Institute of History, Polish Academy of Sciences
Marek Słoń, Mikołaj Słoń, Adam Zapała

Indexing of Tax Registers from 16th Century

TUC, Wien, 8-9.11.2018

Historical Atlas of Poland - history of the project

1880 - general idea: reconstruction of the settlement network and administrative structures in the 16th cent.

1953 - founding of the atlas team and start of the series

2020 - project and series completion

Zaborowo

Krzyczko Wielkie

Krzycko Małe

Jezierzyce

Gołanice

Lgiń

Wilkow

100

120

Zbarzewo

Hetmanice

Piotrowice

80

Niechłód

Tamelsee

Buchwald

SW

Wygnańczyce

Nowa Wieś

Osowa Sień

Trzebinia

100

Długie
Stare

Ogrody

Łysina

Dębowa
Łęka

Lasocic

Tylewice

Długie
Nowe

WSCHOWA

Przyczyna
Wielka

Przyczyna
Mała

Przybyszewo

Tax registers

Central Archives of Historical Records, Archives of State Treasury, division I, vol. 6, p. 484v, vol. 13, p. 139v, vol. 11, p. 1516

Data from tax registers

The **character** of settlement, its name, taxed categories

Data from tax registers

The character of settlement, its **name**, taxed categories

Data from tax registers

The character of settlement, its name, **taxed categories**

Map

GIS
Datbase

Footnotes

Atla

Tabela haseł Tabela zapisek Skany rękopisu Mapa   Wschowska księga gro

Rękopis   ✕

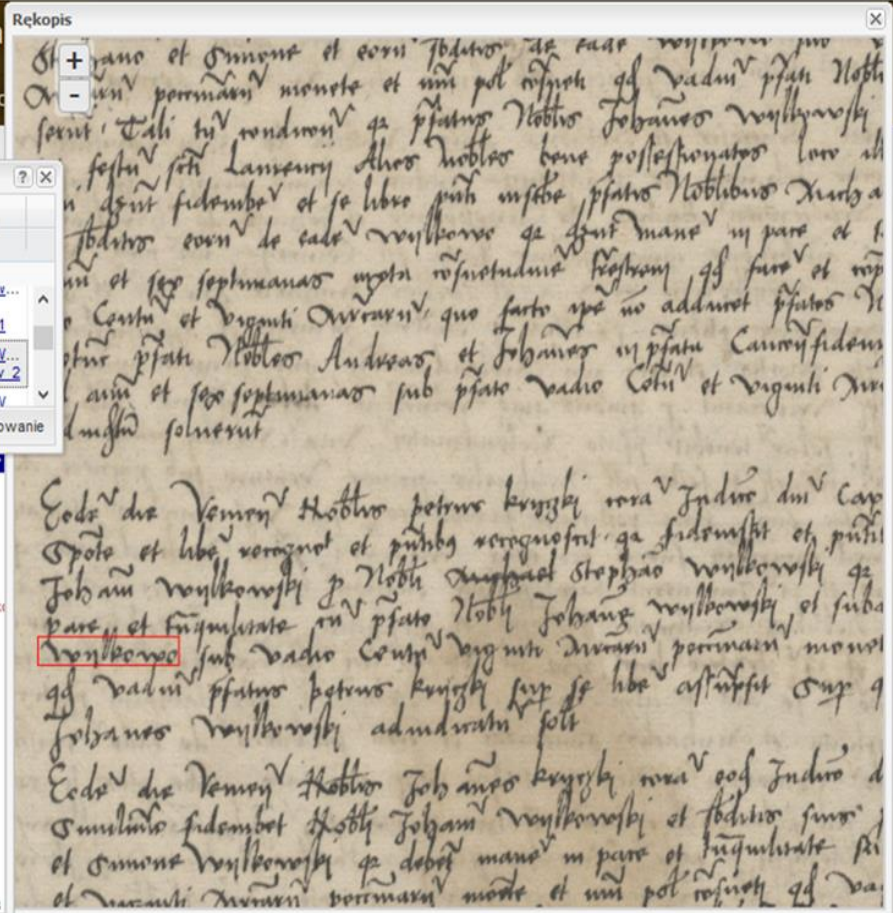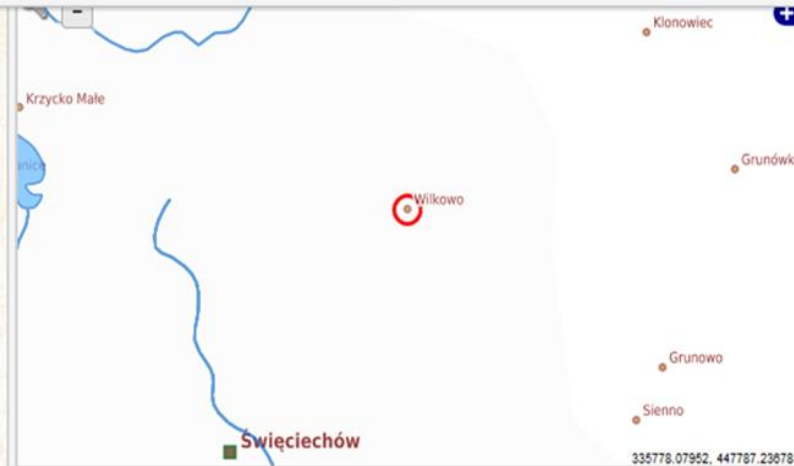**Hasła**   Kliknięcie powiększa (z klawiszem Ctrl tylko przesuwa) skan rękopisu   ? ✕

Id	Osoba - źródło	Osoba - informacje	Osoba	Osoba - Nazwisko	Osada	Osada - źródło	Hasło rzeczo...	Uwagi	SGH	Zapiska
🔍	🔍	🔍	🔍	🔍	🔍	🔍	🔍			
7.3	nob. Casper Rotemburkg Zbarzewski		Kasper Rotemberg Zbarzew...	Rotemberg Zbarzew...						APPzn W... Gr. 1_100v_1
8.1					Wilkowo	Wylkowo				APPzn W... Gr. 1_73v_2
9.1	honst	mąż	Jadwiga	Ulrich				nazwisko		APPzn W

◀| ◀ Page 1 of 103 ▶ ▶| ⟳   Usuń filtrowanie

Klonowiec

Krzycko Małe

Grunówk

Wilkowo

Grunowo

Sienno

Święciechów

335778.07952, 447787.23678

c.a. 25 000 settlements

c.a. 70 000 pages of manuscripts

hundreds of different hands and writing styles

c.a. 250 000 headlines with settlements names

Material used in the model

Characteristics

Changing of the hands in one registers volume

Variation of the language (polish, latin, polish/latin)
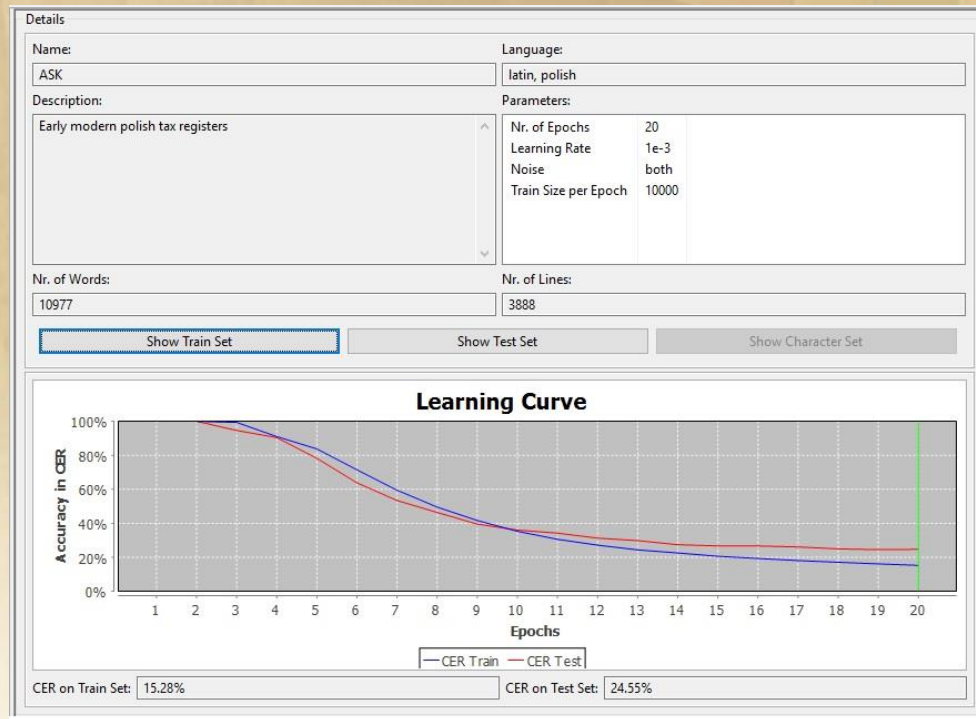
Irregular abbreviations

Different styles of writing in the headlines and in the main text

HTR(CITlab) Model

Nr. of Words: 10 977

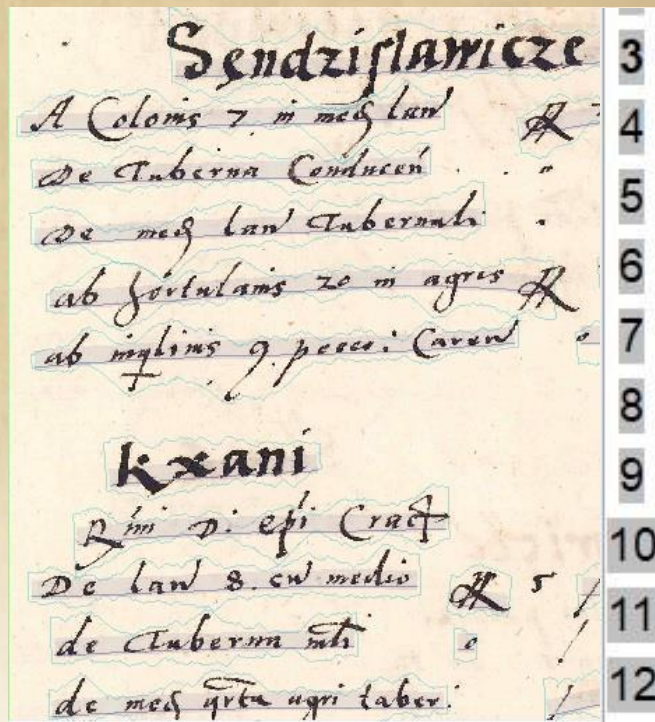CER on Train Set: 15.28%

CER on Test Set: 24.55%

An example of the automatic text recognition

Proper recognition of frequently used words and abbreviations

Worse recognition of rarely used words and abbreviations

Very bad recognition of headlines written with different style.
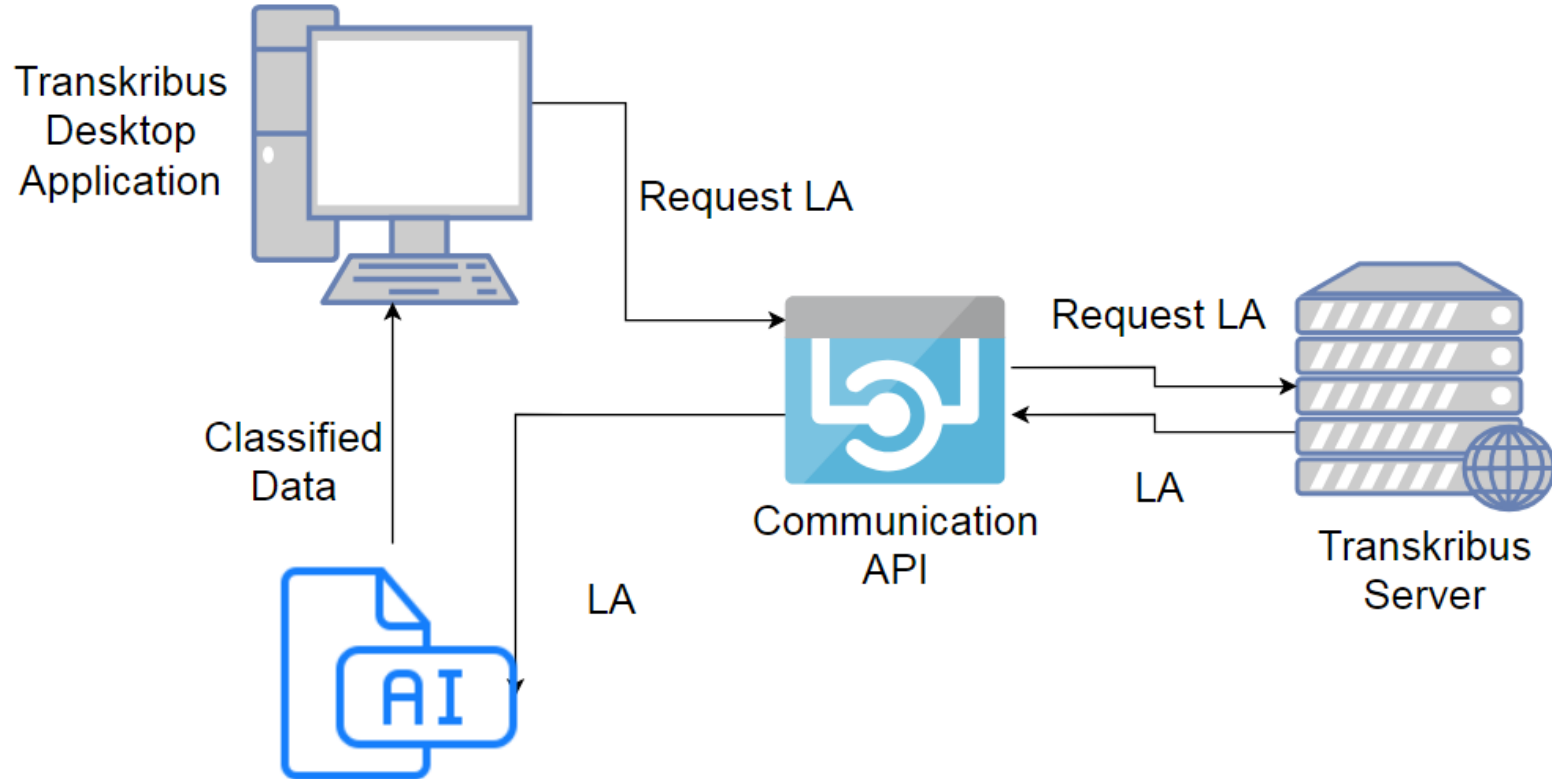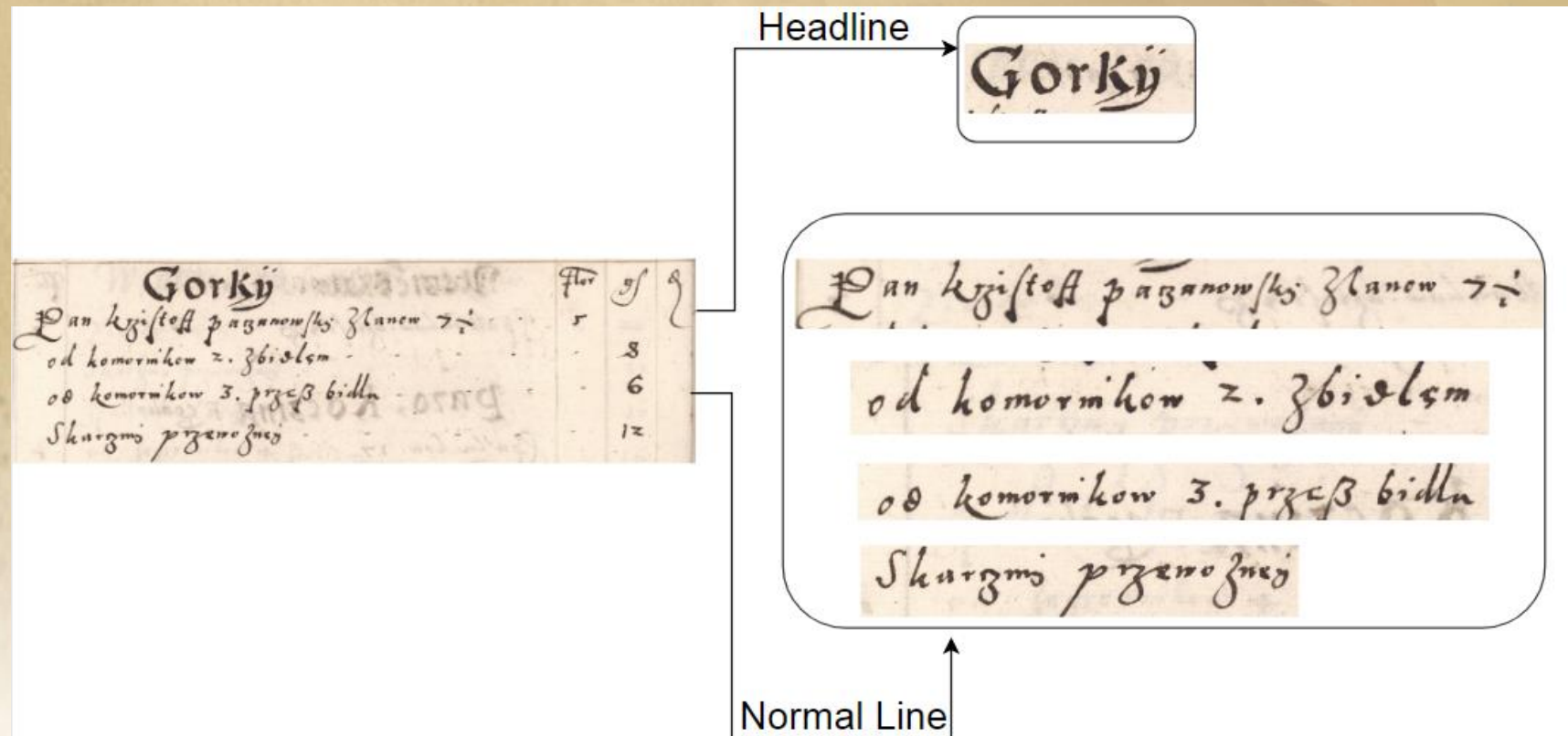
Classification tool for headlines

Classification tool for headlines

Classification tool for headlines

Headline

Gorky

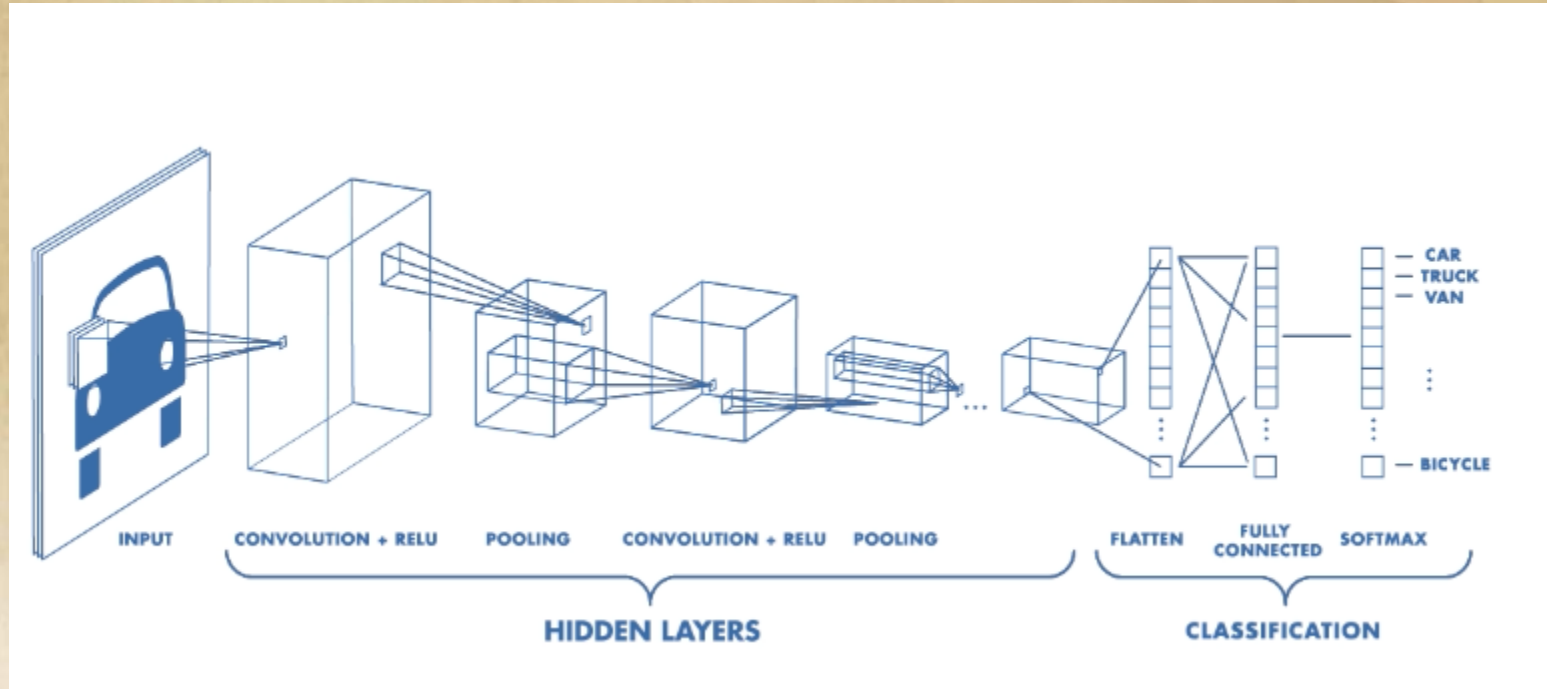Normal Line
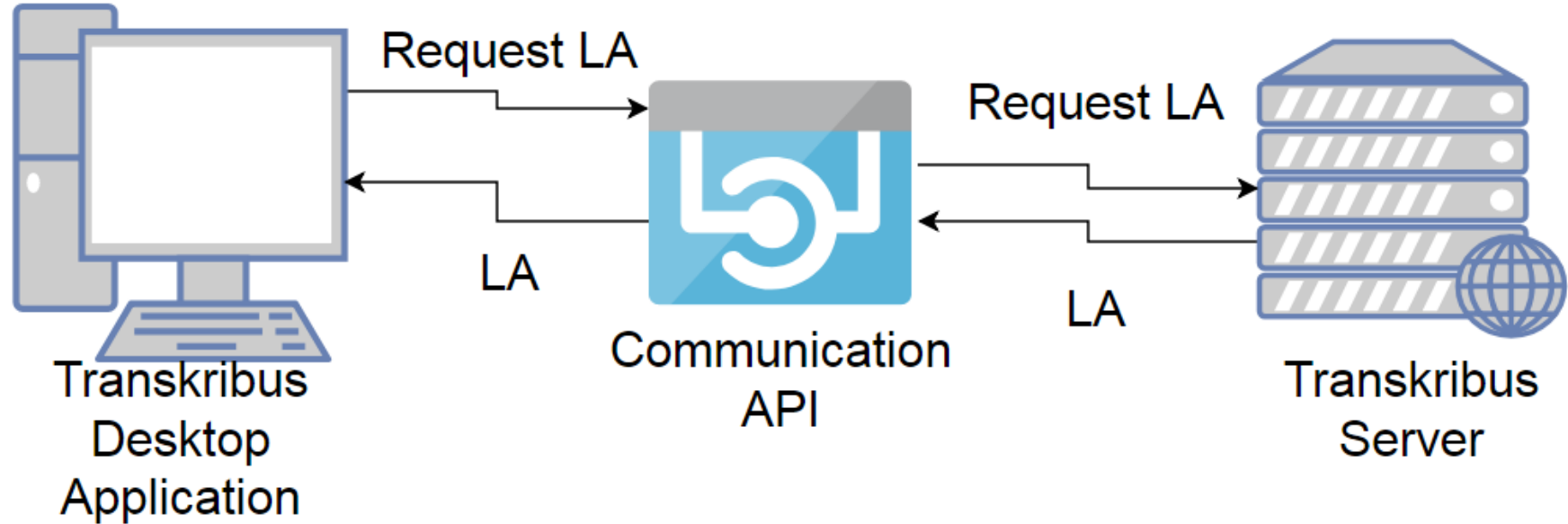
Tool selection - Convolutional Neural Network

Image source: https://medium.freecodecamp.org

Challenges - Image normalization

Institute of History, Polish Academy of Sciences
Marek Słoń, Mikołaj Słoń, Adam Zapała

Indexing of Tax Registers from 16th Century

www.ihpan.edu.pl
www.atlasfontium.pl

TUC, Wien, 8-9.11.2018