

From: **impresso project** info@impresso-project.ch
Subject: [Test] Indian Summer Release of the impresso interface
Date: 7 September 2020 at 20:59
To: info@impresso-project.ch



Now online: The Indian Summer Release of the *impresso* interface

Media Monitoring of the Past
Search through 200 years of newspapers

search for ...



REQUEST FULL ACCESS TO THE CORPUS

As announced in our previous email, the Indian Summer Release of the impresso interface is now available for testing. This has been a major upgrade on both front- and backend so not everything is running 100% smoothly yet:

- searching for similar images via file upload is temporarily disabled
- article titles for the Neue Zürcher Zeitung are not currently displayed in the interface

We are working hard on resolving these problems and will keep you updated.

The interface is available for testing here: <https://impresso-project.ch/app/>

Integration of Luxembourgish newspapers

FILTER BY NEWSPAPER TITLES

check one or more newspaper to filter results

- Luxemburger Wort (1,129,183 results)
- L'indépendance luxembourgeoise (733,601 results)
- Obermosel-Zeitung (551,764 results)
- Escher Tageblatt (281,040 results)
- d'Letzeburger Land (263,898 results)
- Luxembourg (1935) (141,396 results)
- Bürger- und Beamten-Zeitung (86,373 results)
- L'Union (77,130 results)
- Courrier du Grand-Duché de Luxembourg (71,292 results)
- D'Unio'n (63,774 results)
- Der Wächter an der Sauer (35,979 results)
- Volksfreund (33,772 results)
- L'avenir (33,179 results)
- De Letzeburger (28,100 results)

- De Letzebuerger (20,103 results)
- Arme Teufel (20,365 results)
- D'Wäschfra (14,546 results)
- Luxemburger Zeitung - Journal de Luxembourg (13,349 results)
- Diekircher Wochenblatt (7,475 results)
- De Mitock e Wocheblad fir Jux an Zodi (5,925 results)
- Ons Jongen (5,552 results)

We are very happy to report a significant expansion of our corpus: We can finally give access to large parts of newspaper collection of the **National Library of Luxembourg**. You can now use **topics, named entities and word embeddings** alongside the keyword search to explore newspapers from both countries.

Please note: We are currently experiencing problems with the newspaper *Luxemburger Wort* and can not guarantee that all articles saved to collections will be stored permanently. We will resolve this problem in the near future and keep you updated.

Word Embeddings

The screenshot shows a web application interface with a modal dialog box titled "Find similar words in corpus". The dialog has a search bar containing the word "nation" and a "FETCH !" button. To the right of the search bar is a language dropdown menu currently set to "FRENCH". Below the search bar, a list of similar words is displayed in a grid-like format:

- nation
- nalion
- émocratie
- patrie
- monarchie
- indivisible
- plutocratie
- nation-
- civilisée
- nationalité'

Word embeddings can help to **find relevant keywords** for a search. Simply enter a first keyword and explore the list of similar terms based on the impresso newspaper corpus.

The suggestions will likely include **spelling variations and OCR mistakes** which can be added to a query to retrieve more articles.

Introducing Named Entities

CURRENT SELECTION **CURRENT SEARCH**

Joe Strummer PERSON

1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012

apply current search filters
32 results from 1987 to 2012

ADD TO CURRENT SEARCH FILTERS

Joe Strummer (Q310052)
HUMAN Ankara, 1952 - Broomfield, 2002

British musician, singer, actor and songwriter
SOURCE: WIKIDATA /Q310052

MORE...

ALL PEOPLE MENTIONED ▾

Search one of 269 available choices

269 results

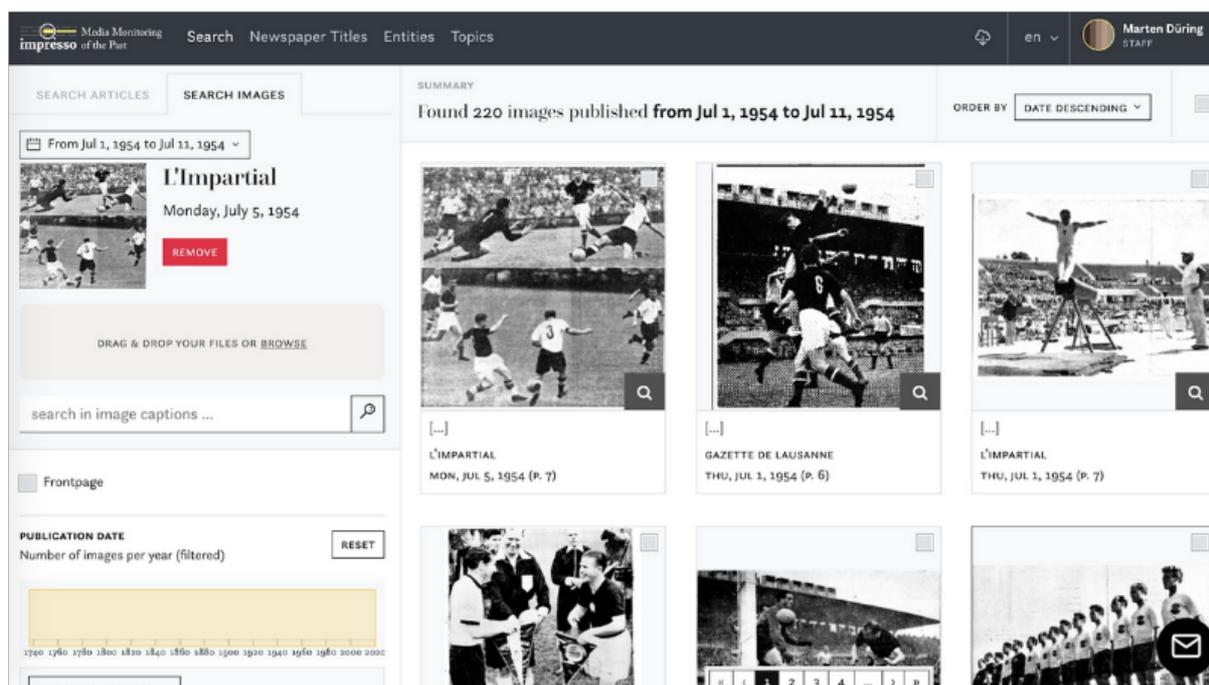
- Joe Strummer (32)
- Jim Jarmusch (8)
- Tom Waits (5)
- James Brown (4)
- Screamin' Jay Hawkins (4)
- Jane Birkin (3)
- Mick Jones (The Clash) (3)
- Nicoletta Braschi (3)
- Paul Simonon (3)
- Ray Gange (3)

« < 1 2 3 4 ... > »

We have added the first experimental version of **linked named entities** to the interface. This means that you can for example **retrieve all articles** in which the musician Joe Strummer has been detected and see with whom he was mentioned.

Please note: Named entities have been detected and linked automatically which is not always easy when dealing with historical newspapers so not all links will be accurate. Quality will improve in future releases.

New ways to search for images in the corpus



We have **expanded** the corpus of images significantly: You can now search for similar images across selected Swiss and Luxembourgish newspapers and **filter your search by newspaper**. We are working on making the full corpus of newspapers searchable.

We now offer three ways to search for similar images:

- **Use dates or newspapers as filters**, then select an image and retrieve similar ones
- **Upload** an image from your computer and see if it appears in the corpus
- **Search image captions** for keywords (experimental)

Do try it out and let us know how it works.

Frequently Asked Questions

FAQ

The corpus

Which newspapers do you have in your corpus?

Why can't I see everything? How do I get access to the full corpus?

Can I download everything? Do you have an API?

The impresso interface has the ambition to put **Natural Language Processing technologies** to good use for historians but also to provide them with the information they need to **critically assess** the value of the corpus they are working on.

We launched a first version of **Frequently Asked Questions** to give you some background information on the impresso interface and how it works.

Deadline approaching: CfP Workshop "Digitised newspapers"



This is to remind you that you have until 15th September to submit your abstract for the upcoming workshop on Digitised Newspapers. You will find all details [here](#).

Use your credentials to login and explore the new release. If you would like to gain access to the impresso interface, email us at info@impresso-project.ch

We are curious to hear from you

We are keen to explore with you how impresso's advanced search and discovery features work and to see how they relate to your own project. Do send us an email to info@impresso-project.ch and we will be happy to follow up!

Outlook: Future releases

Upcoming releases will include the following features and additions:

- **search for recurrent text segments throughout the corpus** to trace how e.g. a press release has been altered and how it spread over time and across newspapers
- **improvements in the quality of character recognition** depending on the newspaper font style (Gothic/Antiqua). This in turn will improve the quality of our text processing and yield for example better keyword searches, named entity recognition and finally more legible text

- **information on the quality of entity recognition** - the models used to extract entities will be evaluated against a ground truth which will cover several newspapers and 20 time periods. We will attempt to add confidence scores for each extracted entity mention in the interface
- a tool for the **visual comparison of collections** which will help you to detect where two collections overlap and differ
- a **recommendation service** will be available which will suggest articles with similar content
- tools which allow you to **track semantic shifts of words over time** by taking into account how their word-neighbours change

Following the 2019 autumn release we will begin working on custom visualisation case studies. Here we seek to support scholars with their projects by providing customised data visualisations which match their specific research interests. [Do get in touch](#) if you are interested in this.



Copyright © 2020 impresso project, All rights reserved.
You are receiving this email because you opted in via our website.

Our mailing address is:
impresso project
EPFL CDH DHI DHLAB
INN 140 (Bâtiment INN) / Station 14
Lausanne CH-1015
Switzerland

Add us to your address book

Want to change how you receive these emails?
You can update your preferences or unsubscribe from this list.

