AGEPI AGENȚIA DE STAT PENTRU PROPRIETATEA INTELECTUALĂ A REPUBLICII MOLDOVA

INFOINVENT

18th International Specialized Exhibition
22-24 November 2023
www.infoinvent.md

MINISTERUL EDUCAȚIEI ȘI CERCETĂRII · MINISTERUL CULTURII · ANCD · WIPO WORLD INTELLECTUAL PROPERTY ORGANIZATION · Europäisches Patentamt European Patent Office Office européen des brevets

HeDy – A PLATFORM FOR HETEROGENEOUS DOCUMENTS PROCESSING

Authors: T. Bumbu, L. Burtseva, S. Cojocaru, A. Colesnicov, L. Malahov

Institution: Vladimir Andrunachievici Institute of Mathematics and Computer Science of USM

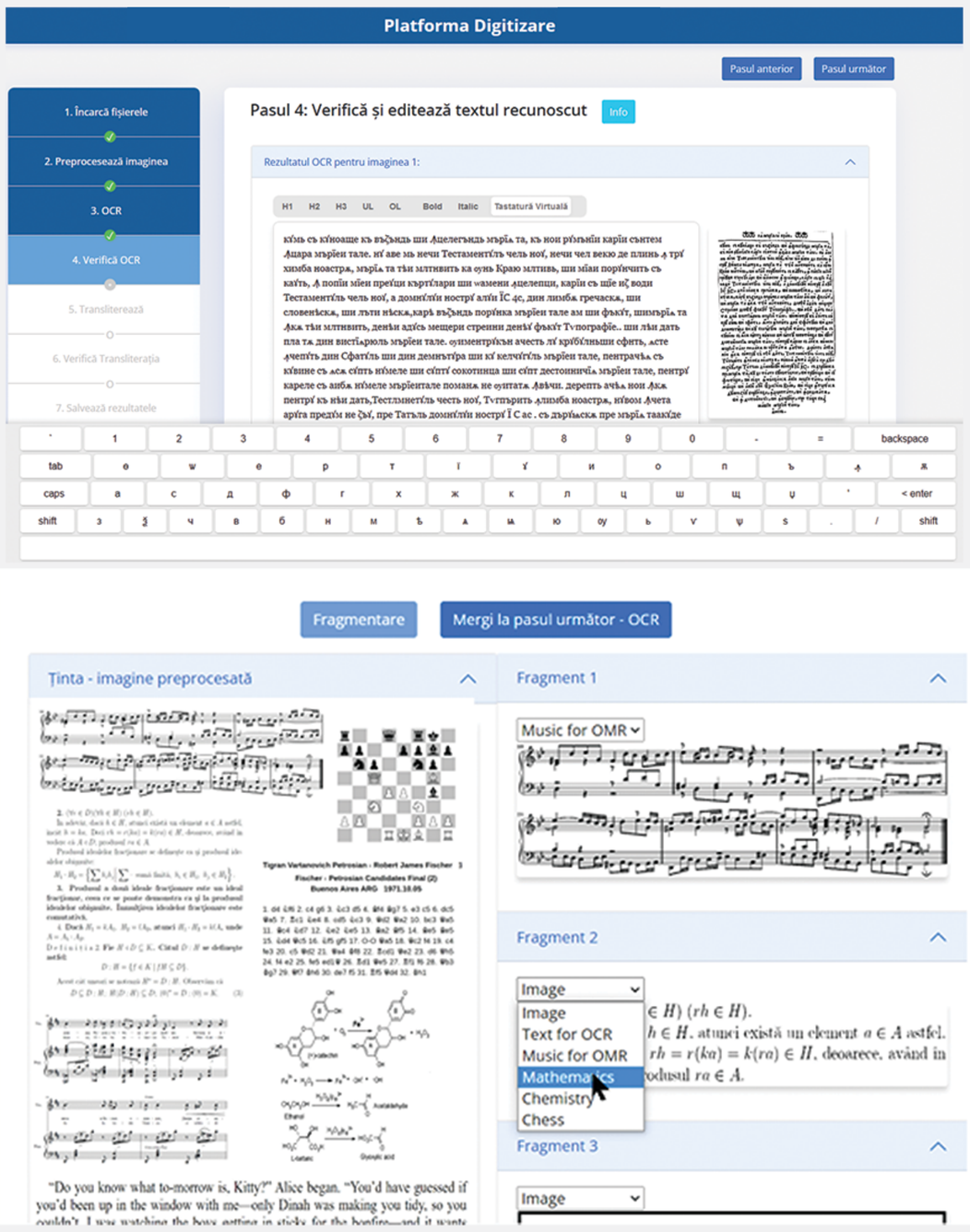
ABOUT

The HeDy platform implements the "one-stop-shop" concept for digitization and transliteration of printed documents, including historical ones. The platform integrates several OCR models, targeting different time periods and font classes. The font classification algorithm is based on a neural network having 96% accuracy. Transliteration from Cyrillic to modern Latin is based on sets of rules (free and context-dependent) achieving 98% correctness. In addition to the text, the digitization platform also processes documents containing other elements such as formulas, images, diagrams, etc. It can be accessed online: https://digitizare.math.md/

TECHNICAL IMPORTANCE

The Web platform makes an important contribution to the revitalization and valorization of the national literary-historical heritage, offering users - researchers in various fields, publishers, librarians, archivists, etc. - an efficient technology for automating the processing of old and contemporary documents, including heterogeneous ones, into editable formats. The tool is also useful to the general public by facilitating access to old and rare editions. The proposed technology has already been used in the re-editing (or preparation for re-editing) of several books, articles, and archival documents, ranging from scientific literature in various fields to fiction, periodicals, etc.s a platform for new NLP applications.

USER INTERFACE



EXAMPLE: REPUBLISHING OF A MATHEMATICAL BOOK

1979 2017

Vladimir Andrunachievici Institute
of Mathematics and Computer Science, USM
str. Academiei 5, tel. 022 72 59 82
www.math.md

Contact: tudor.bumbu@math.md

Note: The application was developed within the project 20.80009.5007.22