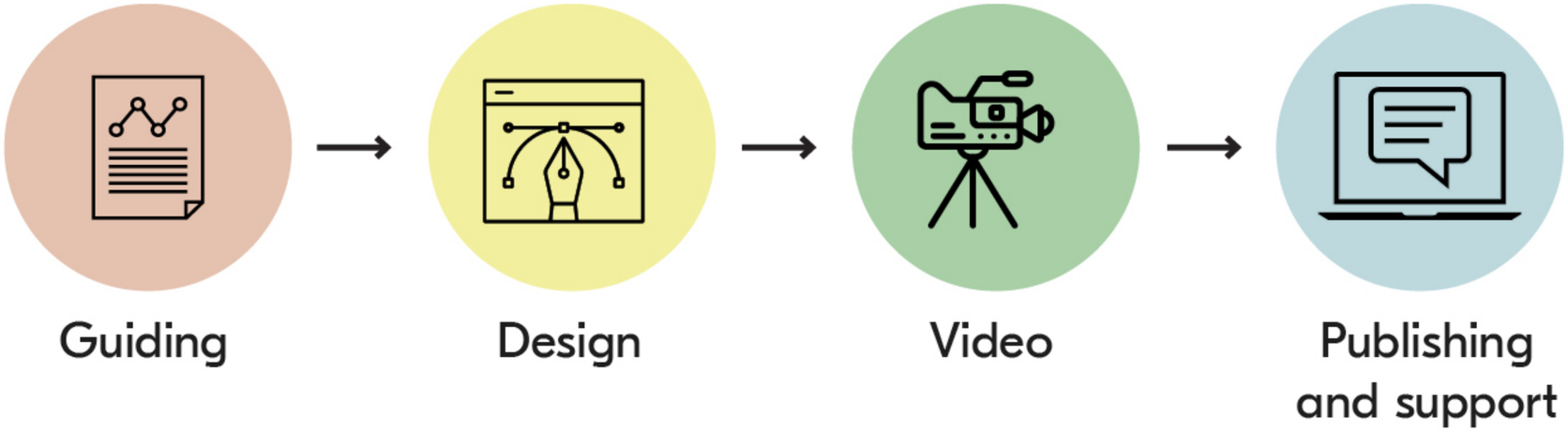
The University has more than **15 major laboratories** and **26 resource centers**, which are part of the country's leading scientific park. The unique scientific and pedagogical staff as well as the developed infrastructure allow unique interdisciplinary research to be conducted and contribute to the development of new principles and advanced ideas in various scientific fields.

Our Mission

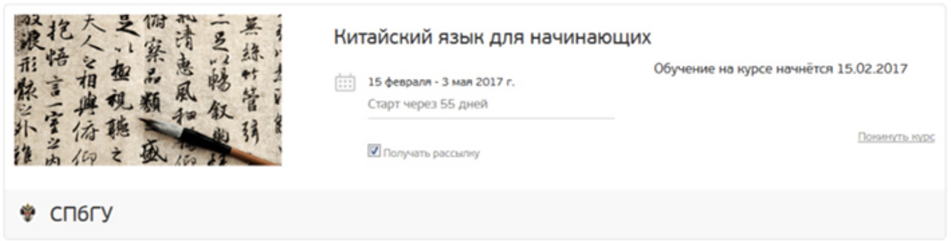
One of the University's aims and its mission are to **educate and provide access to knowledge**. In its role as a knowledge translator, the mission and objectives of the University are directly consistent with one of the goals set by the Russian Prime Minister D.A. Medvedev, i.e. reducing the lack of high quality educational content by creating special resources, particularly online courses.

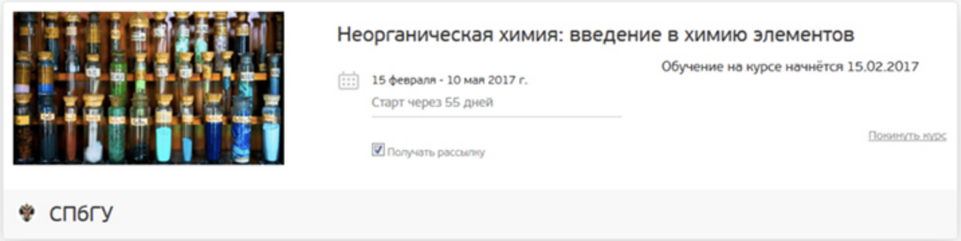
Online education and MOOC's (Massive Open Online Courses)

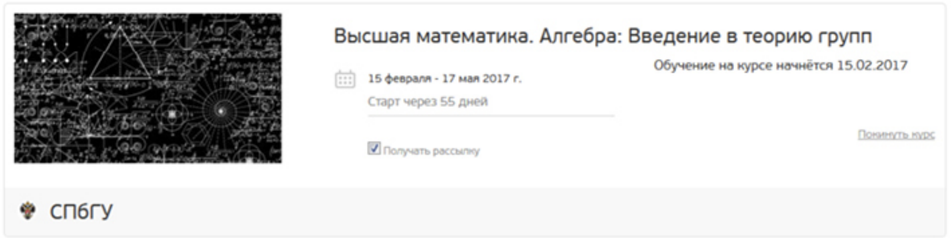
Online education is one of the main trends in the development of a modern education system and the largest of the emerging education services markets. Last year, a new division, the Center of the Development of Electronic Educational Resources, was created at St. Petersburg State University. The main aim of the Center is to provide access to knowledge for everyone interested. The Center has already created about fifty courses in various fields.

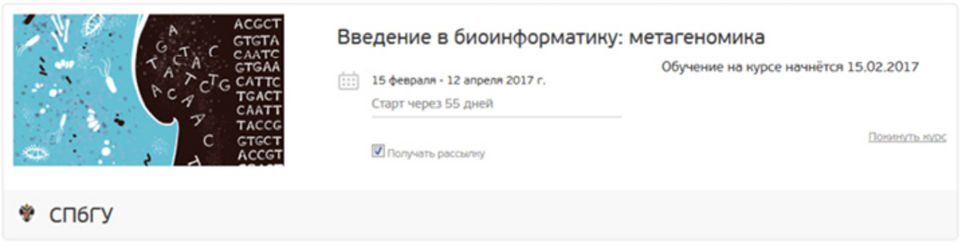


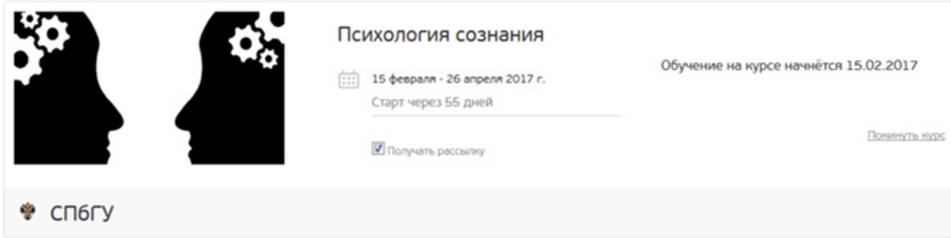
More than 70 MOOC's

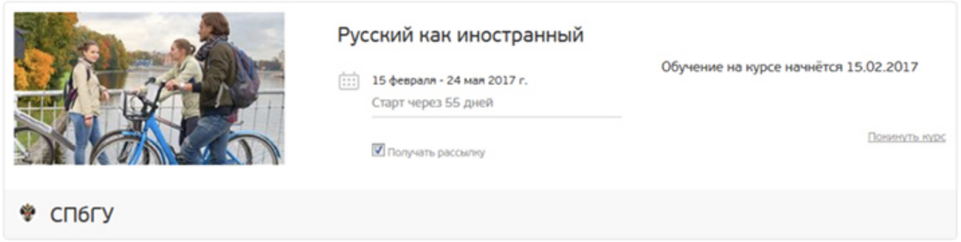






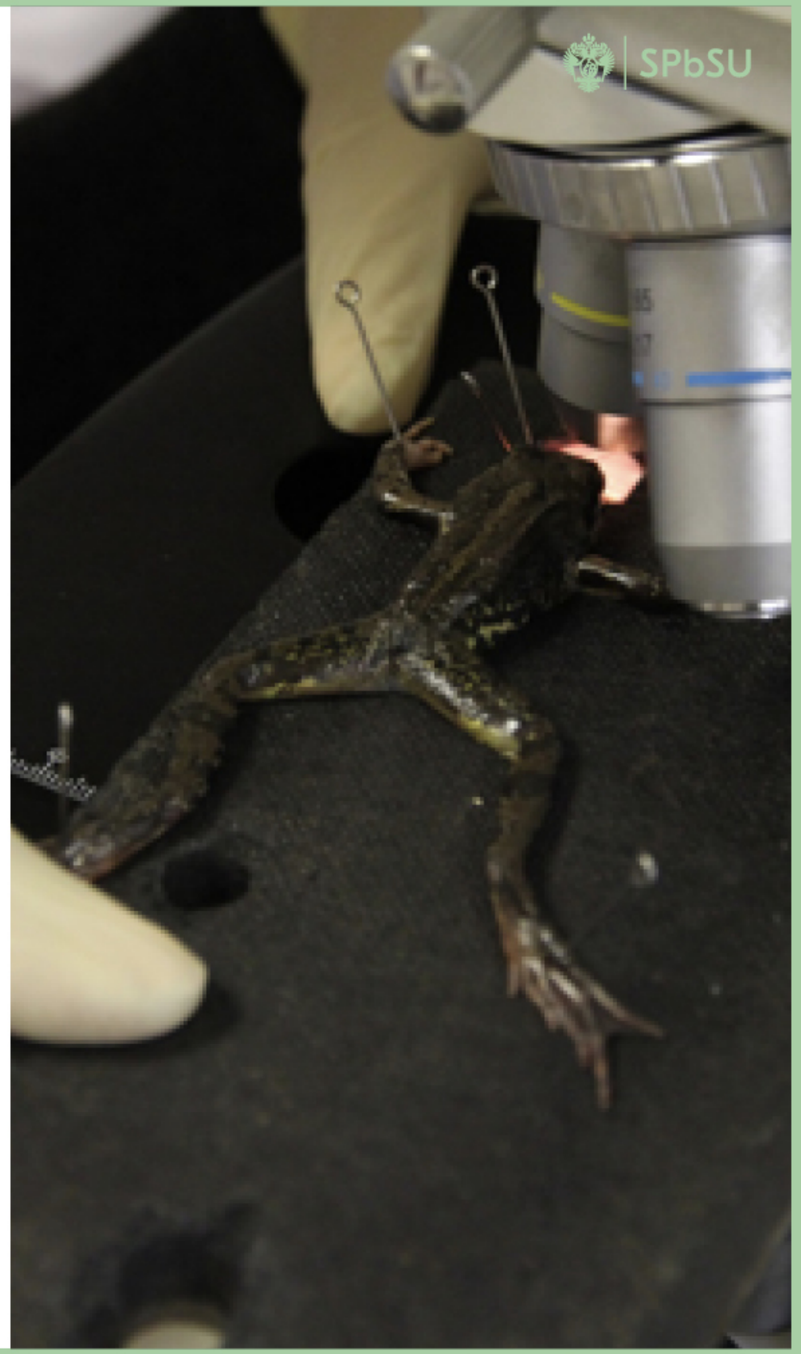






Course "Introduction to Pathophysiology"

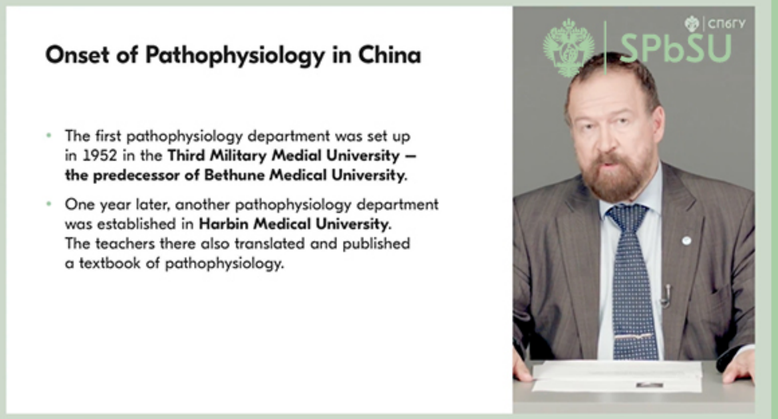
One of the courses was "Introduction to Pathophysiology," which will soon be launched on the educational platform Coursera. This course is intended for medical school students as well as for students majoring in biology and chemistry who would like to learn more about human pathology.

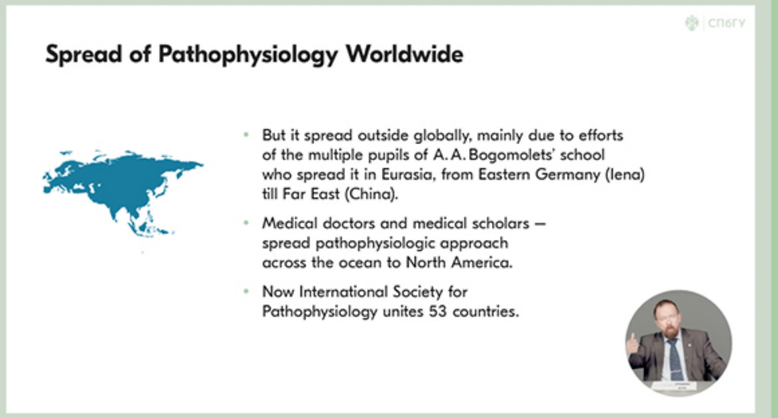


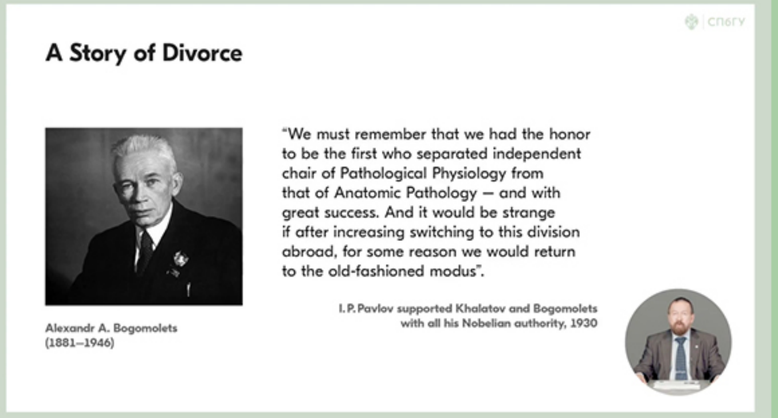
Course "Introduction to Pathophysiology"

Additionally, the course is useful for researchers involved in the development of biomedical projects or people without formal medical education who work in medical research centers or research institutes. It provides basic training in biology, chemistry, and engineering to students who need to know the causes, mechanisms and modeling of diseases and pathological processes.

Our idea was to give access to pathophysiological knowledge to students and self-studying intellectuals everywhere, to support those who needs combined interdisciplinary education and those who strives to improve his competencies in Medicine. It can be used for distant refreshing and postgraduate learning of medical doctors and medical teachers.







Russian-Chinese training practices

Currently, SPbSU plans to offer its online courses to the Chinese market and also to present the best Chinese online training practices to Russian users. This will strengthen mutually beneficial cooperation between our countries in the field of education and innovative educational technologies will drive the sustainable development of Sino-Russian relations

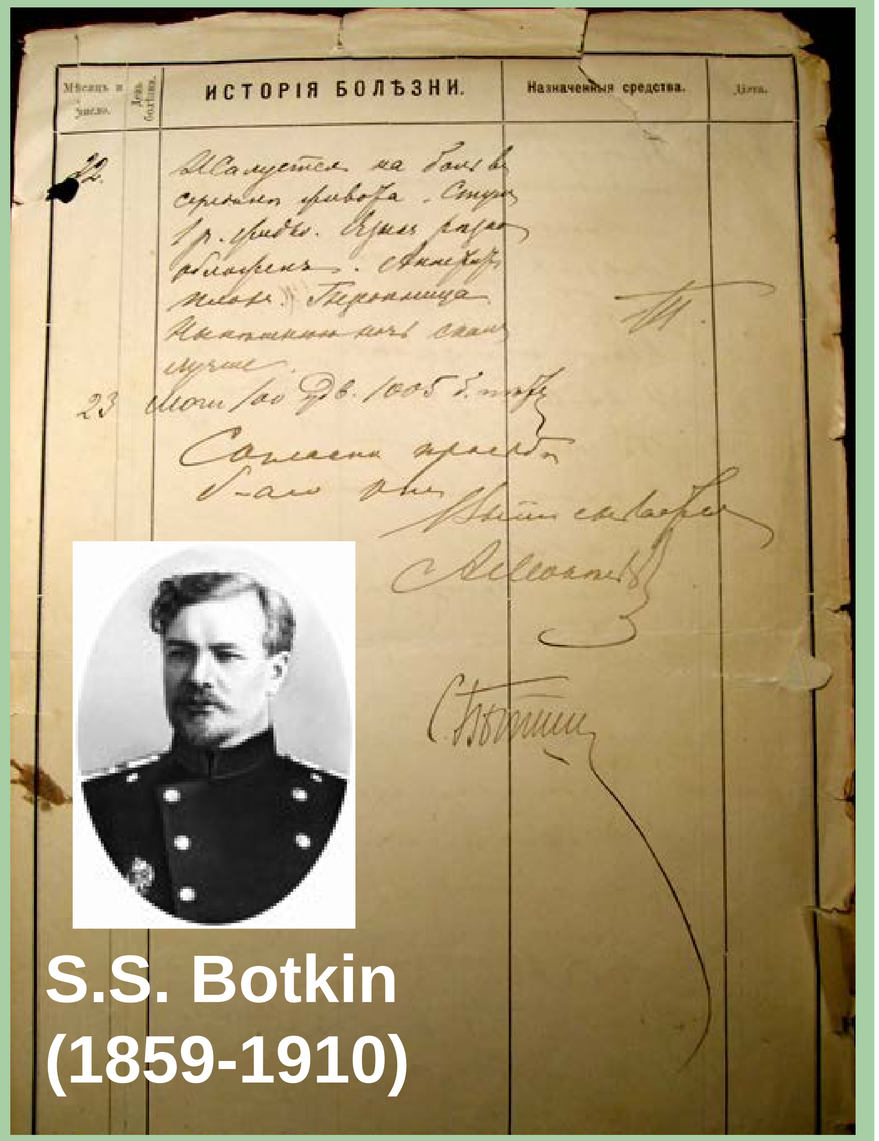
China currently is leading in biomedical research and has more Pathophysiology Depts & Labs, more potential pathophysiological students and professional pathophysiologists - than any other country of the world.

Renown Chinese pathophysiologist **Prof. Chang-Qing Xu** has agreed to work on Chinese subtiters for this Pathophysiology MOOC

Having long fruitful collaboration with Harbin Medical University and Rector Acad. Yang Baofeng, we hope for help of Chinese pathophysiologists in this project.

NOT ONLY SPREAD KNOWLEDGE DISTANTLY – ALSO COLLECT MEDICAL INFO DISTANTLY

Saint Petersburg State University

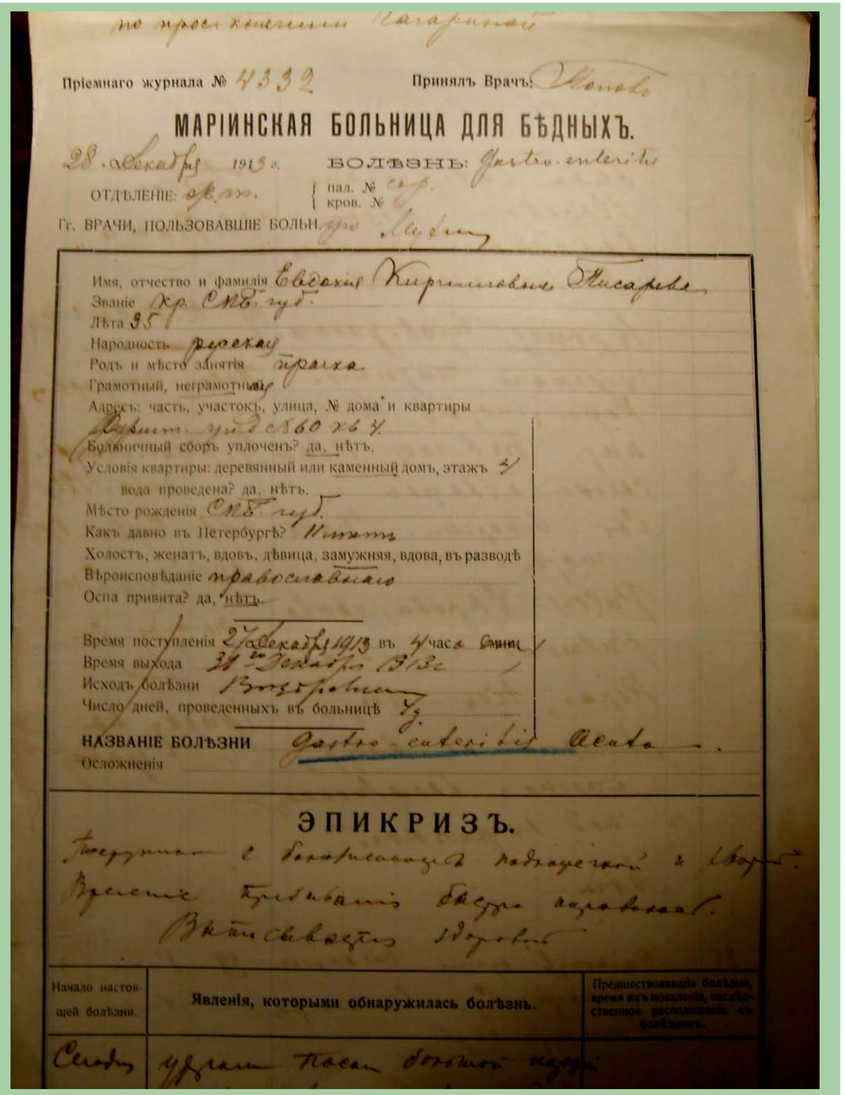


S.S. Botkin (1859-1910)

The main source of pathophysiological knowledge besides lab experiments are clinical observations of physicians.
Case History or Outpatient Registry is a most valuable scientific, practical and judicial document.
Our archive contains world largest database of clinical records from 15000 patients with autoimmune thyroid diseases.
But PC entered into routine health care just 30 years ago.
For 200 years case histories were recorded MANUALLY. Valuable old archives are still not digitalized.
The main obstacle is legendary poorly understandable handwriting of many medical doctors

Handwritten medical records recognition project

Saint Petersburg State University



Пріемнаго журнала № 4332 Принялъ Врачъ:

МАРІИНСКАЯ БОЛЬНИЦА ДЛЯ БѢДНЫХЪ.

28. декабря 1913 г. БОЛѢЗНЬ:

ОТДѢЛЕНІЕ: пал. № кров. №

Гг. ВРАЧИ, ПОЛЬЗОВАВШІЕ БОЛЬН.

Имя, отчество и фамилія Евдокія Кирилловна Тисарева

Званіе Кр. С.-П.б. губ.

Лѣта 35

Народность русская

Родъ и мѣсто занятія прачка

Грамотный, неграмотный

Адресъ: часть, участокъ, улица, № дома и квартиры

Больничный сборъ уплоченъ? да, нѣтъ.

Условія квартиры: деревянный или каменный домъ, этажъ

вода проведена? да, нѣтъ.

Мѣсто рожденія С.-П.б. губ.

Какъ давно въ Петербургѣ? Всегда

Холостъ, женатъ, вдовъ, дѣвица, замужняя, вдова, въ разводѣ

Вѣроисповѣданіе православная

Оспа привита? да, нѣтъ.

Время поступленія 27 декабря 1913 въ 4 часа

Время выхода

Исходъ болѣзни

Число дней, проведенныхъ въ больницѣ

НАЗВАНІЕ БОЛѢЗНИ Gastro-enteritis acuta

Осложненія

ЭПИКРИЗЪ.

Начало настоящей болѣзни.	Явленія, которыми обнаружилась болѣзнь.	Предшествующія болѣзни, время ихъ появленія, насколько состоящія въ связи съ болѣзнью.
Сегодня		

Right: case history of 1913 from our collection;
Left: case history of 1942 written during Siege of Leningrad on backsides of military topographic maps

Motivation

We've got tens of thousands medical records thoroughly accumulated for several decades, all written by hand.

We can't access them for studying without the risk of endangering valuable original handwritten papers.
We can't build indices for faster search and analysis of the data.
We can't build Probabilistic Graphic Models based upon them.

Problems

The recognition task isn't quite straightforward:
•Handwriting quality is low, especially for general purpose recognition software.
•Doctors are concerned about privacy of their patients, so the work couldn't be done as a part of some global community driven digitization initiative such as Project Gutenberg.
•The amount of data is tremendous. The problem can't be cracked in short terms. In long run, people tend to change occupation and the project could lose their individual touch.

Encouraging aspects

On the other hand,

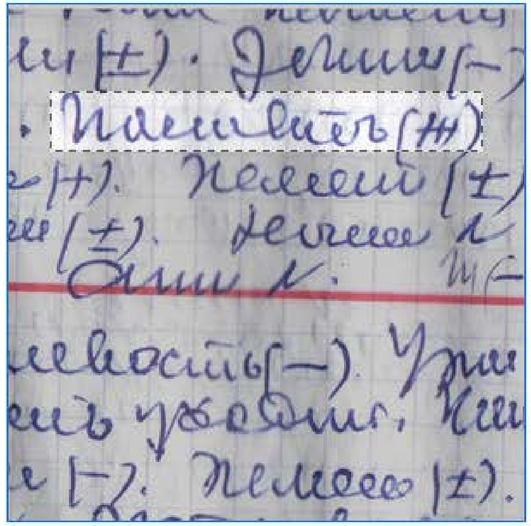
•Text semantic is quite simple. The most important parts are digits, symptoms, and patient's identity.

•Documents have certain format.

Proposed solution - I

The only realistic solution is to perform this work gradually over time with human operator's supervision. With this approach, the data will diverge to three domains:

1) Fully validated.
2) Partially validated, only important parameters have been checked and edited if result of computer vision wasn't precise in each case.
3) Recognized by software, never validated.



Region 12.

x:2098, y:2098, width:575, height:81

Text Потливость (+++)

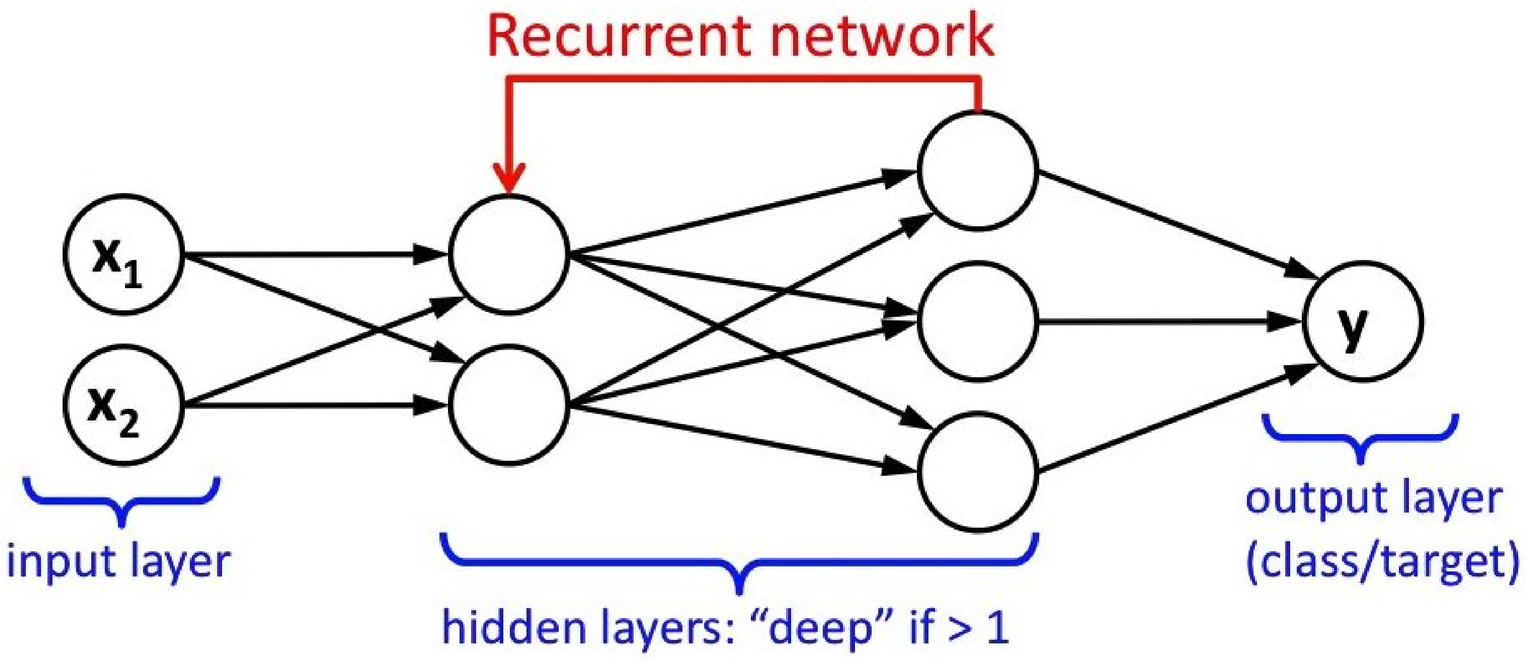
Status
- ○ Created interactively by Human operator
- ○ Created by Machine Learning algorightm / Low estimated probability of error
- ○ Created by Machine Learning algorightm / High estimated probability of error
- ◉ Validity of this data is verified by Human Operator

Tags symptom × + Tag

Because large array of data will inevitably go out of sync and diverge to multiple branches relevant only to particular researcher's current task, it is proposed to store original input bitmaps and recognition results in the same database and access both of them simultaneously with a consistent web interface which would encourage researchers to contribute to this long-term work instead of splitting it to dozens of incompatible pieces.

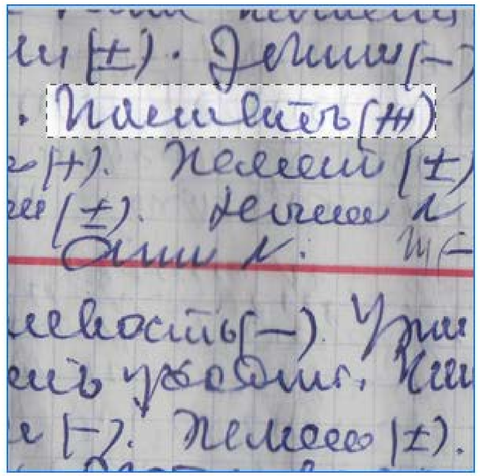
Proposed solution - III

Because input documents have certain format, content of text fragments is dependent on their function within documents. On the other hand, a function of a text fragment could be deduced from text itself. So we have recurrent recognition process here.



Proposed solution - IV

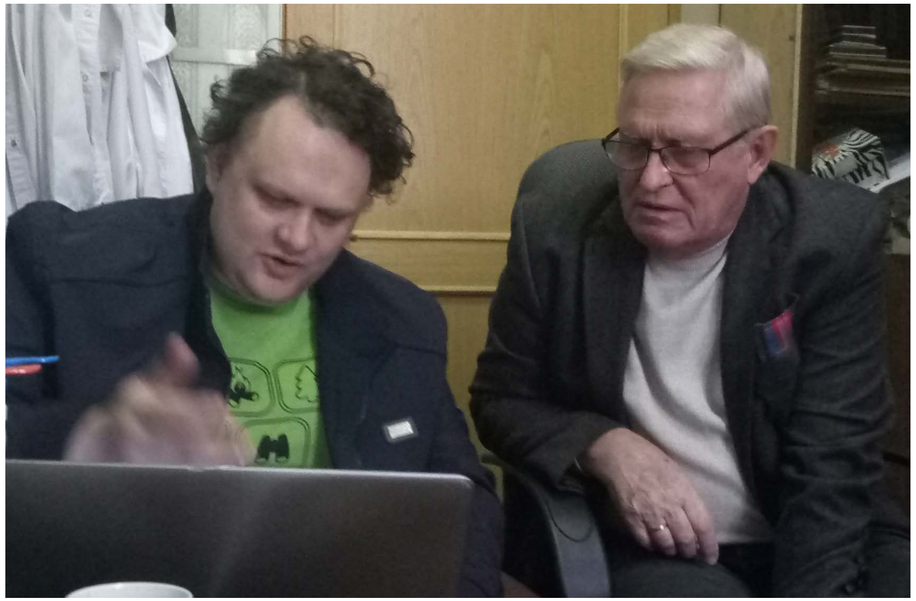
Function of a text fragment is important to software user as well, for example users could make more precise data slices for further data analysis using such metadata**. So, the software should allow to view and edit function of each text fragment using the same interface.**



Region 12.

x:2098, y:2098, width:575, height:81

Text	Потливость (+++)
Status	Verified by Human Operator
Tags	symptom



LET US BRIDGE OUR EFFORTS. WE ARE OPEN FOR COLLABORATION!

Harbin, PRC. Bridge across 松花江 river constructed in 1901 by Russian engineer Adam Lentovsky