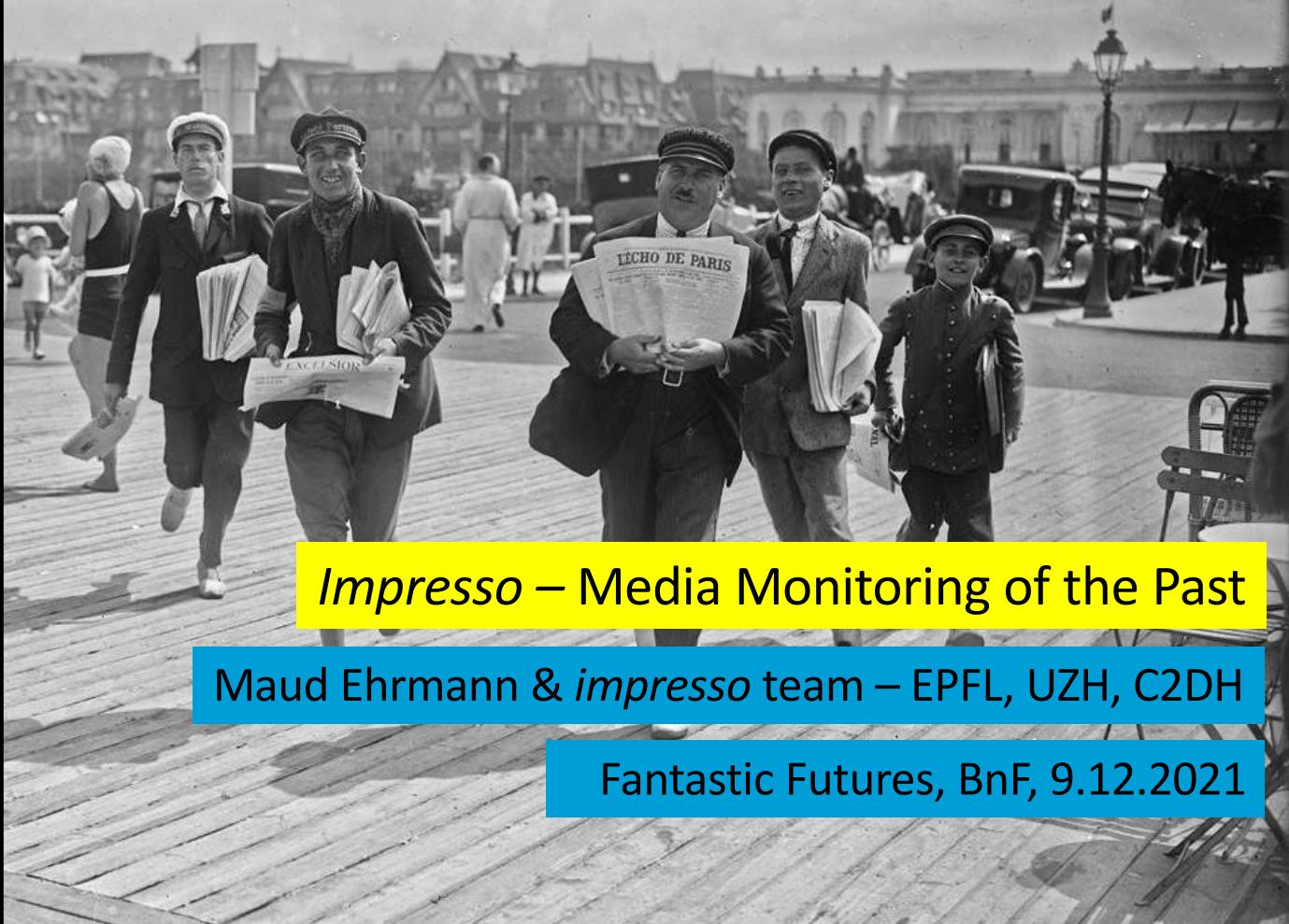




Media
Monitoring
of the Past



Impresso – Media Monitoring of the Past

Maud Ehrmann & *impresso* team – EPFL, UZH, C2DH

Fantastic Futures, BnF, 9.12.2021

NEWSPAPERS





Collaborative work with (in alph. order): Estelle Bunout, Simon Clematide, Marten Duering, Maud Ehrmann, Andreas Fickers, Daniele Guido, Frédéric Kaplan, Peter Makarov, Matteo Romanello, Gerold Schneider, Paul Schroeder, Benoit Seguin, Phillip Stroëbel, Martin Volk, Thijs van Beek, Lars Wieneke.



Parce que le cumul du mandat législatif avec les fonctions publiques, en concentrant tous les pouvoirs en quelques mains, contribue à former un caste privilégiée et crée ainsi un danger pour la démocratie;

Enfin, parce que le principe des incompatibilités, consacré depuis plus demi-siècle par la Constitution fédérale, est en vigueur également dans la plupart de nos cantons confédérés et qu'il fonctionne surtout à la satisfaction générale, pour ne pas parler ici des Etats étrangers qui, les uns après les autres, l'ont introduit dans leurs lois.

Ces Constituantes et ces Confédérés!

C'est une question de parti politique. C'est une simple question de bons sens et de moralité politique. Le proverbe a raison :

On ne peut pas être à la fois au four et au moulins!

Il faut choisir. Le fonctionnaire *salarié par l'Etat*, c'est à dire *par nous tous, pauvres contribuables que nous sommes*, doit son temps & ses fonctions et le magistrat, qui rend la justice, doit rester à l'écart des luttes des partis.

Un juge, homme politique, risque toujours de provoquer la suspicion des justiciables. Or, il faut que le juge, comme la femme de César, ne puisse pas même être soupçonné.

Vous voterez donc le projet de loi constitutionnelle sur les incompatibilités et vous repousserez l'arrêté du Grand Conseil qui l'écarte.

Tous attendent.

Oui pour les incompatibilités.

Non pour l'arrêté du Grand Conseil.

Vive la Suisse! Vive Genève!

Au nom de l'Association démocratique des Eaux-Rives, (Suivent 75 signatures d'électeurs de la commune.)

Les maisons du peuple en Belgique. — C'est le titre de la causerie que M. Emile Vandervelde et Mme Vandervelde feront samedi 30 mars, à cinq heures après midi, au casino de St-Pierre.

Kekpek? — A la demande générale, il sera donné dimanche 31 mars, à huit heures et demie du soir, dans la salle des fêtes de l'institution, une répétition dernière représentation de la tragédie de l'Idiot Borodine, dont le succès s'affirme de plus en plus. Il y aura foule dimanche aux Amis de l'instruction.

Le spectacle commencera par le *Commissaire est bon enfant*, la désopilante comédie de G. Courteille et Lévy.

RÉUNIONS.-CONVOCATIONS.-CONCERTS

Samedi 30 mars

Panorama international, rue du Rhône, 41. Amérique du Sud : le Pérou.

5 h. Casino de St-Pierre. Causerie de M. et Mme Emile Vandervelde, sur les maisons du peuple en Belgique.

8 h. 8 a. de l'Ecole d'horlogerie. Assemblée publique de la Société des anciens élèves de l'Ecole des mines de Dusseldorf sur les cheminées de fer. — 8 h. 15 a. à l'Amphithéâtre de l'Université. — 8 h. 30 a. à l'Amphithéâtre de l'Université. — 8 h. 45 a. à l'Amphithéâtre de l'Université. — 8 h. 45 a. à l'Amphithéâtre de l'Université.

8 h. 45 a. à l'Amphithéâtre de l'Université. — 8 h. 45 a. à l'Amphithéâtre de l'Université.

9 h. 8 a. Cafè Frey, rue de Rive, 7. Assemblée générale de la Société de tir de campagne « le Toli ».

9 h. 8 a. Salons Bohrer, Terrassière. Bal de la Section fédérale de gymnastique de Plainpalais.

M. Berthoud fait observer, dans le même ordre d'idées, que la loi n'entend frapper que celui qui ne paie pas, alors qu'il l'aurait pu. La contravention à la loi est donc un délit réel, commis par faute ou négligence, et qui ne saurait plus être annulé, une fois la sentence prononcée.

M. Usteri ajoute, dans le même sens, que le droit de grâce pourra intervenir dans tous les cas où le jugement condamnatoire consacrerait une iniquité manifeste.

M. Lachenal, en revanche, appuie les vues de M. Richard. Ilien ne nous empêche de décréter que le paiement de la taxe constitue l'exécution finale de la peine. L'Etat, tout en étant soucieux de ses intérêts, ne doit pas être inexorable, il ne doit pas vouloir la mort du pêcheur.

M. Python, qui n'a jamais été enthousiaste de cette loi, serait, au contraire, disposé à voter tous les allégements admissibles. Mais celui-ci ne l'est décidément plus. Le délinquant qui « paie après coup ne répare pas sa faute ».

Avant la proposition de la commission, de rayer la clause en question, et celle définitivement, l'emporte par 21 voix contre 8.

Un juge, homme politique, risque toujours de provoquer la suspicion des justiciables. Or, il faut que le juge, comme la femme de César, ne puisse pas même être soupçonné.

Vous voterez donc le projet de loi constitutionnelle sur les incompatibilités et vous repousserez l'arrêté du Grand Conseil qui l'écarte.

Tous attendent.

Oui pour les incompatibilités.

Non pour l'arrêté du Grand Conseil.

Vive la Suisse! Vive Genève!

Au nom de l'Association démocratique des Eaux-Rives, (Suivent 75 signatures d'électeurs de la commune.)

Les maisons du peuple en Belgique. — C'est le titre de la causerie que M. Emile Vandervelde et Mme Vandervelde feront samedi 30 mars, à cinq heures après midi, au casino de St-Pierre.

Kekpek? — A la demande générale, il sera donné dimanche 31 mars, à huit heures et demie du soir, dans la salle des fêtes de l'institution, une répétition dernière représentation de la tragédie de l'Idiot Borodine, dont le succès s'affirme de plus en plus. Il y aura foule dimanche aux Amis de l'instruction.

Le spectacle commencera par le *Commissaire est bon enfant*, la désopilante comédie de G. Courteille et Lévy.

Le séance est levée à 11 heures.

Par dépêche télégraphique

Berne, 29 mars.

CONSEIL NATIONAL. — Séance à 8 h. 1/4. — Présidence de M. Bühlmann.

Perception de la taxe militaire. — La loi sur la perception de la taxe militaire revient devant une seule divergence, mais dans la séance de ce matin, il a décidé de la passer au vote. — 8 h. 30 a. à l'Amphithéâtre de l'Université. — 8 h. 45 a. à l'Amphithéâtre de l'Université. — 8 h. 45 a. à l'Amphithéâtre de l'Université.

8 h. 45 a. Séance rédactionnelle. — *Histoire de Maedelein.* — Projections lumineuses. Quatorze mixte. Cordiale invitation aux jeunes gens.

8 h. 45 a. Cafè Frey, rue de Rive, 7. Assemblée générale de la Société de tir de campagne « le Toli ».

9 h. 8 a. Salons Bohrer, Terrassière. Bal de la Section fédérale de gymnastique de Plainpalais.

Membres de la direction générale 37,500 fr.; secrétariat 15,500 fr.; département des finances 10,250 fr.; département du commerce 17,650 francs; département de l'exploitation 27,700 fr.; département des travaux 13,500 fr.; département du contentieux 6100 fr.; enfin, frais de bureau 57,300 fr. (oyer, ameublement, acquisition de matériel, etc.).

Ce budget est adopté en bloc et sans discussion.

Affaires de chemins de fer. — Le Conseil décide rapidement, sur les rapports de MM. Amsler (Zürich) et Lüthi, quelques affaires déjà établies. — 8 h. 30 a. à l'Amphithéâtre de l'Université de la chimie de l'air. — *Change Paris.* — 8 h. 30 a. à l'Amphithéâtre de l'Université de l'air. — *Change Berlin.* — 8 h. 30 a. à l'Amphithéâtre de l'Université de l'air.

Motion Hitler. — Dans la session de décembre, M. Birter et quelques-uns de ses collègues avaient déposé une motion invitant le Conseil fédéral à étudier la question du colportage à domicile, et de la concurrence déloyale. Ils demandaient des lois fédérales sur ces deux objets. M. Birter développe sa motion.

Il fait d'abord l'histoire de la jurisprudence fédérale en matière de colportage. Les législations cantonales subsistent et forment un foulis de dispositions contradictoires. L'orateur cite aussi un coup d'œil sur la législation étrangère. Déjà en soi-même (*an und für sich*) cette variété est un mal. Mais le colportage des brasseurs, des alus, M. Birter n'en demande pas l'interdiction. Il voudrait une loi fédérale qui le distingue nettement entre le colportage et le petit commerce. Le colportage devrait souffrir sous une pénalité fédérale, dont le produit se répartirait entre les cantons.

Quant à la concurrence déloyale, la définition est plus difficile. Elle existe lorsqu'un négociant cherche, par des moyens répréhensibles, à détrousser ses concurrents. L'orateur ne peut entrer dans le fond de la question. Au point de vue formel, le Conseil fédéral estime qu'il y aurait lieu d'abord de faire la question de la révision de la Constitution. Cette étude a été faite par la commission de l'Assemblée nationale pour défendre le palais fédéral contre les anarchistes. Moyennant quoi, le Suisse jouira de la liberté suprême.

L'auteur de la pétition engage les Chambres à bien examiner sa pétition et à décider ensuite si l'on convient de la mettre en prison ou dans une maison de santé.

Le rapporteur, M. Isler, excuse de l'incompétence des Chambres de trancher ce dilemme et propose de passer à l'ordre du jour. Adopté.

Le séance est levée à 11 heures.

Sur ces explications, M. Birter retire sa motion.

M. Gschwind (Bâle-Campagne) prend acte des promesses de M. Deucher.

M. Wulschlegel (Schaffhouse) estime que la motion aurait pu être meilleure. L'orateur répond que les deux tentatives de l'ordre pour régler ce problème ont échoué. Il croit que les cantons sont mieux placés pour régler cette matière. Par contre l'orateur a trouvé de meilleur effet une loi fédérale sur la concurrence déloyale, sans en attendre pour cela des résultats bien positifs. Mais l'étude du sujet devrait être portée sur toutes les faces du problème et non pas se limiter à la protection des petits commerçants. On ne peut pas aller contre les lois de l'économie moderne.

La discussion est close.

Le séance est levée.

Demain, séance de procès-verbal à 8 heures.



British Library National Newspaper Archive - Photo by Christopher Furlong / Getty images

SEARCH COLOCOLORADO HISTORIC NEWSPAPERS COLLECTION

Search Colorado's historical newspaper collection.

SEARCH

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

View Images

PRINT

SEARCH COLOCOLORADO HISTORIC NEWSPAPERS COLLECTION

SEARCH COLORADO'S HISTORICAL NEWSPAPER COLLECTION.

SEARCH

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

SEARCH COLOCOLORADO HISTORIC NEWSPAPERS COLLECTION

SEARCH COLORADO'S HISTORICAL NEWSPAPER COLLECTION.

SEARCH

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

SEARCH COLOCOLORADO HISTORIC NEWSPAPERS COLLECTION

SEARCH COLORADO'S HISTORICAL NEWSPAPER COLLECTION.

SEARCH

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

SEARCH COLOCOLORADO HISTORIC NEWSPAPERS COLLECTION

SEARCH COLORADO'S HISTORICAL NEWSPAPER COLLECTION.

SEARCH

SEARCH BY DATE
SEARCH BY NAME
SEARCH BY SUBJECT

SEARCH THE COLLECTION

The screenshot shows the homepage of the Bibliothèque nationale de Luxembourg. At the top, there's a search bar with the placeholder "Rechercher un article". Below it are two smaller search bars: one for "Rechercher un livre" and another for "Rechercher un auteur". The main content area features a large image of a classical building, likely the library itself, with text overlaying it about the "Muséum d'ethnographie en ligne". At the bottom, there's a section titled "Futur(e) bibliothèque" with a green checkmark icon.

LE TEMPS

RECHERCHE RECHERCHE AVANÇÉ AIDE À PHOTOS

Rechercher dans les archives de la Gazette de Lausanne et du Journal de Genève

Rechercher

en cours (contenu en cours depuis 2002)

Accès

Nos partenaires


ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE


UNIVERSITÉ
DE LAUSANNE

The screenshot shows the homepage of the Georgia Historic Newspapers website. At the top left is the "DIGITAL LIBRARY OF GEORGIA" logo. At the top right is the "GALILEO" logo. The main title "GEORGIA HISTORIC NEWSPAPERS" is prominently displayed in large, bold, black letters. Below the title is a navigation bar with links for "HOME", "SEARCH", "ABOUT", and "CONTACT". A search bar is located below the navigation bar. The central feature is a large map of Georgia with county boundaries. Overlaid on the map are several newspaper mastheads from different historical periods. To the right of the map is a sidebar titled "Recently Added Titles" which lists five new titles added to the collection. At the bottom left is a link to an "Interactive map of Georgia".

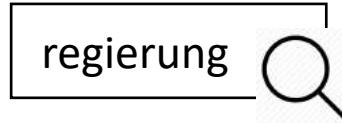
The screenshot shows the homepage of The European Library. At the top, there's a navigation bar with links for 'Home', 'About', 'Search', 'Help', and 'Contact'. Below the navigation is a search bar with placeholder text 'Search the European Library' and a magnifying glass icon. A large blue button labeled 'SEARCH' is positioned below the search bar. To the right of the search bar, there are two sections: 'Recent arrivals' showing thumbnail images of books and 'Newspapers' with a thumbnail of a newspaper page. On the left side, there's a sidebar with a 'Log in' link and a 'Help' section. The main content area features a large banner with the text 'The EBL (eBook Library) has been created by EBSCO Publishing. Once the user logs in, he will be granted a unique set of electronic permissions of EBSCOhost products. This means that the user can access the full-text of the articles in the databases, the user can read them, print them, copy them, and even download them. The user can also be granted the permission to print or download the entire issue of a journal. The user can also be granted the permission to print or download the entire issue of a journal.' Below the banner, there are several search filters: 'Subject', 'Author', 'Title', 'Language', 'Year', 'Format', and 'Source'. A 'Search' button is located at the bottom of the filter section. Further down, there are sections for 'Search Newspapers' (with a 'Search' button) and 'Find This Day in History' (with a 'Search' button). The footer contains links for 'About', 'Help', 'Feedback', 'Privacy Policy', and 'Terms of Use'.

A screenshot of the Delpher website. The header features the word "Delpher." on the left and a search bar with a magnifying glass icon on the right. Below the header is a decorative banner composed of numerous small, overlapping historical documents or maps. A navigation menu is visible on the left side of the main content area, which includes sections like "Nieuws" and "Geschiedenis".

The screenshot shows the Trove search interface. The search bar at the top contains the query "Drowned newspaper and more". Below the search bar, there are two dropdown menus: "Source type" set to "Newspaper" and "Format" set to "Text". A "Search" button is located to the right of the dropdowns. The main content area displays a list of search results. The first result is highlighted in blue and titled "Drowned newspaper and more". It includes a thumbnail image of a newspaper page with the headline "DROWNED NEWSPAPER AND MORE". The list of results also includes other items like "Drowned newspaper and more" (2), "Drowned newspaper and more" (3), "Drowned newspaper and more" (4), and "Drowned newspaper and more" (5). At the bottom of the page, there is a footer with links to "About Trove", "Help", "Feedback", and "Contact us".

Help me

... “expand my query based on common OCR mistakes and spelling variants” + multilingual synonyms



Eegierung
Begierung
Regieruug
Kegierung
Regiexung

similar words
+ edit
distance

Behorden
Finanzdirektion
Staatsrath
Eegierungen
Behörde
Reichregierung
Staatsrat
Staatsrathes

similar words

→ lexical processing

Help me

editorial

... “find the frequency of word X in
articles of certain type or thematic”

title banner and ‘ears’



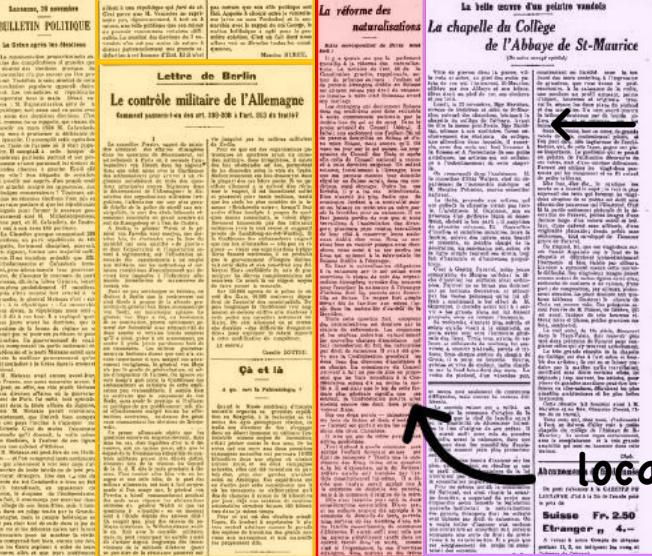
... “distinguish ‘real’ occurrences
from false positives emerging from
ads, stock market or meteo pages”

feuilleton

→ segmentation (CV) and classification (NLP)

... “find the frequency of word X in

culture



local news



Help me

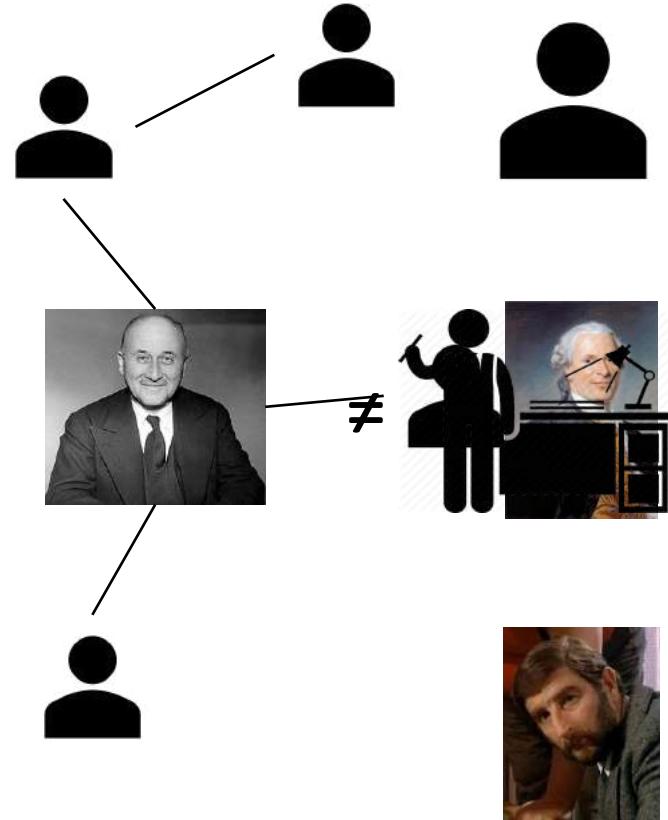
... “ find all articles mentioning Jean Monnet.

And the right one!”

... “ know with whom the name of X is
associated.”

... “ find articles mentioning person X as
author vs. as main subject of the article”

→ named entity processing



Help me

- ... “discover similar items I might be interested in”
- ... “appreciate the high-level thematic(s) of the corpus
- ... “build and interact with my own **collection**”



How to enable semantic indexing
and exploration of large collections
of historic newspapers?

EPFL



University of
Zurich UZH



infoclio.ch

Unil



NZZ LE TEMPS



Critical content mining
of 200 years of
historical
newspapers



Media Monitoring
of the Past

CONTACT-POINT

NLP



Simon
Clematide,
PhD

(digital) historian



Marten
Düring, PhD

NLP/ DH



Maud
Ehrmann,
PhD

co-
designer/web dev



Thijs van
Beek

(digital) historian



Estelle
Bunout, PhD

designer/web dev



Daniele
Guido, M.Sc.

NLP/ DH



Matteo
Romanello,
PhD

designer/web dev



Paul
Schroeder

NLP



Phillip
Ströbel, M.A.

SUPERVISOR
(digital) historian



Andreas
Fickers, Prof.

robotics/DH



Frédéric
Kaplan, Prof.

NLP



Martin Volk,
Prof.

designer/DH



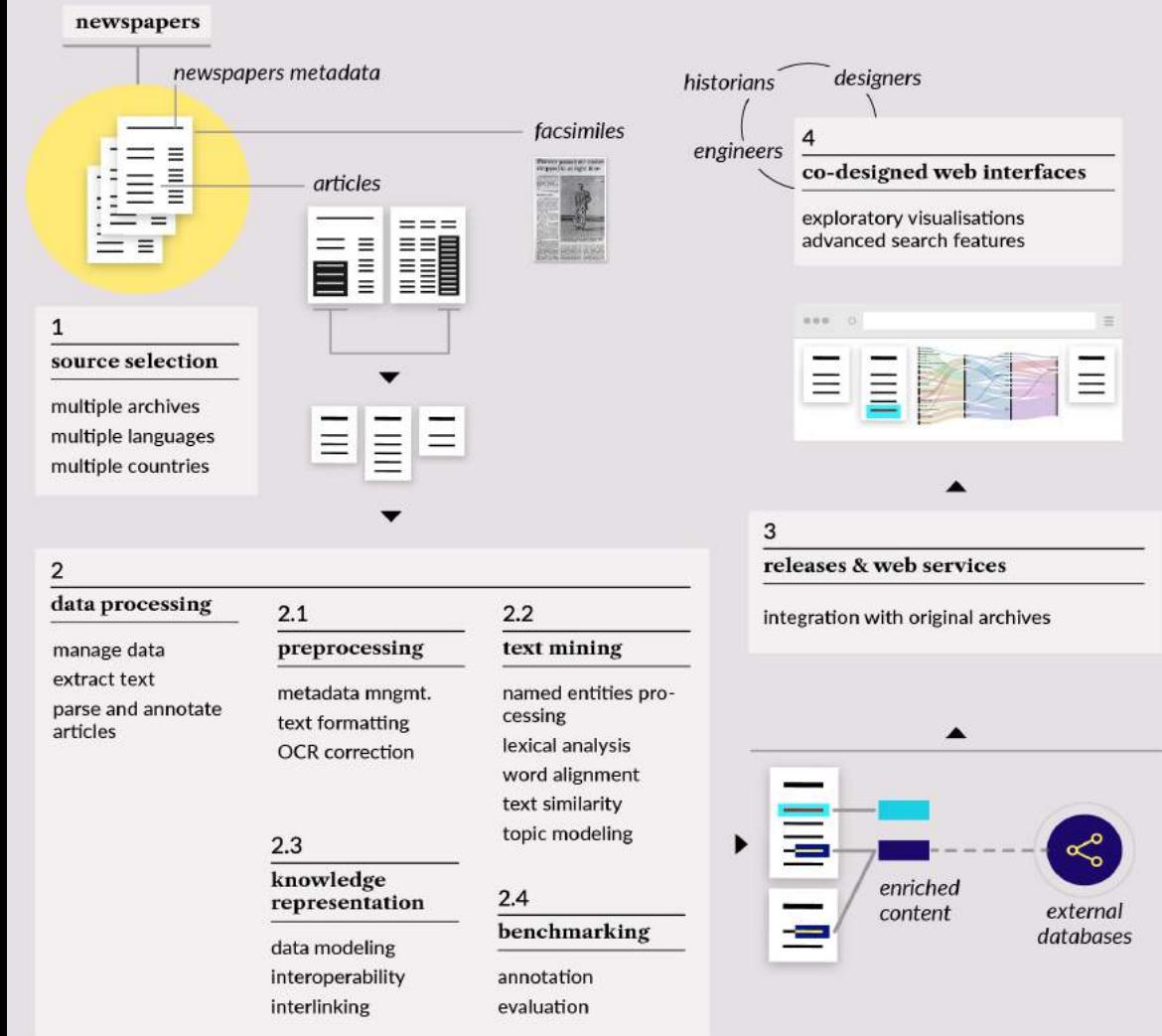
Lars
Wieneke,
PhD

the team

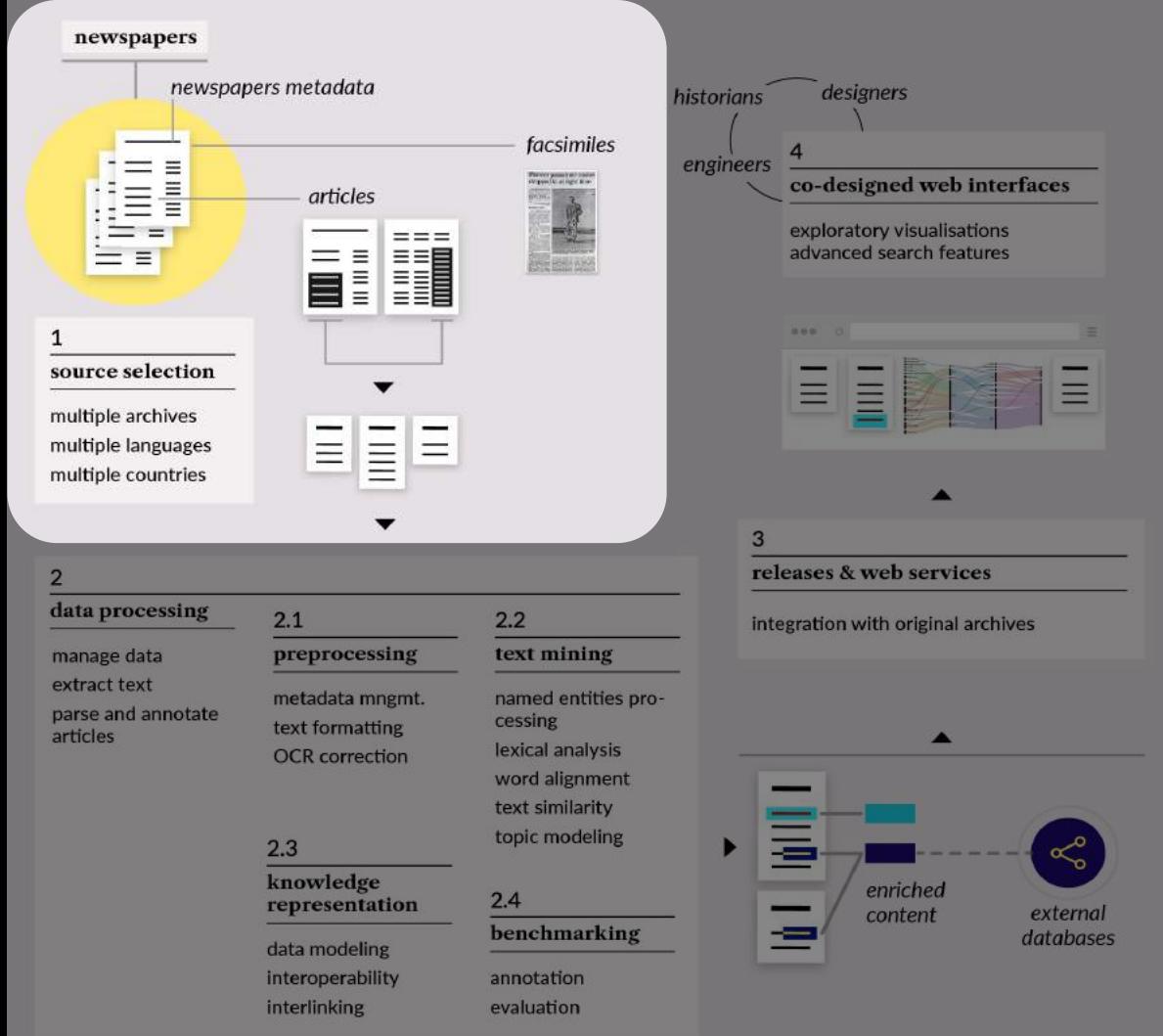
*+ a team of
historical advisors*

*+ a team of
associated historians*

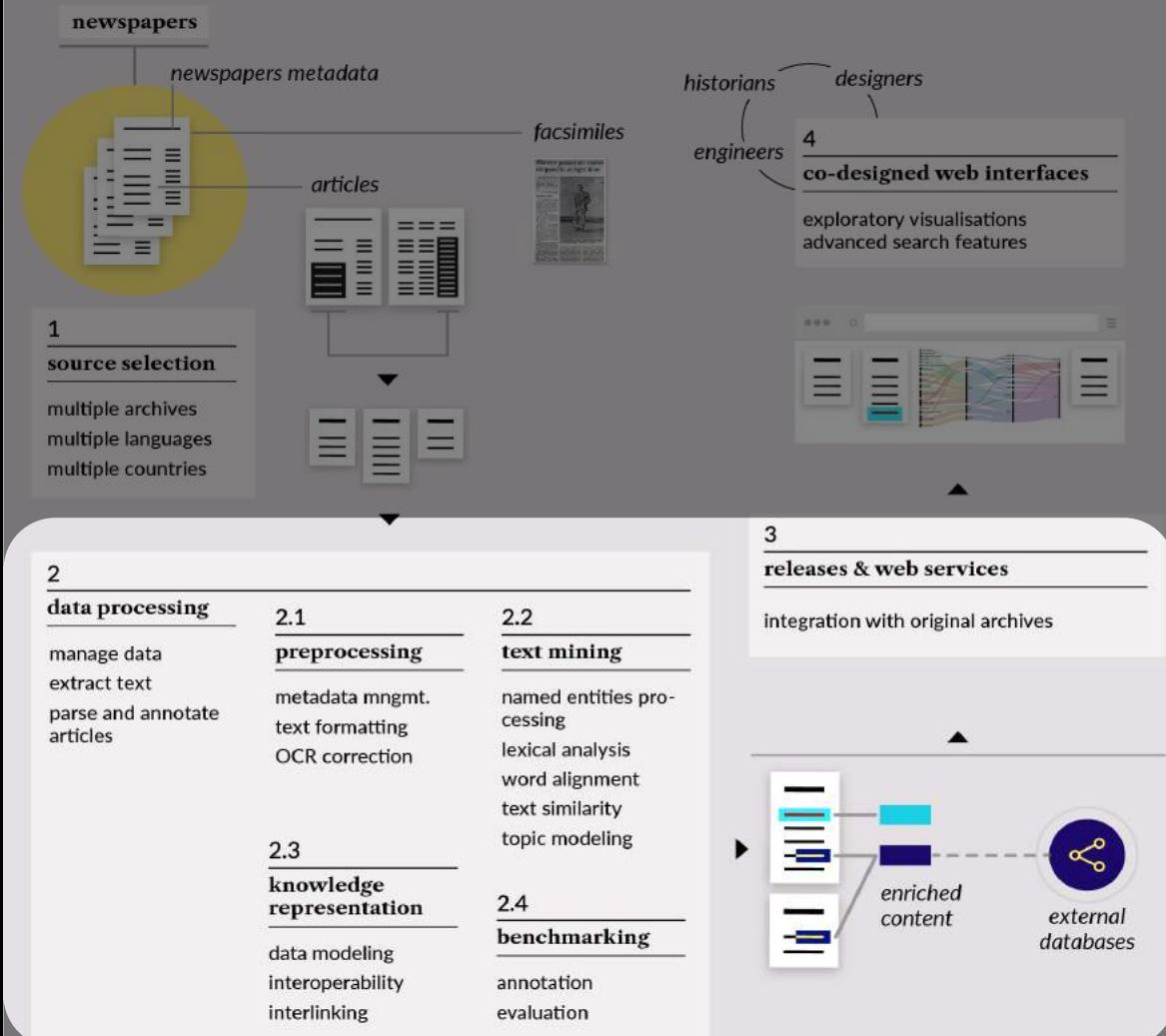
...and our plan:



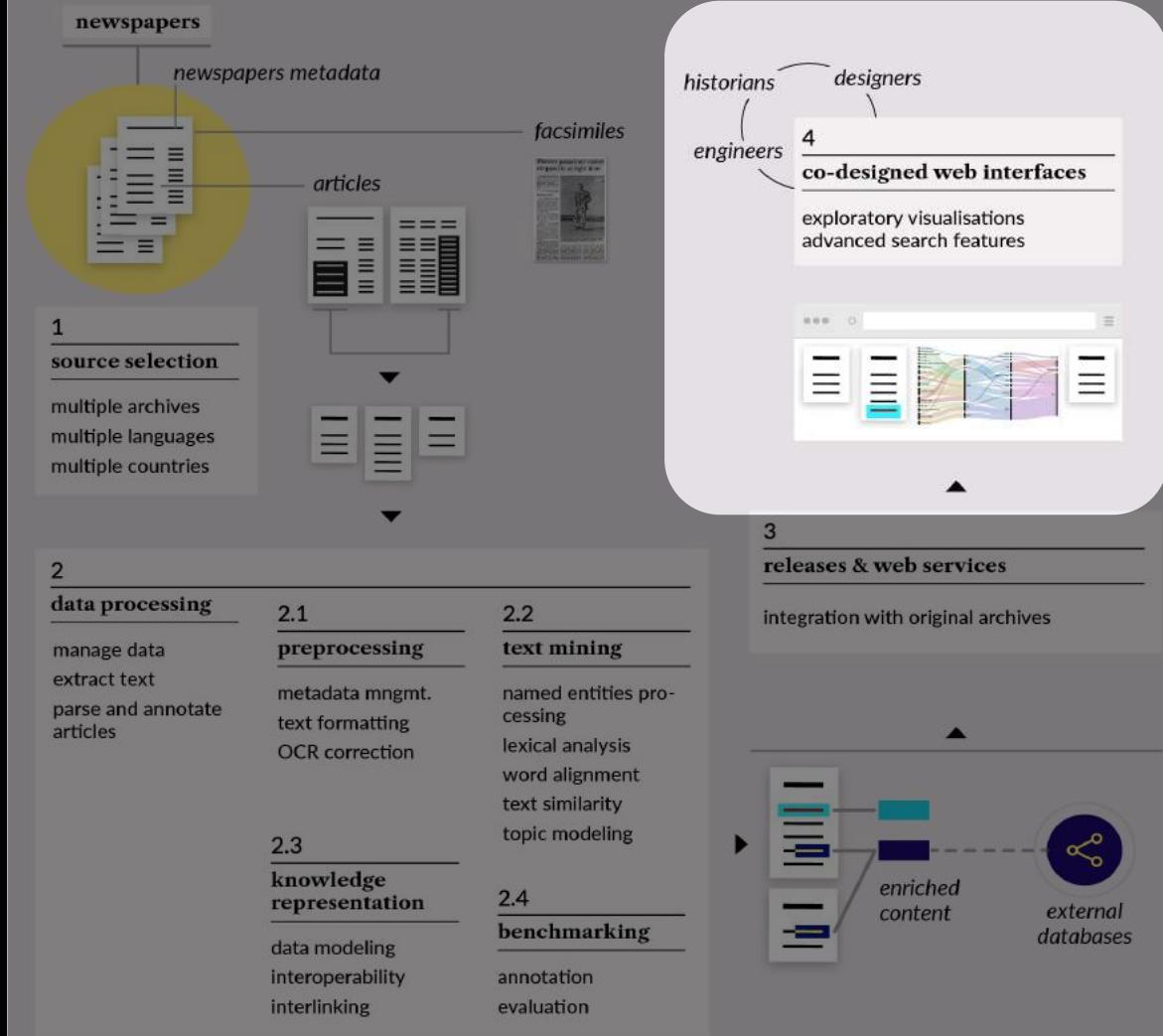
...and our plan:



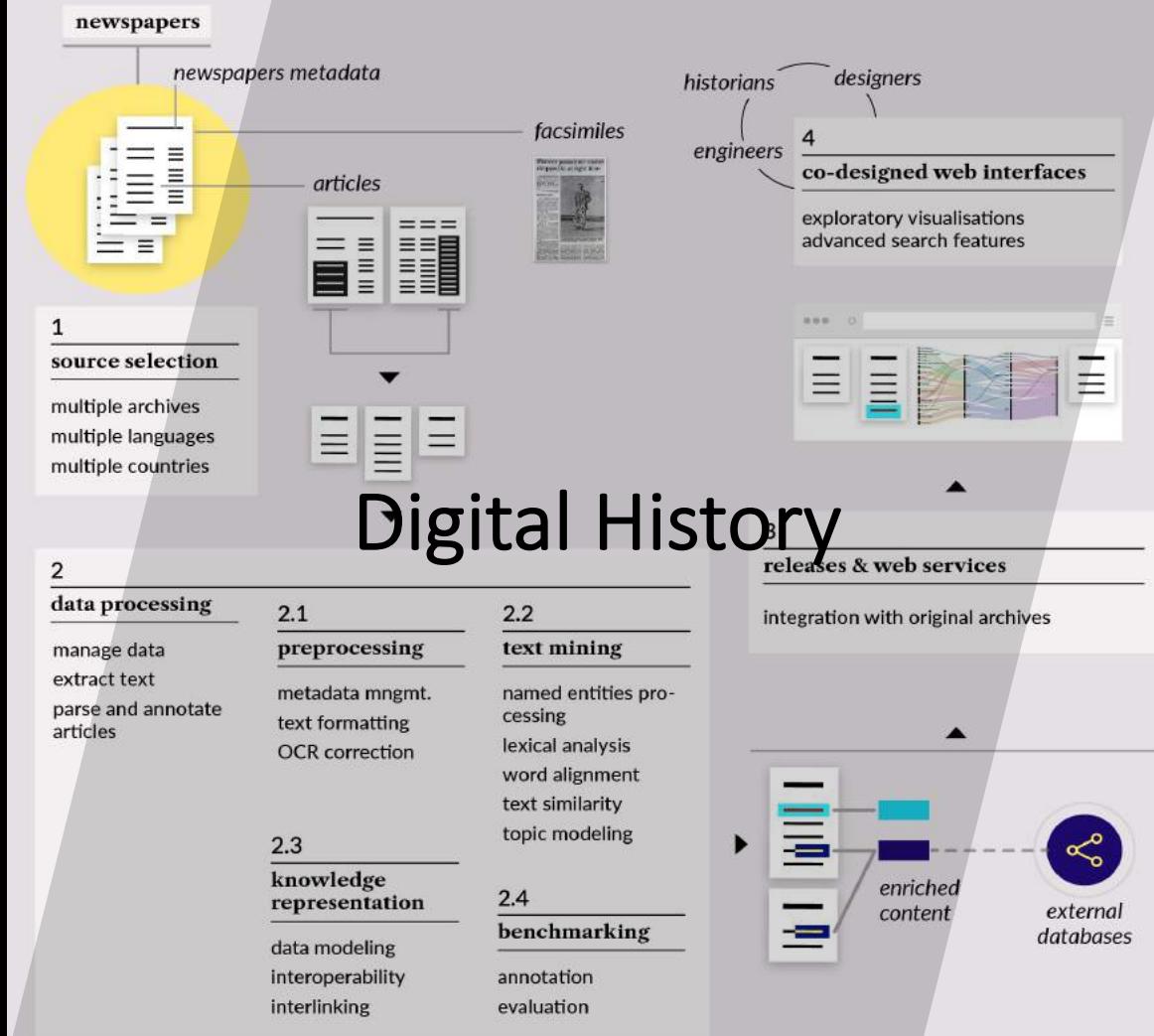
...and our plan:



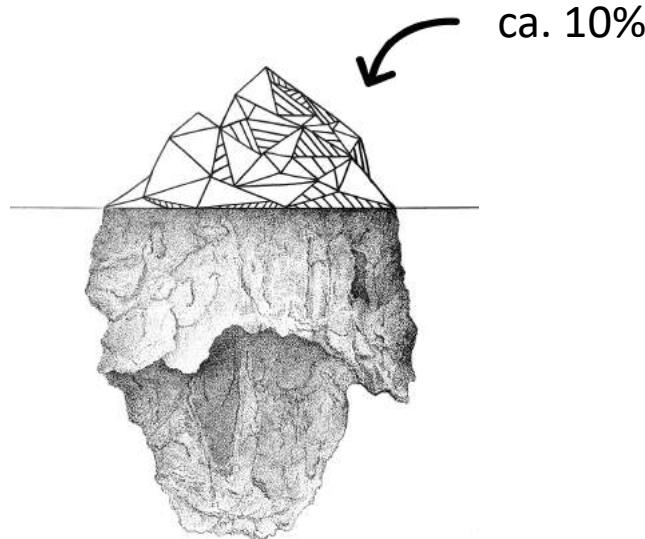
...and our plan:



...and our plan:



5 challenges

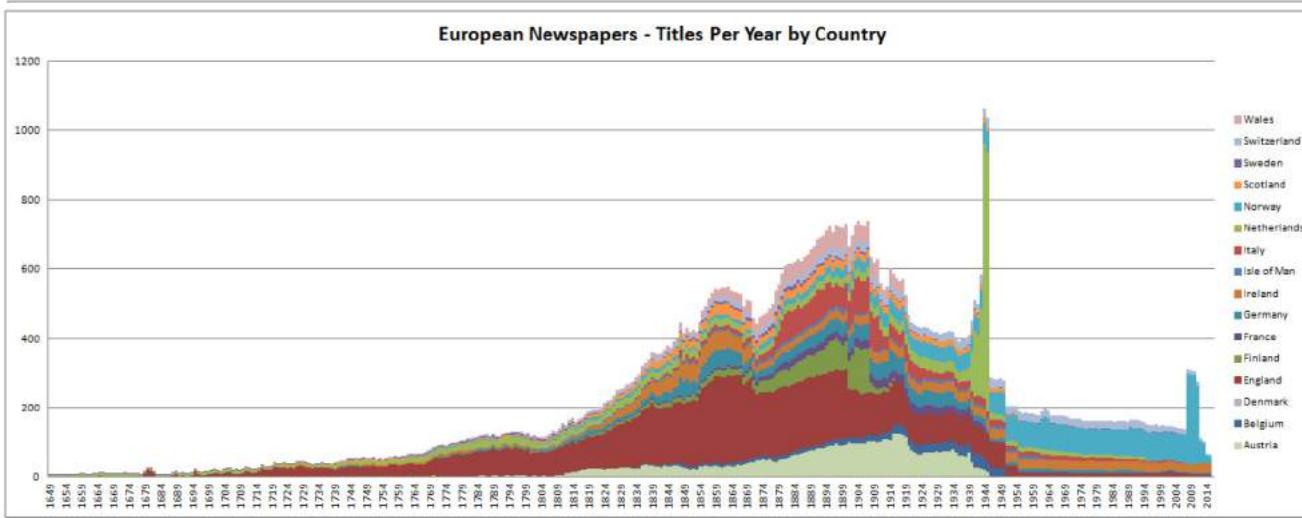


1

Digitized
newspaper silos

Source: Center for Research Libraries. The State of the Art. A Comparative Analysis of Newspaper Digitization to Date, 2015

5 challenges

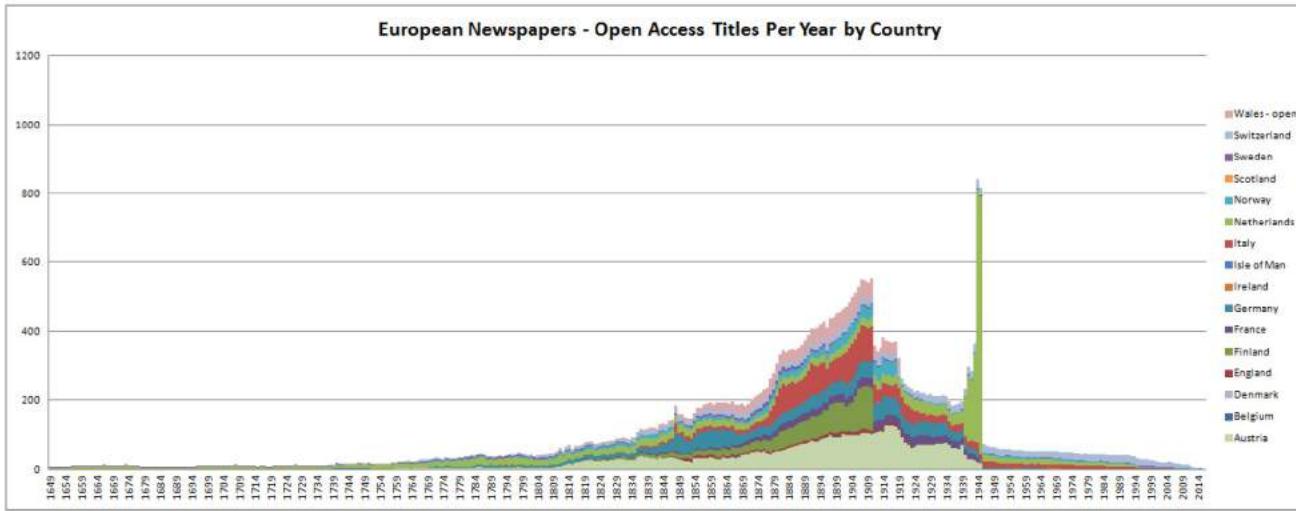


1

Digitized
newspaper silos

Source: Center for Research Libraries. The State of the Art. A Comparative Analysis of Newspaper Digitization to Date, 2015

5 challenges

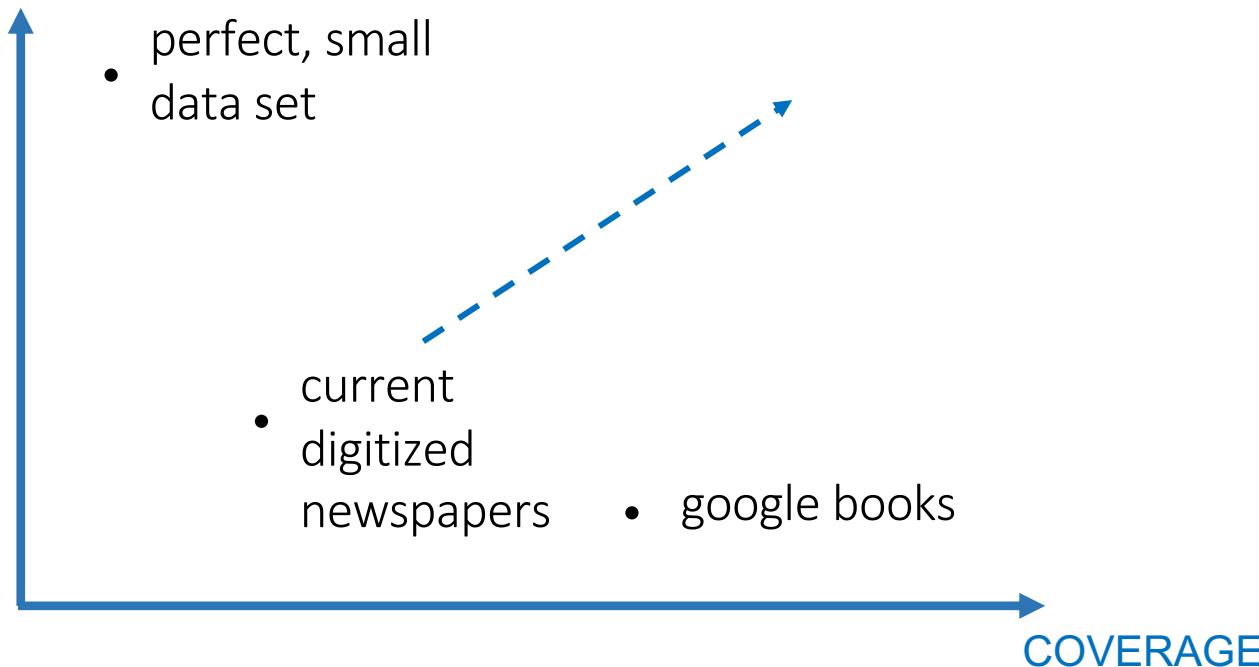


Digitized
newspaper silos

Source: Center for Research Libraries. The State of the Art. A Comparative Analysis of Newspaper Digitization to Date, 2015

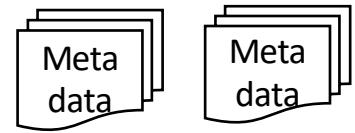
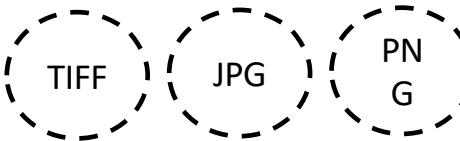
Shape of the (newspaper) digital landscape

HOMOGENEITY



5 challenges

*digital
archives*



*processable
data*

{ JSON }

JPEG2000

1

2

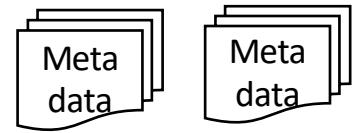
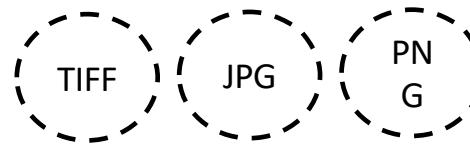
Digitized
newspaper silos

Big, messy
data



5 challenges

*digital
archives*



normalize normalize normalize normalize normalize normalize normalize normalize

*processable
data*

{ JSON }

JPEG2000



1

2

Digitized
newspaper silos

Big, messy
data

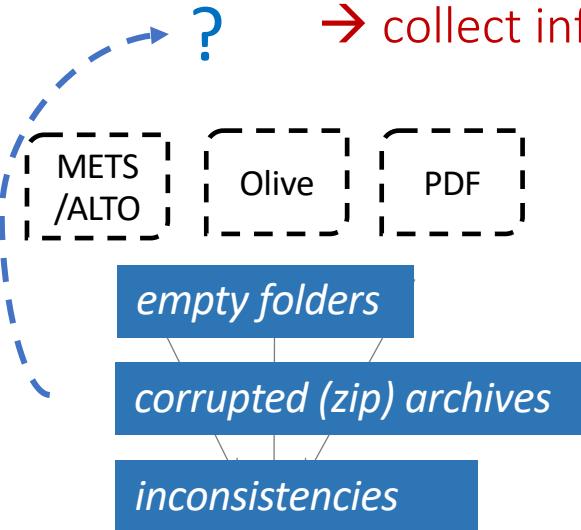
5 challenges

digitisation



→ collect information

*digital
archives*



*processable
data*

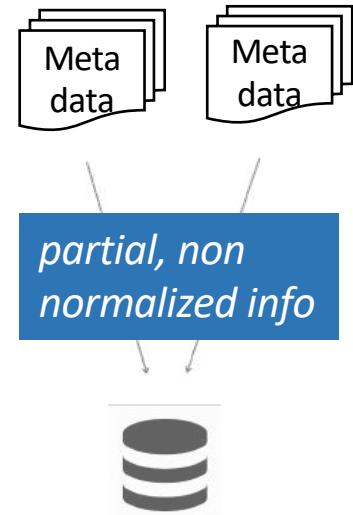
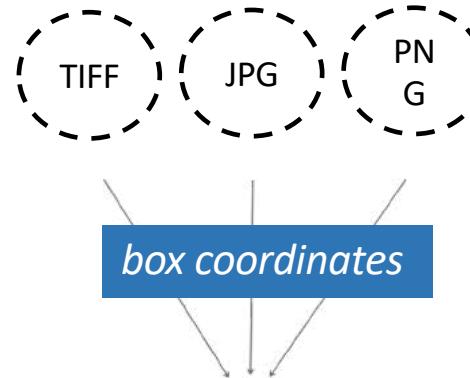
{ JSON }

1

2

Digitized
newspaper silos

Big, messy
data



5 challenges

Gazelles in Switzerland?



1

Digitized
newspaper silos

2

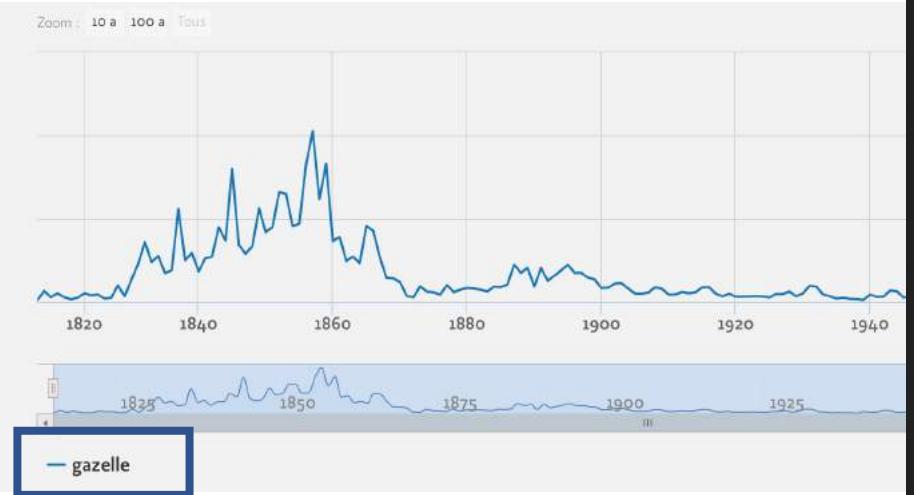
Big, messy
data

3

Noisy,
historical text

5 challenges

Gazelles in Switzerland?



1

Digitized newspaper silos

Big, messy data

Noisy, historical text

5 challenges

Improving OCR quality

OCR-2005

N °. 86 . Zürcher Sam stag , t
7 9 3 & deN26 .
Weinmon . MMMM ? urck da
« franiösische Voll Gouverain
geworden , mit dem 3 « nnte
in volitischen Verhältnissen » . »
Wenige Minuten später schreibt er:
-Measure: 0.147
...nittn sen ; «
... siecht zufrieden damit , im
Empersts "t

F-Measure: 0.147



Digitized newspaper silos

OCR-2017

Prozeß der Marie Antoinette. Nachdem dieselbe am 1. g. Weinm. alten Stvls, oder am rz. des ersten Monat« im 2. em Jahre der Republik neuen Kaleuderstpls, in den Audien-
F-Measure: 0.684

F-Measure: 0.684



Big, messy data

HTR-2018

Prozeß der Marie Antoinette.
Nachdem dieselbe am 15. Weinm.
alten Styls, oder am 23. des ersten
Monats im 2ten Jahre der Republik
neuen Kalenderst.
Audi: F-Measure: ~0.95 de, und
klassem hatte, fragte sie der
Präsident: Wie sie heis- se? „ Ich
nenne mich, antwortete sie, Marie
Antoinette von Lotharingen-

F-Measure: ~0.95

Noisy, historical text

5 challenges

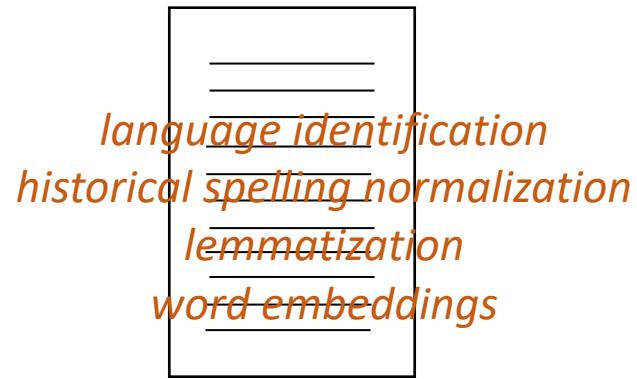
OCR quality assessment

Goal: allow users to **estimate** the completeness of their keyword search results

- **focus** on content words (exit stop words and punctuation)
- **easy** to grasp by non-technicians

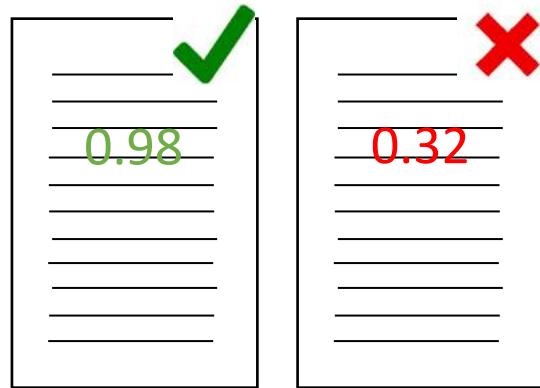


Lexical processing



OCR quality assessment

Lexical
processing



Lexical processing

OCR quality assessment

Topic modeling



kunst konzert bild theater
künstler oper musikalisch
spielen werk künstlerisch
orchester aufführung

deutschen britisch
deutsch truppe krieg
russisch angriff flugzeug
feindlich schwer front
armee kampf london

zimmer lage vermieten
verkaufen wohnung
garten groß auskunft
villa hotel schön zürich
haus

Lexical processing

OCR quality assessment

Topic modeling

Named entity processing refers to a family of tasks that are part of the larger domain of Information Extraction. IE has been defined and formalized in the 90's with the Message Understanding Conferences (MUC) organized by the US DARPA. Besides the recognition of proper names of types e.g. PERS, ORG and LOC, named entity processing also implies disambiguation and relation detection.

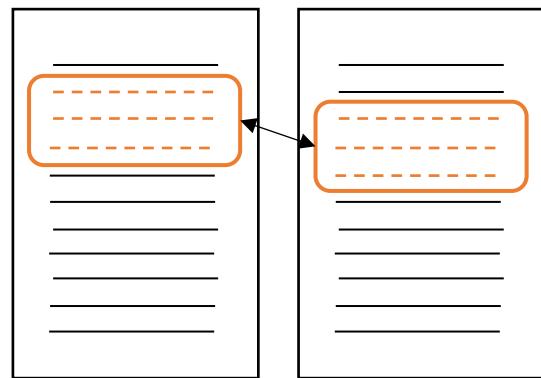
PERSON, ORGANISATION, LOCATION, EVENT, DATE, NLP-TASK

Named entity processing

Lexical
processing

OCR quality
assessment

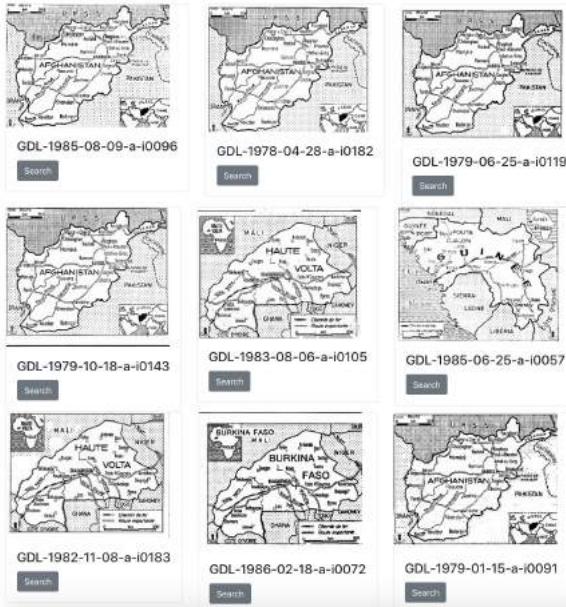
Topic
modeling



Text
re-use

Named entity
processing

OCR quality assessment



Lexical processing

Image similarity

Topic modeling

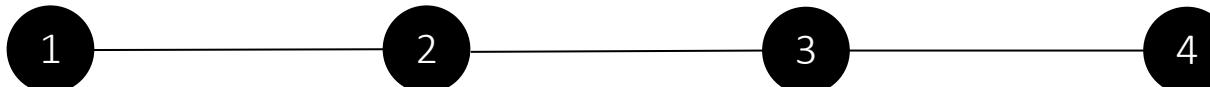
Named entity processing

Text re-use

[Visit the impresso project youtube channel for more detailed videos](#)

5 challenges

How best accommodate text analysis research tools and their usage by humanities scholars?

- 
- 1 Digitized newspaper silos
 - 2 Big, messy data
 - 3 Noisy, historical text
 - 4 Visualization and exploration

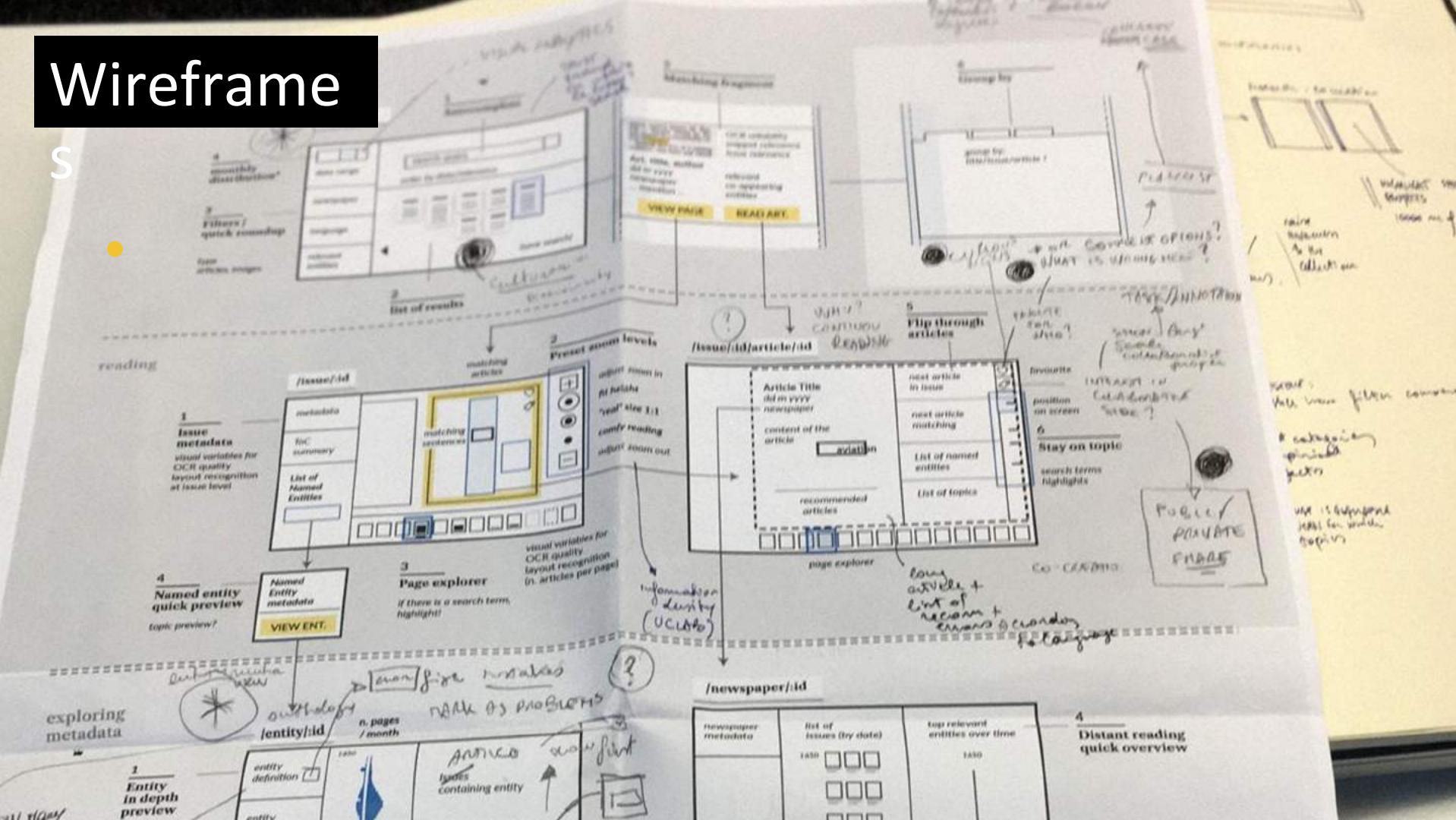
Co-design and fast prototyping



Continuous exchanges to learn from each other, experiment, evaluate

- Workshops
- Associated researchers
- *Impresso* community calls
- One-to-one collaborations
- Survey of existing interfaces

Wireframe



s search for articles **c** collecting
R reading contents **o** organising

M exploring metadata
E visual experiments

Search

code for the SWOT

component description

S.1

search autocomplete: based on input, suggests words, date ranges, named entities or article categories

red text: uncompleted features

S.2

timeline n. results per year. This acts as "date range" filter

S.3

metadata filters / quick roundup how many search results per newspapers titles, languages, page content tags, format tags

explore ▾

YOUR SEARCH PAST SEARCHES

FULL TEXT SEARCH REFINE ...

_ type a text or a date or to start... X

PUBLISHED IN [DATE RANGE] TIMELINE OF

N. ARTICLES N. ISSUES

1800 1850 1900 1950 2000 2050 2100

1745 select a date range to show articles published

NEWSPAPER TITLES

53625 La Gazette de Lausanne quotidien suisse de langue française édité à Lausanne

43864 Le Journal de Genève quotidien suisse qui a paru du 1826 au 28 février 1998

ARTICLE TYPE MORE ...

TOP NAMED ENTITIES

112 partisans 103 Napoleon

50 satirique 50 Zurich, Suisse location

MORE ...

LANGUAGE

112 French

MORE ...

Search

GROUP RESULTS BY ISSUE PAGE ARTICLES SENTENCES

SEARCH SUMMARY Find 53625987 articles in our collection.

1 of 112 Les recherches d'INTERNATIONAL NEWS Gazette de Lausanne Les recherches d'une possibilité JEAN-PHILIPPE de multiples...

C.1 Suggest Public Tags (vocabulary)

C.2 "Forum" newspaper or specific regions

C.3 Contents related to the content

C.4 Use and use their users' public tags

C.5 Explore news by tag (coming)

C.6 suggest new tag

C.7 of the data dimension chosen in [Or.11]

C.8 Freely tune the data "dimensions" from the available metadata

component index

R.14 list of the article and usage regions state by state

R.15 selected favorite collections, with quantities of articles and pages

R.16 add tags to the article and add the article to your collections

R.17 List all the named entities in this article

R.18 Get the list of sentences for the named entity in this article

R.19 named entity to a collection!

R.20 structured overview of your collection, export citations or metadata

R.21 list of items included in your collection

R.22 assign item to more than one of your collections

R.23 "How did you find this item?" list of search results and link to the original query

R.24 switch type of visualization: lists by grid, vs altuve diagram, Choose within a wide range of visual models

R.25 visual feedback

R.26 Freely tune the data "dimensions" from the available metadata

Strength & opportunities

R.4 convenient and aesthetically pleasing summary of search - I also tested using the back button in the browser to go back to the previous page and that works as well (as a user, I am more inclined to use the back button rather than a "return to results" button)

R.6 marginalia is great feature

R.10 visualization of data can be very effective

S.1 search bar draws on common practice of searching for something, which people are comfortable with

R.5 this feature is wonderful, saves time by being able to view articles without having to open them

Weakness

when you select "download as," can you specifically download the marginalia perhaps graph la...

altuve diagrams are difficult to parse, even for data very professionals

how else can the front page introduce users to the collection? eg. featured collection, visual search?

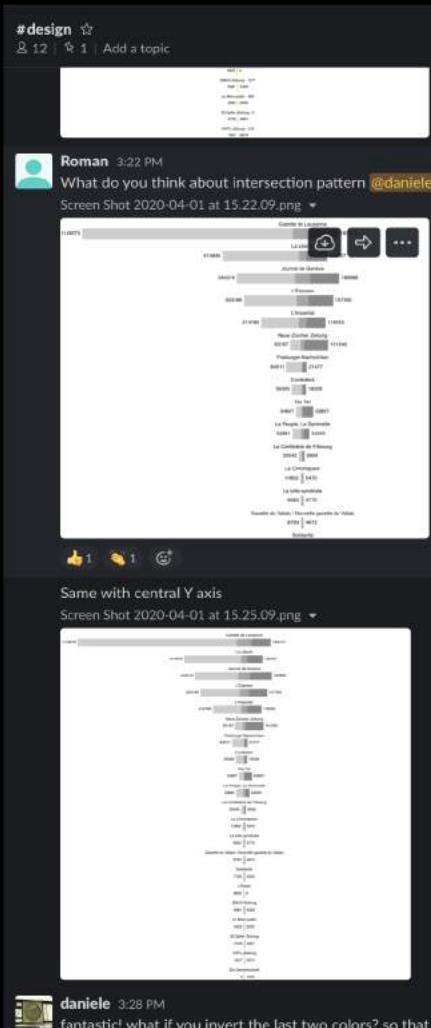
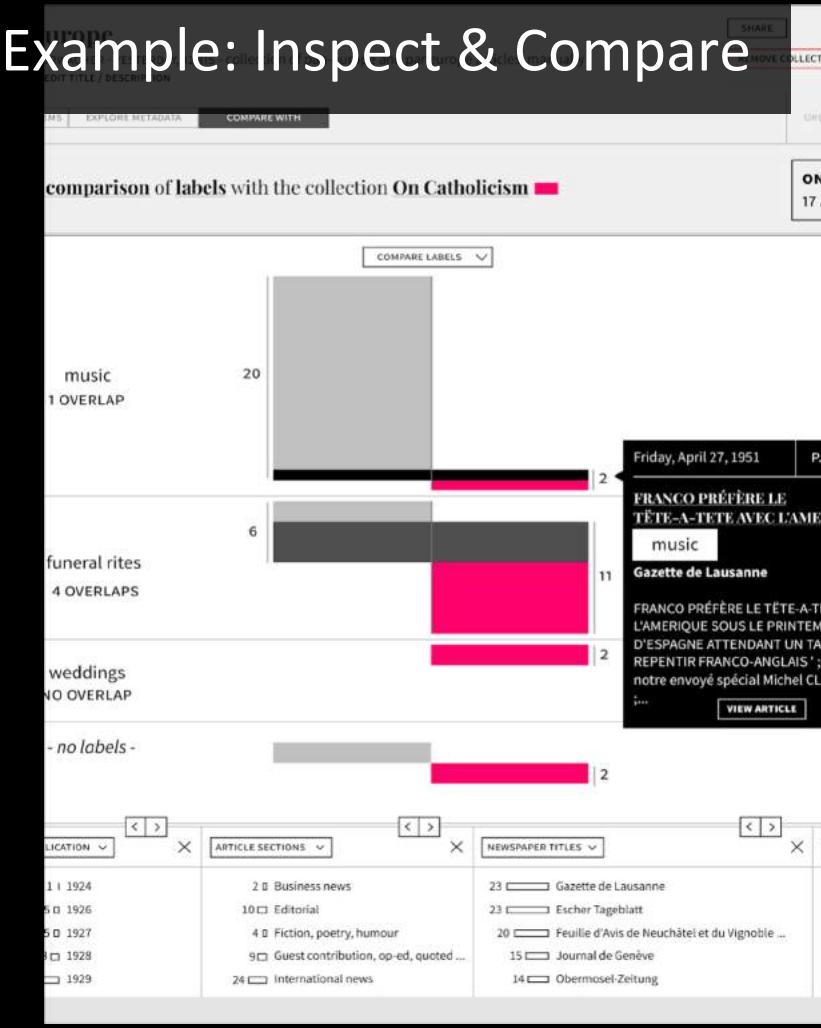
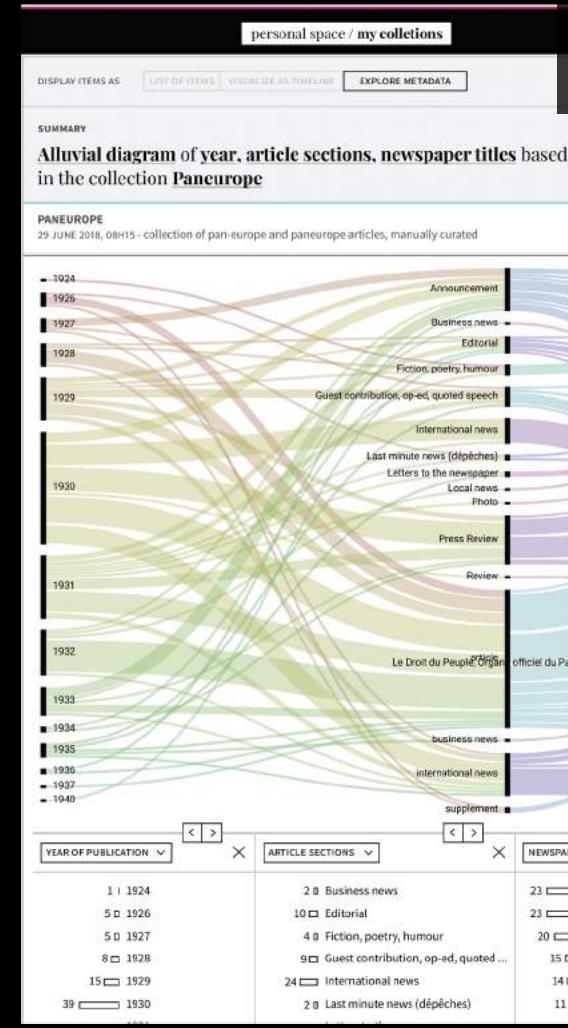
current search is designed for subject matter experts only

Overall, info server loads somewhat slowly

Feedback

Workshops





5 challenges

How to work with algorithmically generated,
changing, and certainly imperfect data?



1

Digitized
newspaper silos

2

Big, messy
data

3

Noisy,
historical text

4

Visualization
and exploration

5

Digital
scholarship

Transparency matters

beauty ≠ perfection

imperfection ≠ useless

Transparency matters

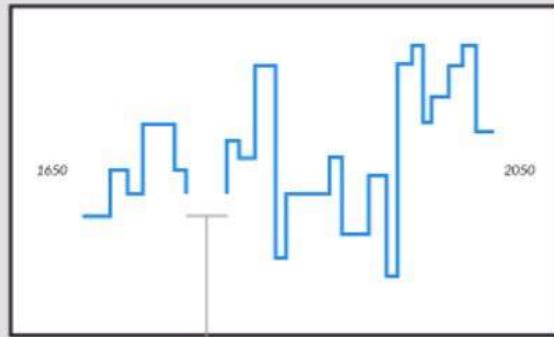
We **inherit, introduce and propagate** biases and variances in our corpus.

How do they affect content exploration? What exactly do historians need to understand to be able to deal with them?

What is the most effective way to **communicate** these biases and variances?

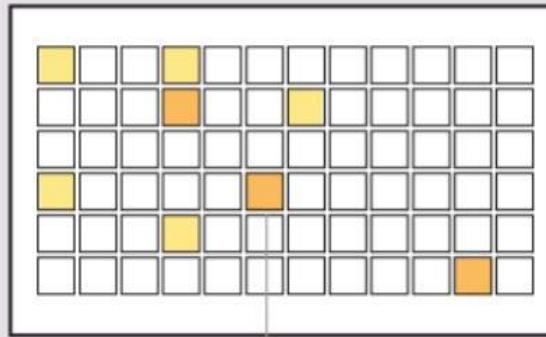
Implementing transparency

/analyse/:query



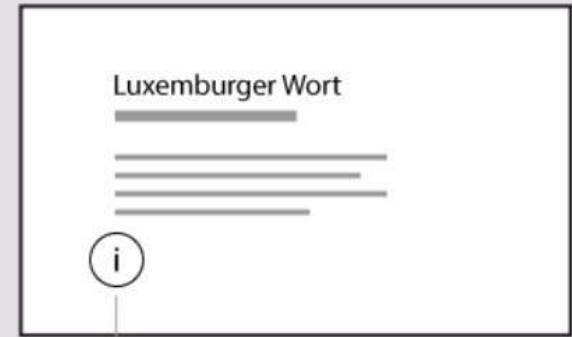
missing
data

/newspaper/:id



OCR quality
information

/newspaper/:id



indication of origins / source

5 challenges

Short demonstration



1

Digitized
newspaper silos

2

Big, messy
data

3

Noisy,
historical text

4

Visualization
and exploration

5

Digital
scholarship

For more information

The screenshot shows the homepage of the impresso website. At the top left is the logo 'impresso' with a small icon above it. To the right is a vertical sidebar with a decorative pattern and icons for Home, Project, App, Blog, Publications, Interview, and Consortium. Below the sidebar is a button 'JOIN OUR MAILING LIST'. The main content area features a large banner with the text 'Media Monitoring of the Past' and 'Mining 200 years of historical newspapers'. Below the banner is a sub-section titled 'How can newspapers help understand the past? How to explore them?' followed by a horizontal line. The central text discusses the project's goal of mining historical newspaper archives using a technological framework. At the bottom of the page is a footer with the text 'impresso website (update soon)'.

[impresso website](#) (update soon)

[zenodo](#) Search

Recent uploads

Recent uploads: [Recent uploads - Media Monitoring of the Past](#)

Extended Overview of CLEF HPC 2020: Named Entity Processing on Historical Newspapers

by [Clefmedia](#) • 0 comments • 0 likes

This document provides an extended overview of the task of named entity processing on historical newspapers, a pioneering shared task dedicated to the evaluation of named entity processing on historical newspapers in news, cultural and legal contexts. Special attention is devoted to the introduction of the task, its history, the main features, the data sets, the evaluation metrics, the results and the conclusions.

View

CLEF-HPC-2020 Named Entity Recognition and Linking on Historical Newspapers (204KB)

by [Clefmedia](#) • 0 comments • 0 likes

This document provides an extended overview of the task of named entity processing on historical newspapers, a pioneering shared task dedicated to the evaluation of named entity processing on historical newspapers in news, cultural and legal contexts. The proposed task has been developed by the CLEF-HPC community during the CLEF 2020 conference.

View

Driving deep learning enhanced digitized newspapers. A hands-on session with the CLEF-HPC 2020 shared task

by [Clefmedia](#) • 0 comments • 0 likes

Method for a hands-on introduction of the impressive workflow, dedicated to public release, for mining of newspaper articles from historical newspapers. This proposed task has been developed by the CLEF-HPC community during the CLEF 2020 conference.

View

Introducing the CLEF 2020 HPC Shared Task: Named Entity Recognition and Linking on Historical Newspapers

by [Clefmedia](#) • 0 comments • 0 likes

This document provides an extended overview of the task of named entity processing on historical newspapers, a pioneering shared task dedicated to the evaluation of named entity processing on historical newspapers in news, cultural and legal contexts. The proposed task has been developed by the CLEF-HPC community during the CLEF 2020 conference.

View

Want your publication to appear in this community?

- CiteZenodo: add a Zenodo record for your publication

data on [Zenodo](#)

code on [GitHub](#)

A screenshot of the Impresso Project YouTube channel page. The channel has 5 subscribers and 1 video. The video thumbnail is titled "Media Monitoring of the Past". The video description includes: "Impresso clip", "228 views • 4 months ago", "This video introduces the Impresso - Media Monitoring of the Past project.", "How can newspapers help understand the past? How to explore them?", and "The objective of the Impresso project is to enable critical text reading more". Below the video, there are sections for "Created playlists" showing "CEP-HPR-2020" and "SHED2020-Workshop", and "Upcasts" showing "Play All". A yellow box highlights the channel name "impresso youtube channel".

impresso [youtube](#) channel

**Media Monitoring
of the Past**

SEARCH ARTICLES SEARCH NEWSPAPERS Topics Impact & Company Test release

SEARCH ARTICLES SEARCH NEWSPAPERS NEWSLETTER

search for –

historical newspapers
76 newspapers collected,
800 issues,
5,650 pages scanned,
473,848 content items identified,
3,497,753 images,
12,413,263 words.

2 countries of publication
870,016 named entities disambiguated

More? Check our blog

[info@impresso-project.ch](#) [www.impresso-project.ch](#)
 [impresso](#)
 [@impresso_impresso](#)

[HOME PAGE](#) [SIGN UP](#)

Mining 200 years
of historical newspapers 

How can newspapers help understand the past? How to explore them?

The legal ownership of all content is available to Open Access.
To gain full access

 [IMPROVE YOUR RESEARCH WITH IMPRESSO](#)

and return the signed form to info@impresso-project.ch

impresso application

impresso application



EPFL



University of
Zurich^{UZH}



FNSNF

SWISS NATIONAL SCIENCE FOUNDATION

Visit impresso-project.ch

some backup screenshots if demo does not work

MENU

SEARCH PILL

HUMAN READABLE QUERY SUMMARY

PERSONAL COLLECTIONS

SEARCH RESULTS

FILTERS

This screenshot shows a search interface for historical news articles. At the top, there's a navigation bar with links like 'Search', 'Newspapers', 'Topics', 'Inspect & Compare', 'Text reuse', and 'Collections'. On the right, there's a user profile for 'Maud Ehrmann STAFF' and options for 'FAQ', 'ORDER BY RELEVANCE', 'DISPLAY AS LIST / TILES', 'COMPARE...', 'SAVE / EXPORT...', and a search history dropdown.

The main search area has fields for 'SEARCH ARTICLES', 'SEARCH IMAGES', 'NGRAMS', and 'GROUP BY ARTICLE'. A search term 'arnhem*' is entered, resulting in 4,081 articles found containing 'arnhem'. Below this, there are sections for 'Frontpage' and 'FIND SIMILAR WORDS'. A chart shows the number of articles per year from 1750 to 2010, with a prominent peak around 1945. There are also filters for 'PUBLICATION DATE' (summarized as 'SUM') and 'CONTENT LENGTH' (ranging from 0 to 9,000).

ARTICLE LIST:

- Arnhem nettoyé** (Le Peuple, La Sentinelle, TUESDAY, APRIL 27, 1945 - p.4)
- Arnhem nettoyée** (Le Peuple, La Sentinelle, FRIDAY, OCTOBER 6, 1944 - p.8)
- Nouvelle offensive britannique sur Arnhem** (Le Peuple, La Sentinelle, WEDNESDAY, FEBRUARY 8, 1984 - p.4)
- sun LES vous NAVIGABLES pE nom...** (La lutte syndicale, WEDNESDAY, APRIL 27, 1983 - p.4)

ARTICLE PREVIEW:

Arnhem nettoyé (Le Peuple, La Sentinelle, TUESDAY, APRIL 27, 1945 - p.4)

Des détachements de reconnaissance ont été envoyés dans la ville d'Arnhem. Terres, près de l'Arnhemse de la campagne, a également été attaqué et Le village de Zedelgem a été détruit. Un train allemand le long de Zedelgem a été coulé.

Arnhem nettoyée (Le Peuple, La Sentinelle, FRIDAY, OCTOBER 6, 1944 - p.8)

On annonce officiellement que les Allemands ont été éliminés de l'Arnhem. Les dernières forces de l'ennemi ont été évacuées de la province de la Gueldre, dont elle est le chef-lieu. Arnhem compte 90,000 habitants.

Nouvelle offensive britannique sur Arnhem (Le Peuple, La Sentinelle, WEDNESDAY, FEBRUARY 8, 1984 - p.4)

Après la bataille de Grammichele, les dernières positions des Allemands ont été évacuées de la ville de la Gueldre. La plus grande ville de la province de la Gueldre, dont elle est le chef-lieu. Arnhem, compte 90,000 habitants.

sun LES vous NAVIGABLES pE nom... (La lutte syndicale, WEDNESDAY, APRIL 27, 1983 - p.4)

sun LES vous NAVIGABLES pE nom... 23-28 avril' » Fila Rhénis à Premier jour, lundi Bâle—Arnhem. Avec le TEE « Rheingold » jusqu'à Arnhem. Dîner au wagon-restaurant. Base Dordrecht.

FILTER BY CONTENT LENGTH ⓘ

check one or more content length to filter results

663

FILTER BY LANGUAGE OF ARTICLES (3 OPTIONS) ⓘ

check one or more language to filter results

- French (3,705 results) ⓘ
- German (3,129 results) ⓘ
- Luxembourgish (6 results) ⓘ

FILTER BY NEWSPAPER TITLES (52 OPTIONS) ⓘ

check one or more newspaper to filter results

- Neuf Zürcher Zeitung (1,483 results) ⓘ
- L'Impartial (907 results) ⓘ
- L'Express (742 results) ⓘ
- La Liberté (680 results) ⓘ
- Journal de Genève (607 results) ⓘ
- Freiburger Nachrichten (579 results) ⓘ
- Gazette de Lausanne (401 results) ⓘ
- Die Tat (383 results) ⓘ
- Luxemburger Wort (275 results) ⓘ
- Le Peuple, La Sentinelle (125 results) ⓘ
- L'indépendance luxembourgeoise (121 results) ⓘ
- Obermosel-Zeitung (77 results) ⓘ
- SMUV-Zeitung (65 results) ⓘ
- Confédéré (61 results) ⓘ
- Escher Tageblatt (56 results) ⓘ
- Bürger- und Beamten-Zeitung (48 results) ⓘ
- d'Letzeburger Land (30 results) ⓘ
- Die Gewerkschaft (21 results) ⓘ
- L'Union (16 results) ⓘ
- La lutte syndicale (17 results) ⓘ

MORE... (32 MORE OPTIONS)

FILTER BY ARTICLE TYPE (7 OPTIONS) ⓘ

check one or more article types to filter results

- article (4,575 results) ⓘ
- page article (1,483 results) ⓘ
- advertisement (715 results) ⓘ
- unclassified content (60 results) ⓘ
- obituary (4 results) ⓘ
- tables (2 results) ⓘ

FILTER BY COUNTRY OF PUBLICATION (2 OPTIONS) ⓘ

check one or more countries to filter results

- Switzerland (6,151 results) ⓘ
- Luxembourg (689 results) ⓘ

FILTER BY ACCESS RIGHT (4 OPTIONS) ⓘ

filter results based on access right

- Personal use (4,606 results) ⓘ
- Personal use (no export) (1,083 results) ⓘ
- Public domain (127 results) ⓘ
- not specified (no export) (121 results) ⓘ

FILTER BY ARCHIVE (4 OPTIONS) ⓘ

check one or more data provider to filter results

- Swiss National Library (4,285 results) ⓘ
- NZZ (1,483 results) ⓘ
- Luxembourg National Library (689 results) ⓘ
- Migros (383 results) ⓘ

link toward snippet views

FILTER BY PERSON (8,970 OPTIONS) ⓘ

check one or more persons to filter results

- Willem II (football club) (295 results) ⓘ
- Roda JC (161 results) ⓘ
- SK Sturm Graz (155 results) ⓘ
- Bayer 04 Leverkusen (133 results) ⓘ
- Fortuna Sittard (103 results) ⓘ
- Les Alliés (90 results) ⓘ
- S.L. Benfica (79 results) ⓘ
- Hertha BSC (78 results) ⓘ
- Paris Saint-Germain F.C. (69 results) ⓘ
- Rudolf Mosse (69 results) ⓘ

MORE... (8,960 MORE OPTIONS)

FILTER BY LOCATION (5,395 OPTIONS) ⓘ

check one or more locations to filter results

- Switzerland (1,543 results) ⓘ
- Paris (1,113 results) ⓘ
- Italy (1,094 results) ⓘ
- Netherlands (1,022 results) ⓘ
- Auch (1,021 results) ⓘ
- Berlin (1,017 results) ⓘ
- France (984 results) ⓘ
- Eindhoven (910 results) ⓘ
- Basel (804 results) ⓘ
- Lage (776 results) ⓘ

MORE... (5,385 MORE OPTIONS)

FILTER BY TOPIC (227 OPTIONS) ⓘ

check one or more topics to filter results

- fr match · équipe · ligue · club · saison (1,333 results) ⓘ
- de teil · zeitung · frage · negierung · weise (715 results) ⓘ
- fr front · armée · guerre · ennemi · nord (706 results) ⓘ
- de berlin · münchen · wien · mark · frankfurt (587 results) ⓘ
- de schweiz · frankreich · deutschland · italien · schweden (583 results) ⓘ
- fr match · jeu · but · équipe · minute (515 results) ⓘ
- fr bat · finale · tour · tournoi · match (491 results) ⓘ
- de stadt · schweiz · bern · genf · teil (480 results) ⓘ
- de london · armee · angriff · luftwaffe · general (412 results) ⓘ
- de don · and · min · sage · uli (385 results) ⓘ
- fr armée · colonel · service · major · chef (372 results) ⓘ
- fr suisse · pays · asile · étranger · canton (370 results) ⓘ
- fr août · juillet · septembre · juin · mois (362 results) ⓘ
- de mann · hand · kopf · nacht · gesicht (356 results) ⓘ
- fr monde · concours · record · sport · suisse (343 results) ⓘ
- fr avril · mai · mars · lieu · juin (328 results) ⓘ
- de regierung · berlin · pol · partei · wien (325 results) ⓘ
- de italien · rom · türkei · mailand · griechenland (317 results) ⓘ
- de april · mai · märz · samstag · mittwoch (315 results) ⓘ
- de regierung · paris · konferenz · frankreich · frage (311 results) ⓘ

MORE... (207 MORE OPTIONS)

METADATA

NAMED ENTITIES

TOPICS

Viewer – Article view

Add article to my collection

Media Monitoring Search ... Newspapers Topics Inspect & Compare Text reuse Collections

Impressum of the Part

TABLE OF CONTENTS
117 articles in 8 pages (Personal use (no export))
Add filters from your current search query
1 search filter can't be applied.

Search words...
show only matching articles (no results)

Der „Verräter von Arnhem“ gestorben

LOCATIONS: London, Germany
PEOPLE: Marika Rökk, Kurt Singer
TOPIC:
26.4% DE film · woche · regie · kino · john ·
9.9% DE regierung · afp · sovjetunion · reuter · moskau ·
3.2% DE arzt · mittel · krankheit · gesundheit · wirkung ·
15.6% DE london · armee · angriß · luftwaffe · general ·
3.6% DE gericht · gefängnis · polizei · angeklagt · fall ·
2.5% DE familie · beerdigung · trauer · herr · hirschfeld ·
10.8% DE regierung · berlin · pol · parlament · wien ·
3.6% DE polizei · brand · feuer · nacht · innen ·
2.1% DE juli · juni · august · mai · sommer ·
PROFICIENCY TEST →

London, 27. Juli. (Reuter.) Der Korrespondent des „Daily Express“ in Amsterdam berichtet seiner Zeitung, daß Christian Lindemann, genannt „King-Kong“, der geschickteste Spion, der den Deutschen den Plan über Landung alliierter Luftlandetruppen bei Arnhem verschafft hat, im Spital eines holländischen Gefängnisses gestorben ist, da er eine zu starke Dosis von Schlafmitteln eingenommen hatte. Seine Pflegerin soll ebenfalls unter den Nachwirkungen des Giftes leiden, doch befindet sich ihr Leben nicht in Gefahr. Man glaubt, daß diese Krankenschwester mit Lindemann verlobt war und ihm zur Flucht verholfen hatte. Mit Hilfe eines anderen Gefangenen soll sie ihn durch das Fenster

Annotations relatives à l'article

Grunderkenntnis gegen den Barbarossa-Plan
zur Kontrolle der Atomenergie
New York, 27. Juli. (AFP.) In einer Sitzung der
Atomkommission der Vereinten Nationen
article | short test p.8

Continue reading: TRANSCRIPT

LOCATIONS: New York, Ruth Landes, Washington,
D.C., Cambridge
PEOPLE: John Hancock, John B. Anderson

Amerika über die
Wirtschaftskatastrophe in Ungarn
Washington, 27. Juli. Das amerikanische
Außenministerium gab gestern abend bekannt:
Amerika hat Redi article | short test p.1

Continue reading: TRANSCRIPT

Fac-simile and OCR

Indication of text re-use

“i-button”:
explanation on
processing

What does text reuse mean?
Text reuse detection aims at finding automatically repeated text passages in a corpus, such as multiple copies of the same article published in different newspapers.
[MORE INFO →](#)

Contrast close and distant views

