iPBL – supplementing literary bibliography with internet sources. Collaborative cataloguing

Koper, Beata

beata.koper@uni.opole.pl
University of Opole, Poland

Rosiński, Cezary

cezary.rosinski@ibl.waw.pl
Institute of Literary Research of the Polish Academy of Sciences, Poland

Wachek, Barbara

barbara.wachek@ibl.waw.pl
Institute of Literary Research of the Polish Academy of Sciences, Poland

Maryl, Maciej

maciej.maryl@ibl.waw.pl
Institute of Literary Research of the Polish Academy of Sciences, Poland

Umerle, Tomasz

tomasz.umerle@ibl.waw.pl
Institute of Literary Research of the Polish Academy of Sciences, Poland

This poster reports on the ongoing R&D project *iPBL*[1] aimed at expanding the source catalogue of Polish Literary Bibliography (PBL) and supplementing it with information about online materials such as electronic literature, literary blogs, websites devoted to the Polish culture, and social media profiles. Literary bibliographies focus mostly on printed and digitised materials (Umerle et al. 2022), thus adding online sources poses a two-fold challenge:

1. **Methodological challenge -** how to select and describe data on the internet culture and reconcile it with the records on print culture within the same database
2. **Technological challenge -** how to obtain metadata of electronic texts and secure their accessibility for future researchers.

Expanding the database

The Polish Literary Bibliography is a large database containing literature, theatre, and film information. The PBL contains approx. 880 thousand bibliographic records in a database form, and approx. 1,5 mln records in a form of scanned pages of bibliography. It covers the period from 1944 to 2012 (the database covers the years 1989-2012). The collected data includes more than 62000 authors, 12200 literary works, nearly 19000 cultural events, more than 25000 theatrical and radio plays, literary programmes and 2000 films. PBL data are aggregated with national bibliographies in a European Literary Bibliography.

The selective and analytical bibliographic database will include approx. 150.000 records from 200 Internet sources, of which 40.000 will be enriched with autopsy descriptions, such as subject headings, genres, and annotations. The digital index will include approx. 4000 data sources. The scope of the database will cover nationwide cultural websites and services, online magazines, blogosphere, and literary works published on the Internet. The revolution we propose is to make available, for any interested party, an even more detailed database, which is highly accessible, free and compatible with international standards.

Workflow

Creating the bibliography of internet sources challenges traditional documentation practices, as the workflow for collecting, curating, and sharing data that will consider, from the very beginning, the human and non-human stakeholders. The project's workflow (Fig. 1.) consists of the following operations:

1. **Sources selection**, based on the analysis of the available sources: the National Library of Poland's Internet Resources Database and Bibliographic Guide, the ISSN portal, and the Internet Archive.
2. **Automated acquisition.** The information will be obtained due to dedicated scraping algorithms and then converted into the internationally approved bibliographic formats (MARC21, Bibtex, RIS); the resulting bibliographic resource will be enriched with Linked Open Data structures (references to international controlled vocabularies and authority databases, e.g. including VIAF, LCSH) and automatically classified using NLP.
3. **Optional manual adjustment.** If the results of the automated process are unsatisfactory, the records will be subject to manual adjustment for the genre, subject heading and content annotation.
4. **Archiving the source in a digital repository.** The sources will be automatically submitted to the Web Archive to ensure their future availability.
5. **Publishing metadata record in the database.** Final record will be incorporated into the PBL database.
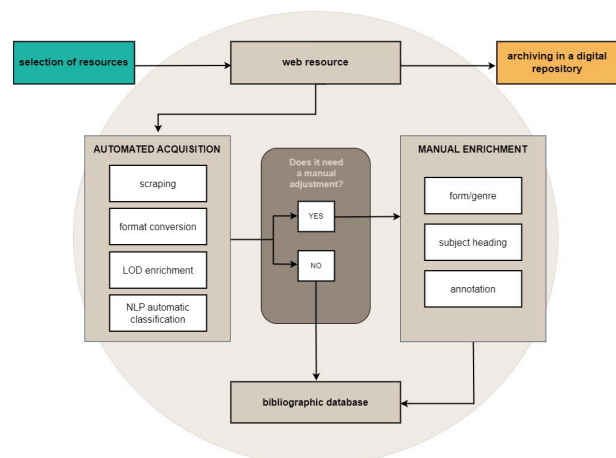
iPBL workflow for managing internet sources

Research potential

New internet data will enable research into Polish digital culture, and even more important, save it from oblivion. There are known cases of complete disappearance of cultural content, 20% of cited web resources become inaccessible (Klein et al.), which makes it impossible to study (Kozioł 2015; Buchwald et al. 2015; Maryl 2016). It is necessary to record data to allow any future research at all (Jones 2001) and the broaden PBL database will be the point where any academic collaboration may begin.

Digital culture is a part of the Polish cultural heritage. This heritage is largely made up of documents created in a digital form and functioning primarily on the Internet: on websites, portals, online magazines, blogosphere, and fora, as well as on websites or social profiles of institutions and people of culture. Recording data from online sources is thus a natural step in the PBL database evolution.

Notes

1. Bibliography of the Polish digital culture online with a catalogue of sources and an archive. Supplement to Polish Literary Bibliography, National Programme for Development of the Humanities, 2023-2025, grant number: NPRH/DN/SP/495736/2021/10.

Bibliography

Buchwald, Dorota / **Żurawski, Mateusz** / **Dziewulska, Małgorzata** / **Soszyński, Paweł** (2015) "Kolekcja tea-tralna: wysoka śmiertelność teatru", in: *Dwutygodnik.com* 12 https://www.dwutygodnik.com/artykul/2221-kolekcja-tea-tralna-wysoka-smiertelnosc-teatru.html [28.04.2023].

Jones, Steve (2001): *Doing Internet research: Critical issues and methods for examining the Net*, Thousand Oaks: Sage Publications.

Klein, Martin / **Sompel Van de, Herbert** / **Sanderson, Robert** / **Shankar, Harihar** / **Balakireva, Lyudmila** / **Zhou, Ke** / **Tobin, Richard** (2014): "Scholarly context not found: One in five articles suffers from reference rot", in: *PloS one* 9.

Kozioł, Paweł (2015): "Lata fermentu", in: *Dwutygod-nik.com* 4 https://www.dwutygodnik.com/artykul/5851-lata-fermentu.html [28.04.2023].

Maryl, Maciej (2016): "Startup literacki. Blog a powieść w odcinkach na przykładzie pierwowzoru *Cwaniar* Sylwii Chutnik", in: Maryl, Maciej / Dąbrówka, Andrzej / Wójtowicz, Aleksandra (eds.): *Teksty kultury uczestnictwa*, Warszawa: Instytut Badań Literackich PAN.

Umerle, Tomasz / **Colavizza, Giovanni** / **Herden, Elżbieta** / **Jagersma, Rindert** / **Király, Péter** / **Koper, Beata** / **Lahti, Leo** / **Lindemann, David** / **Łubocki, Jakub Maciej** / **Malínek, Vojtěch** / **Milanova, Alexandra** / **Péter, Róbert** / **Rißler-Pipka, Nanette** / **Romanello, Matteo** / **Roszkowski, Marcin** / **Siwecka, Dorota** / **Tolonen, Mikko** / **Vimr, řej** (2022): *An Analysis of the Current Bibliographical Data Landscape in the Humanities. A Case for the Joint Bibliodata Agendas of Public Stakeholders*. Zenodo. DOI: 10.5281/zenodo.6559857.