

Data acquisition Data processing Interface

 Media Monitoring
impresso of the Past

Workshop Session 3

impresso team

Lausanne, EPFL, 5-6 July 2018

Program

8h50 / 9h15	Data acquisition & management	11h00 / 13h00	Presentation of interface (mockups) and work in groups
9h15 / 9h45	OCR correction, lexical normalisation, semantic indexing	13h00 / 13h40	Lunch break
9h45 / 10h30	Historians' use cases	13h45 / 14h15	Collecting and organising
10h30 / 11h00	Coffee break	14h15 / 14h45	Conclusion and outlook

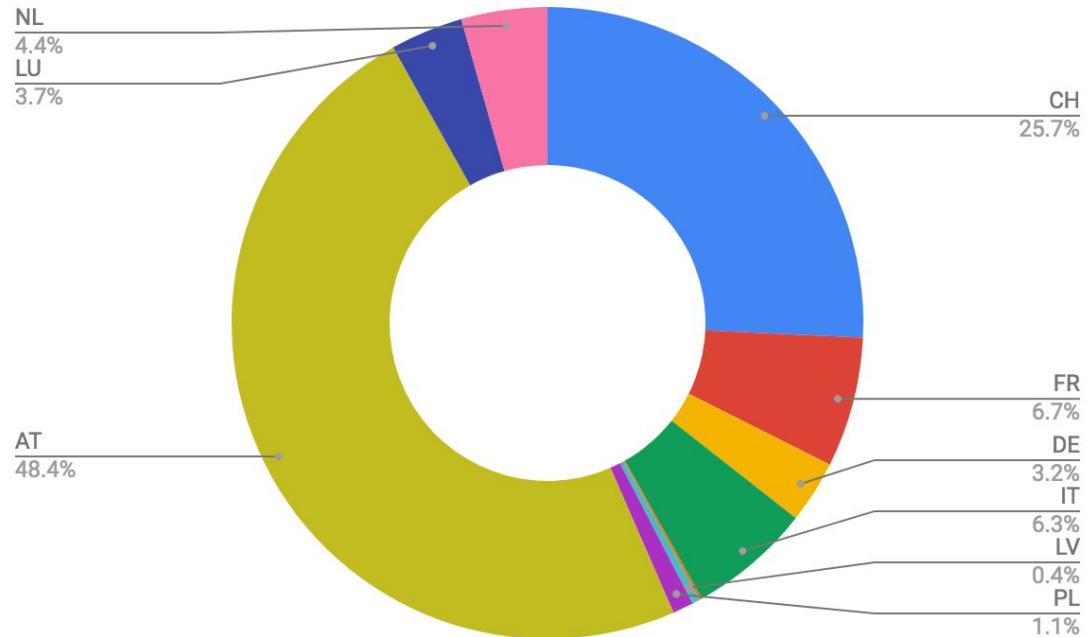
Where are we with newspapers
acquisition?

Potential data providers - distribution by country

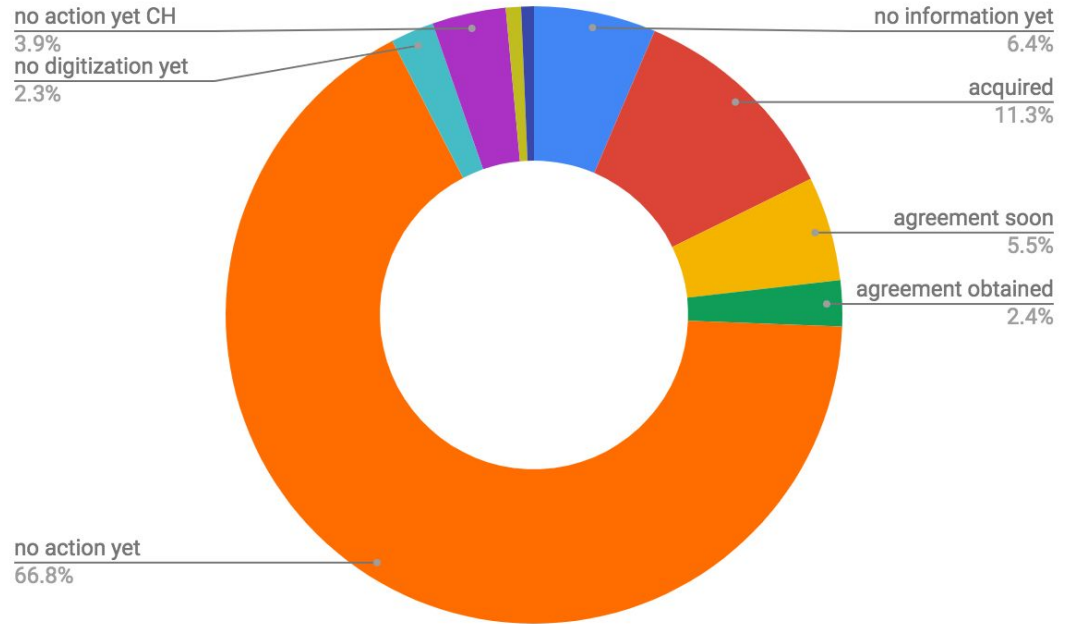
impresso first media inventory

- Fr and De
- 193 titles from CH
- 28 titles from LU
- 529 from Europeana

LFT (Bolzano), NLL (Latvia), NLE (Estonia),
 NLP (Poland), ONB (Austria), SBB (Berlin),
 SUBHH (Hamburg)



Acquisition status - all

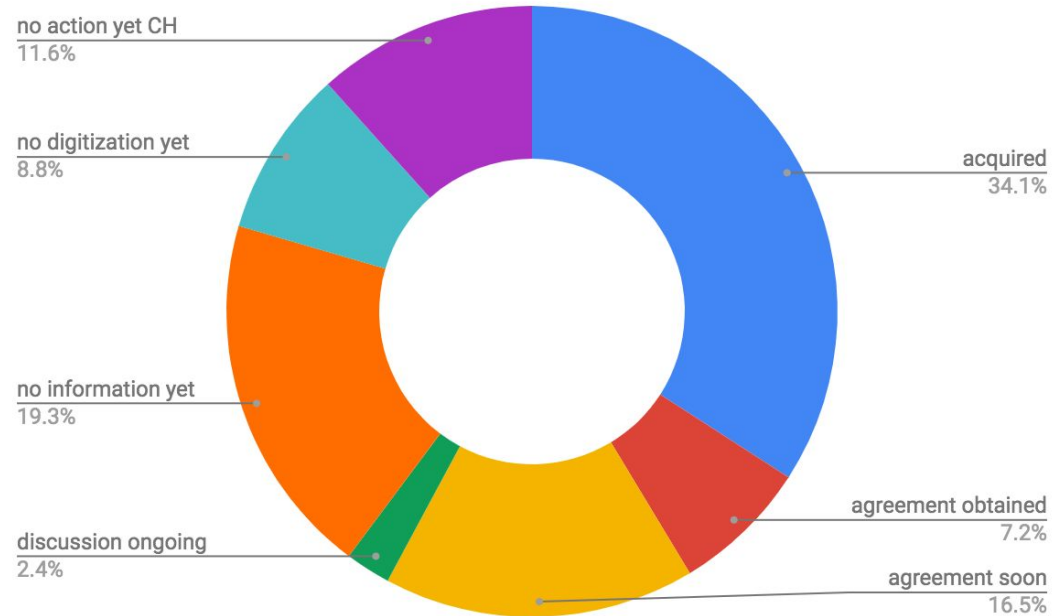


Action taken: 44%

No action taken: 66%

Acquisition status - CH and LUX

Total: 221 titles



cf. [blog post](#) on impresso collection's state

Acquisition status - CH and LUX

Convention signed:

BNL (28), Le Temps (3), NZZ (1)
BCU Fribourg, for public domain titles (9)
Journal Le Confédéré (1)
Socialist party of Neuchâtel (1)
ArclInfo group (2)
Library La Chaux de Fonds (0)
Library Neuchâtel (0)
KB for French titles (32)
SWA (6)
UNIA (Swiss union) (28)

Discussion ongoing:

BCUL
Rhône Médias
SYNA (Union)

No digitization yet:

BV Genève
BC Jura

No action yet:

Alemannic cantons, Migros

Non-disclosure agreements

Different categories of data accessibility

public domain no restriction

semi-closed internal use only, non-transferrable, visible via the public interface, some functionalities require NDA

closed internal use only, non-transferrable, not visible via the public interface, user NDA required

NDA signed at 2 levels:

- at consortium level: data providers ↔ project partners (C2DH, DHLAB, ICL)
- at individual level: project partners ↔ individual users

User access

Level 1: No restriction, full openness

Level 2: Restricted access, user has to send ID card and signed NDA

→ different access rights have to be handled at a technical level.

Next steps

1. closing ongoing acquisitions
2. opening discussions with Alemannic cantons
3. opening discussions with other international institutions

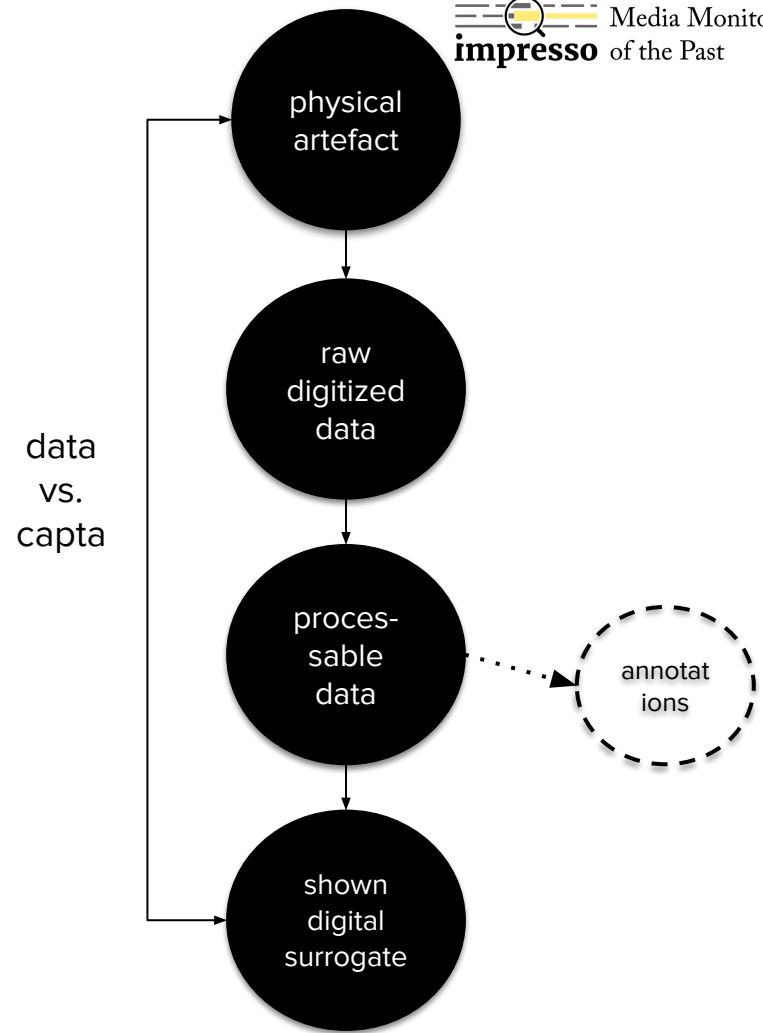
What do we do once we have the data?

A backstage tour

- beyond digitization, primary sources are further *prepared*
- many “hidden” conversion steps happens

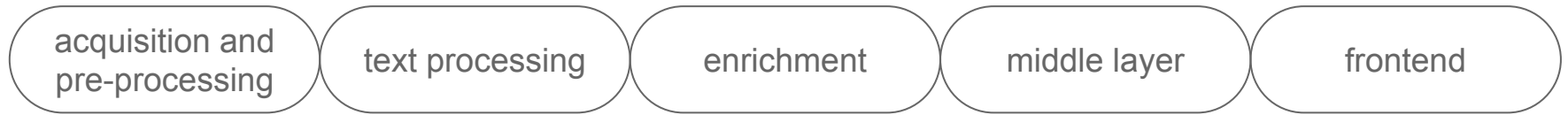
Here, **transparency** as a way to:

- raise awareness about and make comprehensible what happens to sources
- do justice to annoying tasks :)



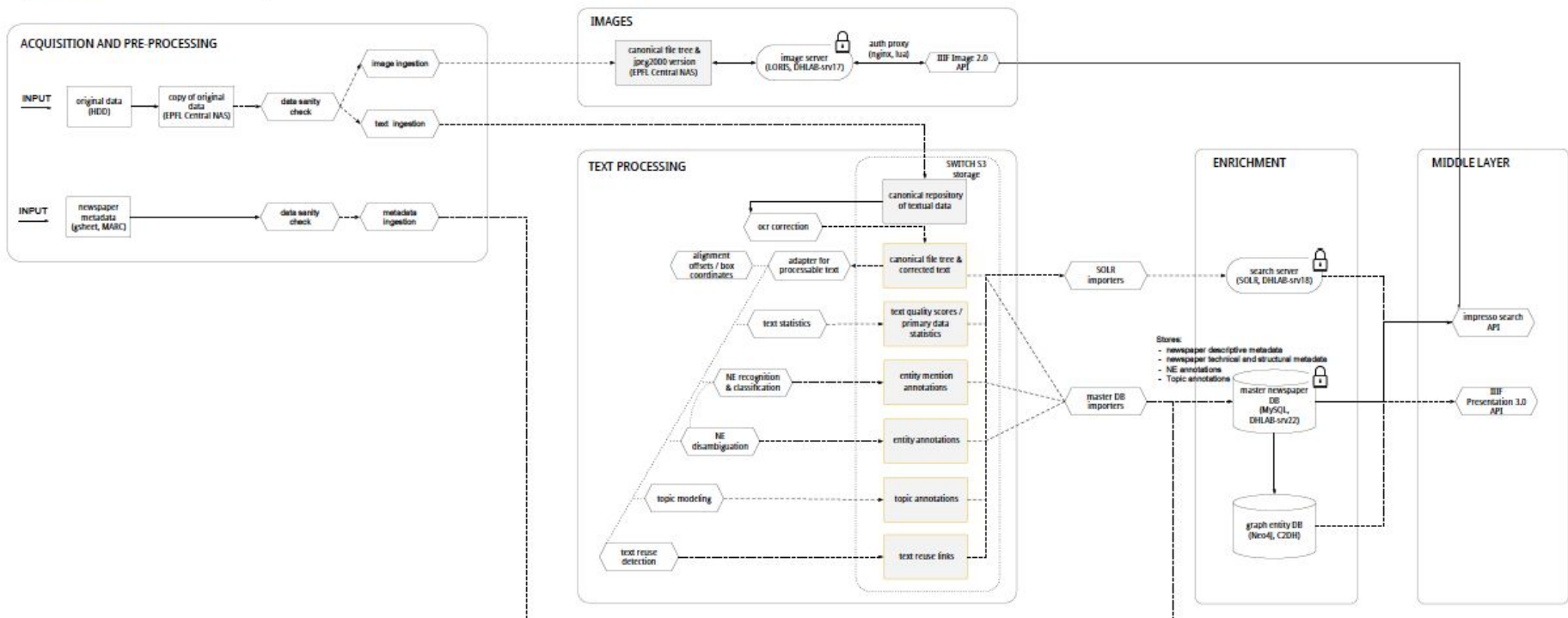
System architecture

How do we organize data, its processing, and visualisation ?



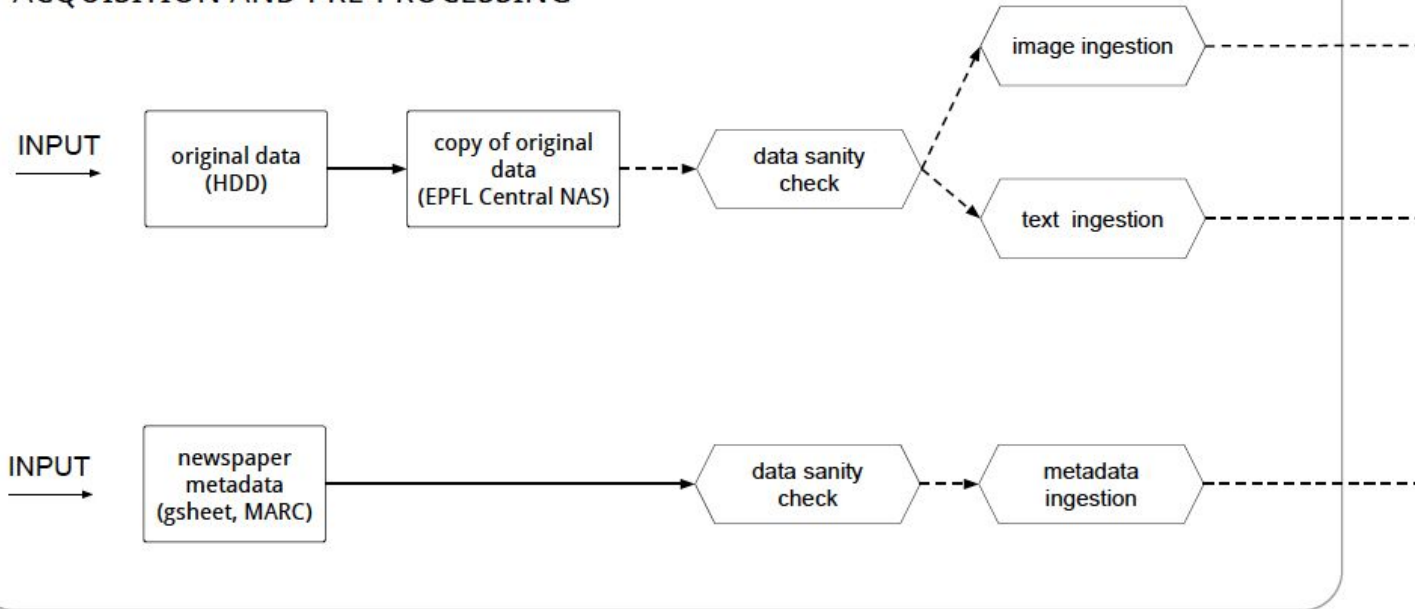
<https://docs.google.com/drawings/d/1puphbUKoOes6n2sb2DA-JrQtw9GPLn0mVfWx6xoNZ6l/edit?usp=sharing>

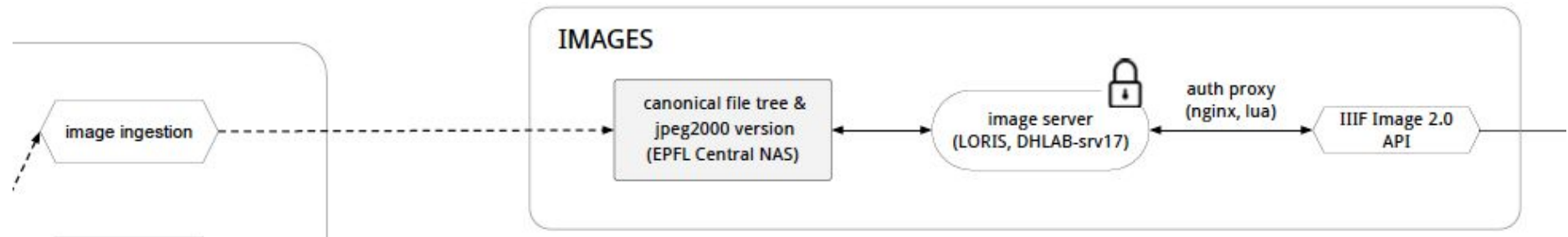
Overview

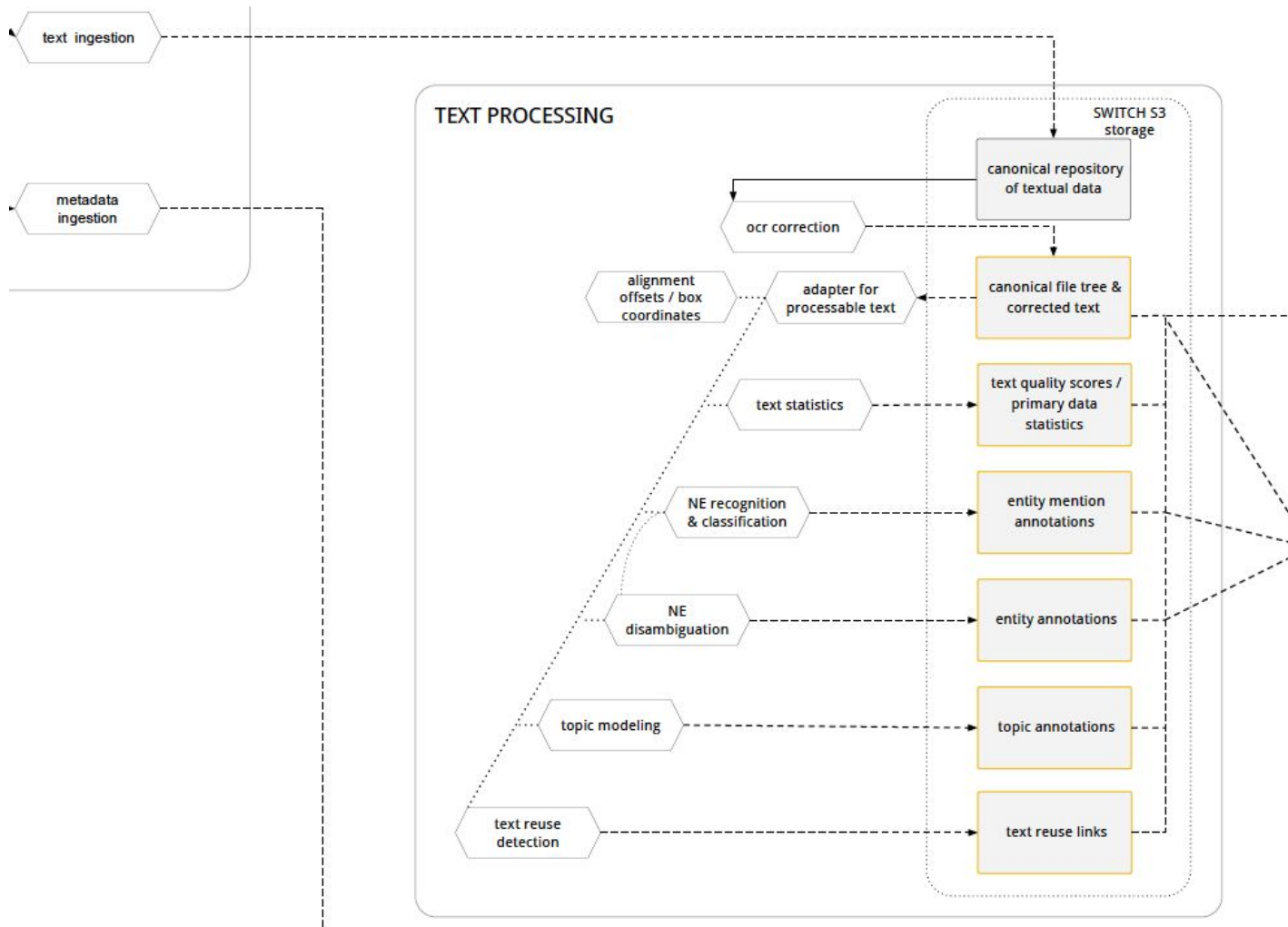


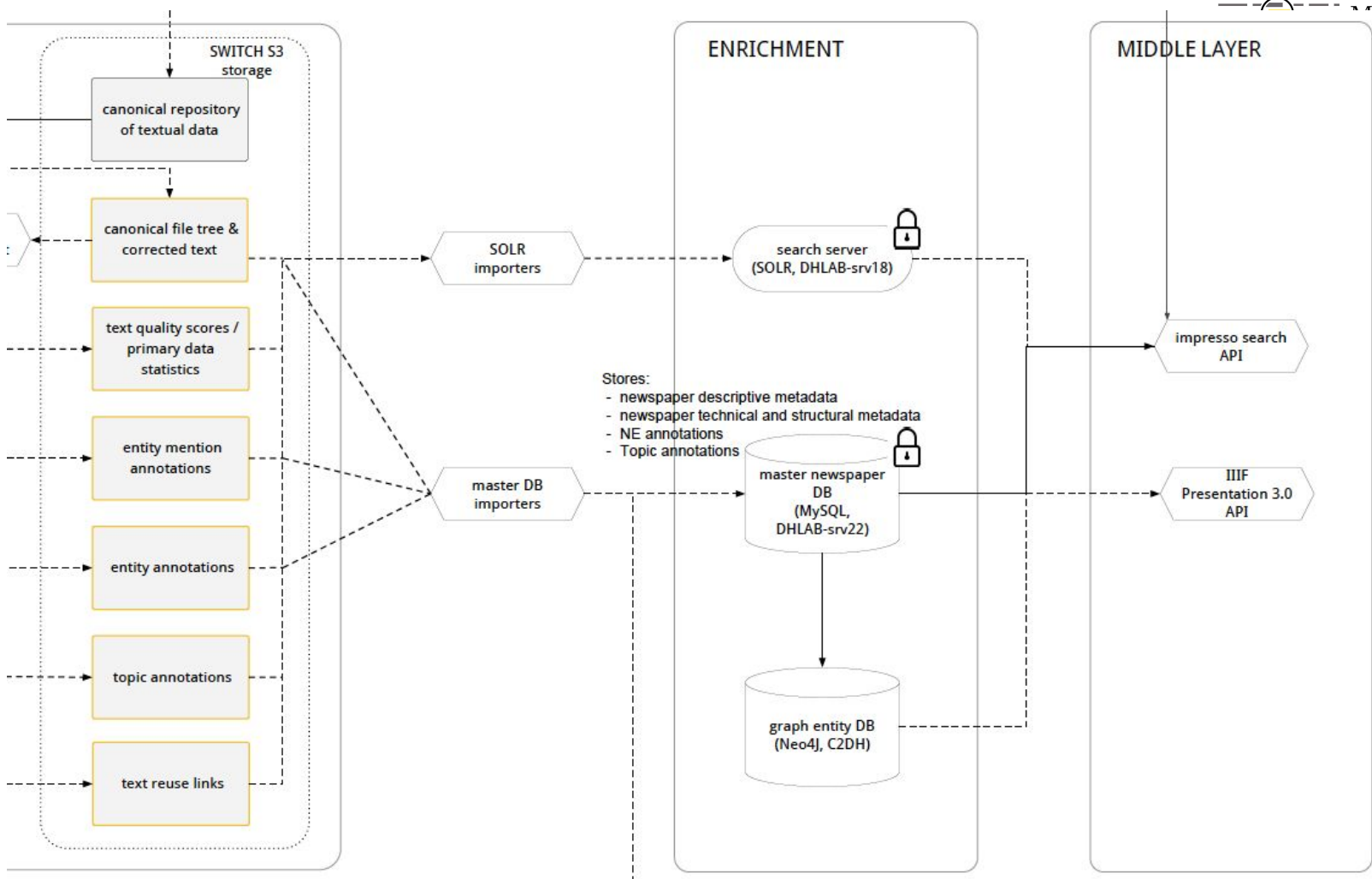


ACQUISITION AND PRE-PROCESSING









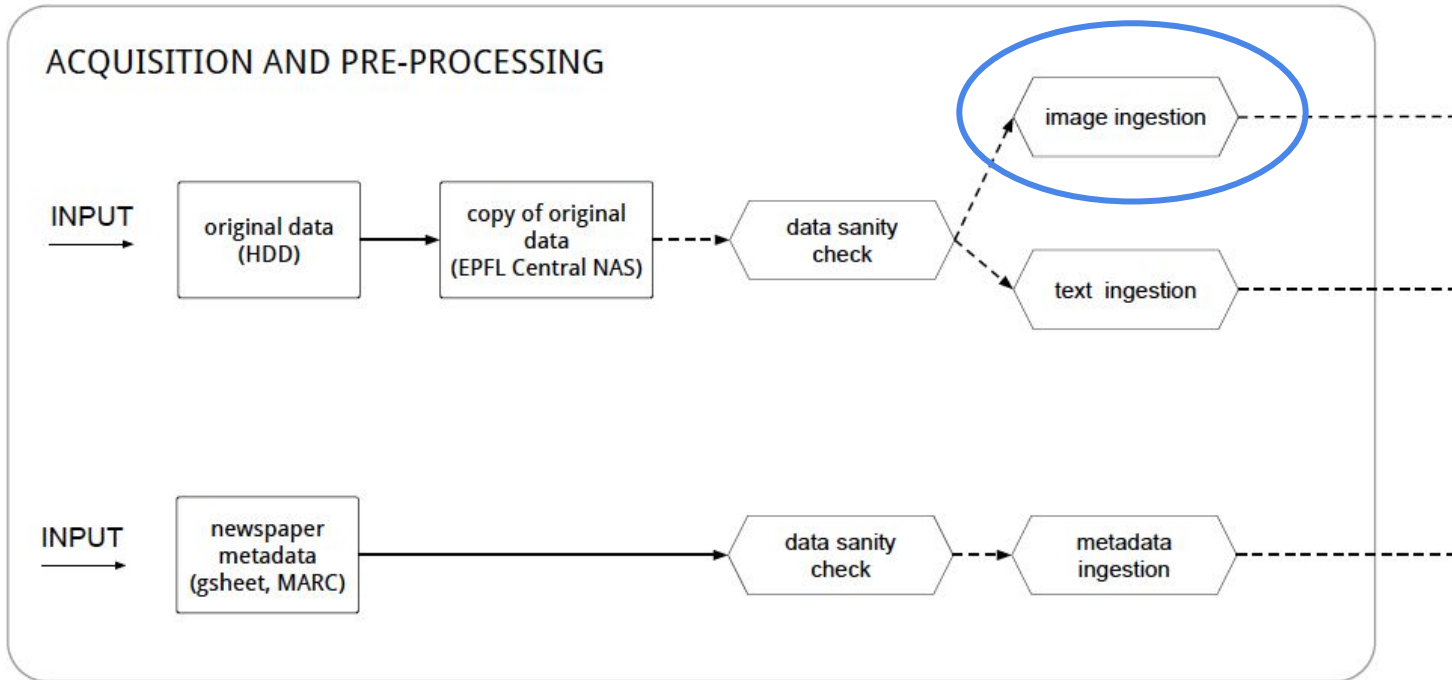


Image processing

given that...	...therefore
not all CH institutions provides IIF APIs	we need to get images on our servers
original files (tiff or jpeg) are heavy and therefore costly	we need to minimize storage cost by reducing image quality
we might need to re-OCR some material	we need to keep a certain image quality
users will intensively visualize images	the viewing experience must be the best possible

What is the best compromise between storage cost, user experience, and OCR performances ?

Image processing

Choice: JPEG2000

- better compression/quality ratio
- encoding of tiles, which can be delivered quickly by image servers

Evaluation

- dataset: 7 pages from GDL 1850/1900/1950 manually transcribed
- test: different images format with different compression ratio
- best tradeoff between image size and OCR performance:
JPEG2000 with compression ratio -10, 70% size gain, >1% OCR performance loss.

Image processing - Difficulties

in one archive, image files:

- can be in different formats (tiff, jpg, png) → different conversions steps
- can have different resolutions → recomputation of box coordinates
- can be missing → recovering strategies, keeping counts

IIIF - International Interoperable Image Framework

A common language for computers to talk about images.

Enable to easily access any part of any page:

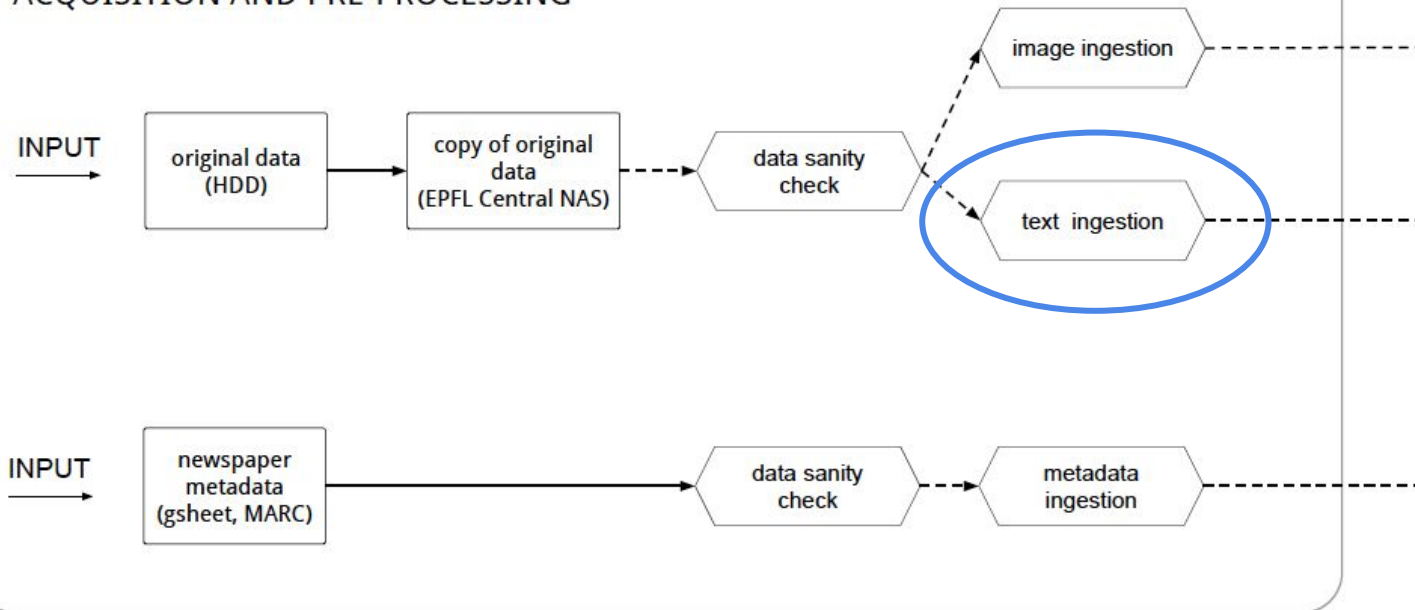
https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1950-04-27-a-p0007/198,5240,131,33/full/0/default.jpg

https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1950-04-27-a-p0007/117,5090,596,863/full/0/default.jpg

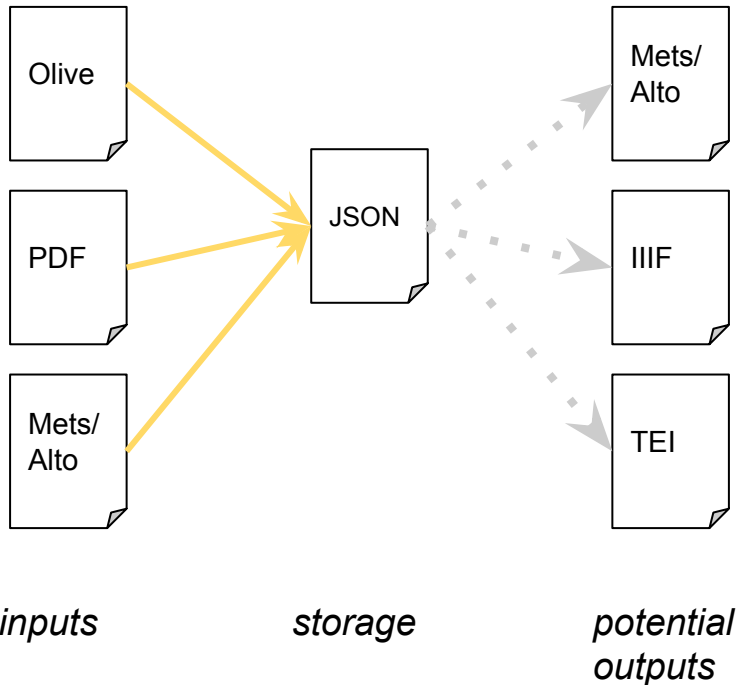
https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1950-04-27-a-p0007/full/full/0/default.jpg



ACQUISITION AND PRE-PROCESSING



Text acquisition



original-canonical-data

Object Count:	1039970
Size:	732.7 GB
Date Created:	Jun 6, 2018
<input type="checkbox"/> Public Access:	Disabled

- 1826
 - 01
 - 02
 - 03
 - 04
 - 05
 - 06
 - 07
 - 04
 - a
 - GDL-1826-07-04-a-issue.json
 - GDL-1826-07-04-a-p0001.json
 - GDL-1826-07-04-a-p0002.json
 - GDL-1826-07-04-a-p0003.json
 - GDL-1826-07-04-a-p0004.json
 - GDL-1826-07-04-a-p0005.json
 - GDL-1826-07-04-a-p0006.json

Impresso canonical JSON

N^o. 54. MARDI 4 Juillet 1826.

GAZETTE DE LAUSANNE ET JOURNAL SUISSE.

-AMERIQUE MÉRIDIONALE.

BOGOTA 9 avril. Le vice-président Santander a adressé au libérateur Bolivar, président de Colombie, la lettre suivante:

« Le vice-président a l'honneur de vous communiquer une nouvelle qui ne peut vous surprendre. Les satellites de la république vous appellent de nouveau, presque à l'unanimité, à la présidence de l'état. La première occasion dans laquelle le peuple de la Colombie a exercé la faculté inappréciable de nommer ses agens, a offert une preuve de sa reconnaissance, de son bon sens et de sa justice. Celui qui a créé la félicité politique de ce peuple devait en être le conservateur.

« Vous êtes appelé à compléter, dans la paix, l'œuvre que votre génie a fondé dans la guerre, et sans vous Colombie ne croit pas qu'elle puisse être élevée au faite de la prospérité et du bonheur.

« Le vice-président de Colombie réunit ses vœux à ceux de ses concitoyens pour vous engager, non-seulement à accepter la présidence, mais encore à voter dans nos bras. Votre présence est importante dans tous les pays; partout votre nom est la terreur des ennemis publics et l'églide des institutions libérales; vous le savez et vous l'avez éprouvé; mais votre pays, le pays pour lequel vous avez prodigué d'innombrables sacrifices, ce pays que vous avez élevé depuis son berceau et soutenu dans ses plus grandes crises, vous appelle et a besoin de vous. »

On parle beaucoup d'un traité qui aurait été conclu entre les états-unis de l'Amérique du nord et les nouvelles républiques du Sud. Ce traité, qui est parvenu en Europe, a pour but, dit-on, d'enlever à l'Angleterre l'influence politique et la suprématie commerciale qu'elle avait acquise dans ces nouveaux états par le moyen des emprunts et de l'exploitation des mines. Il paraît que c'est la connaissance de cette transaction politique qui a provoqué à Londres la baisse des fonds américains et même celle du tiers consolidé.

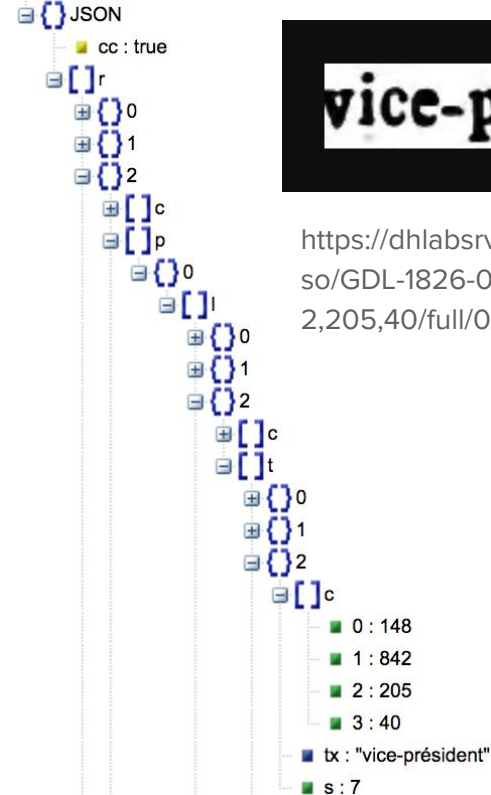
BRÉSIL.

cette nouvelle mesure, qui a déjà coûté la vie à deux sultans, peut réussir sans obstacles, elle sera d'une grande importance pour le salut de l'empire ottoman. L'exemple du vice-roi d'Égypte et les succès obtenus en Morée par des troupes disciplinées ont levé toutes les difficultés que quelques grands de l'empire opposaient encore à ce système, et comme d'ailleurs les privilèges des janissaires et des soldats de la marine leur seront conservés, il paraît qu'il sera généralement admis dans l'empire.

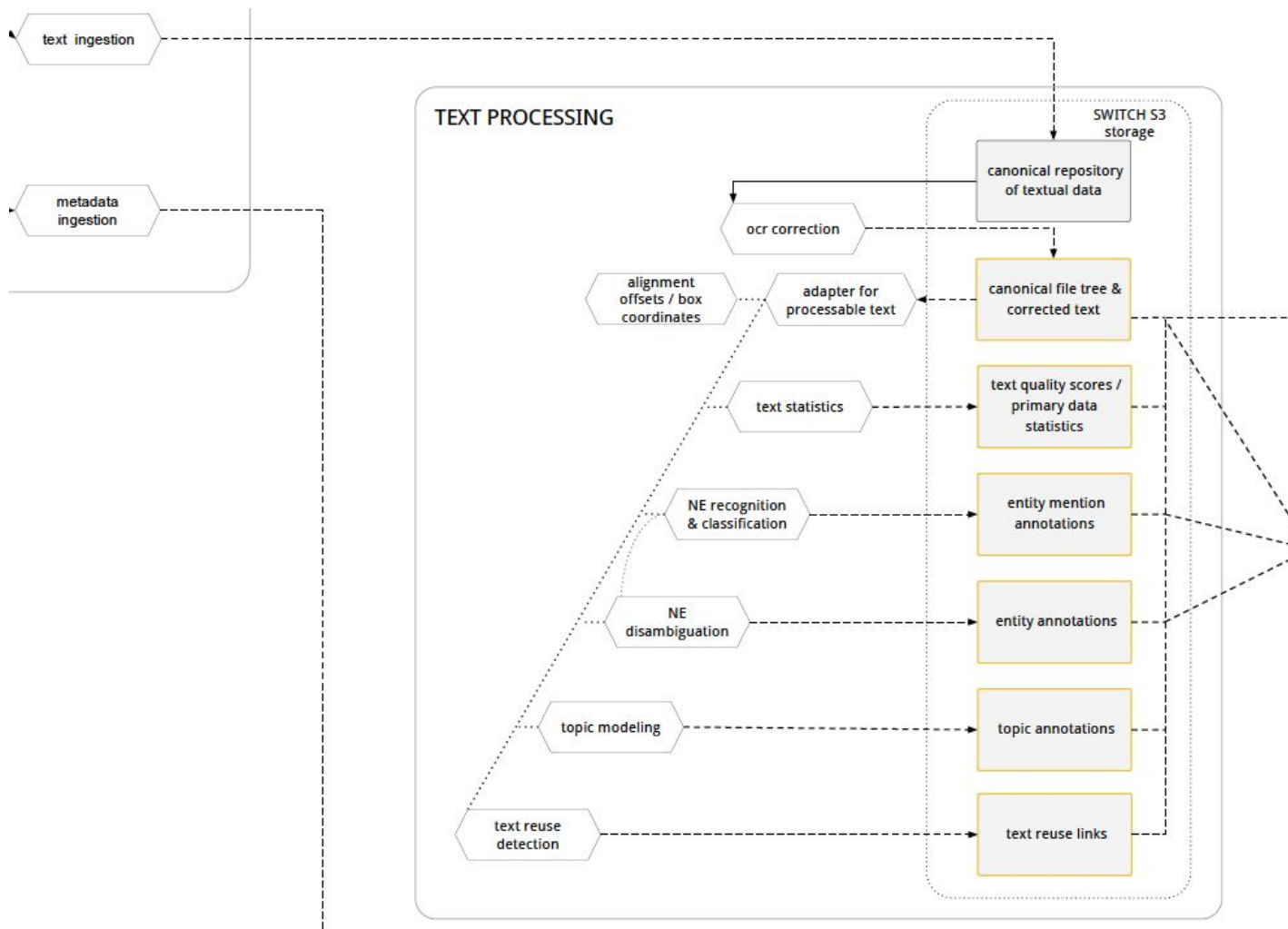
Sir Stratfort-Canning a fait encore une démarche en faveur des grecs, mais en vain. Il offrait, dit-on, au nom de l'assemblée nationale réunie à Epidaure, une soumission conditionnelle et limitée, à l'instar de celle de la Valachie et de la Moldavie, sous un prince que les grecs eussent choisi eux-mêmes. Les turcs n'ont pas même voulu prendre ces ouvertures en considération, quoique M. Canning les eut beaucoup modérées au désavantage des grecs. Il n'y a donc pas ombre d'espoir d'obtenir de cette puissance barbare des conditions tant soit peu tolérables. La soumission absolue ou le massacre général de tous les insurgés, est la seule alternative que laisse la Porte ottomane, tant elle est enfiée par la sécurité que lui donne l'acceptation de sa réponse à l'ultimatum de Pétersbourg.

FRONTIÈRES DE SERVIE 20 mai.

On sait qu'une conspiration tramée contre le gouverneur Mikolosh Obrenovics, fut découverte il y a quelques tems, en Servie. Le fils du fameux Czerny Georges en était le principal chef, et plusieurs individus avaient été arrêtés comme complices. Ceux-ci viennent de subir un supplice qui fait frémir l'humanité. L'instituteur Berissawlewich, Pierre Radossawlewich de Pallank

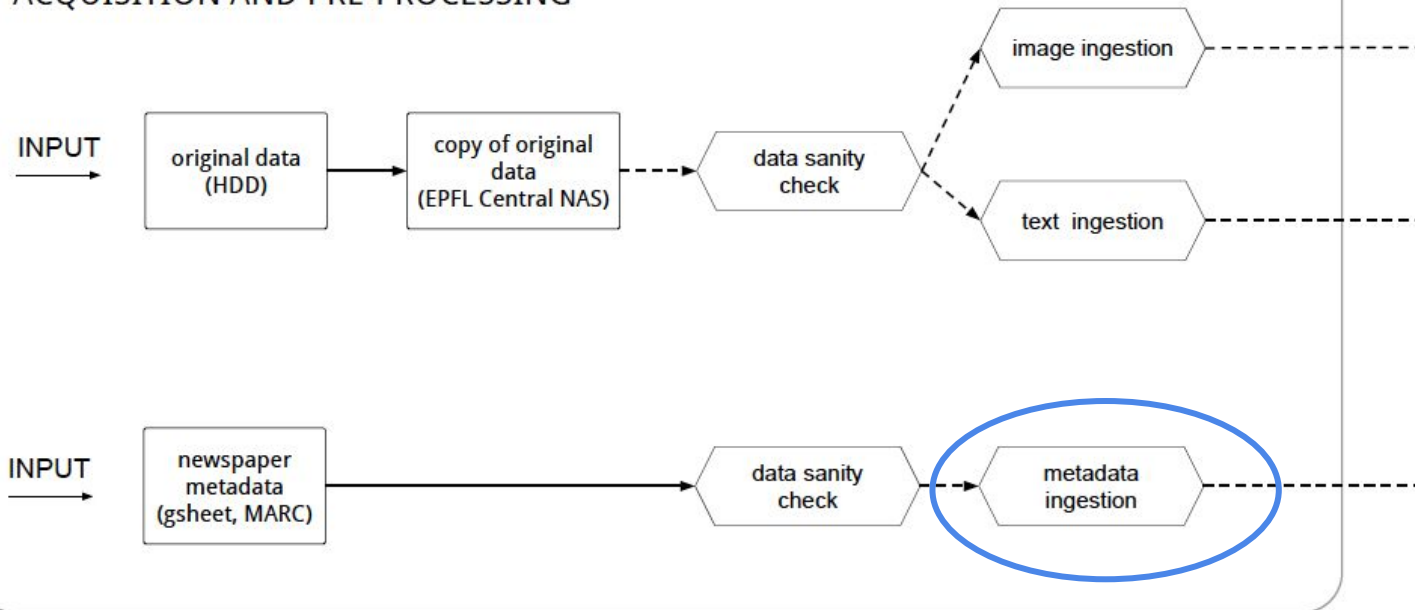


https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1826-07-04-a-p0001/148,842,205,40/full/0/default.jpg





ACQUISITION AND PRE-PROCESSING



Newspaper metadata

What is it useful for?

1. filter
2. provide contextual knowledge

What is it?

What	Example	Provenance	Disputable
hard facts	title, long title, other title, dates, print run, editor, founder, periodicity, etc.	expert knowledge (librarians, historians)	no
soft facts	political orientation, press type, etc.	expert knowledge	yes
computational hard facts	number of issues, of pages, evolution of volume through time	computation	no

Newspaper metadata

Questions

- which information is useful as a filter vs. as contextual knowledge?
- how to collect expert knowledge? what is easy to get?

Current decisions

- library metadata as authority information
- filtering is enabled for most important not disputable facts
- we cannot provide everything but: the infrastructure to enrich/edit/visualize

Newspaper metadata

Current impresso input:

- google spreadsheet, manually filled in
- MARC21 notices from Swiss national library
- BNL notices (to come)

Who would be willing to contribute to edit and discuss metadata?

(e.g. print run, publisher, printer)

```
Quotidien: 17
Wöchentl.: 10
Hebdomadaire: 2
2x par semaine: 4
3x par semaine: 1
2x wöchentl.,: 6
Tägl.: 6
Settimanale: 4
3x par semaine,: 2
6x par semaine: 3
Bi-hebdomadaire: 1
quotidien,: 1
trois fois par semaine,: 2
Quotidien,: 1
quotidien: 2
Wöchentl.,: 1
Mensile: 1
2x all'anno,: 1
Quotidiano,: 2
Quotidiano: 1
12x all'anno: 1
3x la settimana: 1
Halbwöchentl.: 4
3x wöchentl.: 1
Ca. alle 1-2 Jahre: 1
2x monatl.: 1
2-3x wöchentl.: 1
6x wöchentl.: 2
2x wöchentl.: 1
Annuel: 2
Ca. 6x jährl.: 1
Jährl.: 2
```

Archival holes

External factors

1. publication periodicity
2. the journal stopped for some years/months/days

Preservation factors

3. paper copies were lost and never digitized
4. digitization happened but digital copies were lost or damaged

Transparency about

1 & 2: requires external historical knowledge, is often encoded as metadata

3: requires preservation history knowledge, is rarely encoded as metadata

4: detection possible via quality control of digitization process, is almost never encoded.

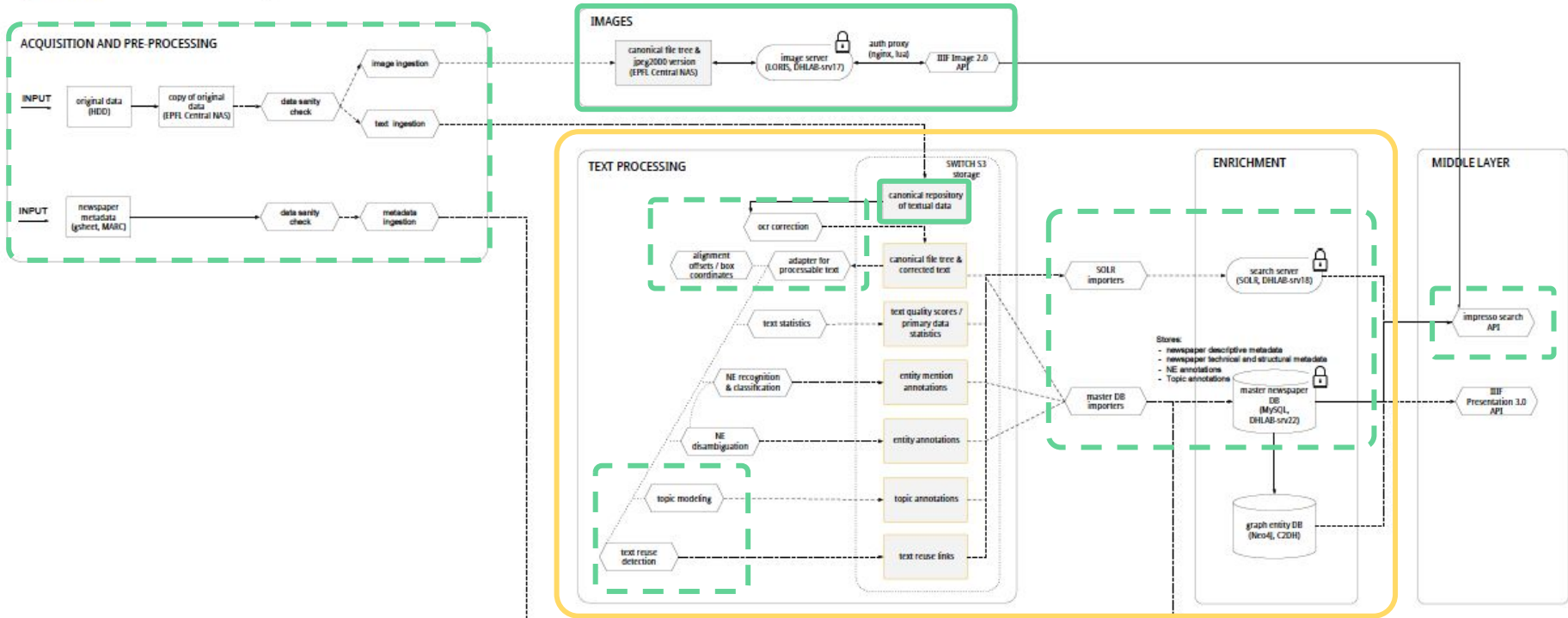
Archival holes

Impresso will encode:

- expected page/issue numbers (what the institution think she has)
- delivered numbers (what exists on the hard disk)
- ingested numbers (what was possible to process)

→ users know a bit better what they see

To sum up: done / ongoing / next

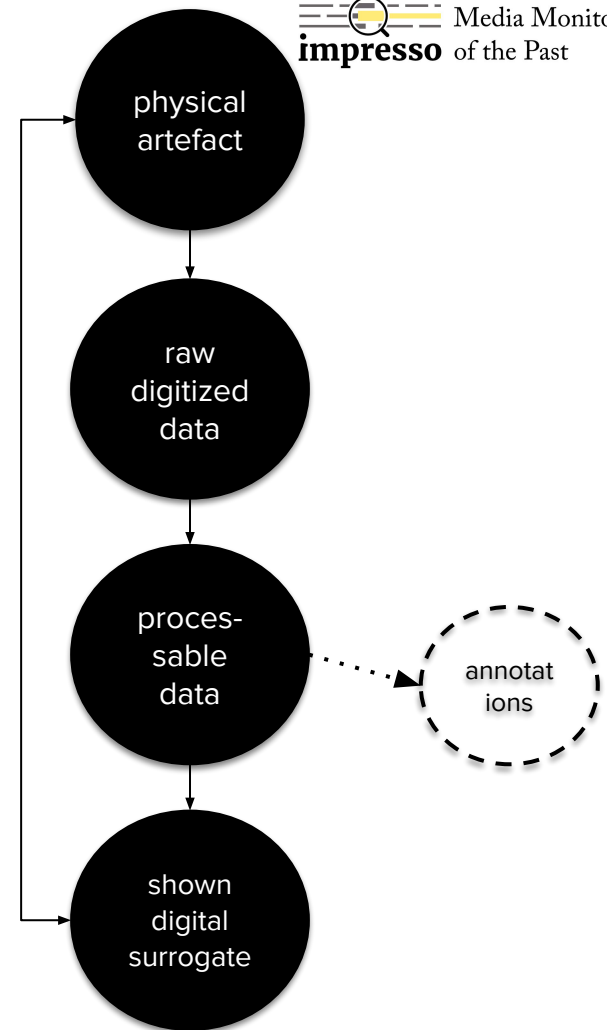




Data vs. Capta

“Differences in the etymological roots of the terms data and capta make the distinction between constructivist and realist approaches clear. Capta is “taken” actively while data is assumed to be a “given” able to be recorded and observed. From this distinction, a world of differences arises. Humanistic inquiry acknowledges **the situated, partial, and constitutive character of knowledge production, the recognition that knowledge is constructed, taken, not simply given as a natural representation of pre-existing fact.**”

J. Drucker (2011), “Humanities Approaches to Graphical Display” in *DHQ* 5(1),
, <http://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html#p3>



Questions?

OCR Quality Assessment

Institute of Computational Linguistics, UZH

The Case of the NZZ



- 2.24 TB (PDFs and XML)
- 1780 - 1950
- ~ 600K pages
- until 1946 in black letter
- digitised from microfilm in 2005
 - = **OCR-2005**

Quality of Images and Original OCR by NZZ

Nr. 4. Zürich Mittwoch



Nr. 37. 1811. Zürcher Zeitung

Nr. 89. Neue Zürcher Zeitung. Dienstag den 29. März 1852.

sagte mir: .Ha. die hauen eben alle ab; ich weide auch bald vrilchwinden! .. Leipzig haltr das festlich!!, «leih an,ir ll>;ss!. daH es für die ?,>!?!<;>; z«r Vrnlnma, hallr, »nd c» war n!ch! schicchi sscardoilrl. !!iil üt'l', "fah»rx!»ch in diesel Farbe lax» man in den Naufhäxrlu eihaltcx u»d viel OrN» ilxii da. Tie (<wststäl!e! l^rnde!?» nür^ nms sie zur U>;e!lNs!»!l!s>; lmtle», und l>:!r Lebr,!?i!»!tel<;;!<;onl>;! für die !llcssl sl,<;fte l

Paris, vom 3. über, wie man sie son der Halle, da Region an Voraban eine heperische Billi stücklichen Verordni gemüßigt. Da dies aus den Worten des Stände dieselben in hört, um so mehr, das Journal de Pa lution einen so glück behaupteten.

Den folgenden D jener Armen Vorkelln, welche als Almosen ge wisse Renten von Stände jährlich zu beziehen hat, einen sehr wesentlichen Dienst erteilten. Bisher wurden diese Renten, nach der eingeführten Bes wohnheit, und weil der Buchstaben ihrer Aus zahlung erst unter Lit. 2. zu stehen kam, sehr spät ausbezahlt. Herr Wilson nahm sich in einer sehr rührenden Adresse an die Versammlung dieser Armen an, und machte die Motion, daß die Renten von 1790, welche den dürftigen Leuten gehören, so gleich mit Anfang Jenner's ausgezahlt werden sollten, besonders auch in der nächsten, weil dieser Monat wegen der rauhen Jahreszeit den Mangel an empfindlichsten zu machen pflegt. Da die ganze Summe nur ungefähr 300000 Francs betragt, so könnte eine solche Einigkeit in den zur Zahlung bestimmten Fonds für diese Zeit keine sehr aufwendige Veranordnung erfordern. Die Stände haben diese Seligschheit mit Freude er

und die untertänigen Zeugn sten Orten die Diener des Volkes sind. Das Depart bei den Ständen wegen före. Berichtbarkeit; unter wische. Sanctionsaus und Gelege predigt, sehr nach Gemacht. Die Stände da Departaments der Untersta geben. Aber auch selbst in Mit lung scheuen sich französisch Angehörigen gegen die Staat sehr tragende Art zu bezeug sich darnm eine sehr stütz schaf, von Clermont best glaubte, er würde den de ed leisten. Allen wie gro und der Unwillen der Stän

eine der hauptwichtigsten... Dresden, 16. April. Die Stände... Der Herr... Ein förmliches Ereignis,

Dresden, 16. April. Die Stände... Dresden, 16. April. Die Stände... Dresden, 16. April. Die Stände...

entworfene Frage ist, daß die... Die Stände... Die Stände... Die Stände...

Die Stände... Die Stände... Die Stände... Die Stände... Die Stände...

Die Stände... Die Stände... Die Stände... Die Stände... Die Stände... Die Stände...

How Big is the Problem?

Systematic evaluation needed!

Transkribus v1.3.7 (23_11_2017_14:37), Loaded doc: nzz_gold, ID: 30948, Page 142, file: nzz_19210406_0001_0001.tif [Image Meta Info: (Resolution:474...

Search current document... 142 /167 In Progress

Server Overview Layout

Logout simon.clematide@gmail

Document M User Manag

Versions Jobs

Recent Documents: User activit

Collections:

phillip.stroebel@gmail.com Collectio

ID	Title
35798	nzz_gold_ocr_only
30948	nzz_gold
30742	nzz_test
30741	Test Document

100

Region: 16 /42

TR L BL W ... H V L

№ 506
 Zweites Wtttagblatt.

Di. 11. Saison — Reise — Ver

Touristik und Reklame.

unseres Gesandtschaften, Konsulate, Schweizer-
 vereine usw. sich zur Mtarbett und Unter-
 stützung herbeilassen.

Ein Schweizerischer Verkehrsmanu schreibt
 uns:
 H. S. Die anregenden Ausführungen Ihres
 Korrespondenten V. P. in Nr. 9 der „Saison-
 Beilage“ vom 18. März, geben mir Anlaß, zu
 diesem Thema ebenfalls einige Gedanken zu
 äußern und bitte ich die Redaktion höf-
 lichenfalls meine Ausführungen geif. Aufnahme zu ge-
 wahren.

Als ehemaliger langjähriger Angestellter der
 Firma Thos. Cook u. Son, bei der ich mich in
 verschiedenen ihrer zahlreichen Filialen be-
 tätigte, hatte ich Gelegenheit, mit dem Reise-
 publikum in Kontakt zu kommen. Gleich diesem
 Einander hatte ich es für absolut ausgezeich-
 net, daß ein Unternehmen wie die Firma Thos.

unseres Gesandtschaften, Konsulate, Schweizer-
 vereine usw. sich zur Mtarbett und Unter-
 stützung herbeilassen.

Erste Aufgabe dieser Bureaus müßte es sein,
 auf das reisefähige Publikum des betreffenden
 Landes zugunsten der Schweiz einzuwirken und
 natürlich auch mit dem durchreisenden Publi-
 kum Rührung zu nehmen; ihnen müßte die
 Propaganda in jenem Lande überlassen werden,
 wodurch mancherlei Redcliffe unterbleiben
 und selber erbracht würden, welche in Er-
 mangelung zuberlässiger Auskunft, jetzt un-
 nütz ausgehen werden; die so getreteten Mit-
 tel könnten fruchtbringender für die Schaffung
 eigener Vertretungen verwendet werden.

Die Ausgabe von Fahrtscheinheften in diesen
 Bureaus halte ich nicht für zweckmäßig, weil
 mit den Interessen der Reisebureaus koll-

Eisenbahnverhältnisse im Glarner u.
 alle Zeiten ergibt sich.

Gegen einen Ausbau der Station
 sprechen folgende Argumente: Bei d
 nehmenden Verkehr gegenüber hat sic
 der Station Siegelbrücke schon läng
 ungenügend und geraden gejährli
 Drei bis vier Bäge treffen dort tägli
 auskommen, so daß es recht schwerli
 auf dem äußert beschränkten Raume
 liegenden Reisenden in geeigneter
 leiser Weise aufzufassen, da keine Stai
 vorhanden sind. Nicht viel besser si
 hältisse der Station Wesen. Ein ra
 dem konstant größer werdenden Verte
 Umbau der Station Siegelbrücke ist
 besser fastleute wegen des beschr
 feldungsraumes zwischen dem beiden
 Staatsstraßenübergängen und den b
 brücken rein unmöglich. Auf dieselbe
 leiten müß ein Ausbau der Station S

1 unsere Gesandtschaften, Konsulate, Schweizer-
 2 Vereine usw. sich zur Mtarbett und Unter-
 3 stützung herbeilassen.

Creating a Ground Truth

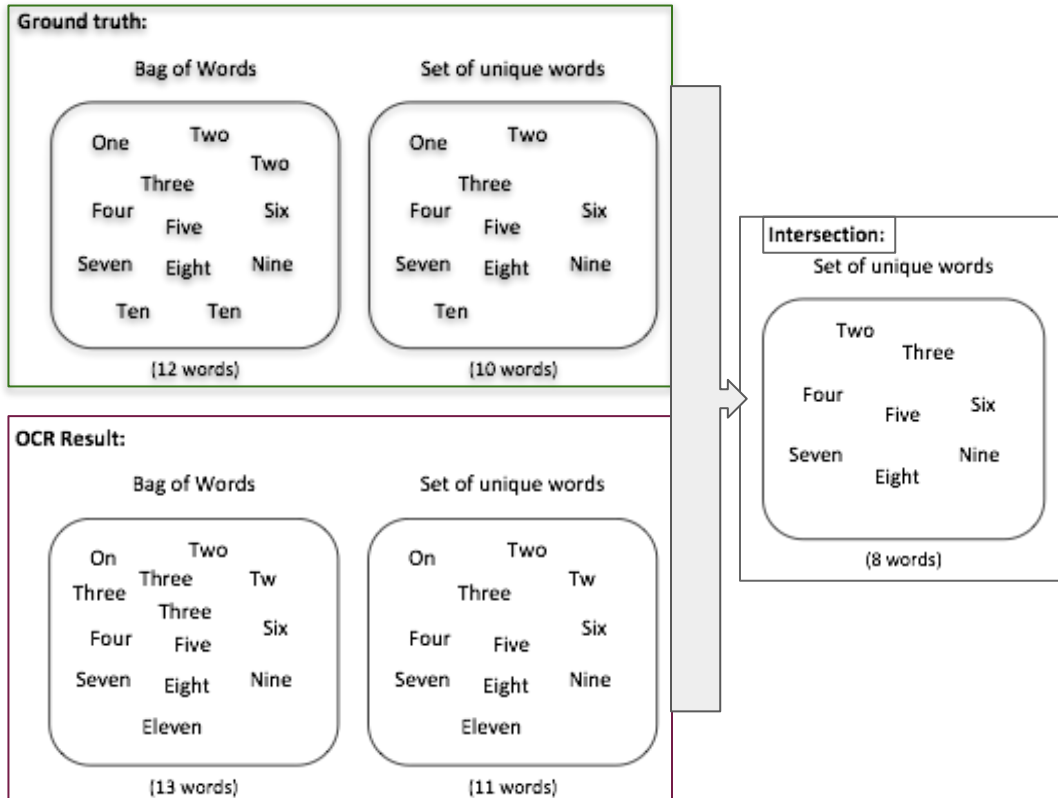
- One randomly selected front page per year (1780-1946)
- Current status: 150 pages
- re-OCR-ised with Transkribus-internal ABBYY Finereader
 - = **OCR-2017**
- Good training material for OCR engines
 - e.g. READ Handwriting Text Recognition (HTR), which we trained from within Transkribus
 - based on Recurrent Neural Networks (RNN)
 - used 112 pages for training
 - performance: approx. 2% Character Error Rate
 - = **HTR-2018**

Compare **OCR-2005** vs **OCR-2017** vs **HTR-2018**

Evaluation with PRima Text Eval 1.4

- Goal: evaluate based on words, not layout
- Edit distance could be severely distorted because of layout recognition issues (e.g. different text box order)
- Solution: evaluation on Bag-of-Words (BoW)
- BoW = how many times does a word appear on a page?

Example



- Recall = $8/12 = 67\%$
- Precision = $8/13 = 62\%$
- F-Measure: harmonic mean of Recall and Precision
 - F-Measure = $2 * R * P / (R + P)$
 - F-Measure = 64%

OCR Quality - The Differences (cont.)

OCR-2017

Prozeß der Marie Antoinette. Nachdem dieselbe am 15. Weinm. alten Stvls, oder am rz. des ersten Monat« im 2, en Jahre der Republik neuen Kaleuderstpls, in den Audienzsaal eingesührt wurde, und sie sich auf den Sessel niedergelassen hatte- fragte sie der Präsident: Wie sie heis-se? „ Ich nenne mich, antwortete sie, Marie Antoinette von Lotharingen, Oestreich in. — Wer seyde ihr ?.. Ich bin dir Wittwer Ludwig Capet«, ehemaligen Königs der Franzosen. — Wie alt? 28 Jahre. — Nun wurde von dem Ge-richtsschreiber die Auklagsakte vorgelesen. Darin« heißt e«,daß aus den dem Tribunale rnhandengestellten Schriften erhellet 'Daß gleich den Messalinen Brunehaut, Fredegoude»nd Medizi«, die man einstKöniainnea von Frankreich ge-nannt habe, und deren verhaßte Namennie au« de» Jahr-büchern der Geschichte werden vertilgt werde» , Marie An-toinette , Louis Capet, Wittwer, seitdem Aufsthaten in Frankreich

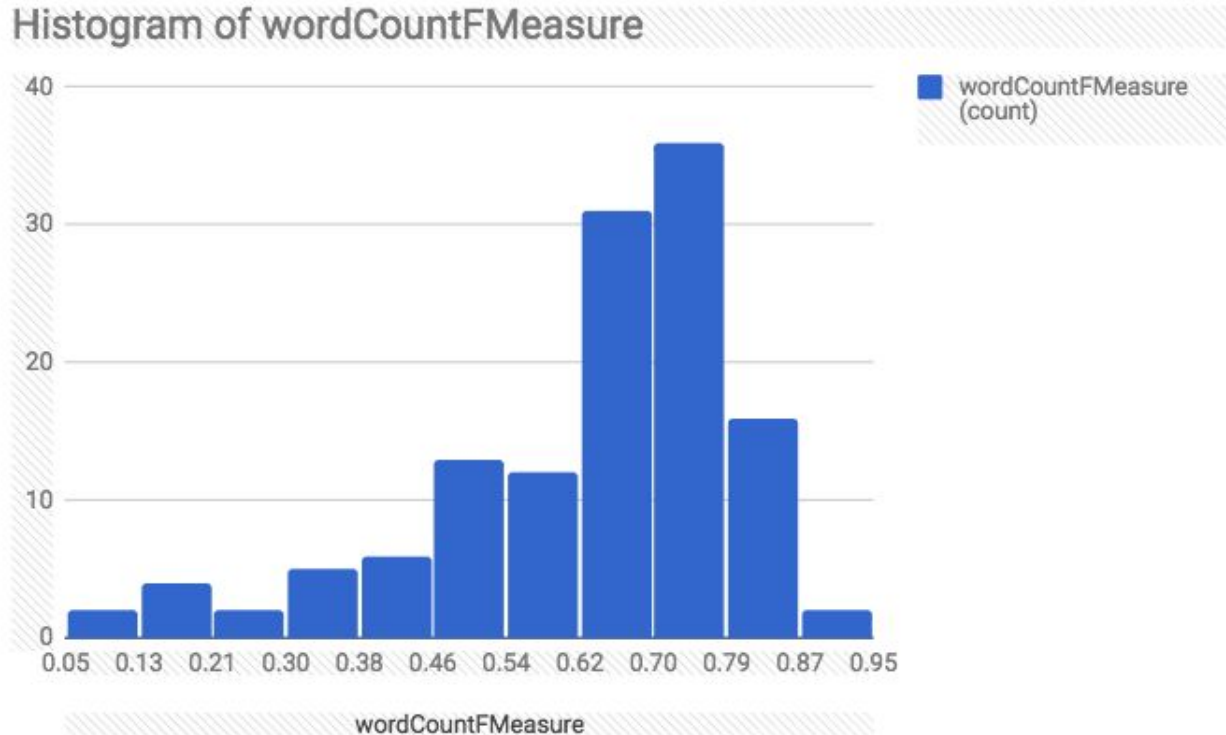
F-Measure: 0.684

HTR-2018

Prozeß der Marie Antoinette. Nachdem dieselbe am 15. Weinm. alten Stvls, oder am 23. des ersten Monats im 2ten Jahre der Republik neuen Kalenderstpls, in den Audienzsaal eingeführt wurde, und sie sich auf den Sessel niedergelassen hatte, fragte sie der Präsident: Wie sie heis-se? „ Ich nenne mich, antwortete sie, Marie Antoinette von Lotharingen, Oestreich in. — Wer seyde ihr ?.. Ich bin dir Wittwer Ludwig Capets, ehemaligen Königs der Franzosen. — Wie alt? 38 Jahre. — Nun wurde von dem Ge-richtsschreiber die Anklagsakte vorgelesen. Darin« heißt es, daß aus en dem Tribunale zuhandengestellten Schriften erhelle: Daß gleich den Messalinen Brunehaut, Fredegonde und Medizis, die man einst Königinnen von Frankreich ge- naunt habe, und deren verhaßte Namen nie aus den Jahr-büchern der Geschichte werden vertilgt werden, Marie Antoinette, Louis Capet, Wittwer, seitdem Aufsthaten in Frankreich

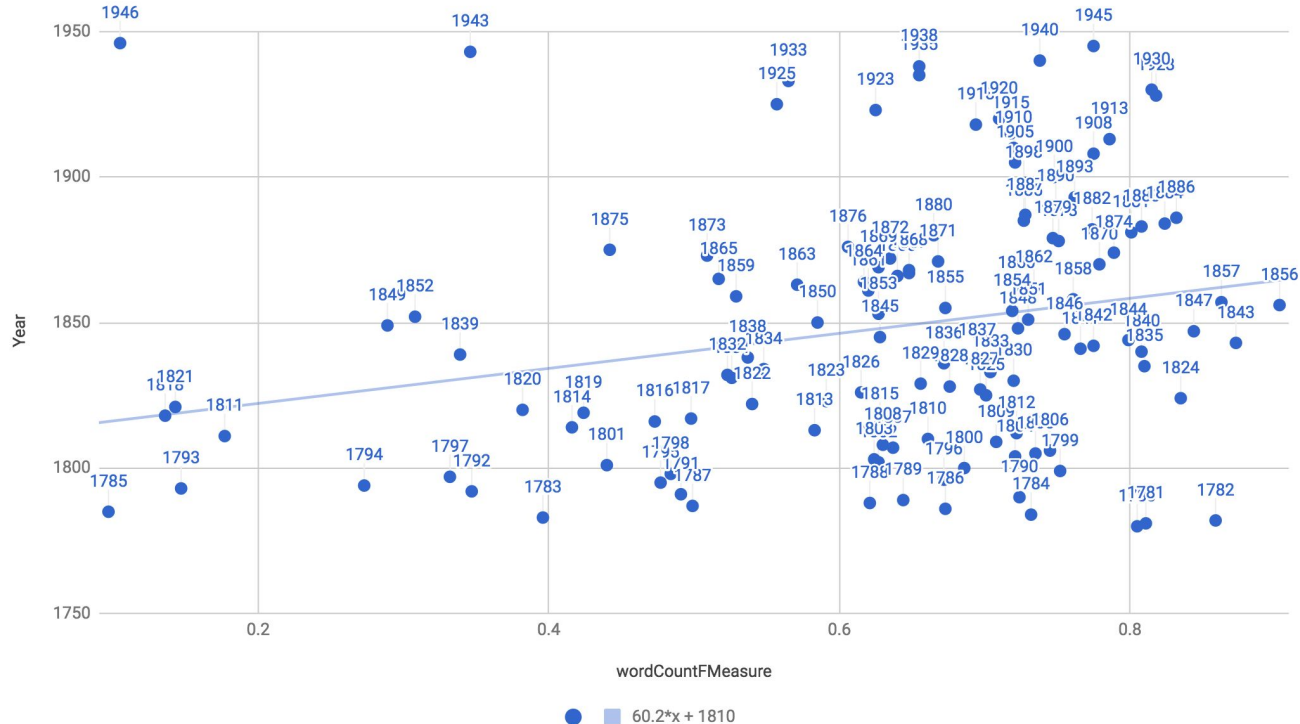
F-Measure: ~0.95

Original OCR Quality (OCR-2005)



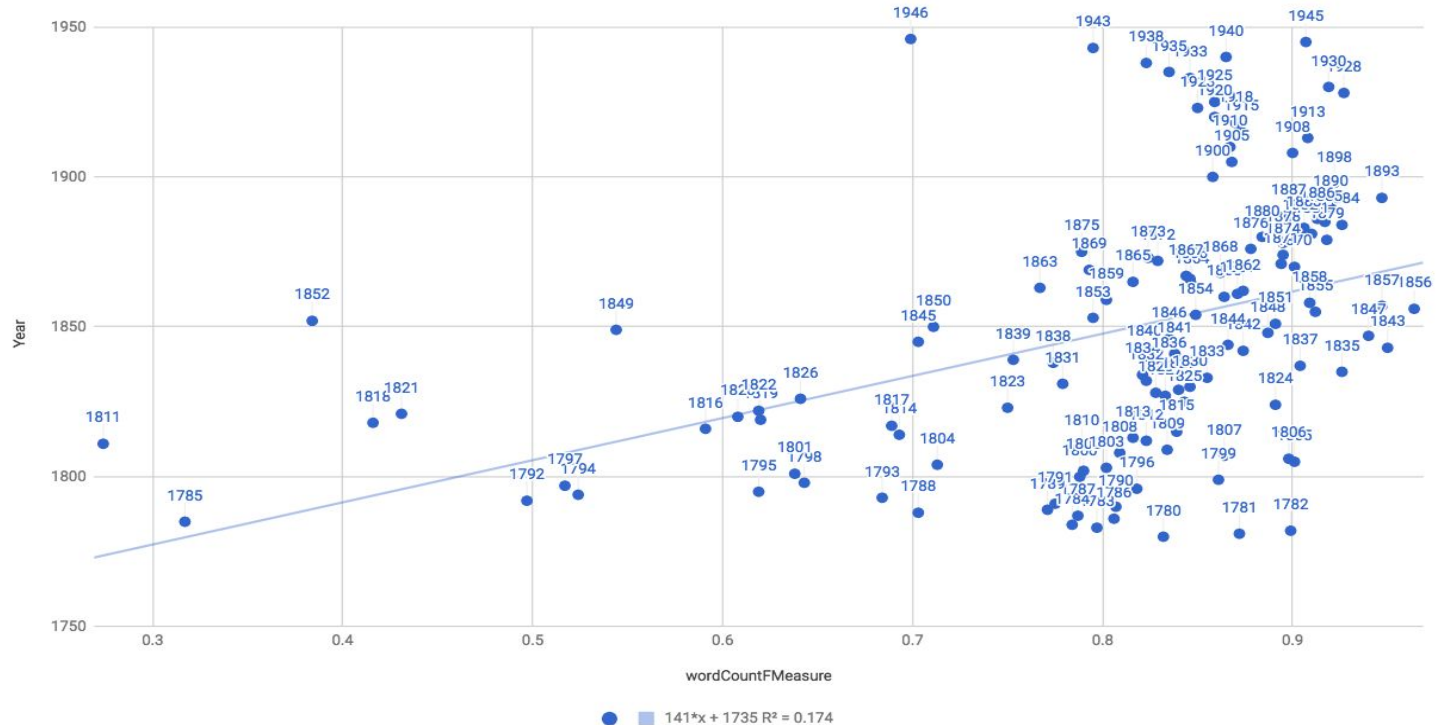
Original OCR Quality (OCR-2005)

Year vs. wordCountFMeasure

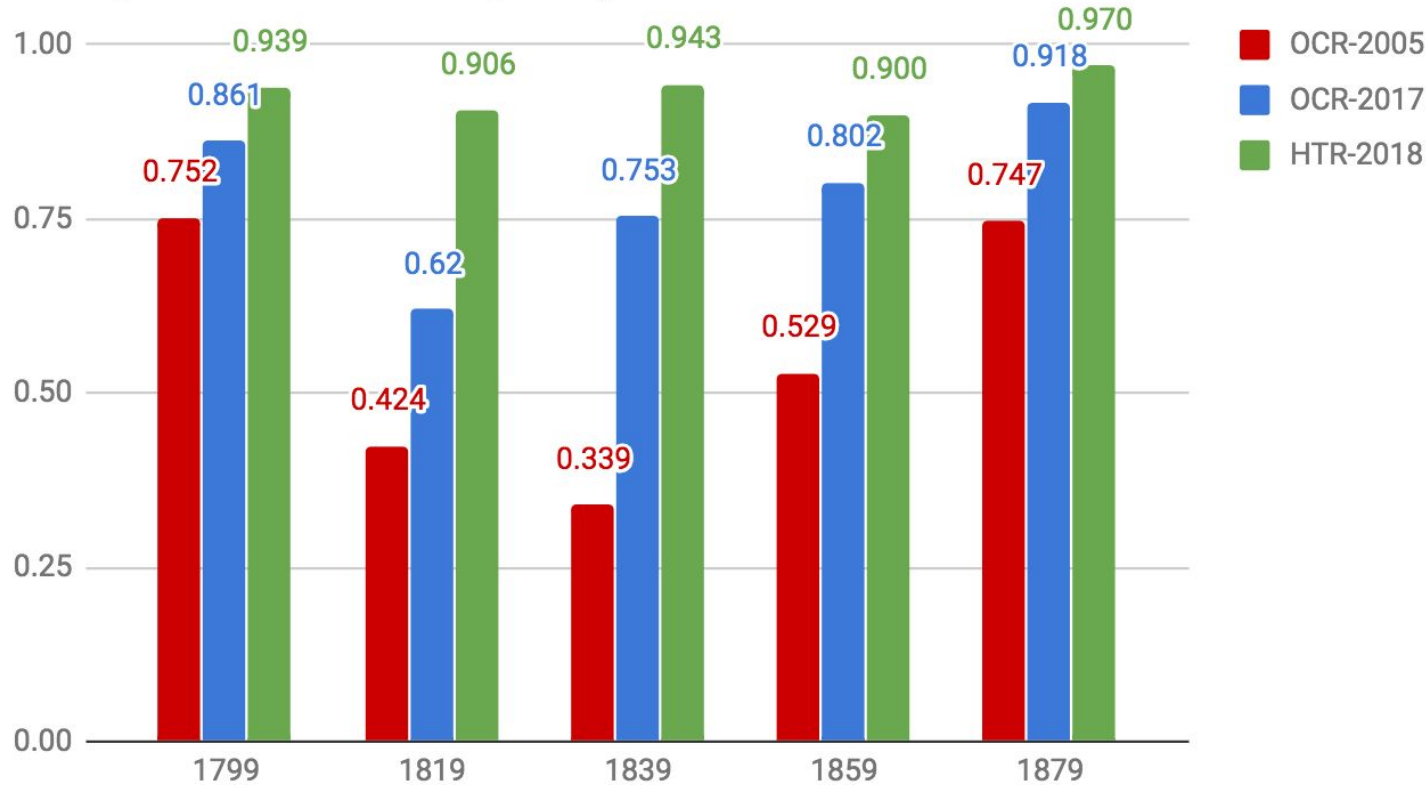


ABBYY OCR-2017

Year vs. wordCountFMeasure



Comparison of OCR Quality



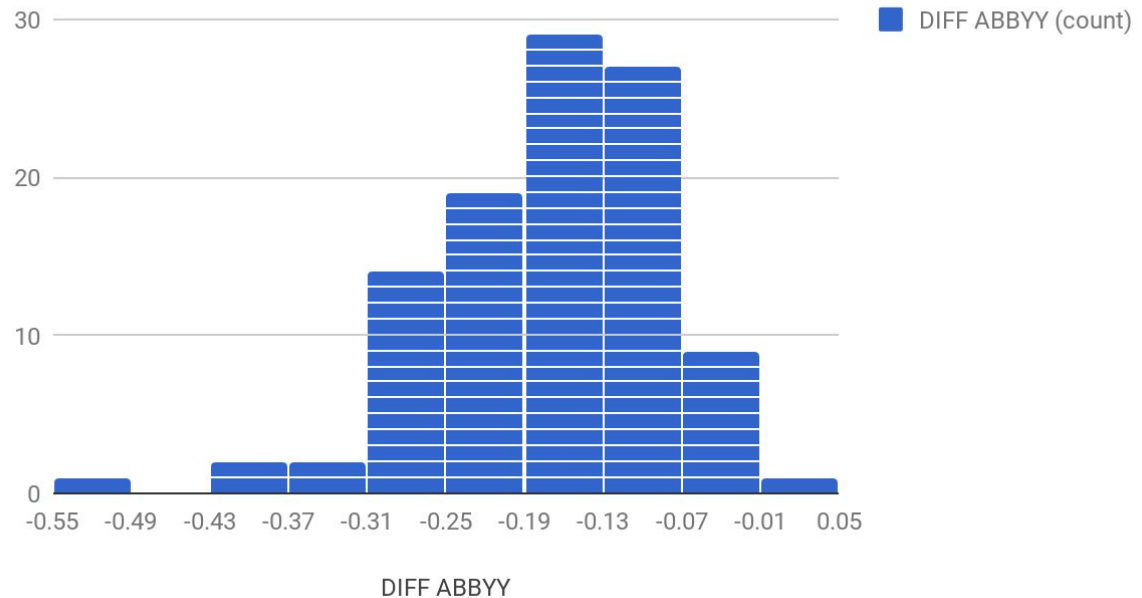
Improvement Re-OCR-ing vs Current Quality

How much better is the
ABBYY OCR-2017 quality vs
the original OCR-2005?

On average:

17 percentage points better!

Histogram of DIFF ABBYY



One Step Further ...

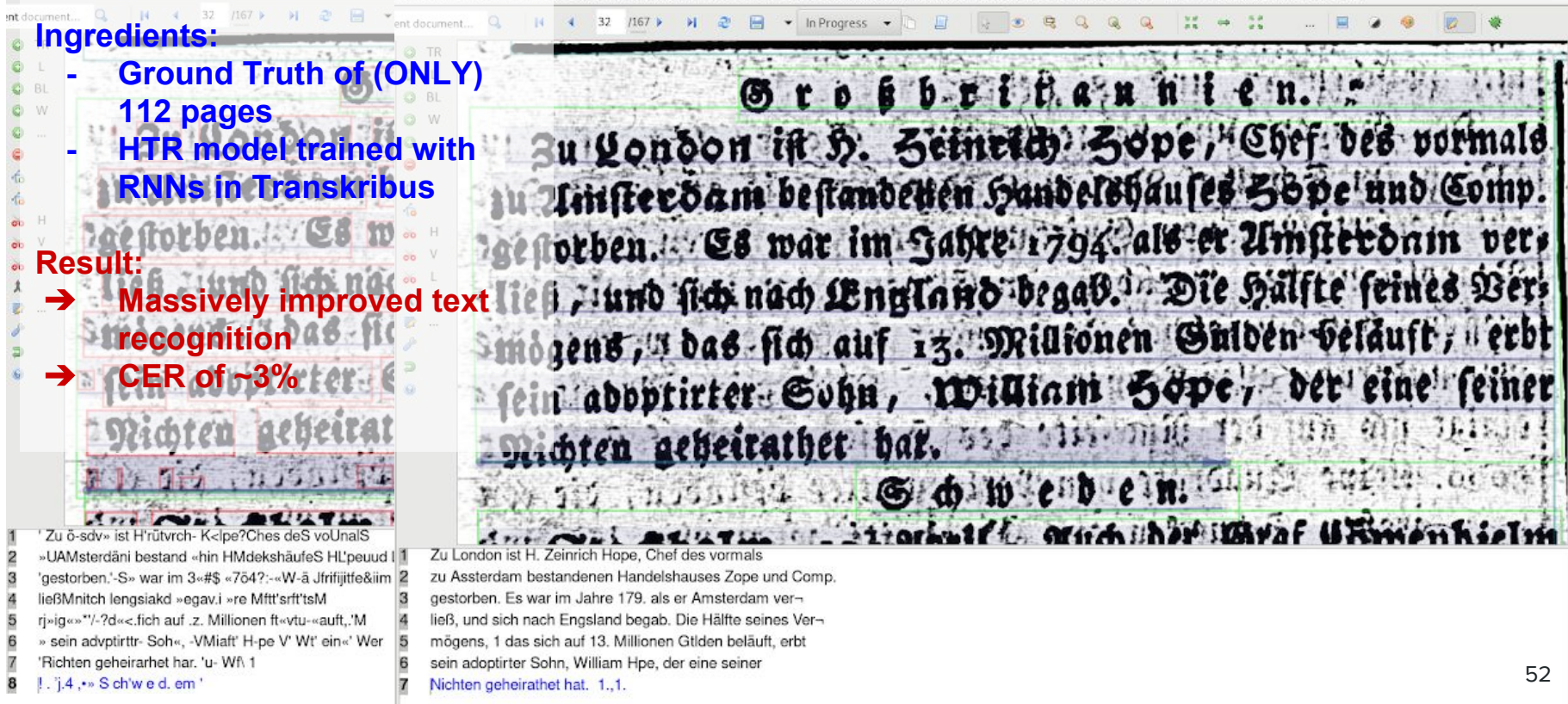
(17), Loaded doc: nzz_gold, ID: 30948, Page 32, file: nzz_18110507_0_0_a_077), Loaded doc: nzz_gold, ID: 30948, Page 32, file: nzz_18110507_0_0_a1_p1_1.tif [Image Meta Info: (Resolution:245.1630052, w*h: 2026 * 2480)] [current line: w*h: 640 * 67]

Ingredients:

- Ground Truth of (ONLY) 112 pages
- HTR model trained with RNNs in Transkribus

Result:

- Massively improved text recognition
- CER of ~3%



The image shows a side-by-side comparison of a document page. On the left is the original, very noisy scan of a page from a historical document. On the right is the same page after being processed by the HTR model, showing significantly improved text recognition. The text is now clear and legible. Below the images, a list of lines from the document is shown, with the original scan on the left and the cleaned-up text on the right. The cleaned-up text is a German snippet about a man named Heinrich Hope, who moved from Amsterdam to London in 1794. The text is: 'Zu London ist H. Zeinrich Hope, Chef des vormals zu Assterdam bestandenen Handelshauses Zope und Comp. gestorben. Es war im Jahre 179. als er Amsterdam verließ, und sich nach England begab. Die Hälfte seines Vermögens, 1 das sich auf 13. Millionen Gtlden beläuft, erbt sein adoptirter Sohn, William Hpe, der eine seiner Nichten geheirathet hat. 1.,1.'

Next Steps ...

- Complete re-digitisation of the NZZ (together with the Zurich Central Library and Swiss National Library)
 - Estimated costs of about 1.2 million CHF (0.6 CHF/page)
 - Including article boundary detection
 - If funded by contributing institutions (incl. Lottery Fund of the City of Zurich), first data packages can be expected in January 2019
- No ground truth for LeTemps
 - yet

Can it get even better?

- NZZ willing to **open** its archive to general public
- a re-digitisation **from paper** is probable
- **article recognition** included

Semantic Indexing of Documents

OCR Errors Dominate Topics

drn verde riner drs nnr lourde
fieren snr znm dri vrr wnrde tir
zürcher sfr hauen bri jür eue
rinr

unc clie icl ncl iic ssc unci icr
clen vnn clc cll lcl icli unss llc
nci clic ncn nnc

vita dtt chm dje dew dep ftn
stm dtn mem hät ftm bch dim
r ßch mam vow pie st

auk äis äor sen nnä äio aut
rä inä aktie van veräen
ligation versen eit bitten tur
ber äon

werben imb beten beii bor finb
biefer alt beut cit bnfi bieten bafi
h beu bcii bcs bitten jii oii

lil eilen lul nil lin ver uni llr
lullen lun lei van vol llo vel lol
vll ren lesen lit

13/50

kenen dea hallen wild lönnce cin
nicken tas lau cinc uicht teil isl
d eue bitten wal dcu eiu mid

nicken bitten vnn icl eit lil
nibt gen liegen clcr clie klein
cit bereit cke eilen cin lum
immer ckcñ

bass werben wild leinen iahen
mehl enn übel wal fehlen non
tann fühlen zul nui zui übn
hatt ihi wied

erden essen gen sel ste
liegen nde lei chen seln eilen
wild ehren slch folgen lan schl
stellen suchen eichen

du leinen tte sep dtt sev
deutsch teu oct diensttag wey
ttt ini ftp freytag dep beu bep
neapel ste

pordenone dez fitch wlld
aeaan tizian tinten abtl
notzen aln chll lwa teuber txn
as iwl novelle hartlieb ninette
del

lar lir iiiii it leihen aii fiir iri lil
ifl imi helfen hui ili iti ifi iif
ullen

Long Story Short ...

- ... before you can get to the interesting stuff, you need to do preprocessing, preprocessing, preprocessing.



Meaning ...

- OCR error (post-)correction
- old spelling normalisation
- lemmatisation

Side Note: Topic Modeling vs Text Classification



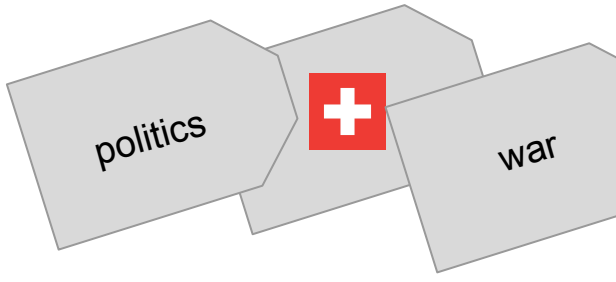
supervised
(labelled training
data needed)

Indexing documents

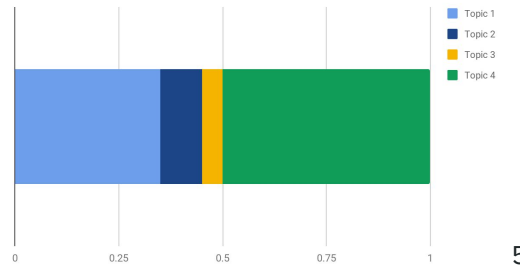
unsupervised

Text classification

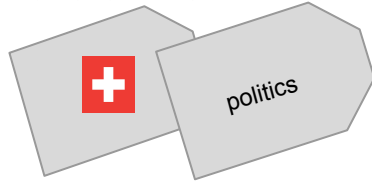
Topic modeling



Topic distribution

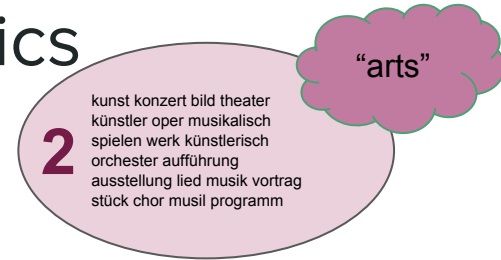
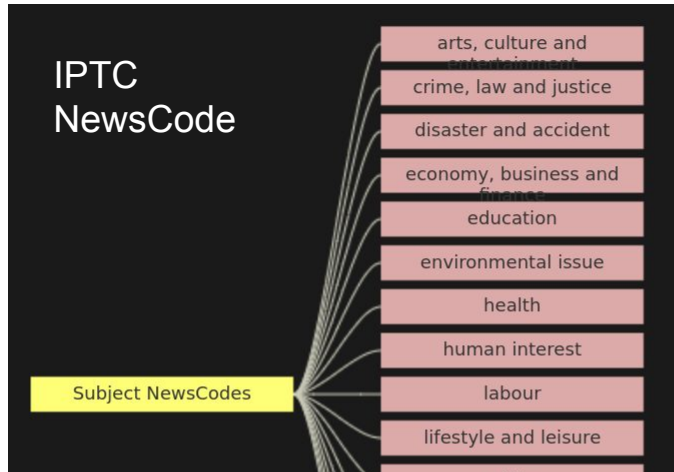


Contrastive View: Classes vs. Topics



classes = defined concepts

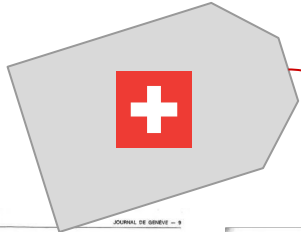
- taxonomy, e.g.



topics = probabilistic word distributions

- ideally semantically coherent
- NO definition
- NO predefined labels
- Should users be able to define their own interpretable labels?

Example: Faceted Search on Page



SUISSE JOURNAL DE GENÈVE - 9

81 85 - Mardi 18 mars 1974

Fortes diminutions de la grippe par rapport à 1973

REVISION PARTIELLE DE LA LOI SUR L'AVS Renvoi au Conseil fédéral pour réexamen mais voté

Banque nationale: M. Fritz Leutwiler succède à M. Stöpper à la présidence

PDC et défense nationale: critiques au Synode 72

Obèques de Mme Pierre Graber hier à Lausanne

Le «swi sur I»

Le trafic routier a rapporté 2 milliards: 60% des recettes douanières à la construction et à l'entretien de routes

Le golfiste en prison

Tagliero - Magagnoli Tagliero

Carte postale GENEVE DE LA RUE CARDAL 16, RUE DE SOURG LAUSANNE

SUISSE JOURNAL DE GENÈVE - 9

81 85 - Mardi 18 mars 1974

Papier plus cher: la hausse est justifiée, estime M. Schürmann

De quel groupe surgira la vérité constitutionnelle?

13e rente AVS: en 1974 seulement

A l'appel du RJ, des Juraissiens de l'extérieur rentrent au pays: l'UPJ dénonce une manœuvre

Militio changeants

Le parlementarisme suisse

Le parlementarisme suisse

Le parlementarisme suisse

Topic distribution

■ Topic 1
■ Topic 2
■ Topic 3
■ Topic 4

SUISSE JOURNAL DE GENÈVE - 9

81 85 - Mardi 18 mars 1974

Longévité: l'écart se creuse entre femmes et hommes

Les Suisses de France ont tenu congrès à Paris Droits politiques, information et AVS

Résultats de la consultation de consultation Entraide judiciaire internationale: de l'approbation à l'abstention

Basketball: sous l'impulsion de M. Leclerc Les «pros» à la conquête de l'Europe

1809, on brûle et sorcières...

Zurich: l'assassin des prostituées a avoué

Nécessaire pour l'obtention des subsides fédéraux L'association Inter-Jura vers le jour

Football: plongez-vous dans la coupe!

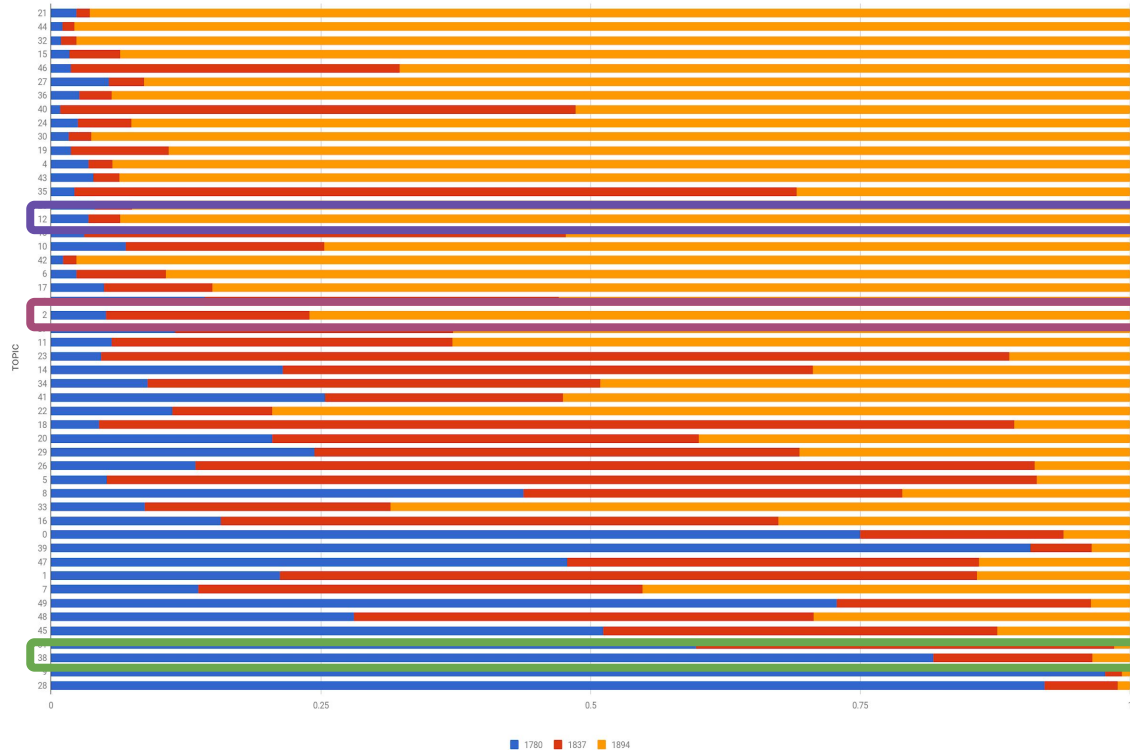
Boxe: fias au Berinois Leo Kakolievic

Basketball: 16e journée du championnat suisse

La coupe de la perplexité

62

Topic Evolution over Time



2

kunst konzert bild theater
 künstler oper musikalisch
 spielen werk künstlerisch
 orchester aufführung
 ausstellung lied musik vortrag
 stück chor musil programm

12

deutschen britisch deutsch
 truppe krieg russisch angriff
 flugzeug feindlich schwer front
 armee kampf london japanisch
 havas amerikanisch feind alliiert
 italienisch

38

truppe general armee mann gen
 feind stadt nachricht tage uhr
 franzose offizier groß kaiser
 regiment gefangen morgen
 befehl befinden ziehen

Insights

- HTR models (or NNs in general) boost OCR quality
- OCR errors have an impact on text mining
- text classification \neq topic modeling

Questions?

Bridging the fields

Bridging the fields: Research scenarios

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Bridging the fields: Research scenarios

Tracking the anti-European posture in the public debate in Switzerland and Luxembourg (1848-1945)

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

Bridging the fields: Research scenarios

Tracking the anti-European posture in the public debate in Switzerland and Luxembourg (1848-1945)

1. Goal: Identify the postures in the debates around the European idea
2. Using people and slogans to collect a broad corpus of articles, across several newspaper titles / cluster them according to manual classification and automatically generated information
3. Quantity and diversity of the collected results / types of incarnation of the European idea.

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> Gerold



The coverage of The Battle of Arnhem in European Newspapers (1944-present)

1. Help me find all articles which cover the events at Arnhem: When does coverage start?
2. Help me discover similarities and differences in articles about the battle 1944 to present
3. Help me understand how “popularity” of the battle is changing over time. How could that be measured?

Estelle >> **Marten** >> Julien >> Solenn >> Enrico >> Tobias >> Gerold

Bridging the fields: Research scenarios

ALIENS IN NEWSPAPERS

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> **Julien** >> Solenn >> Enrico >> Tobias >> Gerold

ALIENS IN NEWSPAPERS

What we want to look for:

- Concomitance with an **event** (astronomical discovery, UFO, release of a movie, publication of a book...)
- "Keyword in **context**"(–history of representations)
 - **Characterization of these non-humans** (judgement, anthropomorphism...)
 - Links with **semantic fields**: war, scare, space, colonization, creatures, intelligence...
 - Mentions of **persons** (e.g. authors, directors, astronomers...)
- A popular subject? (origin of these newspapers, type of the content–fiction, interviews...–size of the articles...)
- Evolutions of representations through time & possible differences between countries

French keywords: *extraterrestre, martien, sélénite, alien...*

First results on *e-newspaperarchives.ch*:

- 1289 results for “extraterrestre” from 1896 to 2012
- 26 results for “sélénite” from 1866 to 2010
- 1097 results for “martien” from 1863 to 2009
- 8725 results for “alien” from 1872 to 2014

Difficulties:

- **Substantivized adjectives:** “extraterrestre” can apply to phenomena and things too
- **Homonyms and similar words:**
 - “martien” often mistaken for “**Martigny**” (Swiss city)
 - “alien” often mistaken for “[Woody] allen”, “allen”, “aliéné” and words with -allen/alien (312 585 results on Retronews !!)
 - “alien”: English words with numerous irrelevant meanings

Bridging the fields: Research scenarios

Funding secondary education in nineteenth century Europe: structuration of a public debate

1. Research question : how educational policies were addressed by journalists and which data were deemed necessary to support the given point of view;
2. Help me find: articles dealing with educational policies, their “type” and the media associated to them ; linked named entities to “follow them”;
3. Expected results/difficulties

Estelle >> Marten >> Julien >> **Solenn** >> Enrico >> Tobias >> Gerold

Bridging the fields: Research scenarios

How do newspapers reflect the history of computing in Switzerland ?

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> **Enrico** >> Tobias >> Gerold

Bridging the fields: Research scenarios

Journalistische Kulturen im Wandel: die NZZ und das JdG im 19. Jahrhundert

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> Enrico >> **Tobias** >> Gerold

Journalistic Cultures during the 19th Century: a Comparison between the *NZZ* and *le JdG*

Main focuses (among others): genre change and meta-discourses

The three most important operationalisations (cf. handout)

- Help me find the lengths of sections (German: *Rubriken*, French: *rubriques*) in both newspapers and the change of these lengths in time
(Handout: 1.1.1.)
- Help me find moments of emergence of sections in newspapers
(Handout: 1.1.2.)
- Help me find/collect words and collocations related to sections
(Handout: 1.2.1.)

Bridging the fields: Research scenarios

The history of thought in medicine: from scholasticism to rationalism

5 mins

1. Research question
2. Operationalisation/tools
3. Expected results/difficulties

Estelle >> Marten >> Julien >> Solenn >> Enrico >> Tobias >> **Gerold**

coffee break!



Interface presentation

 Media Monitoring
impresso of the Past

Workshop Session 3

Lausanne, EPFL, 5-6 July 2018

Interface presentation

Laurel workshop - session 3 - friday 6 july

Overview of the session and co-design package

Interface features:

1. Searching and reading
2. Collecting and organising

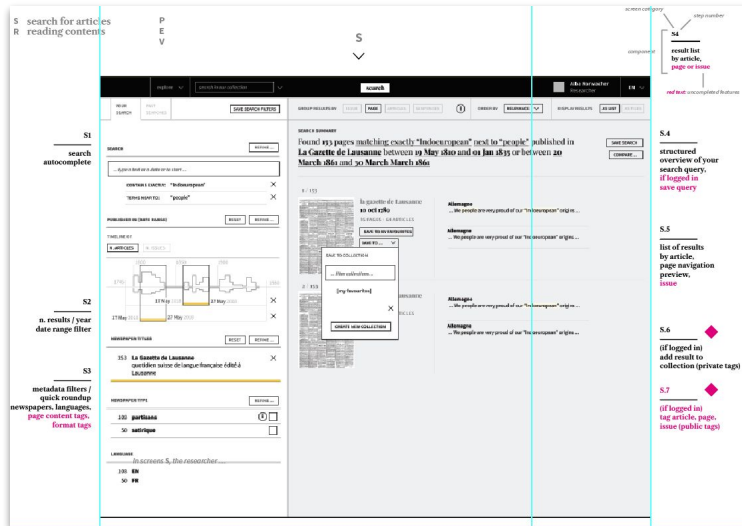
Presentation --> **Mockups**

Interaction --> **Prototype V1**

Feedback --> **Booklet**

Co-design package

Booklet of interface features,
activity by activity



Component index,
by activity

component index

S	R
S.1 search autocomplete	R6 metadata: see tags, entities and collection as marginalia
S.2 n. results / year date range filter	R7 contextual menu for identified regions
S.3 metadata filters / quick roundup newspapers, languages, page content tags, format tags	R9 highlight search results for this issue in the thumbnails
S.4 structured overview of your search query, if logged in save query	R8 add page to user collections

SWOT (feedbacks)

code	Strength & opportunities	weakness & threats	other

Searching and reading: presentation

S (Search activity)

1. Search with autocomplete, include/exclude terms and other search options
2. Metadata filters: *publication date, language ...*
3. Search results: *sort by, group by, display as*

R (Reading activity)

1. Issue viewer
2. Navigation between the list and the objects?

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

YOUR SEARCH | PAST SEARCHES

FULL TEXT SEARCH REFINE ...

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

PUBLISHED IN (DATE RANGE)
TIMELINE OF

N. ARTICLES | N. ISSUES

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

MORE ...

ARTICLE TYPE

112 **partisans**

50 **satirique**

MORE ...

TOP NAMED ENTITIES

103 **Napoleon**

50 **Zurich, Suisse**
location

MORE ...

LANGUAGE

112 **French**

MORE ...

search

Alba Rorwacher
Researcher

EN

GROUP RESULTS BY ISSUE PAGE ARTICLES SENTENCES ORDER BY RELEVANCE DISPLAY RESULTS AS LIST AS TILES

SEARCH SUMMARY
Find 53625987 articles in our collection.



1 of 112
Les recherches d'une science jeune: CELLE DE L'INFORMATION
INTERNATIONAL NEWS

Gazette de Lausanne, Saturday, May 8, 1954, p. 15
Les recherches d'une science jeune CELLE La presse, son rôle, ses possibilités JEAN-PIERRE AGUET A. A. A. A. H est à l'heure actuelle de multiples...

CONFERENCE REPORTS



2 of 112
Les USA et l'AFN
INTERNATIONAL NEWS

Gazette de Lausanne, Tuesday, November 19, 1957, p. 1
Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A la terrasse de café où je suis assis, me défendant tant bien que mal contre l'assaut...

GUERRE FROIDE



3 of 112
FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE
INTERNATIONAL NEWS

Gazette de Lausanne, Friday, April 27, 1951, p. 1
FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE SOUS LE PRINTEMPS D'ESPAGNE ATTENDANT UN TARDIF REPENTIN FRANCO-ANGLAIS'; / Par notre envoyé spécial Michel CLERC '!!!', ' ;...

S.4

structured overview of your search query, if logged in search query can be saved and retrieved later

S.5

list of results by article, page navigation preview, issue

S.6

(if logged in) add result to collection (private tags)

S.7

(if logged in) tag article, page, issue (public tags)

GROUP RESULTS BY

ISSUE

PAGE

ARTICLES

SENTENCES



ORDER BY

RELEVANCE 

DISPLAY RESULTS

AS LIST

AS TILES

REFINE ...



SEARCH SUMMARY

Find **53625987** articles in our collection.

SAVE SEARCH

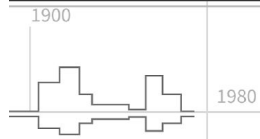
COMPARE ...



1 of 112

Les recherches d'une science jeune: CELLE DE L'INFORMATIONINTERNATIONAL NEWS [TAG AS ...](#)**Gazette de Lausanne**, Saturday, May 8, 1954, p. 15*Les recherches d'une science jeune CELLE La presse, son rôle, ses possibilités JEAN-PIERRE AGUET ^^.-^.^.^ H est à l'heure actuelle de multiples...*[ADD TO FAVOURITES](#)[ADD TO COLLECTION...](#)

CONFERENCE REPORTS

[EXPORT CITATIONS ...](#)[DOWNLOAD AS ...](#)

s published

:aise édité à



2 of 112

Les USA et l'AFNINTERNATIONAL NEWS [COVER PAGE](#) [TAG AS ...](#)**Gazette de Lausanne**, Tuesday, November 19, 1957, p. 1

Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A

[ADD TO FAVOURITES](#)[ADD TO COLLECTION...](#)GUERRE FROIDE 

YOUR SEARCH PAST SEARCHES

GROUP RESULTS BY ISSUE PAGE ARTICLES SENTENCES ⓘ ORDER BY RELEVANCE ▾ DISPLAY RESULTS AS LIST AS TILES

FULL TEXT SEARCH REFINE ...

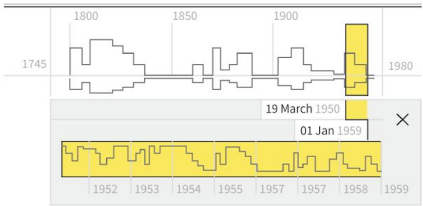
... type a text or a date or to start ...

CONTAINS ▾ EXACTLY ▾ **guerre froide** ✕

CONTAINS ▾ ... ▾ **europe** ✕

PUBLISHED IN (DATE RANGE)

TIMELINE OF N. ARTICLES RESET



ADD RANGE ...

IN NEWSPAPER TITLES RESET

112 **La Gazette de Lausanne**
quotidien suisse de langue française édité à Lausanne ✕

TYPES OF ARTICLES TOP NAMED ENTITIES

112 **international news** 103 **Napoleon**

50 **politics** 50 **Zurich, Suisse**
location

LANGUAGE OF ARTICLES

112 **French**

SEARCH SUMMARY
Found 112 articles matching exactly "guerre froide" AND europe published in La Gazette de Lausanne between Saturday, 19 March 1950 and Friday, 01 Jan 1959 SAVE SEARCH COMPARE ...



1 of 112
Les recherches d'une science jeune: CELLE DE L'INFORMATION
INTERNATIONAL NEWS TAG AS ...

Gazette de Lausanne, Saturday, May 8, 1954, p. 15
Les recherches d'une science jeune CELLE La presse, son rôle possibilités JEAN-PIERRE AGUET ^.^ . ^.^ H est à l'heure de multiples...

en préjugés sur . ses', voisins et souvent en **guerre « froide »**-ou « chaude ». « UNE SEMAINE DANS LE MOND

, est la presse nationale de plu- sieurs pays parmi lesquel les Etats- Unis, huit pays d'**Europe** occidentale

ADD TO FAVORITES ADD TO COLLECTION...

ADD TO COLLECTION
filter or create collection

favorites

✓ **conference reports** ✕

guerre froide

CREATE NEW MANAGE MY COLLECTIONS



2 of 112
Les USA et l'AFN
INTERNATIONAL NEWS COVER PAGE ✕ TAG AS ...

Gazette de Lausanne, Tuesday, November 19, 1957, p. 1
Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A la terrasse de café où je suis assis, me défendant tant bien que mal contre l'assaut...

- nomie . Tant qu'a duré la **guerre froide** , Washington n'a pas eu d'autre souri que de combattre

en chef des forces du Centre-**Europe** , « chef atlantique sûr » , s'employa à accréditer cet- te version

ADD TO FAVORITES ADD TO COLLECTION...

GUERRE FROIDE ✕

EXPORT CITATIONS ... DOWNLOAD AS ...



3 of 112
FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE
INTERNATIONAL NEWS TAG AS ...

Gazette de Lausanne, Friday, April 27, 1951, p. 1
FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE SOUS LE PRINTEMPS D'ESPAGNE ATTENDANT UN TARDIF REPENTIR FRANCO-ANGLAIS ' ; ; / Par notre envoyé spécial MICHEL CLERC ' ! ! , ' ; ; ;

ADD TO FAVORITES ADD TO COLLECTION...

EXPORT CITATIONS ... DOWNLOAD AS ...

S.8
include / exclude options are given for each text string

S.9
zooming on the selected date range

S.10
add article to a collection

S.11
display list of matching text segments in the article

S.12
what you would like to export? Individual articles can be downloaded (text, pdf, images)

YOUR SEARCH

PAST SEARCHES

TEXT SEARCH

REFINE ...

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

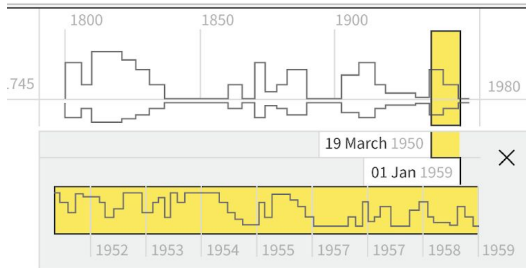
CONTAINS ▾ EXACTLY ▾ guerre froide

CONTAINS ▾ ... ▾ europe

ISHED IN (DATE RANGE)

LINE OF N. ARTICLES

RESET



ADD RANGE ...

NEWSPAPER TITLES

RESET

112 **La Gazette de Lausanne**
quotidien suisse de langue française édité à Lausanne

GROUP RESULTS BY

ISSUE

PAGE

ARTICLES

SENTENCES



ORDER BY

RELEVANCE ▾

DISPLAY RESULTS

AS LIST

AS TILES

SEARCH SUMMARY

Found **112** articles matching exactly **“guerre froide” AND europe** published in **La Gazette de Lausanne** between **Saturday, 19 March 1950** and **Friday, 01 Jan 1959**

SAVE SEARCH

COMPARE ...



1 of 112

Les recherches d'une science jeune: CELLE DE L'INFORMATION

INTERNATIONAL NEWS TAG AS ...

Gazette de Lausanne, Saturday, May 8, 1954, p. 15

Les recherches d'une science jeune CELLE La presse, son rôle possibilités JEAN-PIERRE AGUET ^.^ . ^.^ H est à l'heure de multiples..

en préjugés sur . ses', voisins et souvent en **guerre « froide »** -ou « chaude ». « UNE SEMAINE DANS LE MOND

, est la presse nationale de plu- sieurs pays parmi lesquel les Etats- Unis, huit pays d'**Europe** occidentale

ADD TO FAVOURITES

ADD TO COLLECTION...

ADD TO COLLECTION

filter or create collection

favorites

✓ **conference reports** X

guerre froide

CREATE NEW

MANAGE MY COLLECTIONS



2 of 112

Les USA et l'AFN

INTERNATIONAL NEWS COVER PAGE X TAG AS ...

Gazette de Lausanne, Tuesday, November 19, 1957, p. 1

Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A la terrasse de café où je suis assis , me défendant tant bien que mal contre l'assaut...

- nomie . Tant qu'a duré la **guerre froide** , Washington n'a pas eu d'autre souri que de combattre

ADD TO FAVOURITES

ADD TO COLLECTION...

GUERRE FROIDE X

La Gazette de Lausanne / Lundi 16 Mai 1956

Alba Rorwacher
Researcher

EN

CONTEXT / LIST OF TOPICS HEADLINES OF THE DAY TABLE OF CONTENTS

search in page ...

p 2 of 7

EXPORT CITATIONS ... DOWNLOAD AS ...

OCR QUALITY 82 %

La Gazette de Lausanne
Lundi 16 Mai 1956
6 PAGES - 25 ARTICLES
Archive Du TEMPS

NAMED ENTITIES & TOPICS

ENTITIES TOPICS SAME ISSUE +/-1 WEEK IN DATES ...

IN ALL NEWSPAPERS IN THE SAME NEWSPAPER

n. of articles per day

17 May 1956 27 May 1956

ORDER BY RELEVANCE SHOW PEOPLE ONLY

Napoleon 1 of 454
Napoléon Bonaparte (French: [napoleõ bonapaʁt]; 15 August 1769 – 5 May 1821) was a French statesman and military leader
13545 ARTICLES

Pasquale Paoli 2 of 454
Nationalist Corsican leader.
15 ARTICLES

Marie Louise, Duchess of Parma 2 of 454
(12 December 1791 – 17 December 1847)
Austrian archduchess
15 ARTICLES

SHOW MORE ...

NAMED ENTITIES ON PAGE

ORDER BY RELEVANCE

Pasquale Paoli
Nationalist Corsican leader.

Chiang Kai-shek
political and military leader who served as the leader of the Republic of China

MAX

PAGE TAGS

TAG AS ...

COLLECTION

ADD TO C

ZOOM 22 %

MIN

FULL SCREEN

PREVIOUS RESULT
La Gazette de Lausanne
Lundi 16 Mai 1935

CURRENT SEARCH
Found 112 articles matching exactly "guerre froide" AND europe published in La Gazette de Lausanne between Saturday, 19 March 1950 and Friday, 01 Jan 1959
3 search result matches in this page

NEXT RESULT
La Gazette de Lausanne
Lundi 1 Mai 1945

BACK TO SEARCH RESULT PAGE

R.1
issue context
/ switch to
table of contents

R.2
timeline of
n. articles per day,
focus on the
date of publication of
this issue.
This filters related
entities in R.3

R.3
entities or
topics
mentioned
in the same issue
/ the same day
(according to user's
choice)

R.4
browse through
search results
(if you reached this
page from the
search page)

R.5
zoom on
newspaper
page

R.6
metadata:
see tags,
entities and
collection as
"marginalia"

R.7
highlight search results
for this issue
in the thumbnails

R.2

timeline of
n. articles per day,
focus on the
date of publication of
this issue.

This filters related
entities in R.3

R.3

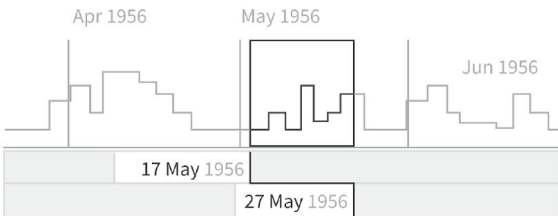
entities or
topics
mentioned
in the same issue
/ the same day
(according to user's
choice)

NAMED ENTITIES & TOPICS

ENTITIES TOPICS SAME ISSUE +/- 1 WEEK IN DATES ...

IN ALL NEWSPAPERS IN THE SAME NEWSPAPER

n. of articles per day



ORDER BY RELEVANCE ▾ SHOW PEOPLE ONLY ▾



Napoleon 1 of 454
Napoléon Bonaparte (French: [napoleõ bõnapart];
15 August 1769 – 5 May 1821) was a French
statesman and military leader

13545 ARTICLES



Pasquale Paoli 2 of 454
Nationalist Corsican leader.

15 ARTICLES



Marie Louise, Duchess of Parma 2 of 454
(12 December 1791 – 17 December 1847)
Austrian archduchess

15 ARTICLES

SHOW MORE ...

R.4



1



2



3



4



ORDER BY RELEVANCE ▾

Pasquale Paoli
Nationalist Corsican leader.

Chiang Kai-shek
political and military leader
who served as the leader of
the Republic of China

Vertical text on the right edge of the page, likely a page number or identifier.



NAMED ENTITIES ON PAGE

ORDER BY **RELEVANCE** ▼

Pasquale Paoli
Nationalist Corsican leader.

Chiang Kai-shek
political and military leader who served as the leader of the Republic of China

Le président de la République a reçu à l'Élysée ce matin le général de Gaulle...

Le président de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée. Le général de Gaulle a été reçu par le président de la République à l'Élysée...

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

Le général de Gaulle

Le général de Gaulle a été reçu ce matin par le président de la République à l'Élysée...

PAGE TAGS

TAG AS ...

COLLECTION

ADD TO CO

ZOOM 22%

LA JOANNE

REVUE FEMME

100 pages de beauté

Beauté, mode, sport

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

100 pages de beauté

R.10
omatic switch
le of contents
en reading an
article

R.11
lected (active)
article

R.12
quickly go to
search results
in this article

CONTEXT / HEADLINES
LIST OF TOPICS / LIST OF THE DAY
TABLE OF CONTENTS

La Gazette de Lausanne Lundi 16 Mai 2018

6 PAGES - 25 ARTICLES

Archive Du TEMPS

1540 issues from 1840 to 1950

LIST OF ARTICLES

ALL ARTICLES

ONLY ARTICLES IN MY COLLECTIONS

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

VI / RUE DES PALMES

p. 2, *novel*

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les ...

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de

PREVIOUS RESULT
La Gazette de Lausanne
Lundi 16 Mai 1935

VI / RUE DES PALMES
p. 2, *novel*

p. 2

VI

RUE DES PALMES

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les stations hivernales.

Dès ces premiers jours de juin, Nice, abandonnée sous la poussière, prenait son aspect de Pompéi ensevelie sous la cendre. Avec les étrangers, la richesse, le mouvement, la vie, se retirait. Hôtels et villas se fermaient successivement. Les rues semblaient vidées. Sur le seuil des magasins ternes, des boutiquiers mélancoliques haletaient entre deux échéances, et la plage, démeublée des pittoresques silhouettes de Russes, d'Anglais et d'Américains, faisait l'effet d'un cadre sans tableau.

Son caractère cosmopolite ainsi perdu, la capitale de l'exotisme se retrouvait simple ville de province; et, du même coup, l'esprit de province y revivait. S'observer et se quereller entre soi restait à peu près la seule distraction de ceux auxquels une raison spéciale n'avait

p. 2

pas encore permis de prendre leur vol.

Les Roquiescize attendaient les vacances des enfants, les du Valbonnet la liquidation de l'héritage, la princesse le départ de son neveu, Mme Guidal l'arrivée de sa fille, la comtesse de Wactembrode, une invitation à la campagne.

Des motifs plus impérieux encore retenaient les Faveroles dans leur petite villa blanche du quartier des Bânettes.

La santé d'Yvonne commandait l'intérêt, et, pendant une quinzaine, un médecin attentif l'avait clouée sur une chaise longue, auprès de laquelle Hubert s'était plu à mouter une garde pressée.

C'avait été un nouvel et charmant épisode de leur vie conjugale que cette réclusion à deux.

Alanguie et nonchalante sous les plis vaporeux d'un élégant peignoir, ses cheveux blonds coquettement ébouriffés, Yvonne faisait la plus gentille malade du monde; et, dans cette jolie

chambre de jeune mariée, où les stores haissaient

CURRENT SEARCH

Found 112 articles matching exactly "guerre froide" AND europe published in La Gazette de Lausanne between Saturday, 19 March 1950 and Friday, 01 Jan 1950

VI

RUE DES PALMES

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les Maître Corbeau, sur un arbre perché,

Tenait en son bec un fromage.

Maître Renard, par l'odeur alléché,

Lui tint à peu près ce langage :

« Hé ! bonjour, Monsieur du Corbeau.

Que vous êtes joli ! Que vous me semblez beau !

Sans mentir, si votre ramage

Se rapporte à votre plumage,

Vous êtes le Phénix des hôtes de ces bois. »

A ces mots le Corbeau ne se sent pas de joie ;

'était bien l'été cette fois, la

morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les Maître Corbeau, sur un arbre perché,

Tenait en son bec un fromage.

Maître Renard, par l'odeur alléché,

Lui tint à peu près ce langage :

« Hé ! bonjour, Monsieur du Corbeau.

Que vous êtes joli ! Que vous me

semblez beau !

Sans mentir, si votre ramage

Se rapporte à votre plumage,

Vous êtes le Phénix des hôtes de ces

bois. »

A ces mots le Corbeau ne se sent pas de

joie ;

Et pour montrer sa belle voix,

Il ouvre un large bec, laisse tomber sa

ADD TAG ...

ADD TO COLLECTION...

NOVELS

POEMS

EXPORT CITATIONS ...

DOWNLOAD AS ...

NAMED ENTITIES ON PAGE

ORDER BY RELEVANCE ▾

Maître Renard
Nationalist Corsican leader.

Maître du Corbeau
political and military leader
who served as the leader of
the Republic of China

text
image
side

R. 14
add 1
and :
your

R.15
List :
entit

n l'été cette fois, la morte-saison
i, nulle part, ne s'éprouve autant
stations hivernales.

emiers jours de juin, Nice, aban-
la poussière, prenait son aspect de
velie sous la cendre. Avec les
richesse, le mouvement, la vie,
Hôtels et villas se fermaient suc-

Les rues semblaient vidées. Sur
magasins ternes, des boutiquiers
s haletaient entre deux échéances,
meublée des pittoresques silhouet-
s, d'Anglais et d'Américains, fai-
in cadre sans tableau.

ère cosmopolite ainsi perdu, la ca-
otisme se retrouvait simple ville
et, du même coup, l'esprit de
vivait. S'observer et se quereller
lait à peu près la seule distraction
uels une raison spéciale n'avait

mis de prendre leur vol.
eize attendaient les vacances des
du Valbonnet, la liquidation de
princesse le départ de son neveu,
l'arrivée de sa fille, la comtesse
ode, une invitation à la cam-

plus impérieux encore retenaient
dans leur petite villa blanche du
oumettes.

l'ovonne commandait l'intérêt, et,
ruinzaine, un médecin attentif

C'était bien l'été cette fois, la morte-saison
des villes qui, nulle part, ne s'éprouve autant
que dans les Maître Corbeau, sur un arbre
perché,

Tenait en son bec un fromage.

Maître Renard, par l'odeur alléché,

Lui tint à peu près ce langage :

« Hé ! bonjour, Monsieur du Corbeau.

Que vous êtes joli ! Que vous me semblez
beau !

Sans mentir, si votre ramage

Se rapporte à votre plumage,

Vous êtes le Phénix des hôtes de ces bois. »

A ces mots le Corbeau ne se sent pas de joie ;

'était bien l'été cette fois, la
morte-saison des villes qui, nulle part,
ne s'éprouve autant que dans les Maître

Corbeau, sur un arbre perché,

Tenait en son bec un fromage.

Maître Renard, par l'odeur alléché,

Lui tint à peu près ce langage :

« Hé ! bonjour, Monsieur du Corbeau.

Que vous êtes joli ! Que vous me

semblez beau !

Sans mentir, si votre ramage

ADD TAG ...

ADD TO COLLECTION...

NOVELS

POEMS

EXPORT CITATIONS ...

DOWNLOAD AS ...

NAMED ENTITIES ON PAGE

ORDER BY

RELEVANCE ▾

Maître Renard

Nationalist Corsican leader.

Maître du Corbeaupolitical and military leader
who served as the leader of
the Republic of China

OCCURRENCES IN ARTICLE

Monsieur du Corbeau (5)

Maître Corbeau (4)

le corbeau (3)

IN OTHER ARTICLES ...

ADD TO COLLECTION...

R. 16

Get the list of
occurrences for the
selected named entity
in this article

R. 17

Add the named entity to
a collection!



Log in impresso-project.ch/v1

Username: **your email**

password : **[not available]**

Search



Go!

impresso-project.ch/v1

Search and reading: interactions (20 min)

1. **Search for articles** in the available corpus *using given options*
2. Use the **Save to favourites*****
3. Write suggestions, feedbacks, impressions ... on **SWOT table...** *for the pictures*

code	Strength & opportunities	weakness & threats	other

Search and read and **save***** (until 12h30)

Suggested queries in the corpus of **Gazette de Lausanne, 1950-1959.**

Culture : Trouver des concerts des 4 Barbus en Suisse (dates et lieux).

Cyclisme : Collecter des informations sur la participation de cyclistes suisses au Tour de France. (Piste: le 29 juillet 1951, Hugo Koblet remporte le Tour de France.)

Politique : Quelles sont les votations fédérales proposées entre 1950 et 1954?

Sciences: trouver des articles sur l'activité du Conseil européen pour la recherche nucléaire (CERN), créé le 29 septembre 1954.

Log in **impresso-project.ch/v1**

Username: **your email**

password : **Laurel2018!**

Searching and reading: feedback

Query results ?

- What articles? What informations?
- What difficulties? What strategies to overcome them?

Searching and reading: outlook

Towards query criticism

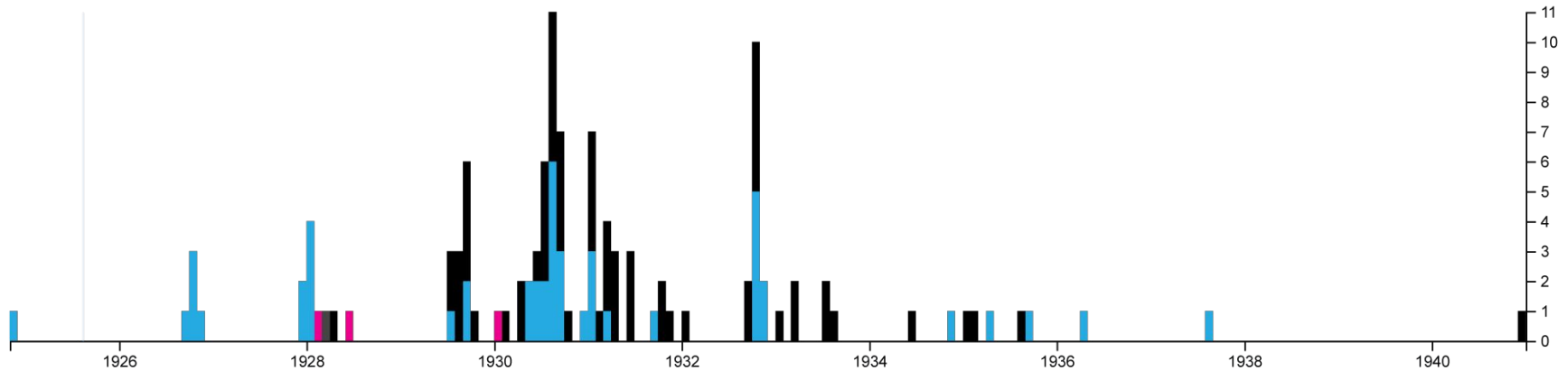
- Traceability of research steps
- Learning by following somebody else's footsteps ?

Exploration of the result lists : tools?

Visual exploration of the result list

Comparing frequencies:

Distribution over time of the queries “Paneurope” / “Paneuropa”



NGRAM VIEWER
OVER YEARS

VISUALIZATION TYPE

BUMP CHART

STREAM GRAPH

SEARCH

paneurope

DIMENSIONS

PER NEWSPAPER

PER COLLECTION

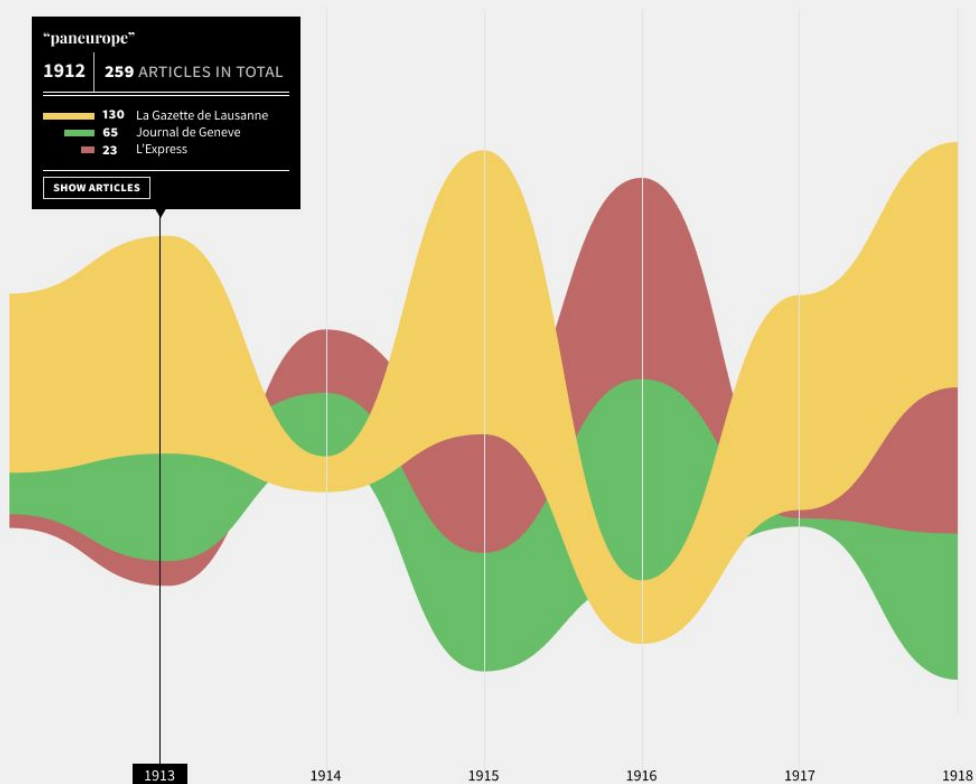
PER ARTICLE TAG

NEWSPAPERS - TOTALS

1230 La Gazette de Lausanne
650 Journal de Geneve
1023 L'Express

VISUALIZATION BRIEF

number of article containing "paneurope" per year. Each curve represents the number of articles per newspaper



Visual exploration of the result list

Applying facets
to the frequency

lunch break



Collecting and organising: how

Available metadata:

Feuille d'Avis de
Vevey

Mardi 4 octobre 1932

p.4

URL

(with query
sometimes)

NOUVELLES SUISSES

La jeunesse au congrès paneuropéen

Lundi s'est tenue à Bâle, dans le cadre du congrès paneuropéen, une assemblée de la jeunesse européenne, au cours de laquelle de nombreux orateurs, notamment le professeur Recasens-Siches, de l'Université de Madrid, le poète Jacob Wassermann, de Vienne, l'écrivain français Alfred Fabre-Luce, le poète bâlois Liehburg et l'un des chefs de la jeunesse européenne, Gerhard Meyer, ont pris la parole. Tous ont été d'avis que, sans une collaboration efficace de la jeunesse du 20^{me} siècle, on n'arrivera pas à une Paneurope. Il faut agir rapidement car la situation est critique et dangereuse.

Cette manifestation de la jeunesse européenne au cours de laquelle des messages provenant du sénateur français Caillaux, de Francis Delaisi, du comte Moltke, ancien ministre des affaires étrangères du Danemark, de M. Grimm, ancien ministre des cultes de Prusse, ont été lus, avait réuni un nombre imposant de participants.

(Manually) Organising the results
into collections:

Generic category (public)
Local news

Section (public)
"Nouvelles Suisses"

User category >> aka. Collection
(private)
"Paneurope Congress-Basel"

Collecting and organising

Features for the creation of research corpora:

- Assigning generic categories to articles (public)
 - Types of article, located in particular sections
 - Types of objects (ears, banners...)
- Assigning user-specific categories to articles (private)
 - Saved in the personal space within specific collections
 - Free text categories
 - Guided categories (autofill suggestions)

>>>> Steps to integrate the queries from the interface into the research workflow.

Collecting and organising: how

Suggested categories:

Format	Content
<p>Advertisements (commercial)</p> <p>Announcement (hard news, non commercial)</p> <p>Article (default category)</p> <p>Banner</p> <p>Box</p> <p>Cartoon</p> <p>Civil registry (obituaries, weddings, births)</p> <p>Column</p> <p>Outline (pictures)</p> <p>Ears</p> <p>Index, table of content</p> <p>Letters to the newspaper</p> <p>Photograph</p> <p>Press review</p> <p>Reviews (books, entertainment...)</p>	<p>Arts, culture and entertainment news</p> <p>Business news</p> <p>Crime, law and justice news</p> <p>Guest contribution, op-ed, quoted speech</p> <p>Fiction, poetry, humour</p> <p>International news</p> <p>Interviews</p> <p>Last minute news (dépêches)</p> <p>Local news</p> <p>News story (faits-divers)</p> <p>National news</p> <p>Science news</p> <p>Sports news</p>

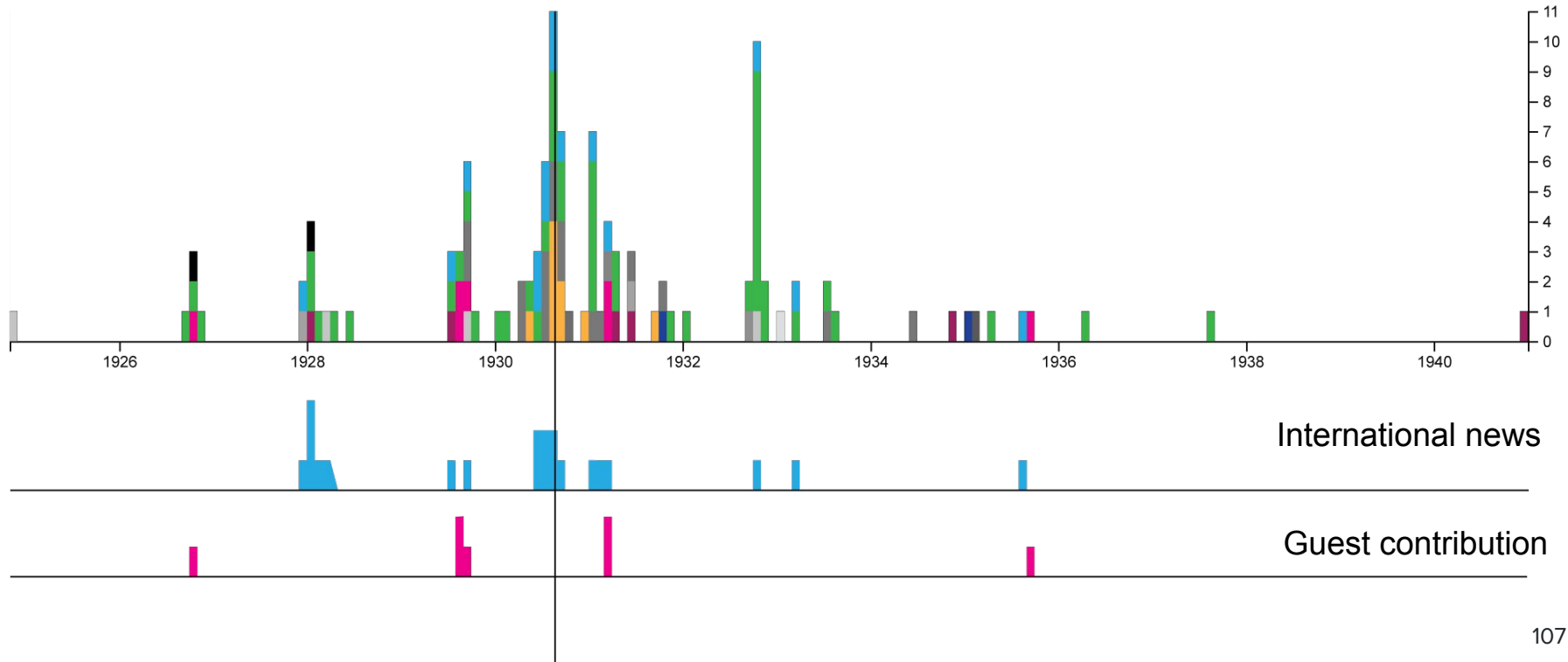
Collecting data from the queries: why

Paneurope use case:

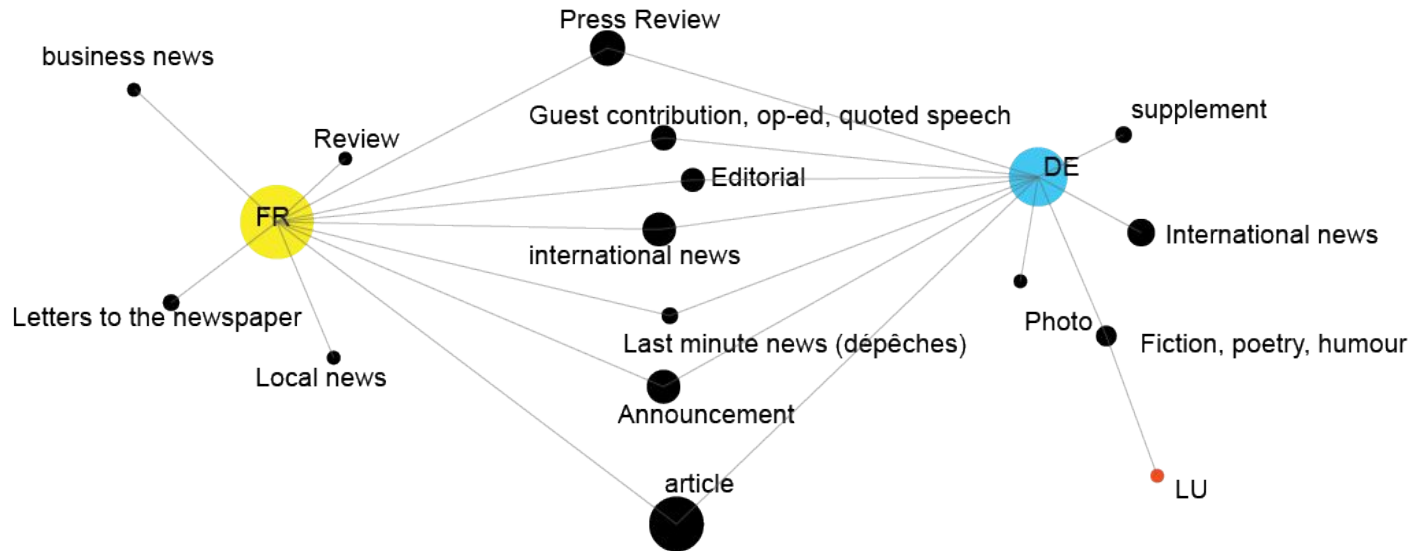
	A	B	D	E	F	G	H	I	J	K	L	N	O	P
1	query	Collection	Metadata/article		Metadata/ Title					Classification /tags				
2	query	Collection	author	title of the article	newspaper title	date-iso	page	URL	nr	section/ typologie des articles	Historical section title	type of newspaper	lang.	KWIC?
6	Paneurope	L'Express/impa	P.B.	"A la Société des Nations. - Un dîner européen"	L'impartial. Journal quotidien et Feu	1929-09-10	8		14926	article	Revue du jour	Feuille d'Avis	FR	
7	paneurope	e-luxemburg	Richard Coudenhove-Kale	"Abessinien und Paneuropa"	Obermosel-Zeitung	1935-09-03	1	<http://www	204	Guest contribution, op-ed, quoted sp		Local newspaper	DE	
9	paneurope	e-luxemburg	s.a.	"Allemagne. M. Coudenhove-Kalergi candidat au pri	L'indépendance luxembourgeoise	1931-11-23	1	<http://www	327	article		Cultural newspaper	FR	
14	Paneurope	Scriptorium (BC s.a.		"Après l'Assemblée des délégués de la Fédération c	La revue	1933-01-19	3		18	Business news	Page Agricole de la	General informatio	FR	
15	Paneurope	L'Express/impa	s.a.	"Association neuchâteloise pour la S.d.N."	Feuille d'Avis de Neuchâtel et du Vi	1935-01-16	4		13	Letters to the ne	La vie de nos socié	Feuille d'Avis	FR	
16	paneurope	e-luxemburg	Richard Coudenhove-Kale	"Atlantide"	L'indépendance luxembourgeoise	1929-08-08	1	<http://www	220	Guest contribution, op-ed, quoted sp		Cultural newspaper	FR	
17	paneurope	e-luxemburg	Richard Coudenhove-Kale	"Atlantide"	L'indépendance luxembourgeoise	1929-08-09	1	<http://www	221	Guest contribution, op-ed, quoted sp		Cultural newspaper	FR	
18	Paneurope	Le temps ar	s.a.	"Au congrès de l'Union Pan-Européenne. M. Briand	Gazette de Lausanne	1930-05-19	6	<http://www	137	Announcement	Dernières Dépêch	General informatio	FR	
19	Paneurope	Le temps ar	s.a.	"Au Congrès paneuropéen de Bâle"	Journal de Genève	1932-10-05	10	<http://www	272	article		General informatio	FR	
20	Paneurope	L'Express/impa	s.a.	"Au Reichstag"	Feuille d'Avis de Neuchâtel et du Vi	1930-06-27	6		146	article	Politique et informa	Feuille d'Avis	FR	
21	Paneurope	Le temps ar	s.a.	"Au sommaire du dernier numéro de la revue PANE	Journal de Genève	1927-12-24	2	<http://www	352	Announcement	Echos	General informatio	FR	
22	Paneurope	Le temps ar	s.a.	"Aussi longtemps que Paneurope n'aura pas dispar	Gazette de Lausanne	1931-01-21	2/3	<http://www	20	Press Review	Dans la presse	General informatio	FR	L'entourage de
23	Paneurope	Le temps ar	s.a.	"BALE-VILLE - Pour l'entente entre les peuples."	Gazette de Lausanne	1933-09-20	2/3	<http://www	262	Announcement		General informatio	FR	
24	paneurope	e-luxemburg	Georges Reynaud, ancien	"La menace de l'Anschluss"	L'indépendance luxembourgeoise	1931-03-31	1	<http://www	90	Guest contributi	Billet extérieur	Cultural newspape	FR	
25	paneurope	e-luxemburg	Georges Reynaud, ancien	"Malaise européen"	L'indépendance luxembourgeoise	1931-03-03	1	<http://www	62	Guest contributi	Billet extérieur	Cultural newspape	FR	
27	paneurope	e-luxemburg	s.a.	"Briands Paneuroparede."	Luxemburger Wort	1930-09-12	1	<http://www		International nev	Paneuropa	General informatio	DE	
28	Paneurope	Scriptorium (BC Mt.		"Bulletin du jour. Vers l'Europe d'après-guerre"	La Tribune de Lausanne	1940-12-19	1-2		353	Editorial		General informatio	FR	
29	Paneurope	Le temps ar	s.a.	"Ce qui pour M. Briand est un but, pour l'Allemagne	Gazette de Lausanne	1930-07-22	2/3	<http://www	200	Press Review	Dans la presse	General informatio	FR	Ces nations sav
31	Paneurope	L'Express/impa	s.a.	"Conférence pour la trêve douanière. M. Schulthess	Feuille d'Avis de Neuchâtel et du Vi	1930-02-21	4		43	article	Politique et informa	Feuille d'Avis	FR	
36	paneurope	e-luxemburg	s.a.	"Coudenhove fordert "Sicherheit"."	Escher Tageblatt	1931-01-21	2	<http://www		article		Local newspaper	DE	

Frequency of the query distributed over NSP sections

Distribution across all newspaper titles.

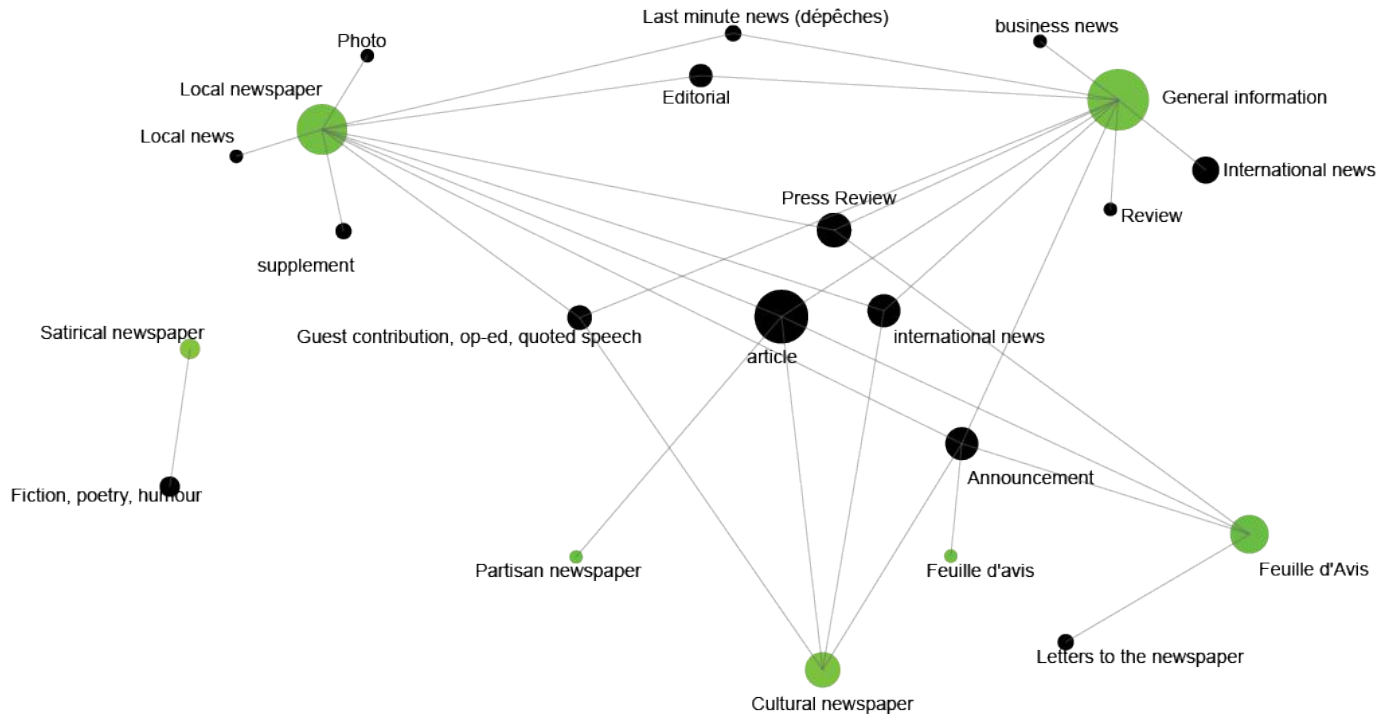


Paneurope use case



Languages/
newspaper
sections

PanEurope use case



Newspaper type/
 newspaper
 sections

Collecting and organising: how

Available metadata:

Feuille d'Avis de
Vevey,
Mardi 4 octobre
1932, p. 4, URL.

Generic category:
Local news

Section: **“Nouvelles
Suisses”**

User category:
**“Paneurope
Congress-Basel”**

NOUVELLES SUISSES

La jeunesse au congrès paneuropéen

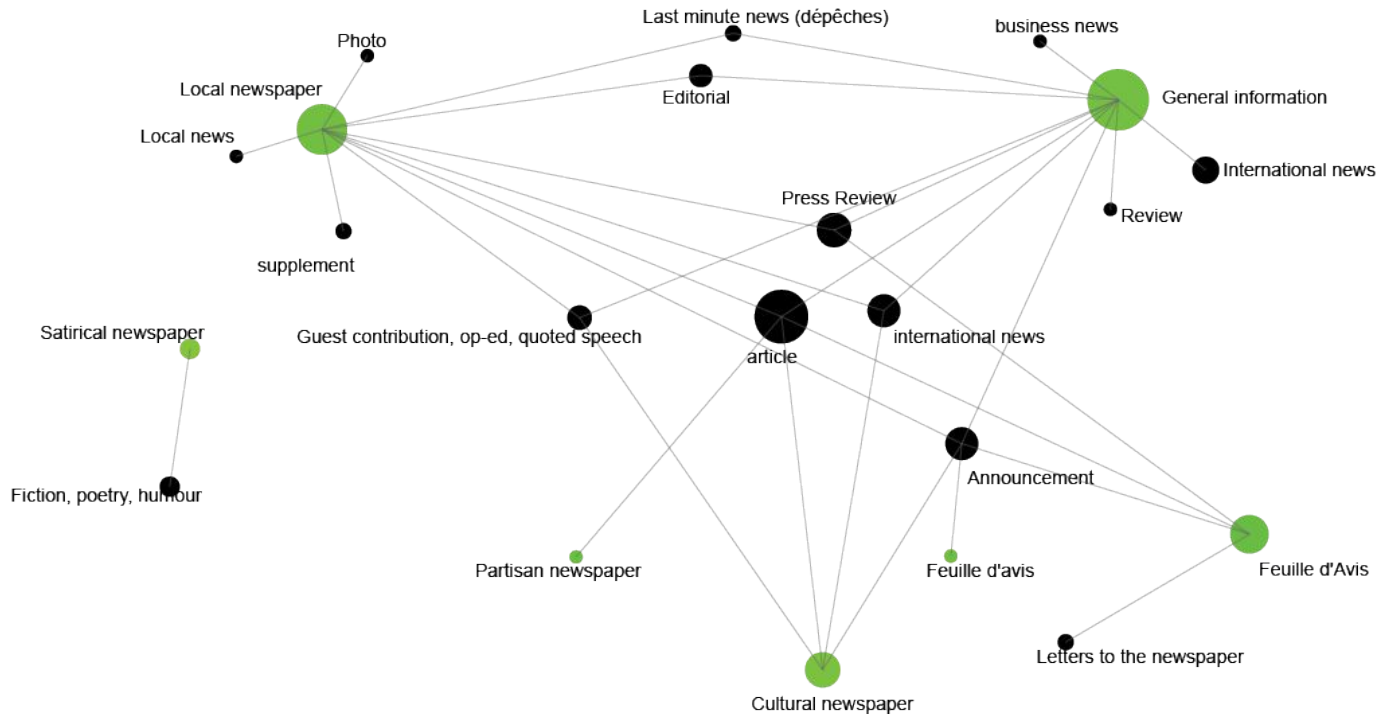
Lundi s'est tenue à Bâle, dans le cadre du congrès paneuropéen, une assemblée de la jeunesse européenne, au cours de laquelle de nombreux orateurs, notamment le professeur Recasens-Siches, de l'Université de Madrid, le poète Jacob Wassermann, de Vienne, l'écrivain français Alfred Fabre-Luce, le poète bâlois Liehburg et l'un des chefs de la jeunesse européenne, Gerhard Meyer, ont pris la parole. Tous ont été d'avis que, sans une collaboration efficace de la jeunesse du 20^{me} siècle, on n'arrivera pas à une Paneurope. Il faut agir rapidement car la situation est critique et dangereuse.

Cette manifestation de la jeunesse européenne au cours de laquelle des messages provenant du sénateur français Caillaux, de Francis Delaisi, du comte Moltke, ancien ministre des affaires étrangères du Danemark, de M. Grimm, ancien ministre des cultes de Prusse, ont été lus, avait réuni un nombre imposant de participants.

Identified named entities:

Basel
Denmark
Paneurope
Liehburg
Moltke
Jacob Wassermann
Recasens-Siches
Francis Delaisi
Caillaux
Gerhard Meyer
Alfred Fabre-Luce
M. Grimm

PanEurope use case

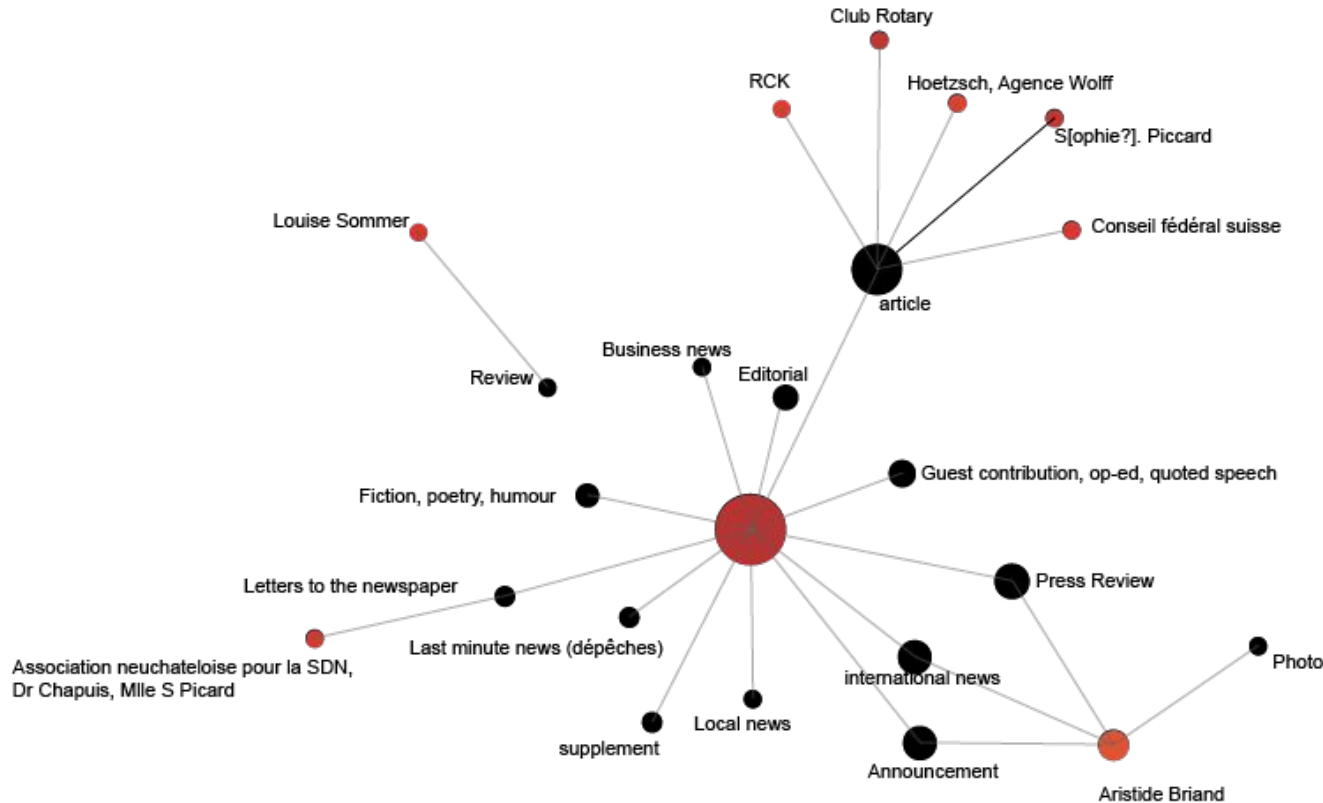


Newspaper type/
 newspaper
 sections

Setting the frame to
 integrate further
 automatically
 generated data, such
 as NER?

A base for
 recommendation?

Paneurope use case



Newspaper type/
newspaper
sections

Setting the frame to
integrate further
automatically
generated data, such
as NER?

A base for
recommendation?

CONTEXT /
LIST OF TOPICSHEADLINES
OF THE DAYTABLE OF
CONTENTSVI / RUE DES PALMES
p. 2, *novel*La Gazette de Lausanne
Lundi 16 Mai 2018

6 PAGES - 25 ARTICLES

Archive Du TEMPS

1540 issues from 1840 to 1950

LIST OF ARTICLES

ALL ARTICLES

ONLY ARTICLES IN MY COLLECTIONS

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

VI / RUE DES PALMES

p. 2, *novel*

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les ...

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

Vous m'avez exhorté à la franchise, la dernière fois que j'ai l'honneur de devise, rien pour rien.

POEMS()

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

CURRENT SEARCH

p. 2

VI

RUE DES PALMES

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les stations hivernales.

Dès ces premiers jours de juin, Nice, abandonnée sous la poussière, prenait son aspect de Pompéi ensevelie sous la cendre. Avec les étrangers, la richesse, le mouvement, la vie, se retirait. Hôtels et villas se fermaient successivement. Les rues semblaient vidées. Sur le seuil des magasins ternes, des boutiquiers mélancoliques haletaient entre deux échéances, et la plage, démeublée des pittoresques silhouettes de Russes, d'Anglais et d'Américains, faisait l'effet d'un cadre sans tableau.

Son caractère cosmopolite ainsi perdu, la capitale de l'exotisme se retrouvait simple ville de province; et, du même coup, l'esprit de province y revivait. S'observer et se quereller entre soi restait à peu près la seule distraction de ceux auxquels une raison spéciale n'avait

p. 2

pas encore permis de prendre leur vol.

Les Roquiescize attendaient les vacances des enfants, les du Valbonnet la liquidation de l'héritage, la princesse le départ de son neveu, Mme Guidal l'arrivée de sa fille, la comtesse de Wartembode, une invitation à la campagne.

Des motifs plus impérieux encore retenaient les Faveroles dans leur petite villa blanche du quartier des Bumesles.

La santé d'Yvonne commandait l'intérêt, et, pendant une quinzaine, un médecin attentif l'avait clouée sur une chaise longue, auprès de laquelle Hubert s'était plu à mouter une garde empressée.

C'avait été un nouvel et charmant épisode de leur vie conjugale que cette réclusion à deux.

Alanguie et nonchalante sous les plis vaporeux d'un élégant peignoir, ses cheveux blonds coquettement ébouriffés, Yvonne faisait la plus gentille malade du monde; et, dans cette jolie chambre de jeune mariée, où les stores baissés

VI

ADD TAG ...

ADD TAG

filter predefined tag ...

FORMAT

CONTENTS

Advertisements (commercial)

Arts, culture and entertainment

Announcement (hard news,
non commercial)

Business news

Banner

Crime, law and justice news

Box

 Fiction, poetry, humour
you and 3 more users selected this

Cartoon

International news

Civil registry (obituaries,
weddings, births)

Last minute news (dépêches)

Column

Local news

Cutline (pictures)

News story (faits-divers)

Ears

National news

Guest contribution, op-ed,
quoted speechReviews (books,
entertainment...)

Index, table of content

Science news

Interviews

Sports news

 Letters to the newspaper

1 user selected this

Photograph

Press review

SUGGEST NEW

EXPLORE TAGS ...

Sans mentir, si votre ramage

Se rapporte à votre plumage,

Vous êtes le Phénix des hôtes de ces bois. »

A ces mots le Corbeau ne se sent pas de joie ;

Et pour montrer sa belle voix,

Il ouvre un large bec, laisse tomber sa

C.1

Suggest
Public Tags
(fixed vocabula

C.2

"Format":
newspaper
sections or spec
page regions

C.3

"Contents":
tag mainly relat
article content

C.4

See and use
other users'
public tags

C.5

Explore
impresso by tag
(coming)

VI

RUE DES PALMES

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les stations hivernales.

Dès ces premiers jours de juin, Nice, abandonnée sous la poussière, prenait son aspect de Pompéi ensevelie sous la cendre. Avec les étrangers, la richesse, le mouvement, la vie, se retiraient. Hôtels et villas se fermaient successivement. Les rues semblaient vidées. Sur le seuil des magasins ternes, des boutiquiers mélancoliques haletaient entre deux échéances, et la plage, démeublée des pittoresques silhouettes de Russes, d'Anglais et d'Américains, faisait l'effet d'un cadre sans tableau.

Son caractère cosmopolite ainsi perdu, la capitale de l'exotisme se retrouvait simple ville de province; et, du même coup, l'esprit de province y revivait. S'observer et se quereller entre soi restait à peu près la seule distraction de ceux auxquels une raison spéciale n'avait

as encore permis de prendre leur vol.

Les Roquiescize attendaient les vacances des enfants, les du Valbonnet la liquidation de héritage, la princesse le départ de son neveu, Mme Guida l'arrivée de sa fille, la comtesse Wartembrode, une invitation à la campagne.

Des motifs plus impérieux encore retenaient s Faveroles dans leur petite villa blanche du quartier des Baumettes.

La santé d'Yvonne commandait l'intérêt, et, pendant une quinzaine, un médecin attentif avait clouée sur une chaise longue, auprès de laquelle Hubert s'était plu à monter une garde

VI

ADD TAG ...

ADD TAG

filter predefined tag ...

FORMAT	CONTENTS
Advertisements (commercial)	Arts, culture and entertainment
Announcement (hard news, non commercial)	Business news
	Crime, law and justice news
Banner	✓ Fiction, poetry, humour ✕ you and 3 more users selected this
Box	
Cartoon	International news
Civil registry (obituaries, weddings, births)	Last minute news (dépêches)
Column	Local news
Cutline (pictures)	News story (faits-divers)
Ears	National news
Guest contribution, op-ed, quoted speech	Reviews (books, entertainment...)
Index, table of content	Science news
Interviews	Sports news
✓ Letters to the newspaper 1 user selected this	
Photograph	
Press review	

SUGGEST NEW EXPLORE TAGS ...

Sans mentir, si votre ramage
Se rapporte à votre plumage,
Vous êtes le Phénix des hôtes de ces

C.1

Suggest
Public Tags
(fixed vocabulary)

C.2

“Format”:
newspaper
sections or specific
page regions

C.3

“Contents”:
tag mainly related to
article content

C.4

See and use
other users'
public tags

VI / RUE DES PALMES

p. 2, *novel*

VI

RUE DES PALMES

C'était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les stations hivernales.

Dès ces premiers jours de juin, Nice, abandonnée sous la poussière, prenait son aspect de Pompéi ensevelie sous la cendre. Avec les étrangers, la richesse, le mouvement, la vie, se retiraient. Hôtels et villas se fermaient successivement. Les rues semblaient vidées. Sur le seuil des magasins ternes, des boutiquiers mélancoliques haletaient entre deux échéances, et la plage, démeublée des pittoresques silhouettes de Russes, d'Anglais et d'Américains, faisait l'effet d'un cadre sans tableau.

Son caractère cosmopolite ainsi perdu, la capitale de l'exotisme se retrouvait simple ville de province ; et, du même coup, l'esprit de province y revivait. S'observer et se quereller entre soi restait à peu près la seule distraction de ceux auxquels une raison spéciale n'avait

pas encore permis de prendre leur vol.

Les Roquescize attendaient les vacances des enfants, les du Valbonnet la liquidation de l'héritage, la princesse le départ de son neveu, Mme Guida l'arrivée de sa fille, la comtesse de Wartembrode, une invitation à la campagne.

VI

ADD TAG ...

ADD TAG

essay

FORMAT

CONTENTS

no content tags matching "essay"

no content tags matching "essay"

CATEGORY

VALUE

FORMAT

CONTENTS

essay

SHORT DESCRIPTION

SUGGEST NEW

EXPLORE TAGS ...

who served as the leader of the Republic of China

était bien l'été cette fois, la morte-saison des villes qui, nulle part, ne s'éprouve autant que dans les Maîtres Corbeau, sur un arbre perché, Tenait en son bec un fromage. Maître Renard, par l'odeur alléché, lui fit à son tour sa leçon.

C.7

“Suggest new”
Add a description
for the tag

YOUR SEARCH

PAST SEARCHES

FULL TEXT SEARCH

REFINE ...

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

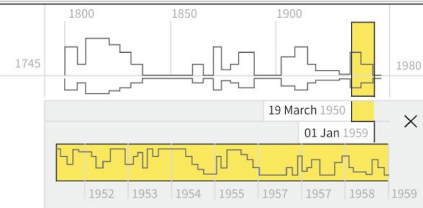
CONTAINS EXACTLY guerre froide

CONTAINS europe

PUBLISHED IN (DATE RANGE)

TIMELINE OF N. ARTICLES

RESET



ADD RANGE ...

IN NEWSPAPER TITLES

RESET

112 **La Gazette de Lausanne**
quotidien suisse de langue française édité à Lausanne

TYPES OF ARTICLES

TOP NAMED ENTITIES

112 international news

103 Napoleon

50 politics

50 Zurich, Suisse
location

LANGUAGE OF ARTICLES

112 French

GROUP RESULTS BY

ISSUE

PAGE

ARTICLES

SENTENCES

i

ORDER BY

RELEVANCE

DISPLAY RESULTS

AS LIST

AS TILES

SEARCH SUMMARY

Found 112 articles matching exactly "guerre froide" AND europe published in La Gazette de Lausanne between Saturday, 19 March 1950 and Friday, 01 Jan 1959

SAVE SEARCH

COMPARE ...



1 of 112

Les recherches d'une science jeune: CELLE DE L'INFORMATION

INTERNATIONAL NEWS TAG AS ...
Gazette de Lausanne, Saturday, May 8, 1954, p. 15
Les recherches d'une science jeune CELLE La presse, son rôle possibilités JEAN-PIERRE AGUET ^{AA} . ^A . ^A . ^{AA} H est à l'heure de multiples...

en préjugés sur . ses', voisins et souvent en guerre « froide »-ou « chaude ». « UNE SEMAINE DANS LE MOND

, est la presse nationale de plu- sieurs pays parmi lesquel les Etats- Unis , huit pays d'Europe occidentale

ADD TO FAVOURITES

ADD TO COLLECTION...

ADD TO COLLECTION

filter or create collection

favorites

✓ conference reports

guerre froide

CREATE NEW

MANAGE MY COLLECTIONS

S.10

add article to a collection



2 of 112

Les USA et l'AFN

INTERNATIONAL NEWS COVER PAGE TAG AS ...

Gazette de Lausanne, Tuesday, November 19, 1957, p. 1
Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A la terrasse de café où je suis assis , me défendant tant bien que mal contre l assaut...

- nomie . Tant qu'a duré la guerre froide , Washington n'a pas eu d'autre souri que de combattre

en chef des forces du Centre-Europe , « chef atlantique sûr » , s'employa à accréditer cet- te version

ADD TO FAVOURITES

ADD TO COLLECTION...

GUERRE FROIDE X

S.11

display list of matching text segments in the article



3 of 112

FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE

INTERNATIONAL NEWS TAG AS ...

Gazette de Lausanne, Friday, April 27, 1951, p. 1
FRANCO PRÉFÈRE LE TÊTE-A-TÊTE AVEC L'AMÉRIQUE SOUS LE PRINTEMPS D'ESPAGNE ATTENDANT UN TARDIF REPENTIR

ADD TO FAVOURITES

ADD TO COLLECTION...

EXPORT CITATIONS ...

S.12

what you would like to export? Individual articles can be downloaded (text, pdf, images)

S.8
include / exclude options are given for each text string

S.9
zooming on the selected date range

Alba Rorwacher Researcher EN

personal space / my collections

MY COLLECTIONS MY SAVED QUERY

DISPLAY ITEMS AS LIST OF ITEMS VISUALIZE AS NETWORK ORDER BY LATEST ADDED

LIST OF MY COLLECTIONS

ORDER BY LATEST MODIFIED

PANEUROPE 1 of 4
 YESTERDAY, 12H15 - collection of pan-europe and paneurope articles, manually curated
 2 ISSUES 8 PAGES 700 ARTICLES

UNITED STATES OF EUROPE 3 of 10
 collection of pan-europe and paneurope articles, manually curated
 8 PAGES 7 ARTICLES

GUERRE FROIDE & EUROPE 3 of 10
 29 JUNE 2018, 12H15 - collection of pan-europe
 1 PAGE 2 ARTICLES

LES CHIMERES 4 of 4
 collection of pan-europe and paneurope articles, manually curated
 2 ISSUES 8 PAGES 700 ARTICLES

CONFERENCE REPORTS
 -
 2 ISSUES 8 PAGES 700 ARTICLES

ADD NEW EMPTY COLLECTION

ALWAYS AVAILABLE COLLECTIONS

All items in all personal collection
 contains every article, issue, page you have ever included in one of your collections
 2 ISSUES 8 PAGES 700 ARTICLES

FAVOURITE ITEMS
 This list contains
 2 ISSUES 8 PAGES 700 ARTICLES

SUMMARY
 710 items in the collection **Guerre Froide & Europe**, ordered by **latest added**
 EXPORT CITATIONS ... EXPORT ITEMS METADATA

GUERRE FROIDE & EUROPE
 29 JUNE 2018, 12H15 - collection of pan-europe
 EDIT TITLE / DESCRIPTION REMOVE COLLECTIONS

Gazette de Lausanne
 Page 1
 Wednesday, June 10, 1959, p. 1

La section Léman de l'Union européenne se préoccupe de l'intégration économique et du problème de Berlin
 INTERNATIONAL NEWS
 Gazette de Lausanne, Wednesday, June 10, 1959, p. 4
 La section Léman de l'Union européenne se préoccupe de l'intégration économique et du problème de Berlin La section Léman de l'Union européenne a tenu son assemblée à l'Hôtel de la ...
 EXPORT CITATIONS ... DOWNLOAD AS ...

En rejetant les propositions alliées pour le règlement du problème autrichien Le Kremlin ne paraît pas près de modifier sa politique en Europe
 INTERNATIONAL NEWS
 Gazette de Lausanne, Monday, August 18, 1952, p. 6
 En rejetant les propositions alliées pour le règlement du problème autrichien Le Kremlin ne paraît pas près de modifier sa politique en Europe (U . P .) Le...

OTHER COLLECTIONS
 CONFERENCE REPORTS ADD TO COLLECTION

KEYWORDS SEARCHED
 au mouvement Paneurope et ses nombreuses publications et conférences sur des sujets européens . Commentant
 VIEW ORIGINAL QUERY

OTHER COLLECTIONS
 ADD TO COLLECTION

KEYWORDS SEARCHED
 d'une modification de la po- litique étrangère de Moscou et d'un règlement prochain dé la « guerre froide
 pas près de modifier sa politique en Europe (U . P .) Le rejet par Le Kremlin des propositions
 VIEW ORIGINAL QUERY

which type of visualisation (list vs graph, cf. Or.9)

Or.5
 structured overview of your collection, export citations or metadata

Or.6
 list of items included in your collection

Or.7
 assign item to more than one collections

Or.1
 list of your "collections" sort by: latest created, latest modified, alphabetically

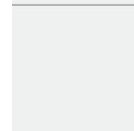
Or.2
 selected (active) collection, with quantity of articles and pages

Or.3
 default collections like "favourites" cannot be deleted or modified

Or.8
 "How did you find this item?" list of search results and link to the original query

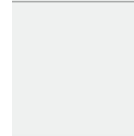


4 of 4



N

have ever



2 of 3

La section Léman de l'Union européenne se préoccupe de l'intégration économique et du problème de Berlin

INTERNATIONAL NEWS

Gazette de Lausanne, Wednesday, June 10, 1959, p. 4
La section Léman de l'Union européenne se préoccupe de l'intégration économique et du problème de Berlin La section Léman de l'Union européenne a tenu son assemblée à l'Hôtel de la ...

[EXPORT CITATIONS ...](#) [DOWNLOAD AS ...](#)

OTHER COLLECTIONS

CONFERENCE REPORTS

[ADD TO COLLECTION](#)

KEYWORDS SEARCHED

au mouvement Paneurope et ses nombreuses publications et conférences sur des sujets européens . Commentant

[VIEW ORIGINAL QUERY](#)



En rejetant les propositions alliées pour le règlement du problème autrichien Le Kremlin ne paraît pas près de modifier sa politique en Europe

INTERNATIONAL NEWS

Gazette de Lausanne, Monday, August 18, 1952, p. 6
En rejetant les propositions alliées pour le règlement du problème autrichien Le Kremlin ne paraît pas près de modifier sa politique en Europe (U . P .) Le...

OTHER COLLECTIONS

[ADD TO COLLECTION](#)

KEYWORDS SEARCHED

d'une modification de la po- litique étrangère de Moscou et d'un règlement prochain dé la « guerre froide

pas près de modifier sa politique en Europe (U . P .) Le rejet par le Kremlin des propositions

[VIEW ORIGINAL QUERY](#)

MY
COLLECTIONSMY SAVED
QUERY

DISPLAY ITEMS AS

LIST OF ITEMS

VISUALIZE AS TIMELINE

EXPLORE METADATA

ORDER BY

LAST ADDED ▾

LIST OF MY COLLECTIONS

ORDER BY **LATEST MODIFIED** ▾**PANEUROPE**

1 of 4

YESTERDAY, 12H15 - collection of pan-europe and paneurope articles, manually curated

135 ARTICLES

UNITED STATES OF EUROPE

3 of 10

collection of pan-europe and paneurope articles, manually curated

8 PAGES 7 ARTICLES

GUERRE FROIDE & EUROPE

3 of 10

29 JUNE 2018, 12H15 - collection of pan-europe

1 PAGE 2 ARTICLES

LES CHIMÈRES

4 of 4

collection of pan-europe and paneurope articles, manually curated

2 ISSUES 8 PAGES 700 ARTICLES

CONFERENCE REPORTS

2 ISSUES 8 PAGES 700 ARTICLES

ADD NEW EMPTY COLLECTION

ALWAYS AVAILABLE COLLECTIONS

All items in all personal collection

contains every article, issue, page you have ever included in one of your collections

2 ISSUES 8 PAGES 700 ARTICLES

FAVOURITE ITEMS

This list contains

2 ISSUES 8 PAGES 700 ARTICLES

SUMMARY

Alluvial diagram of year, article sections, newspaper titles based on 135 items in the collection Paneurope

EXPORT CITATIONS ...

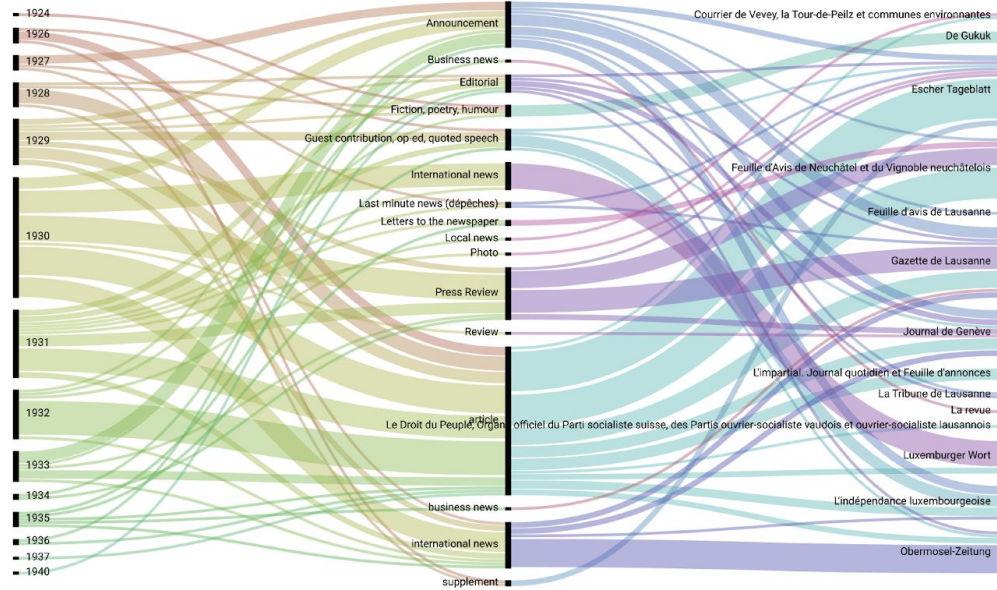
EXPORT ITEMS METADATA

PANEUROPE

29 JUNE 2018, 08H15 - collection of pan-europe and paneurope articles, manually curated

EDIT TITLE / DESCRIPTION

REMOVE COLLECTIONS



YEAR OF PUBLICATION ▾

ARTICLE SECTIONS ▾

NEWSPAPER TITLES ▾

ADD NEW DIMENSION ▾

1 | 1924

5 | 1926

5 | 1927

8 | 1928

15 | 1929

39 | 1930

2 | Business news

10 | Editorial

4 | Fiction, poetry, humour

9 | Guest contribution, op-ed, quoted ...

24 | International news

2 | Last minute news (dépêches)

23 | Gazette de Lausanne

23 | Escher Tageblatt

20 | Feuille d'Avis de Neuchâtel et du Vignoble ...

15 | Journal de Genève

14 | Obermosel-Zeitung

11 | L'indépendance luxembourgeoise

Or.10

Visual feedback
of the data dimension
chosen in [Or.11]

Or.11

Freely tune
the data
"dimensions"
from the
available
metadata

EST MODIFIED ▾

1 of 4

i - collection of pan-europe
articles, manually curated

OF EUROPE 3 of 10

y-europe and paneurope
ly curated

LES

E & EUROPE 3 of 10

415 - collection of pan-europe

ES

y-europe and paneurope 4 of 4

ly curated

5 700 ARTICLES

EPORTS

5 700 ARTICLES

NEW EMPTY COLLECTION

.E COLLECTIONS

ersonal collection

rticle, issue, page you have ever
of your collections

5 700 ARTICLES

MS

5

5 700 ARTICLES

Alluvial diagram of year, article sections, newspaper titles based on 135 items
in the collection **Panurope**

EXPORT CITATIONS ...

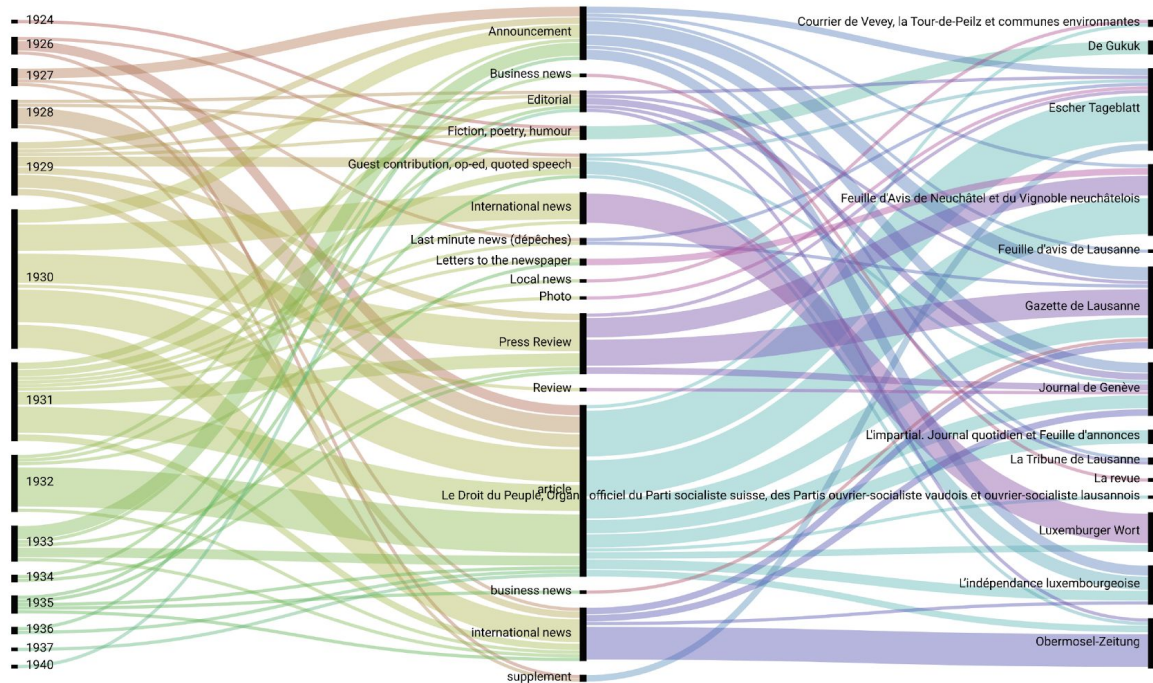
EXPORT ITEMS METADATA

PANEUROPE

29 JUNE 2018, 08H15 - collection of pan-europe and paneurope articles, manually curated

EDIT TITLE / DESCRIPTION

REMOVE COLLECTIONS



YEAR OF PUBLICATION ▾ < > × ARTICLE SECTIONS ▾ < > × NEWSPAPER TITLES ▾ < > × ADD NEW DIMENSION ▾

1 1924	2 Business news	23 Gazette de Lausanne
5 1926	10 Editorial	23 Escher Tageblatt
5 1927	4 Fiction, poetry, humour	20 Feuille d'Avis de Neuchâtel et du Vignoble ...
8 1928	9 Guest contribution, op-ed, quoted ...	15 Journal de Genève
15 1929	24 International news	14 Obermosel-Zeitung
39 1930	2 Last minute news (dépêches)	11 L'indépendance luxembourgeoise

Choose within a
wide range
of visual models

Or.10

Visual feedback
of the data dimension
chosen in [Or.11]

Or.11

Freely tune
the data
"dimensions"
from the
available
metadata

Collect and organise: interaction

Based on the previous searches: how to organise the results?

Culture : Trouver des concerts des 4 Barbus en Suisse (dates et lieux).

Cyclisme : Collecter des articles sur la participation de cyclistes suisses au Tour de France. (Piste: le 29 juillet 1951, Hugo Koblet remporte le Tour de France.)

Politique : Quelles sont les votations fédérales proposées entre 1950 et 1954?

Sciences: trouver des articles sur l'activité du Conseil européen pour la recherche nucléaire (CERN), créé le 29 septembre 1954.

- Advertisements (commercial)
- Announcement (hard news, non commercial)
- Arts, culture and entertainment news
- Business news
- Cartoon
- Civil registry (obituaries, weddings, births)
- Column
- Crime, law and justice news
- Cutline (pictures)
- Fiction, poetry, humour
- Guest contribution, op-ed, quoted speech
- Index, table of content
- International news
- Interviews
- Last minute news (dépêches)
- Letters to the newspaper
- Local news
- News story (faits-divers)
- National news
- Photograph
- Press review
- Reviews (books, entertainment...)
- Science news
- Sports news

Collecting and organising: outlook

How useful/not useful/other categorisations needed?

Further elements to organise the collection of articles:

- Validation of suggested generic categories?
- Linking out/interlinking the categories: Match your categories with e.g. wikidata/via WD to other sources of knowledges?
- **Recommendation** based on the stored elements (e.g. similarity based on NER)

Sharing the collections?

- For teaching?
- For collaborative research?

Side Note: Topic Modeling vs Text Classification



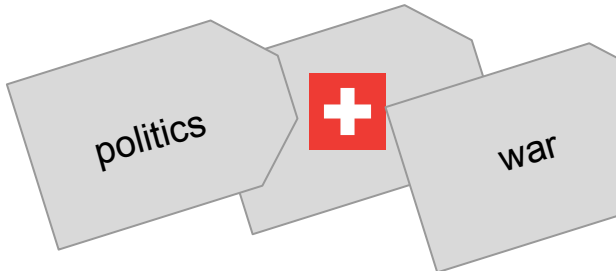
supervised
(labelled training data needed)

Indexing documents

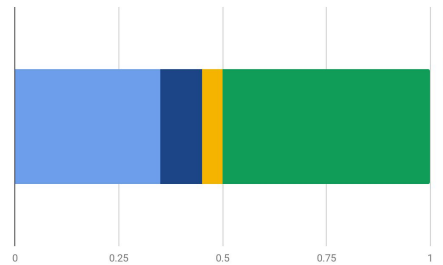
unsupervised

Text classification

Topic modeling



Topic distribution



Topic Modeling vs Text Classification (cont.)

Dienstag, 29 April 1948 Blatt 7 Der Zürcher Zeitung 169. Jahrgang Preis 20 Rp. Abendausgabe Nr. 207

Neue Zürcher Zeitung

Täglich drei Ausgaben und schweizerisches Handelsblatt

Zwei Kumburgen zum 1. Mai

Bei der erste Kumburgen der Zürcher Zeitung im Jahr 1848 in Paris stattfand, die 1. Mai zum Arbeit und Feiertag für die Arbeiter der Schweiz wurde. Die Arbeiter haben sich damals in einem Kampf um die Verkürzung der Arbeitszeit und die Einführung der 8-Stunden-Arbeitszeit. Die zweite Kumburgen fand im Jahr 1904 in Bern statt, wo die Arbeiter um die Verkürzung der Arbeitszeit und die Einführung der 8-Stunden-Arbeitszeit kämpften. Die Kumburgen sind ein wichtiger Bestandteil der Schweizer Arbeiterbewegung.

Waffenstillstand in Jerusalem

Die israelischen Streitkräfte haben einen Waffenstillstand mit den syrischen Streitkräften in Jerusalem vereinbart. Der Waffenstillstand ist ein wichtiger Schritt zur Beilegung des Konflikts in der Region. Die israelischen Streitkräfte haben die Kontrolle über die Stadt Jerusalem wieder erlangt.

Zimmerlage vermieten

Ein Zimmer in der Stadt Zürich ist vermietet. Die Zimmerlage ist schön und modern. Die Zimmerlage ist in der Nähe der Stadtmitte und hat eine tolle Aussicht auf die Stadt. Die Zimmerlage ist für Einzelpersonen oder Paare geeignet.

Verkaufen garten

Ein Garten in der Stadt Zürich ist zum Verkauf. Der Garten ist groß und schön. Der Garten hat eine tolle Aussicht auf die Stadt. Der Garten ist für Einzelpersonen oder Paare geeignet.

auskunft groß

Ein Haus in der Stadt Zürich ist zum Verkauf. Das Haus ist groß und schön. Das Haus hat eine tolle Aussicht auf die Stadt. Das Haus ist für Einzelpersonen oder Paare geeignet.

schön zürich haus

Ein Haus in der Stadt Zürich ist zum Verkauf. Das Haus ist schön und modern. Das Haus hat eine tolle Aussicht auf die Stadt. Das Haus ist für Einzelpersonen oder Paare geeignet.

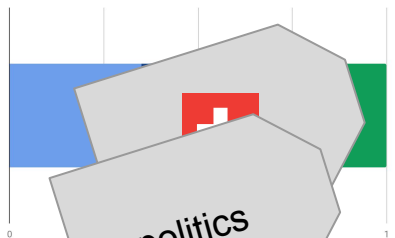
pension komfort preis see

Ein Pension in der Stadt Zürich ist zum Verkauf. Die Pension ist komfortabel und preiswert. Die Pension hat eine tolle Aussicht auf die Stadt. Die Pension ist für Einzelpersonen oder Paare geeignet.

telephon sonnig modern

Ein Haus in der Stadt Zürich ist zum Verkauf. Das Haus ist modern und sonnig. Das Haus hat eine tolle Aussicht auf die Stadt. Das Haus ist für Einzelpersonen oder Paare geeignet.

Topic distribution



politics

1

kunst konzert bild theater
künstler oper musikalisch
spielen werk künstlerisch
orchester aufführung
ausstellung lied musik vortrag
stück chor musil program

“arts”

Topic distribution



regional news

8

zimmer lage vermieten verkaufen
wohnung garten groß auskunft
villa hotel schön zürich haus
pension komfort preis see
telephon sonnig modern

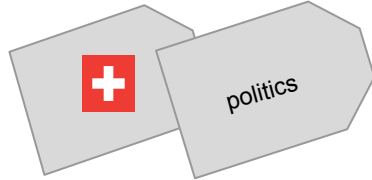
2

deutschen britisch deutsche truppe
krieg russisch englischer luftzug
feindlich schwer front armee
kampf london japanisch havas
amerikanisch feind alliiert
italienisch

“war”

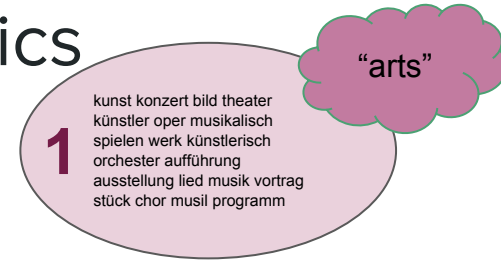
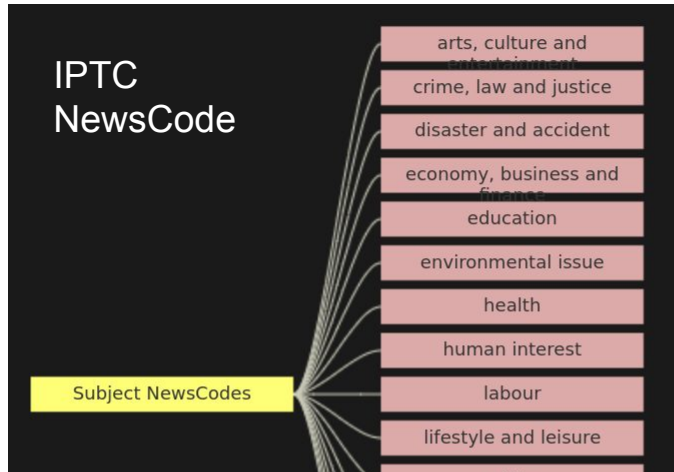
“flat advertisements”

Contrastive View: Classes vs. Topics



classes = defined concepts

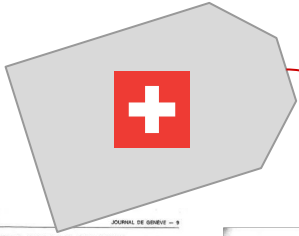
- taxonomy, e.g.



topics = probabilistic word distributions

- ideally semantically coherent
- NO definition
- NO predefined labels
- Should users be able to define their own interpretable labels?

Example: Faceted Search on Page



SUISSE JOURNAL DE GENÈVE - 9
No 63 - Mardi 18 mars 1974

Fortes diminutions de la grippe par rapport à 1973

REVISION PARTIELLE DE LA LOI SUR L'AVS Renvoi au Conseil fédéral pour réexamen mais voté

Banque nationale: M. Fritz Leutwiler succède à M. Stöpper à la présidence

PDC et défense nationale: critiques au Synode 72

Obèques de Mme Pierre Graber hier à Lausanne

Le «swi sur I»

Le trafic routier a rapporté 2 milliards: 60% des recettes douanières à la construction et à l'entretien d'infrastructures

Tagliero - Magagnoli Tagliero

Cartes postales

Cardinal 16, rue de Bourg Lausanne

SUISSE JOURNAL DE GENÈVE - 9
No 63 - Mardi 18 mars 1974

Papier plus cher: la hausse est justifiée, estime M. Schürmann

De quel groupe surgira la vérité constitutionnelle?

Diébet sur la participation au Conseil national

Siègne de l'après-midi à la chambre basse Les tenants des trois propositions continuent à s'affronter

13e rente AVS: en 1974 seulement

Les indépendants déposent une initiative fiscale

M. Hay ne méritait pas cela

Le marché des logements: tournant décisif?

A l'appel du RJ, des Juraissiens de l'extérieur rentrent au pays: l'UPJ dénonce une manœuvre

Militio changeants

Topic distribution

■ Topic 1
■ Topic 2
■ Topic 3
■ Topic 4

SUISSE JOURNAL DE GENÈVE - 9
No 63 - Mardi 18 mars 1974

Longévité: l'écart se creuse entre femmes et hommes

Les Suisses de France ont tenu congrès à Paris Droits politiques, information et AVS

Résultats de la consultation de consultation Entraide judiciaire internationale: de l'approbation à l'abstention

Basketball: sous l'impulsion de M. Leclerc Les «pros» à la conquête de l'Europe

1809, on brûle et sorcières...

Zurich: l'assassin des prostituées a avoué

Football: plongez-vous dans la coupe!

Appel d'offres international

Boxe: fiasco au Berinois Leo Kakolievicz

Basketball: 16e journée du championnat suisse

La coupe de la perplexité

126

Outlook on the personal space

Exportability of the collections?

> The collection is a starting for further exploitation.

>> Goal: output of a *semi-automatically* curated dataset to be used elsewhere.

Exploration of the stored collections?

> Visual explorations of the different personal collection...

Query/collection comparisons?

> Compare the collections in terms of NER, topics...

What's next?

- Workshop debrief in (short) Skype calls with you
- Lavender workshop 11.7. in Belval
- impresso community calls on topics of interest, details tba
- General updates via mailing list

..and finally

What are your thoughts on the direction of the impresso project?

What did you enjoy/not enjoy/miss during the workshop?

What can be improved for future events?

