

Data acquisition Data processing Interface



Media Monitoring
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

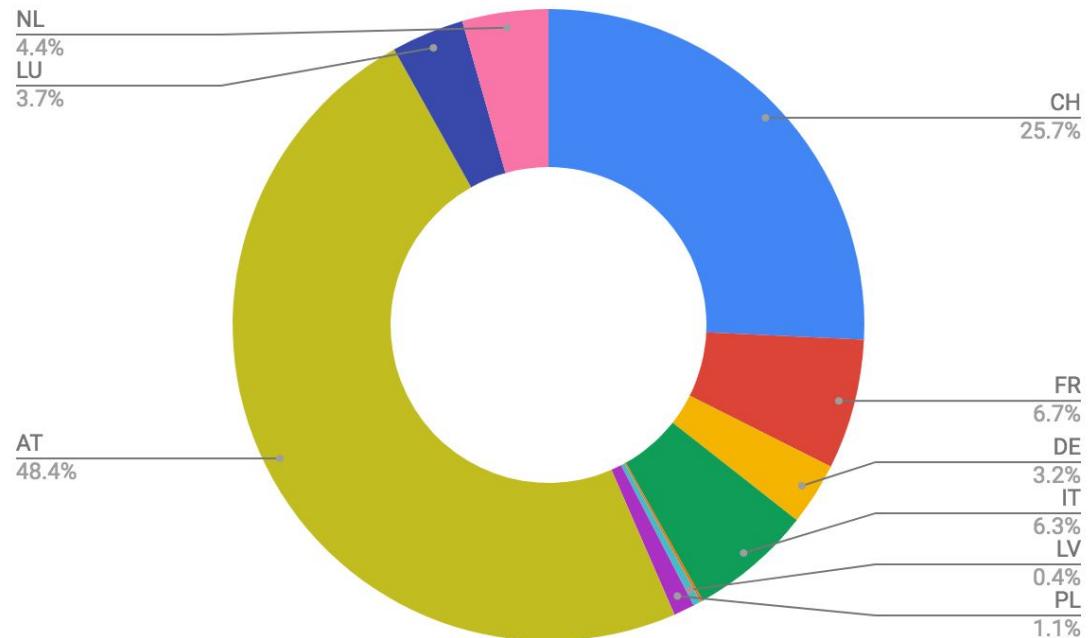
Where are we with newspapers acquisition?

Potential data providers - distribution by country

impresso first media inventory

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

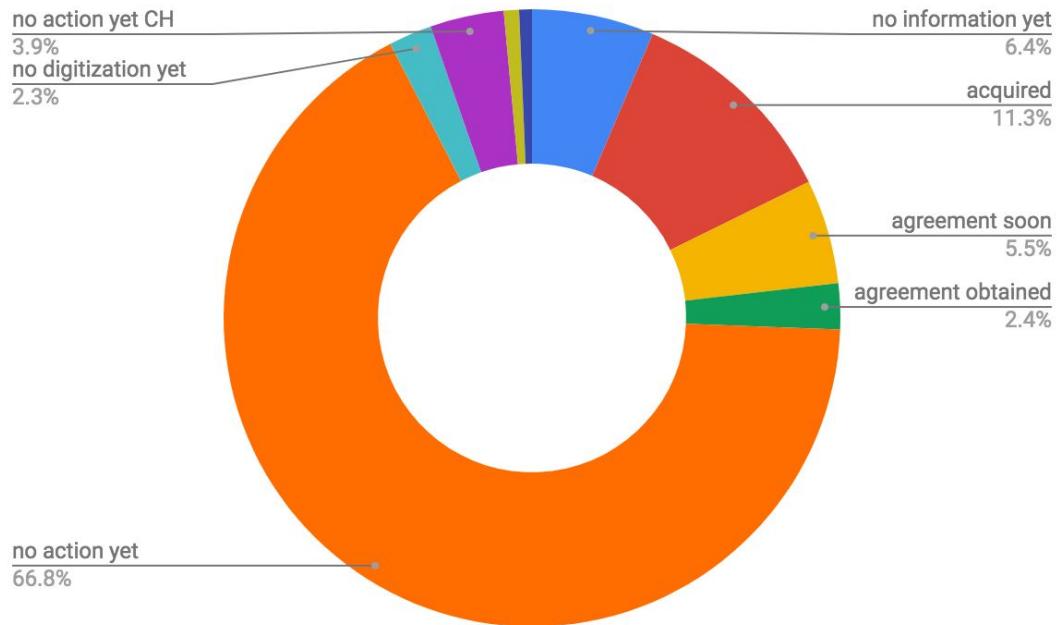
LFT (Bolzano), NLL (Latvia), NLE (Estonia),
NLP (Poland), ONB (Austria), SBB (Berlin),
SUBHH (Hamburg)



Acquisition status - all

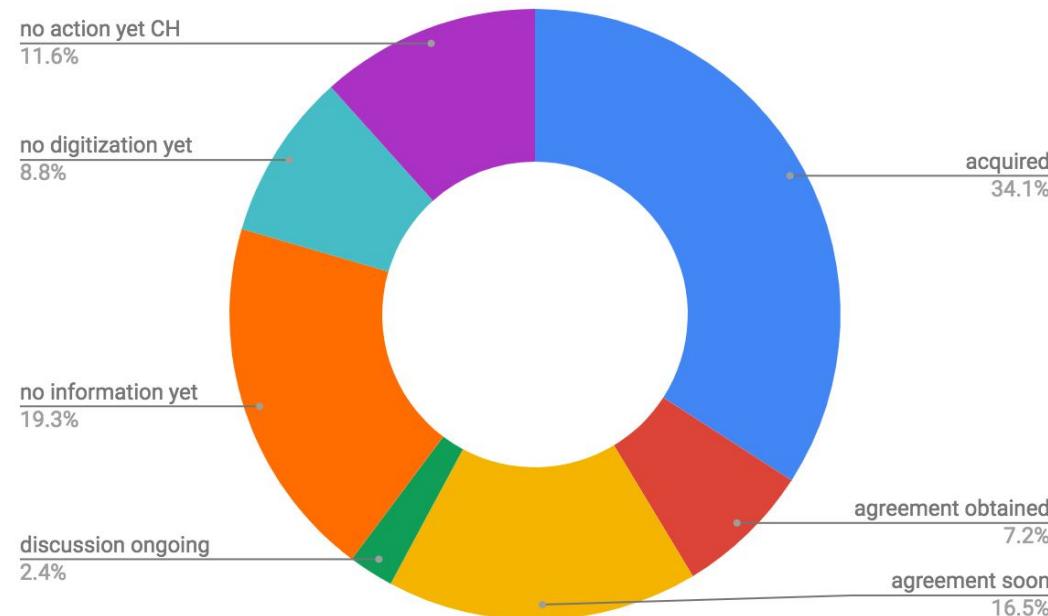
Action taken: 44%

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

NDAs 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

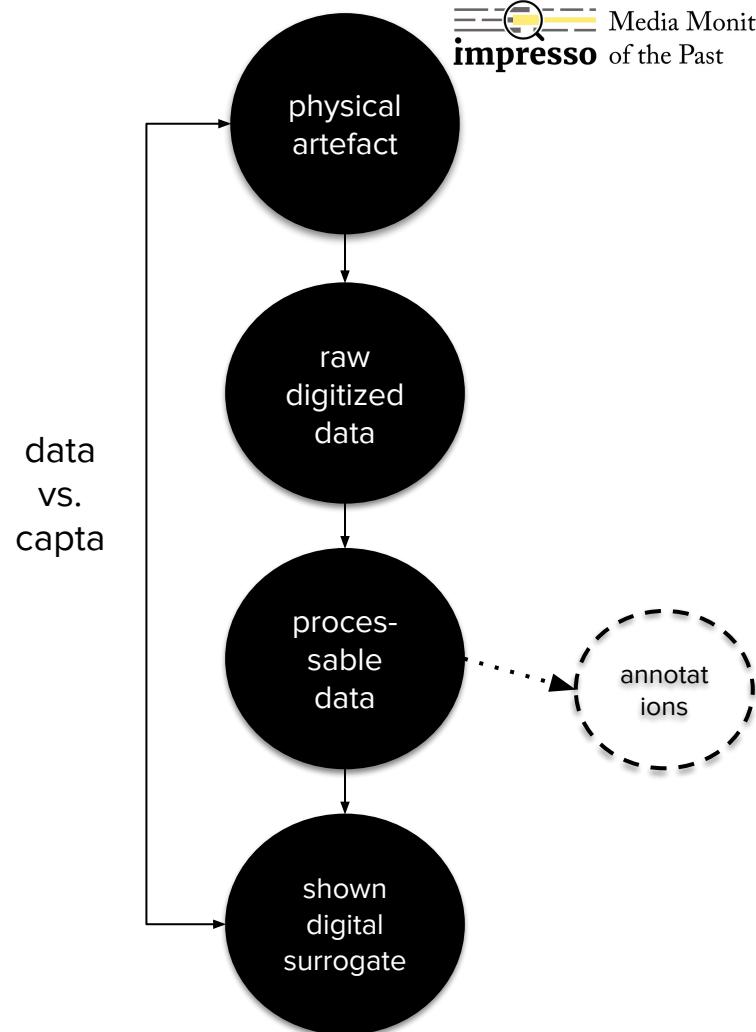
What do we do once we have the data?

A backstage tour

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

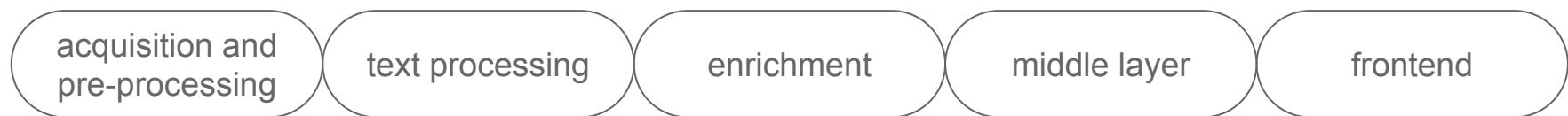
Here, **transparency** as a way to:

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



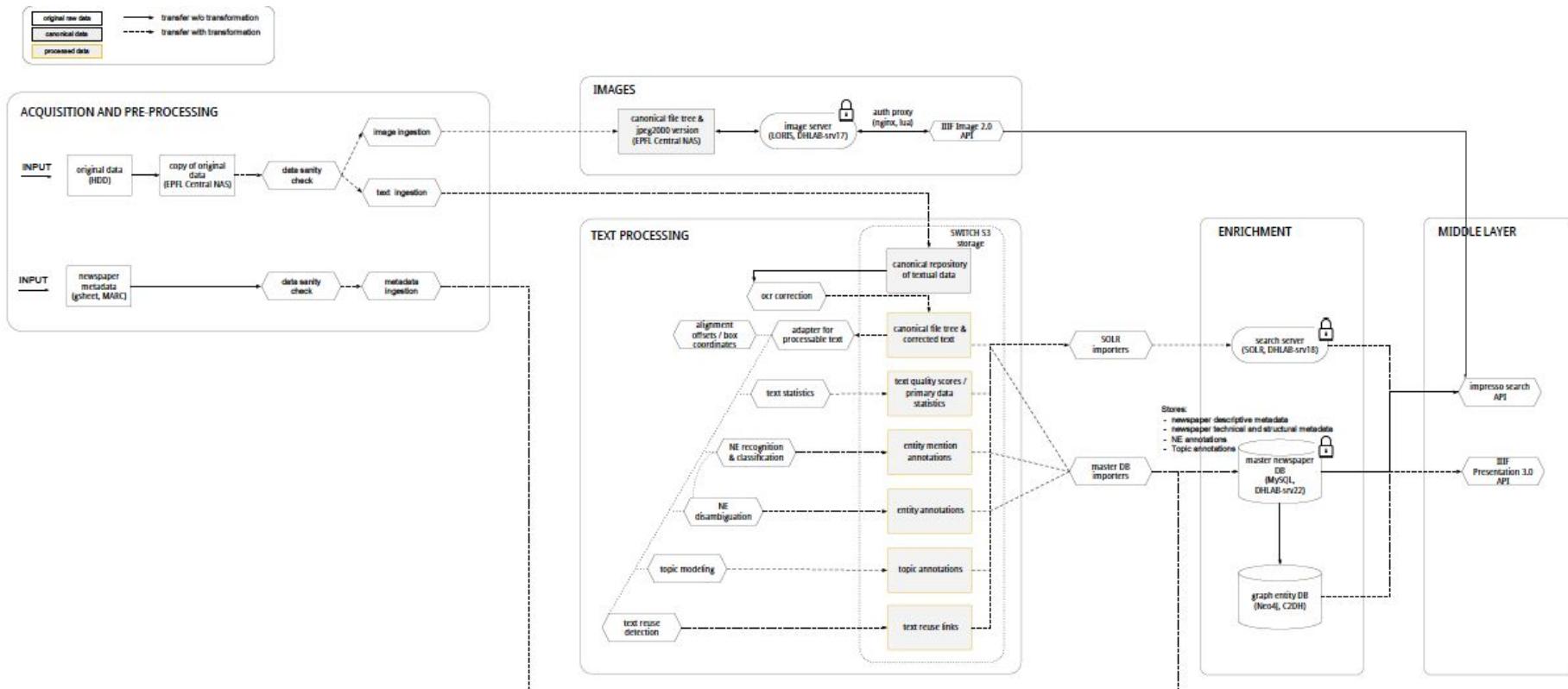
System architecture

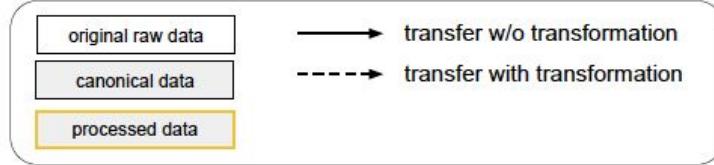
How do we organize data, its processing, and visualisation ?



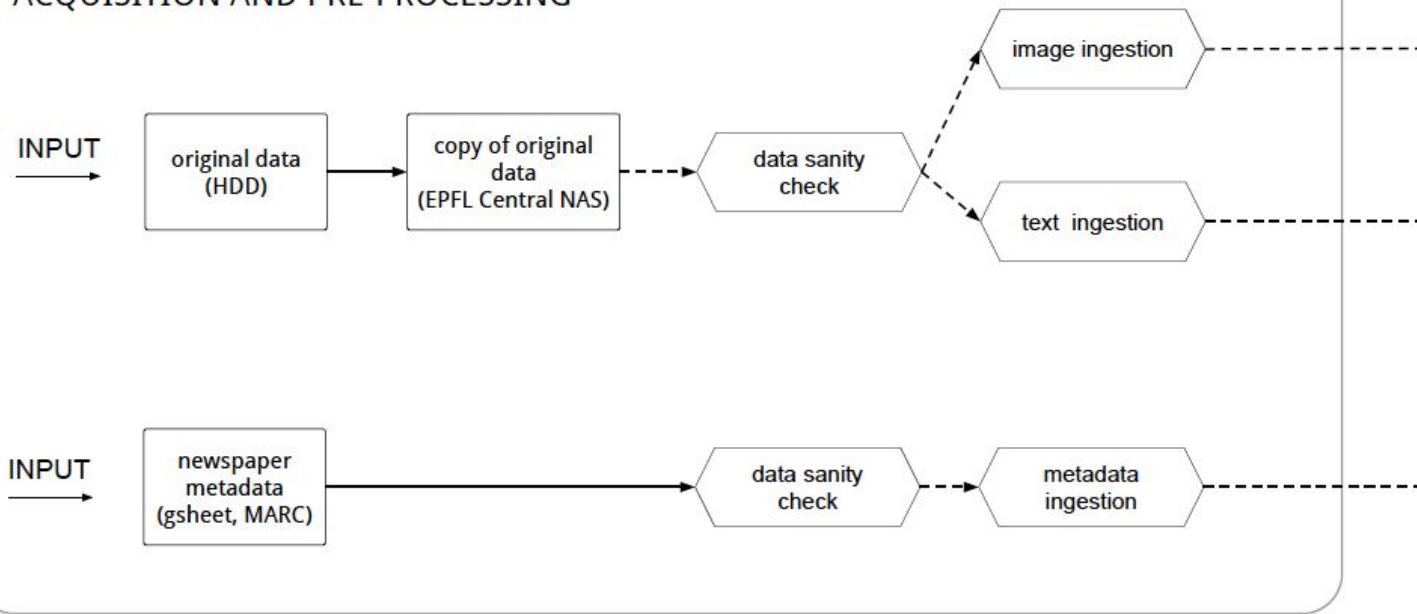
<https://docs.google.com/drawings/d/1puphbUKoOes6n2sb2DA-JrQtw9GPLn0mVfWx6xoNZ6I/edit?usp=sharing>

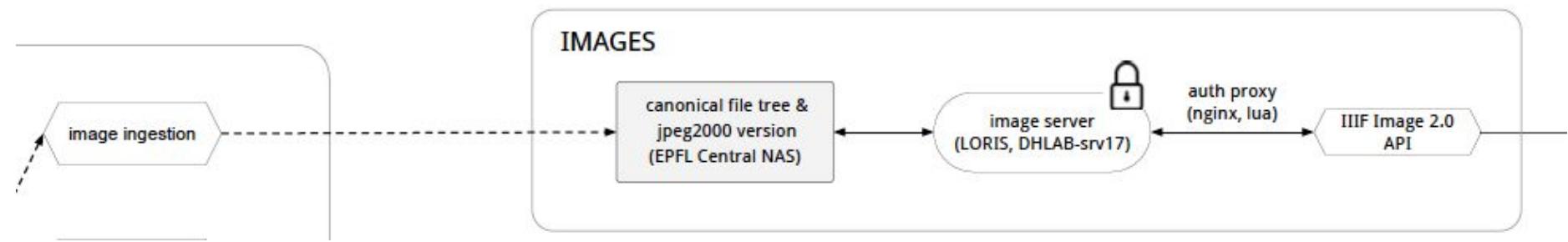
Overview

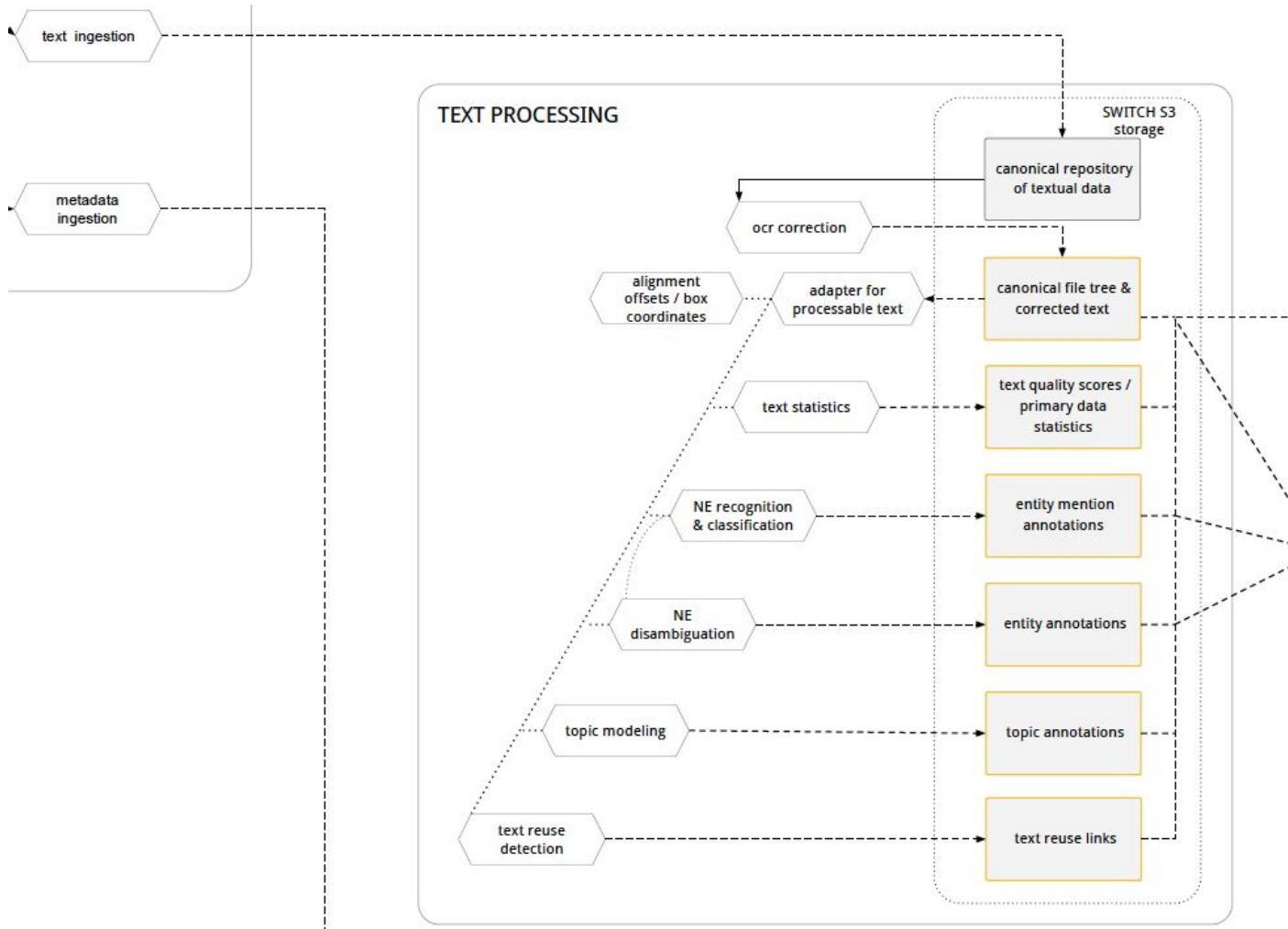


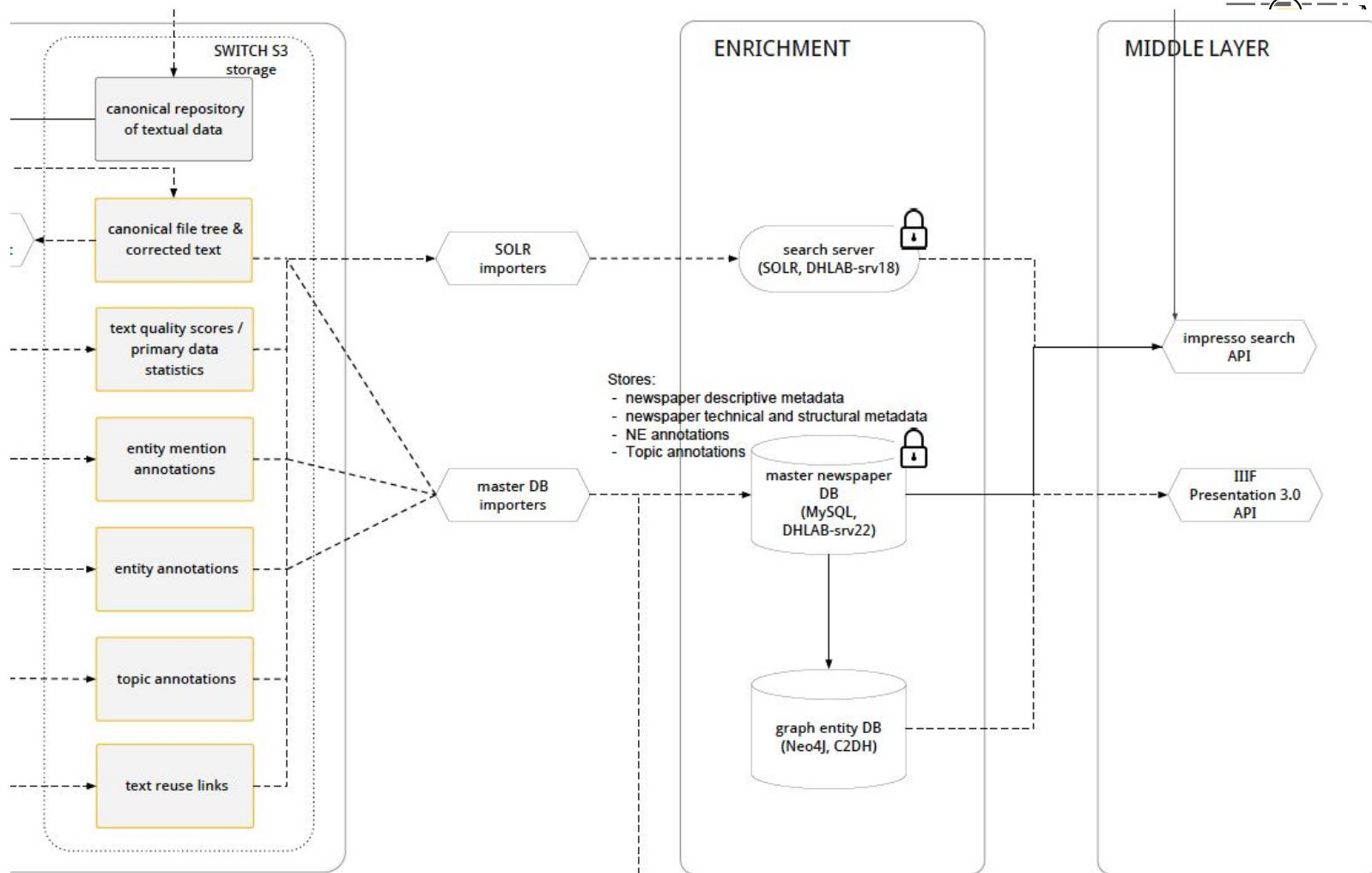


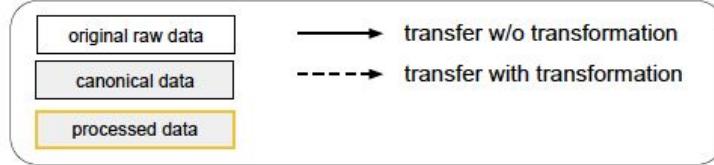
ACQUISITION AND PRE-PROCESSING











ACQUISITION AND PRE-PROCESSING

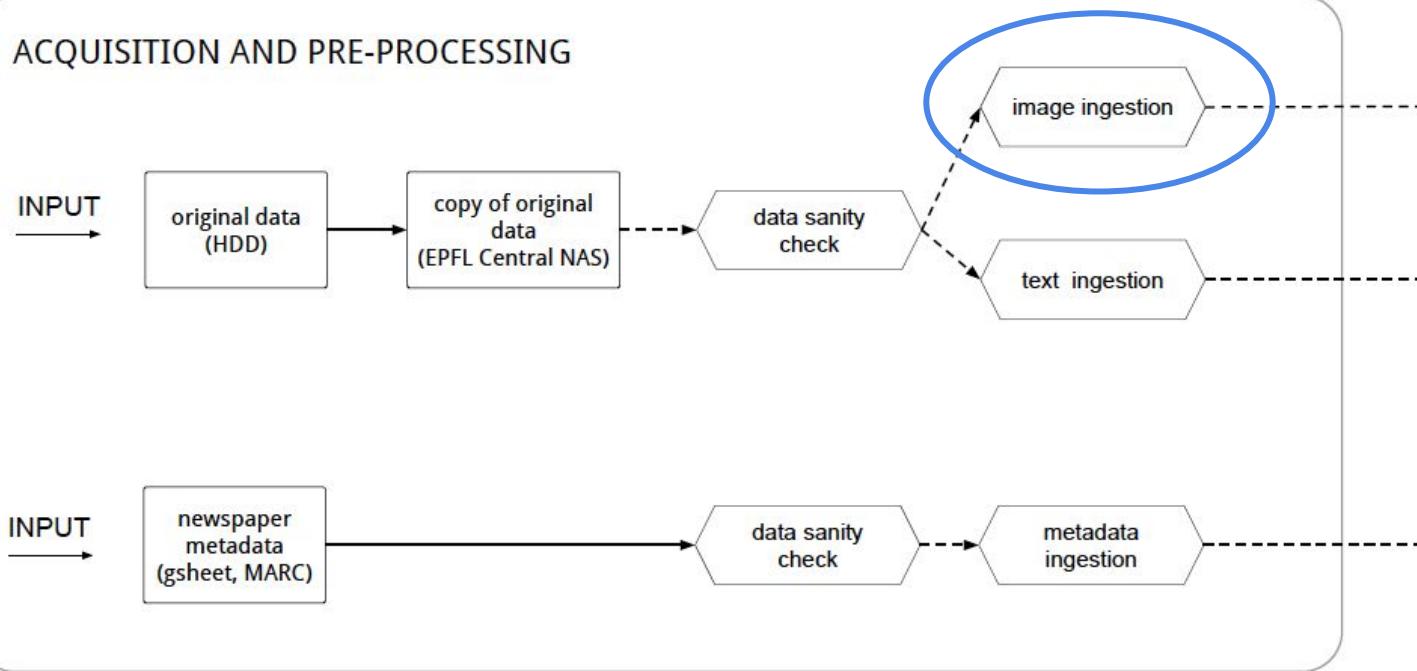


Image processing

given that...	...therefore
not all CH institutions provides IIIF 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

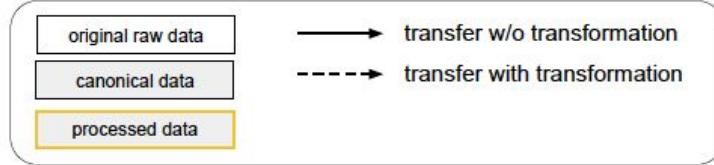
A common language for computers to talk about images.

Enable to easily access any part of any page:

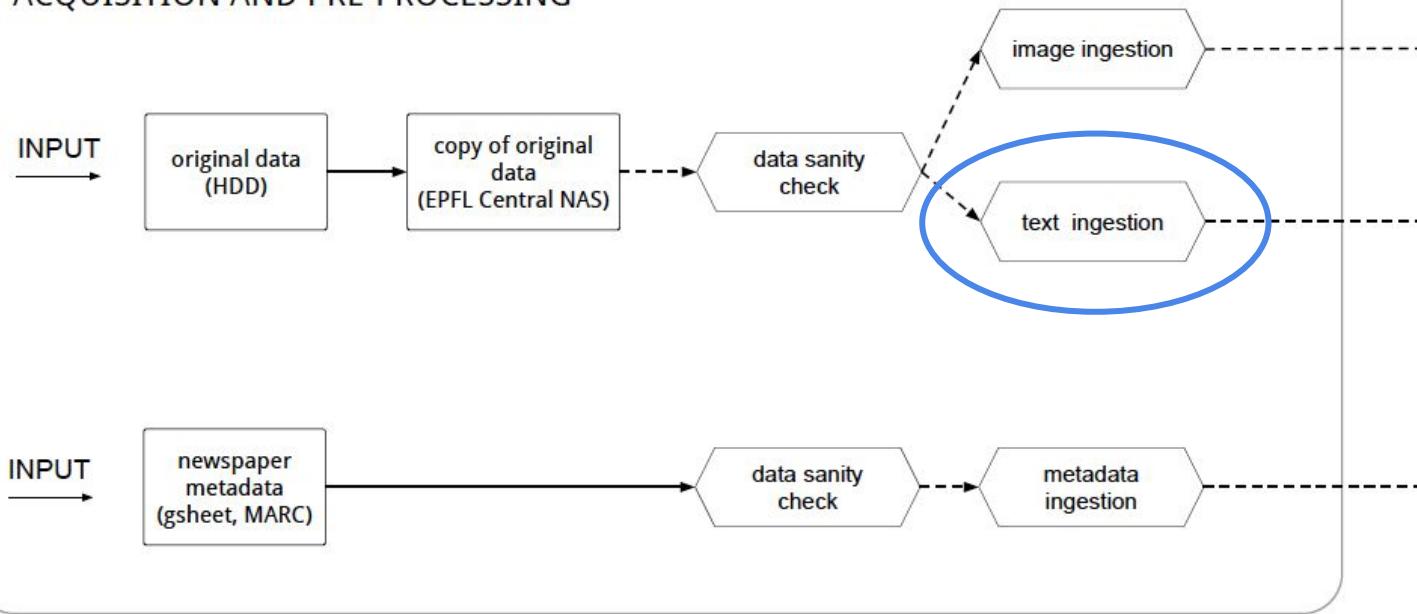
https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1950-04-27-a-p0007/198,5240,131,33/full/0/default.jpg

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

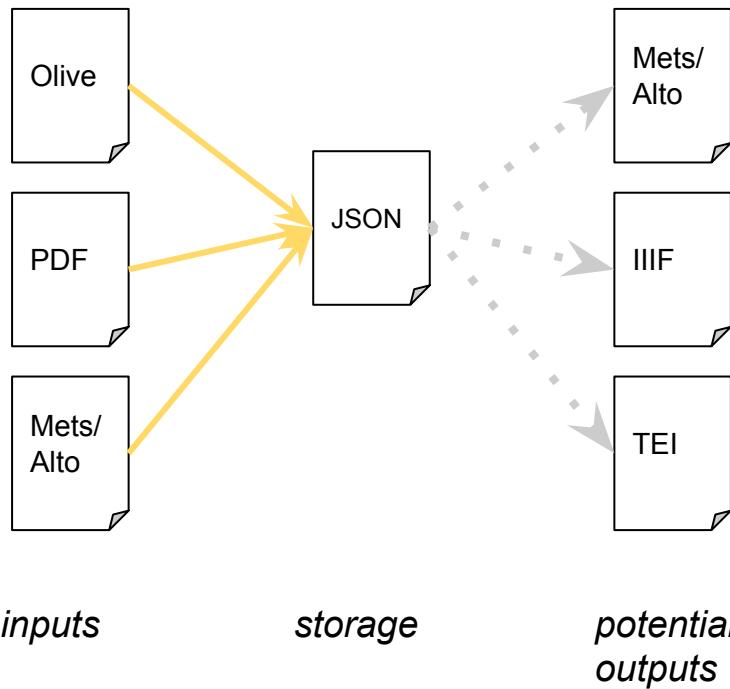
https://dhlabsrv17.epfl.ch/iiif_impresso/GDL-1950-04-27-a-p0007/full/full/0/default.jpg



ACQUISITION AND PRE-PROCESSING



Text acquisition



original-canonical-data	
Object Count:	1039970
Size:	732.7 GB
Date Created:	Jun 6, 2018
Public Access:	Disabled
<ul style="list-style-type: none"> ▼  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°. 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 suffrages de la république vous appellent de nouveau, presqu'à l'unanimité, à la présidence d'état. La première occasion dans laquelle le peuple de la Colombie a exercé la faculté inappréciable de nommer ses agents, a offert une preuve de sa reconnaissance, le son bon sens et de sa justice. Celui qui a créé la félicité politique de ce peuple devait en être le conservateur.

"Vous êtes appels 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.

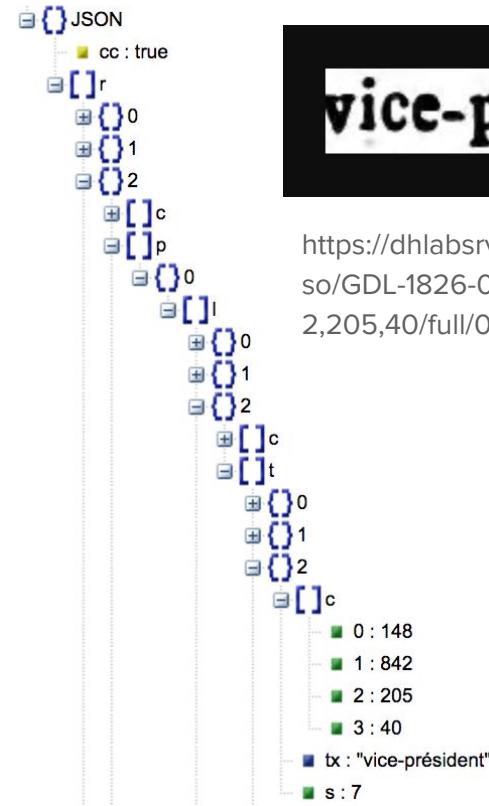
"Le vice-président de Colombie réunit ses vœux à ceux de ses concitoyens pour vous engager, non-seulement à accepter la présidence, mais encore à voter dans nos bras. Votre présence est importante dans tous les pays; partout votre nom est la terreur des ennemis publics et l'église des institutions libérales;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

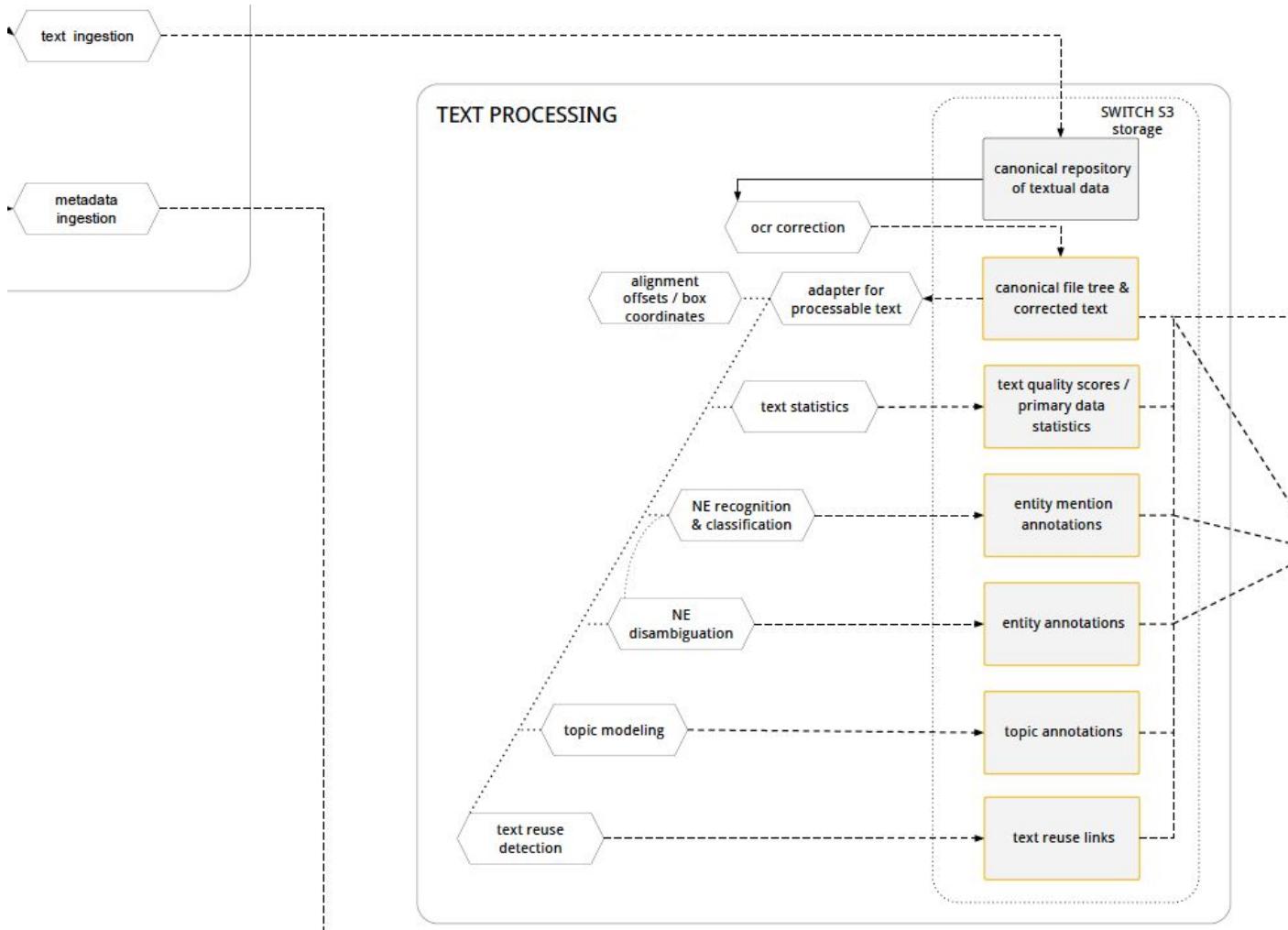
FRONTIERES DE SERVIE 20 mai.

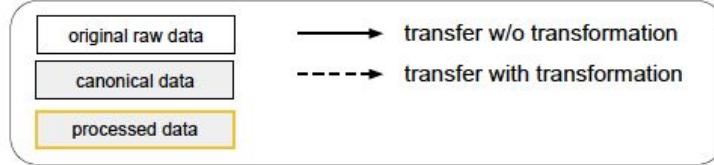
On sait qu'une conspiration tramée contre le gouverneur Milosch Obrenovics, fut découverte il y a quelque tems, en Servie. Le fils du fameux Czerny Georges, en était le principal chef, et plusieurs individus avaient été arrêtés comme complices. Ceux-ci viennent de subir un supplice qui fait frémir l'humanité.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



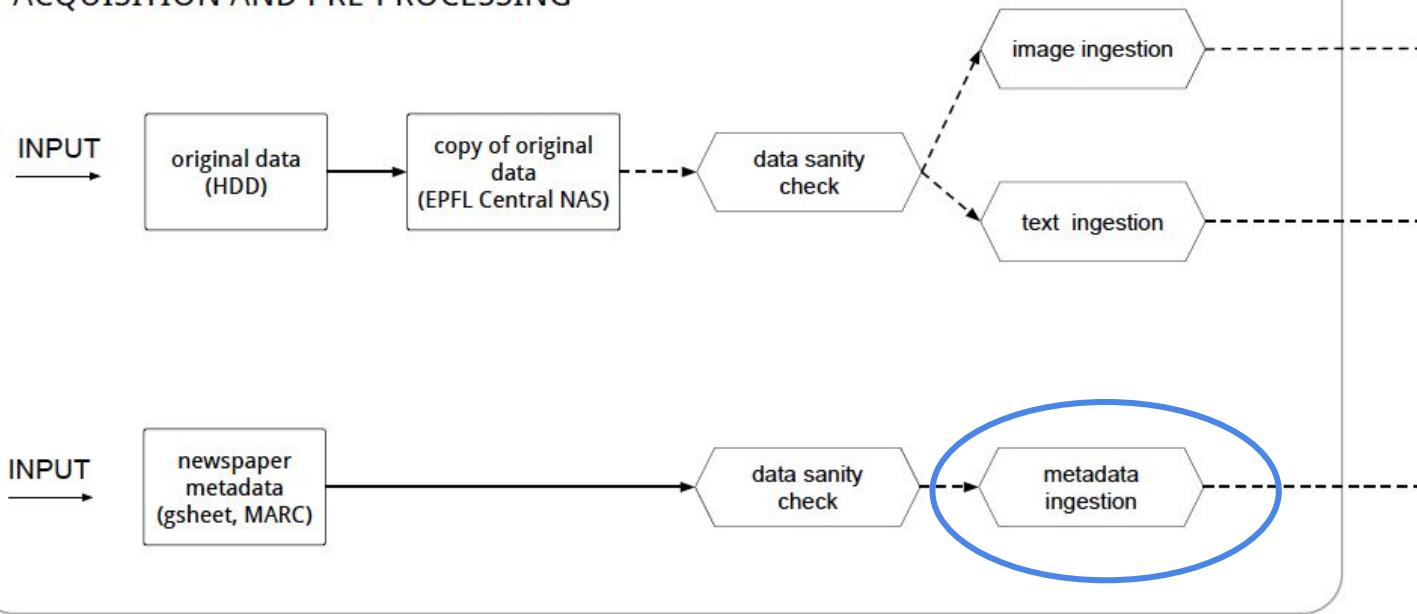
vice-président

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

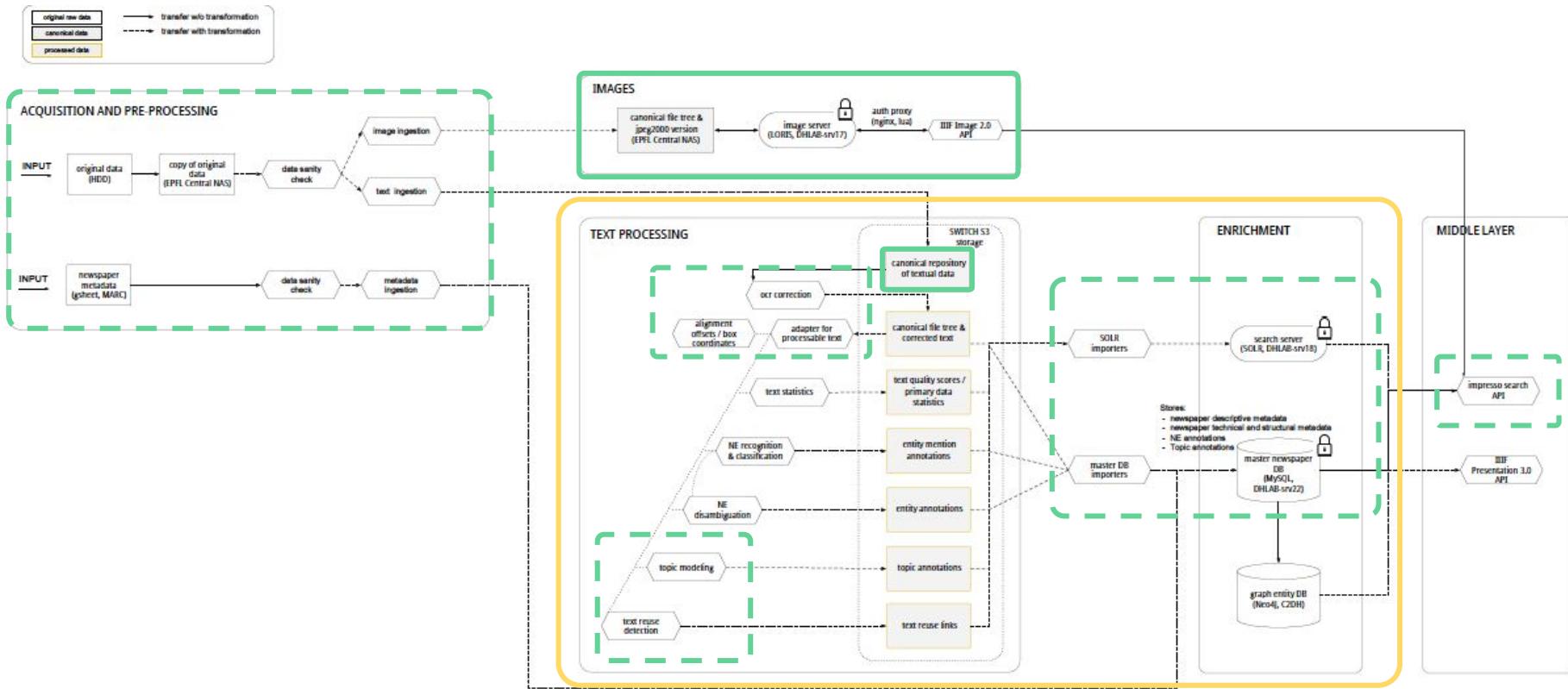
4: detection possible via quality control of digitization process, is almost never encoded.

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

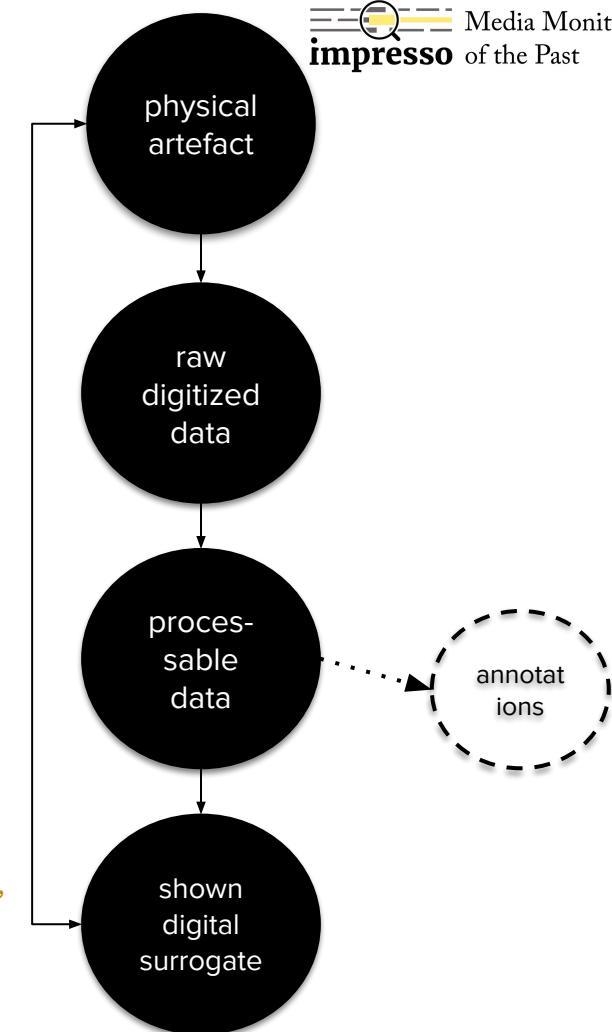
To sum up: **done** / **ongoing** / **next**





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



How Big is the Problem?

Systematic evaluation needed!



The screenshot shows the Transkribus software interface. On the left, the 'Server' tab is active, displaying a list of recent documents: 35798 nzz_gold_ocr_only, 30948 nzz_gold, 30742 nzz_test, and 30741 Test Document. The main window displays a page from the 'Neue Zürcher Zeitung' (NZZ) dated March 11, 1921. The page features a large masthead 'Neue Zürcher Zeitung' and sections like 'Saison — Reise — Ver'. A red box highlights the first column of text under the heading 'Touristik und Reklame.' The text discusses the relationship between tourism and advertising in Switzerland. A red arrow points from the top right towards this text. The bottom of the page contains numbered notes: 1. unsere Gesandtschaften, Konsulate, Schweizer-, 2. Vereine usw. sich zur Mitarbeit 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

Ground truth:

Bag of Words			Set of unique words		
One	Two	Two	One	Two	
Three			Three		
Four	Five	Six	Four	Five	Six
Seven	Eight	Nine	Seven	Eight	Nine
Ten	Ten		Ten		
(12 words)			(10 words)		

OCR Result:

Bag of Words			Set of unique words		
On	Two	Tw	On	Two	Tw
Three	Three	Tw	Three		
Four	Three	Six	Four	Five	Six
Seven	Eight	Nine	Seven	Eight	Nine
Eleven			Eleven		
(13 words)			(11 words)		

- 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



OCR-2005

N° 86 Zürcher Samstag t 793 & deN26 Weinmon MMMM ?

urd
vol
ser
&a
&q
nd
so
&a
3 &
« e
son
« V
» ro

MMMM

Frankreich die Plage war künftig wieder bei uns, daß sie noch vor der glücklichen Revolution, wo zurück das französische Volk Gouverneur geworden, mit den Menschen, welchen nur von Böhmen und Ungarn, unterteilt in politischen Verbündeten standen; die Verbindungen ein' Mittelpunkt waren, — „ah-siecht“ zufrieden damit, im Empfangs-^{er} V. mit Luwig Capets Muadem, und dem <H.^{er} v. N.^{er} studierten Calonne „MBefriedigun“ „M.^{er} de Meklemburg“ zum „nd“ ihrer Besoldung der Agenten lbs. trimmeln ^{zum} „uf“ eine schreckliche Weise die S-^{er}, lsfm-en, drn Schwe- des Volks Verschleudert „ubten“ haben »yck« „A^{er} am « „u^{er} in Zeit Millionen“, den Kai er schickte, „eit“ haben, und noch d^{er} in den, wider Flancret den N^{er} « „u^{er} zusammen „u^{er}, dar durch dieser Sesselordnung verstaubungen der Sonnenblume ganz aufgewor- fen, daß ferner seit der Revolution Capets Witwe nicht mehr Prinzessin aufscheit habe durch ihre vertrauten

e in
nInttn
h^ «
e
em °»°
^ « lu
den , u
Gäst
n bie

insbestndir . en , und well sie '« M4 .. W « inm . üb « I da « ^^ ünd Verstelln »» s ° weit « -triebe » - > ;»» , „ N : Schrltel ,

OCR Quality - The Differences (cont.)

OCR-2017

Prozeß der Marie Antoinette. Nachdem dieselbe am i g. Weinm. alten StylS, 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 i. — Wer seyd ihr ? . Ich bin dir Witwrr Ludwig Capet«, ehemaligen Königs der französen — Seu.— Wie alt? Z8 Jahre. — Nun wurde von dem Ge-richtsschreiber die Auklagsakte vorgelesen. Darin» heißt e»,daß aus den dem Tribunale rnhandengestellten Schriften er-helle 'Daß gleich den Messalinen Brunehaut, Fredegoude»nd Medizi«, die man einstKöniannea von Frankreich ge-nannt habe, und deren verhaßte Namennie au« de» Jahr-büchern der Geschichte werden vertilgt werde» , Marie An-toinette , Ich in Gant Wille es mit ihm . Aufenthalte in Engla

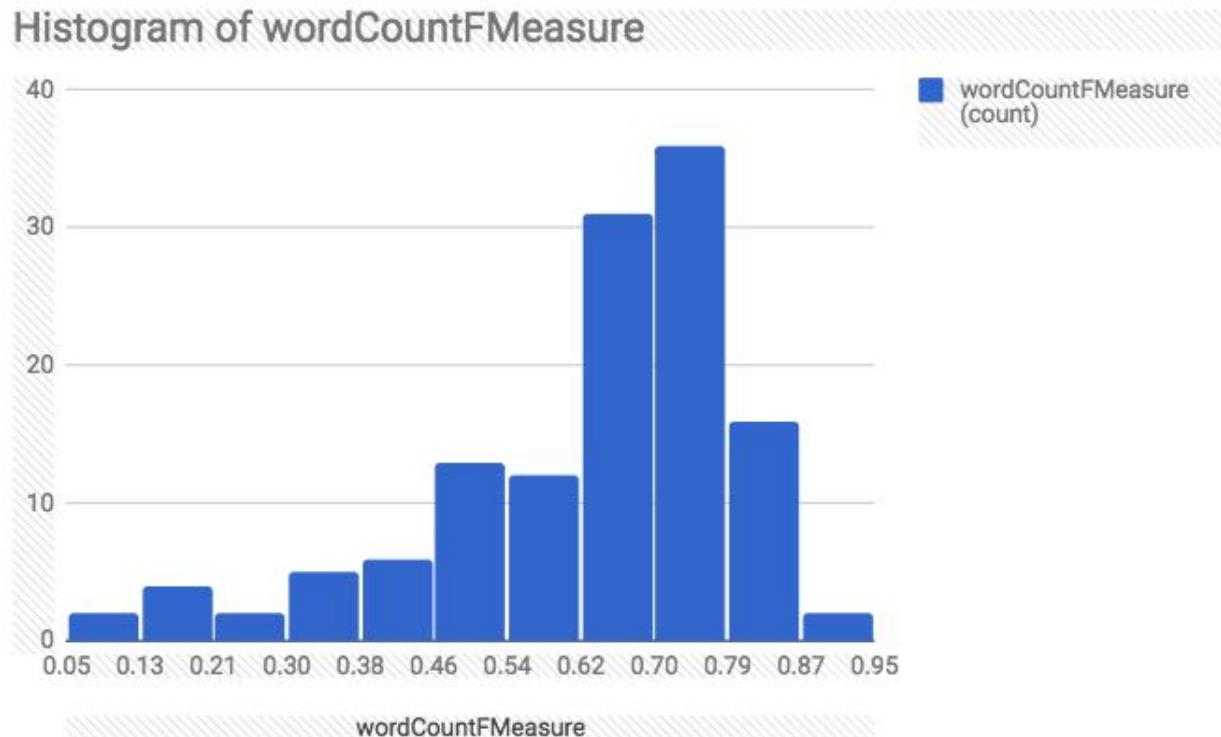
HTR-2018

Prozeß der Marie Antoinette. Nachdem dieselbe am 15. Weinm. alten Styls, oder am 23. des ersten Monats im 2ten Jahre der Republik neuen 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 i. — Wer seyd ihr ? . Ich bin dir Witwrr Ludwig Capet«, ehemaligen Königs der französen — Wie alt? 38 Jahre. — Nun wurde von dem Ge-richtsschreiber die Anklagsakte vorgelesen. Darinn heißt es, daß aus en dem Tribunale zuhandengestellten Schriften er-helle: Daß gleich den Messalinen Brunehaut, Fredegonde und Medizis, die man einst Königinnen von Frankreich ge-nannt habe, und deren verhaßte Namen nie aus den Jahr-büchern der Geschichte werden vertilgt werden, Marie Antoinette, die in Gant Wille es mit ihm . Aufenthalte in Engla

F-Measure: 0.684

F-Measure: ~0.95

Original OCR Quality (OCR-2005)

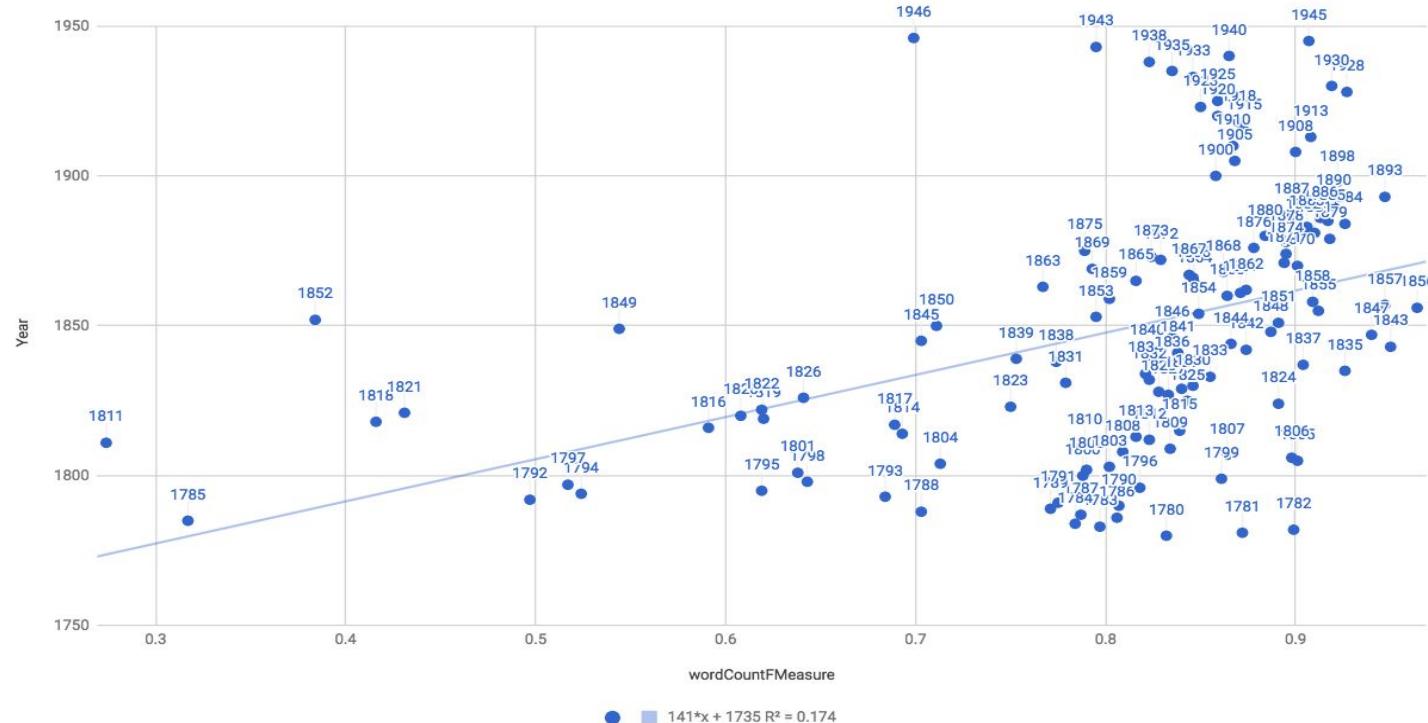


Original OCR Quality (OCR-2005)

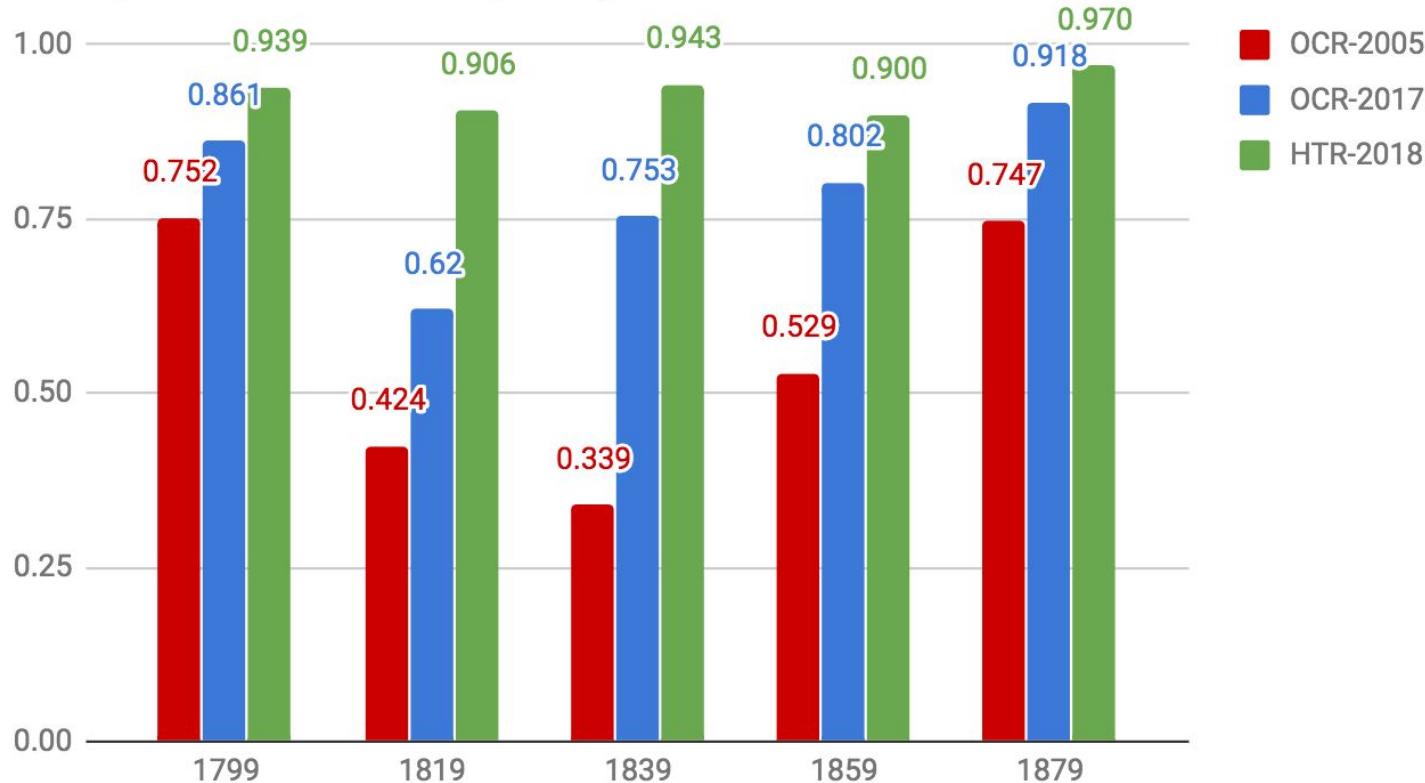


ABBYY OCR-2017

Year vs. wordCountFMeasure



Comparison of OCR Quality

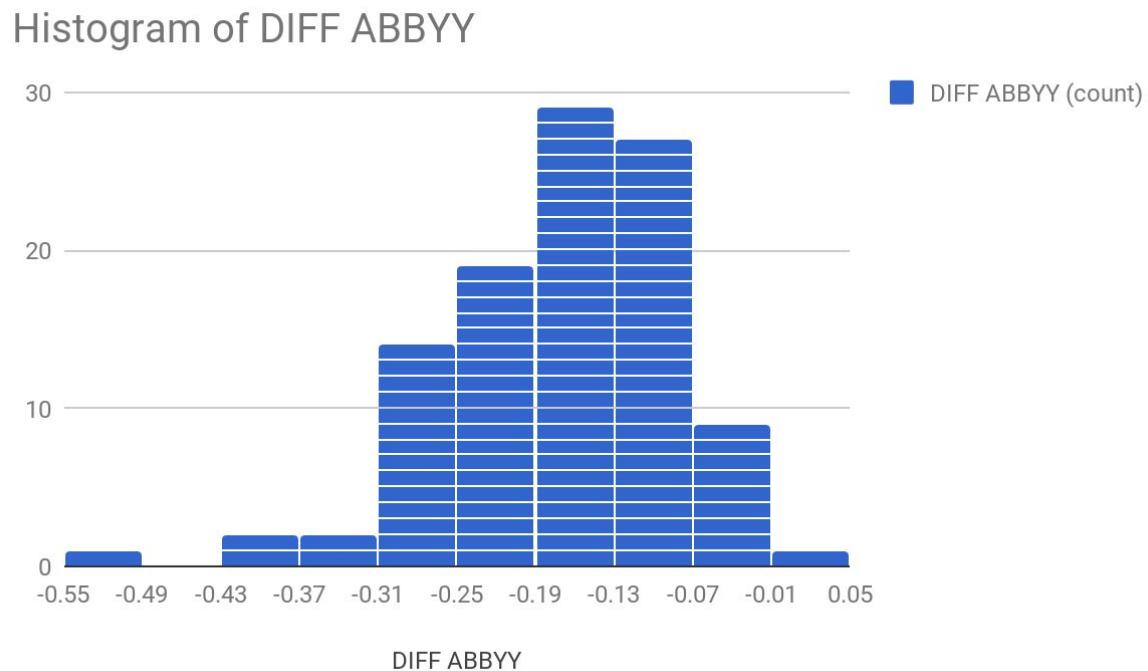


Improvement Re-OCR-ing vs Current Quality

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

On average:

17 percentage points better!



One Step Further ...

[7], Loaded doc: nzz_gold, ID: 30948, Page 32, file: nzz_18110507_0_0_a077), 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]

Screenshot of a digital edition interface showing two versions of a historical document page side-by-side.

Left Panel (Ground Truth):

- Contains a vertical toolbar on the left with icons for L, BL, W, H, V, L, etc.
- The main area shows the original text in a light gray font.
- Annotations are overlaid in red boxes, primarily around the word "Nichten" in the middle of the page.

Right Panel (HTR Model Output):

- Shows the same text as the left panel but with improved recognition.
- The word "Nichten" is correctly identified and highlighted in blue.
- Annotations are present at the bottom of the page.

Annotations at the Bottom:

1 'Zu ö-sdv' ist H'rürvch- K<pe?Ches deS voUaNs
 2 »UAMsterdāni bestand «hin HMdekshäufeS HL'peuud
 3 'gestorben.'-S» war im 3-#\\$? «W-ä Jrfifjifte&iim
 4 ließMnitch lengsiakd »egav.i »re Mftt'srif'tsm
 5 rj»ig «*/-?d«.fich auf z. Millionen ft<vtu-<auft,'M
 6 » sein adptirtrr Soh«, -VMiaft' H-pe V Wt' ein« Wer
 7 'Richten geheirathet har. 'u- Wf 1
 8 ! . j.4 ,• Sch'w e d. em '

Annotations at the Top:

1 Zu London ist H. Zeinrich Hope, Chef des vormals
 zu Assterdam bestandenen Handelshauses Zope und Comp.
 2 gestorben. Es war im Jahre 1794, als er Amsterdam ver-
 ließ, und sich nach Englsland begab. Die Hälften seines Ver-
 mögens, 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?

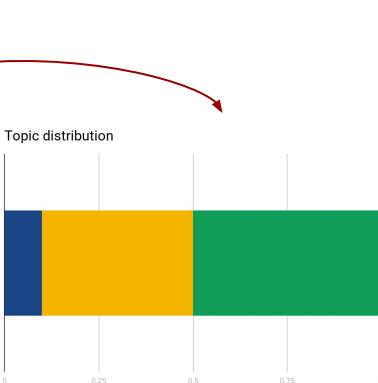
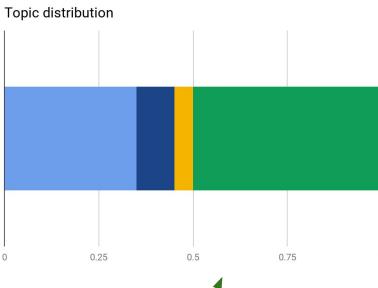
- NZZ willing to **open** its archive to general public
- a re-digitisation **from paper** is probable
- **article recognition** included

Semantic Indexing of Documents

Impact of Bad OCR on Topic Modeling



Zwei Kundgebungen zum 1. Mai



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

2

britisch deutsch truppe
zeug
e
lavas

zimm
wohnun
villa hotel schon zürich
pension komfort preis sec
telephon sonnig modern

8

OCR Errors Dominate Topics

13/50



Long Story Short ...

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

Meaning ...

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



Side Note: Topic Modeling vs Text Classification

supervised
(labelled training
data needed)

Text classification

politics



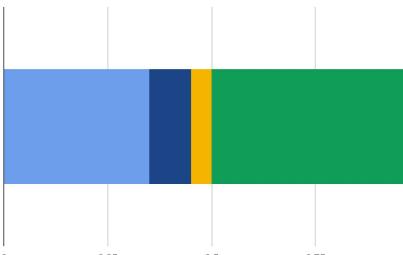
war



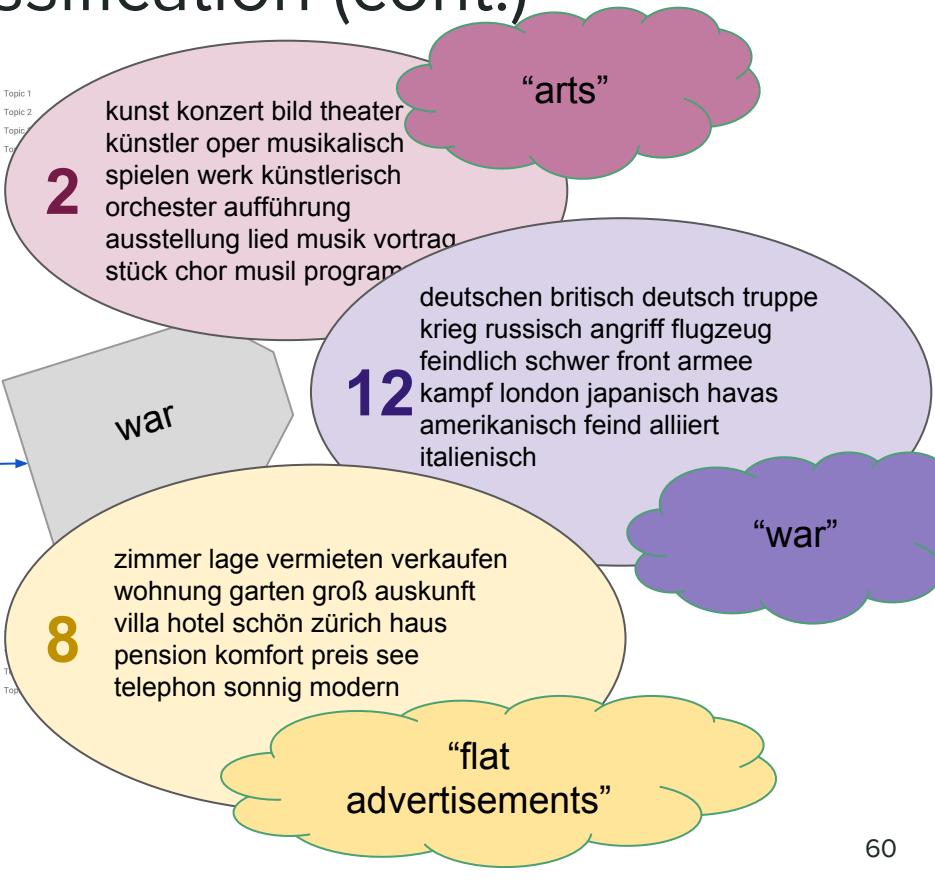
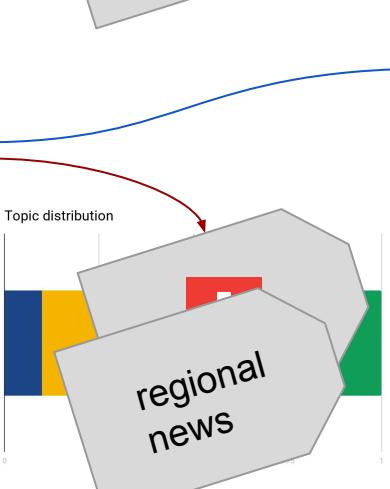
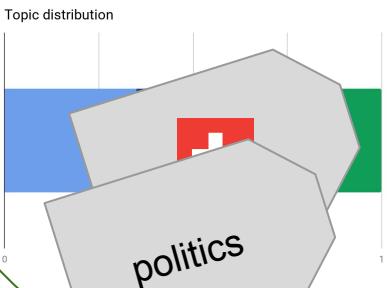
unsupervised

Topic modeling

Topic distribution



Topic Modeling vs Text Classification (cont.)

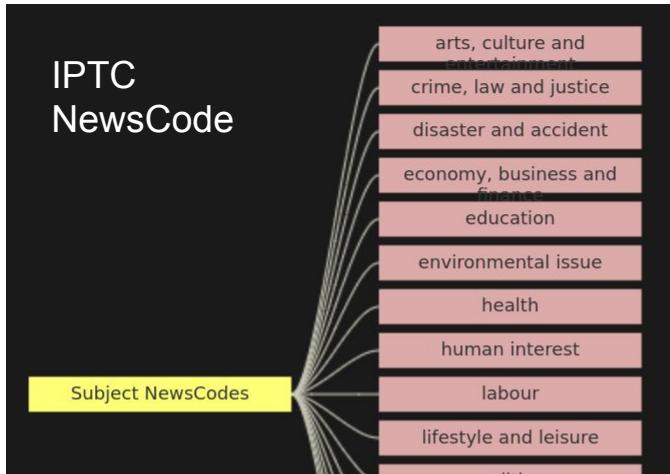


Contrastive View: Classes vs. Topics



classes = defined concepts

- taxonomy, e.g.



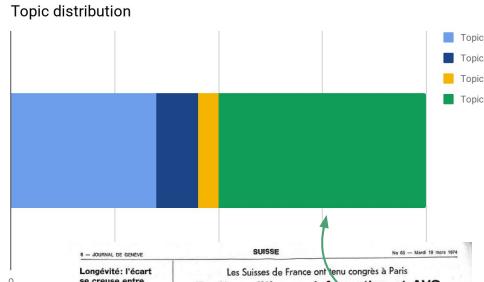
<http://show.newscode.org/index.html?newscode=subj&lang=en-GB&startTo>Show>



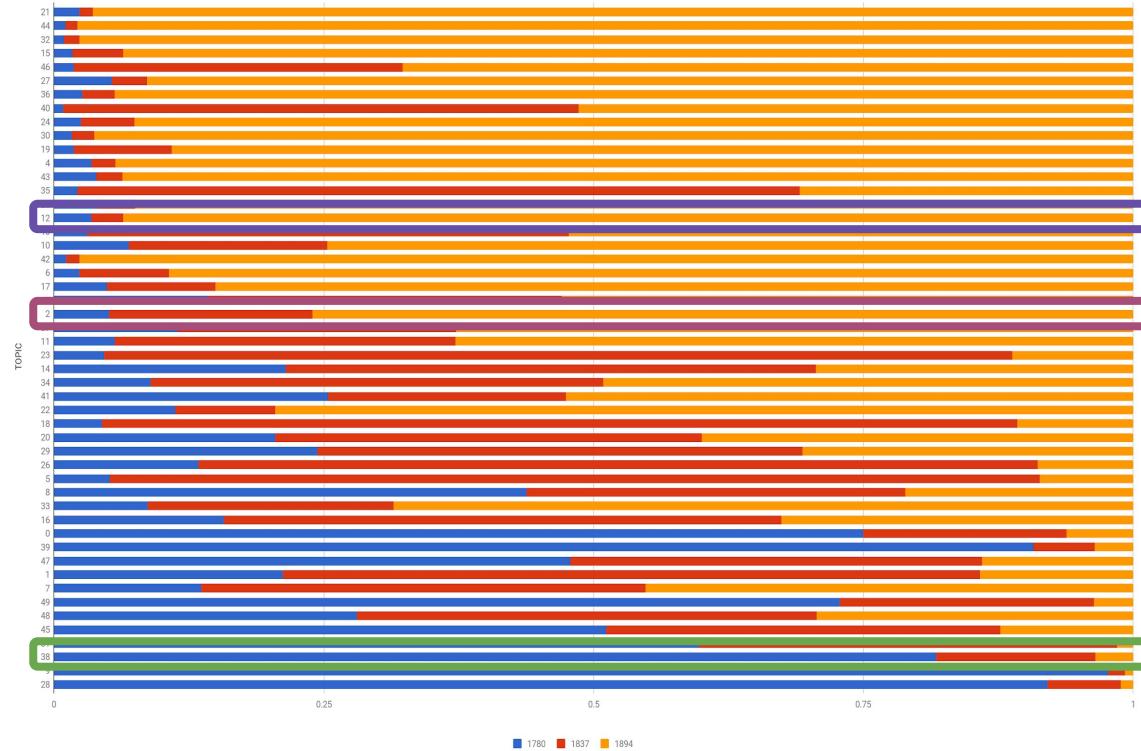
topics = probabilistic word distributions

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

Example: Faceted Search on Page



Topic Evolution over Time



2

kunst konzert bild theater
künstler oper musikalisch
spielen werk künstlerisch
orchester aufführung
ausstellung lied musik vortrag
stück chor musik 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 ≠ 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] alien", "alien", "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 reflects
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
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

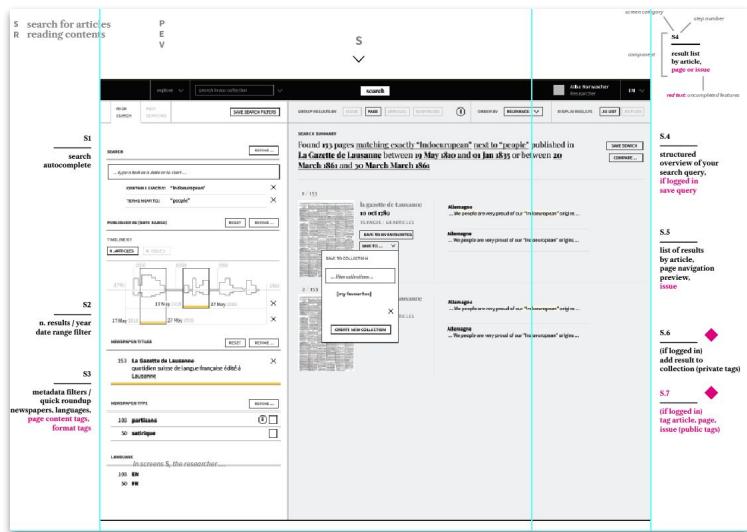
Presentation → **Mockups**

Interaction → **Prototype V1**

Feedback → **Booklet**

Co-design package

Booklet of interface features,
activity by activity



Component index,
by activity

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

S.1

Component
description

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

red text: uncompleted features

S.2

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

S.3

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

explore ▾

search

Alba Rorwacher
Researcher

EN ▾

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 ...

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 Gèneve 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 SUMMARY

Find 53625987 articles in our collection.

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 , ses possibilités JEAN-PIERRE AGUET ^^. ^., ^ H est à l'heure actuelle de multiples...

CONFERENCE REPORTS

EXPORT CITATIONS ...

DOWNLOAD AS ...

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...

GUERRE FROIDE X

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 ' !! , ' ...

EXPORT CITATIONS ...

DOWNLOAD AS ...

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](#)

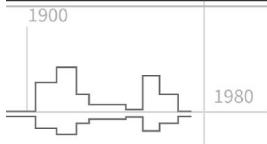
ORDER BY

[RELEVANCE ▾](#)

DISPLAY RESULTS

[AS LIST](#)[AS TILES](#)

SEARCH SUMMARY

[REFINE ...](#)Find **53625987** articles in our collection.[SAVE SEARCH](#)[COMPARE ...](#)

as published

aise édité à



1 of 112

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

INTERNATIONAL NEWS

[TAG AS ...](#)[ADD TO FAVOURITES](#)[ADD TO COLLECTION...](#)

CONFERENCE REPORTS

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



2 of 112

Les USA et l'AFN

INTERNATIONAL NEWS

[COVER PAGE](#) X[TAG AS ...](#)[ADD TO FAVOURITES](#)[ADD TO COLLECTION...](#)

GUERRE FROIDE X

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

Les USA et l'AFN Un reportage de Charles-Henri Favrod en Tunisie A la recherche des effets à long terme de l'effacement de l'Algérie

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)

The screenshot shows a search interface for historical newspaper articles. At the top, there are tabs for 'explore' and 'search'. The search bar contains the query 'guerre froide AND europe'. Below the search bar are sections for 'GROUP RESULTS BY' (Issue, Page, Articles, Sentences), 'ORDER BY' (Relevance), and 'DISPLAY RESULTS' (As List, As Tiles). A 'SEARCH SUMMARY' section indicates 112 articles found between Saturday, 19 March 1950 and Friday, 01 Jan 1959. The main area displays a timeline of articles from 1745 to 1959, with a focus on the period from 1950 to 1959. Each article is represented by a thumbnail image of its front page. To the right of each article thumbnail, there is a summary of the content, including the date, source, and a snippet of text containing the search terms. On the far right, there are buttons for 'ADD TO FAVOURITES', 'ADD TO COLLECTION...', 'CREATE NEW', and 'MANAGE MY COLLECTIONS'. A sidebar on the left provides filters for 'PAST SEARCHES', 'FULL TEXT SEARCH' (with dropdowns for 'CONTAINS' and 'EXACTLY'), 'PUBLISHED IN (DATE RANGE)' (with a timeline from 1800 to 1980 and specific dates 19 March 1950 and 01 Jan 1959), 'IN NEWSPAPER TITLES' (listing 'La Gazette de Lausanne' as the source), 'TYPES OF ARTICLES' (international news, politics), 'TOP NAMED ENTITIES' (Napoleon, Zurich, Suisse), and 'LANGUAGE OF ARTICLES' (French). The bottom of the interface includes buttons for 'EXPORT CITATIONS...' and 'DOWNLOAD AS...'. A dotted line connects the 'ADD TO COLLECTION...' button in the top-right panel to the 'ADD TO COLLECTION...' button in the bottom-right panel.

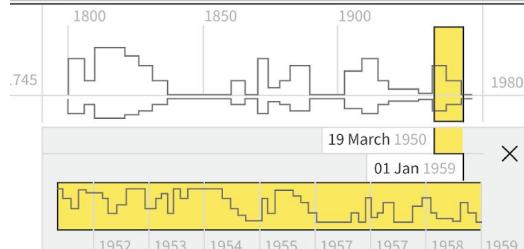
YOUR
SEARCHPAST
SEARCHES

L TEXT SEARCH

REFINE ...

. type a text or a date or to start ... XCONTAINS ▾ EXACTLY ▾ **guerre froide** XCONTAINS ▾ ... ▾ **europe** X

LISTED 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 X
 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 lesquels les Etats- Unis , huit pays d'**europe** occidentale

ADD TO FAVOURITES

ADD TO COLLECTION...

ADD TO COLLECTION

filter or create collection

favourites

✓ conference reports X

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...

ADD TO FAVOURITES

ADD TO COLLECTION...

GUERRE FROIDE X

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



1

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

2



3

PAGE TAGS

TAG AS ...

COLLECTION

ADD TO C

ZOOM 22 %

17 May 1956

27 May 1956

RELEVANCE ▼

SHOW PEOPLE ONLY ▼

Napoleon

1 of 454

Napoléon Bonaparte (French: [napoleō bɔ̃naparɛt]; 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 ...

2



3



4



5

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

BACK TO SEARCH RESULT PAGE

NEXT RESULT

La Gazette de Lausanne
Lundi 1 Mai 1945

LA GAZETTE

SEARCH RESULTS

Search term: guerre froide AND europe

Number of results: 112

Number of pages: 11

Number of results per page: 10

ADD TO CO

ZOOM 22%



MIN

FULL

SCREEN

956



Y ▼

454

[alt];

454

454

2

3

4

5

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

BACK TO SEARCH RESULT PAGE

political and military leader
who served as the leader of
the Republic of China



The highlighted article's title is "VI / RUE DES PALMES". The menu below it includes: READ, ADD TO ..., CITE ARTICLE ..., and MORE ACTIONS.

Other visible menu items include: OFFICE DE LA CENSURE, LA GAZETTE, and LA JOURNAL. The menu also includes options like ADD TO COLLECTION, ADD TO READING LIST, and ADD TO FAVORITES.

NEXT RESULT

La Gazette de Lausanne
Lundi 1 Mai 1945

R.8

contextual menu
for identified layout
regions

R.9

add article
to user collections

RUE DES PALMES

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
volie 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
tait à peu près la seule distraction
quels une raison spéciale n'avait

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

plus impérieux encore retenaient
dans leur petite villa blanche du
surnommes.
l'onne commandait l'intérêt, et,
quinzaine, un médecin attentif

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

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

OCCURRENCES IN ARTICLE

Monsieur du Corbeau (5)

Maître Corbeau (4)

le corbeau (3)

IN OTHER ARTICLES ...

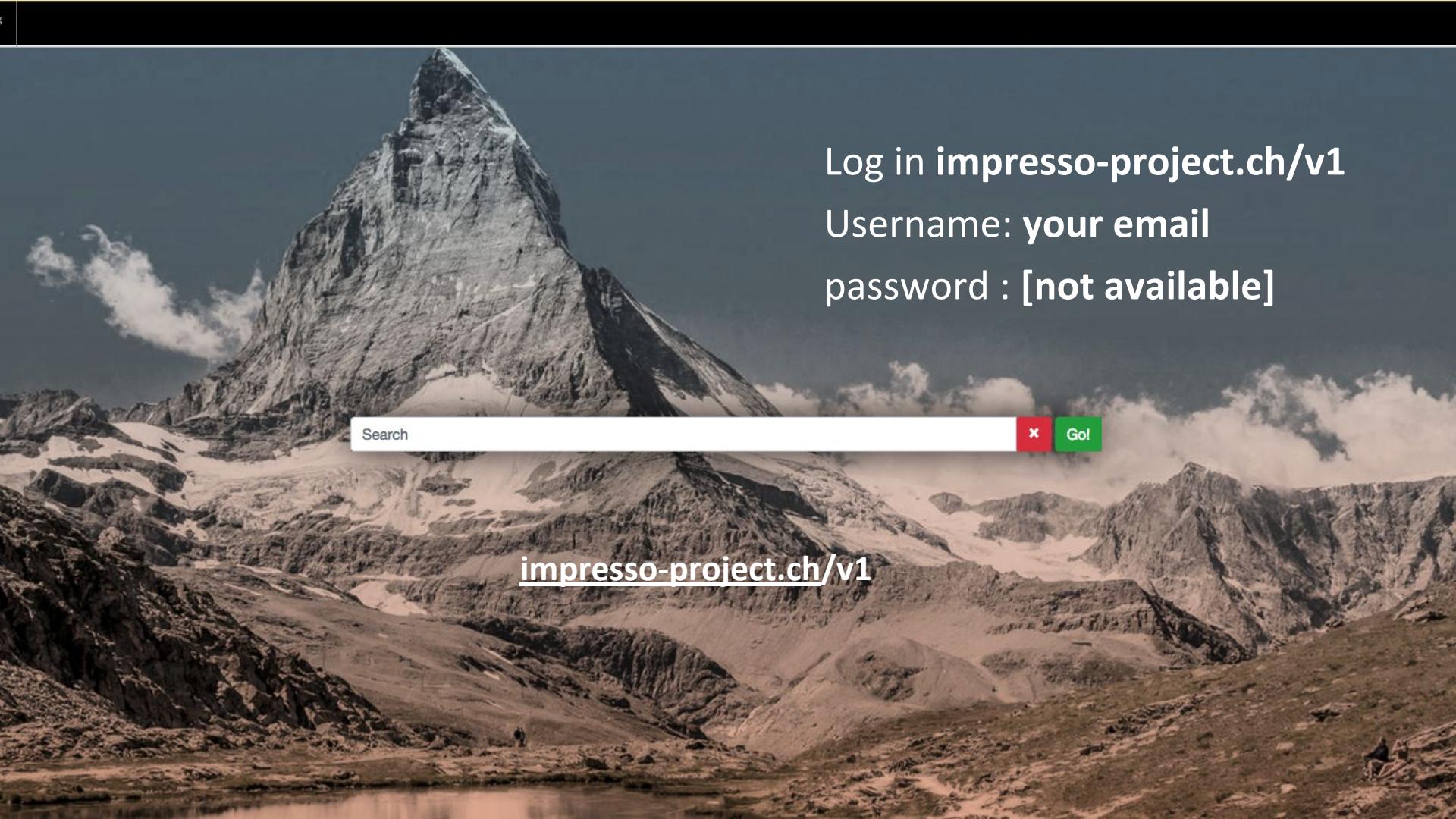
ADD TO COLLECTION...

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

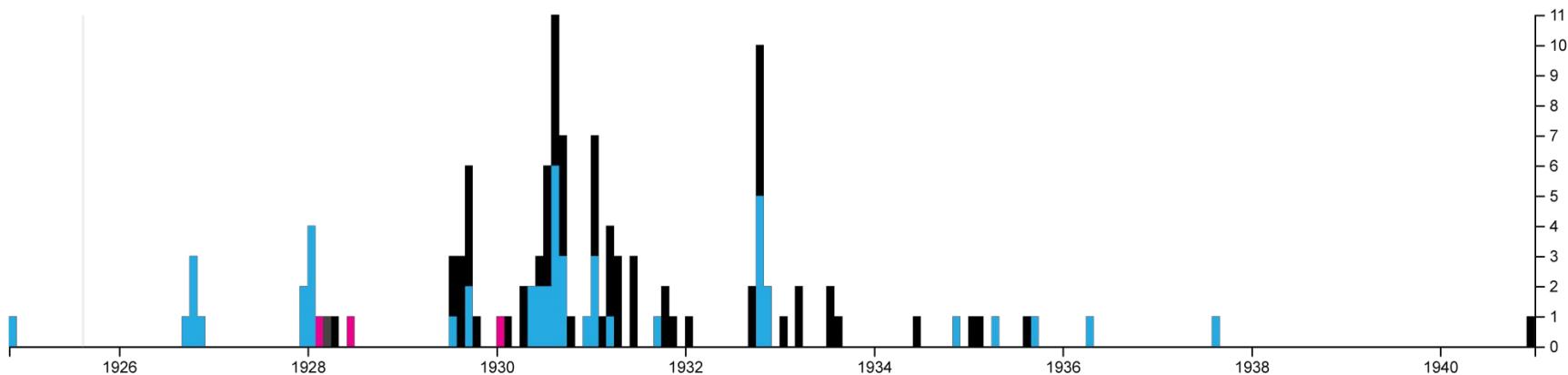
- Traceability of research steps
- Learning by following somebody else's footsteps ?

Exploration of the result lists : tools?

Visual exploration of the result list

Comparing frequencies:

Distribution over time of the queries “Pan-europe” / “Pan-europa”



NGRAM VIEWER
OVER YEARS

VISUALIZATION TYPE

BUMP CHART

STREAM GRAPH

SEARCH

paneurope

DIMENSIONS

PER NEWSPAPER ▾ PER COLLECTION

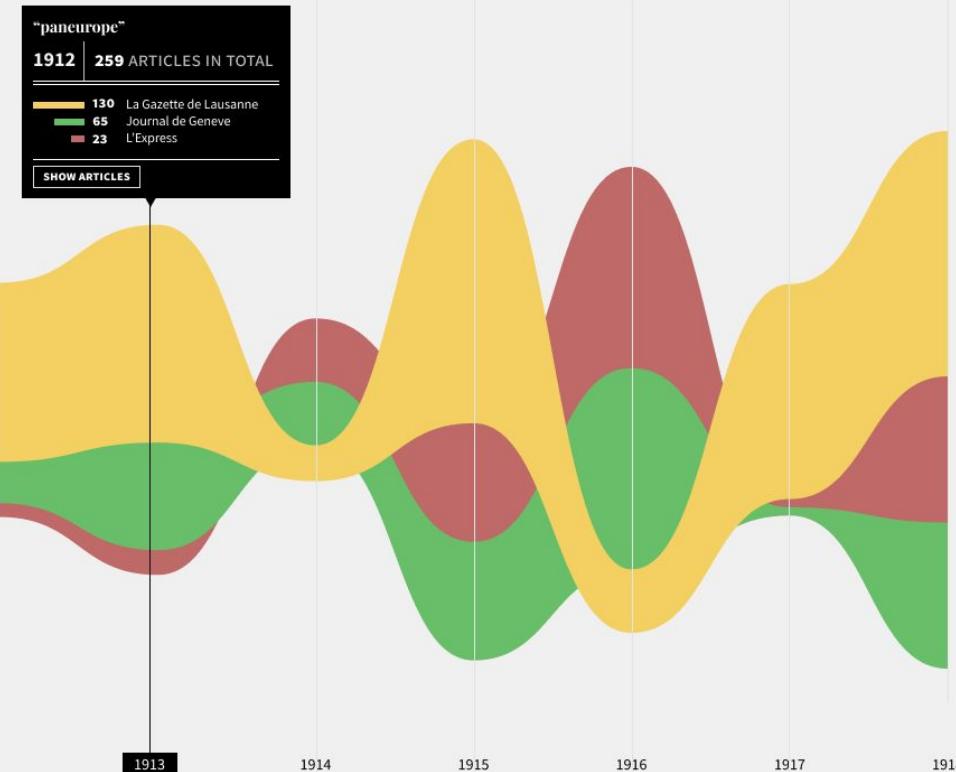
PER ARTICLE TAG

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 20me 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
Advertisements (commercial)	Arts, culture and entertainment news
Announcement (hard news, non commercial)	Business news
Article (default category)	Crime, law and justice news
Banner	Guest contribution, op-ed, quoted speech
Box	Fiction, poetry, humour
Cartoon	International news
Civil registry (obituaries , weddings, births)	Interviews
Column	Last minute news (dépêches)
Cutline (pictures)	Local news
Ears	News story (faits-divers)
Index, table of content	National news
Letters to the newspaper	Science news
Photograph	Sports news
Press review	
Reviews (books, entertainment...)	

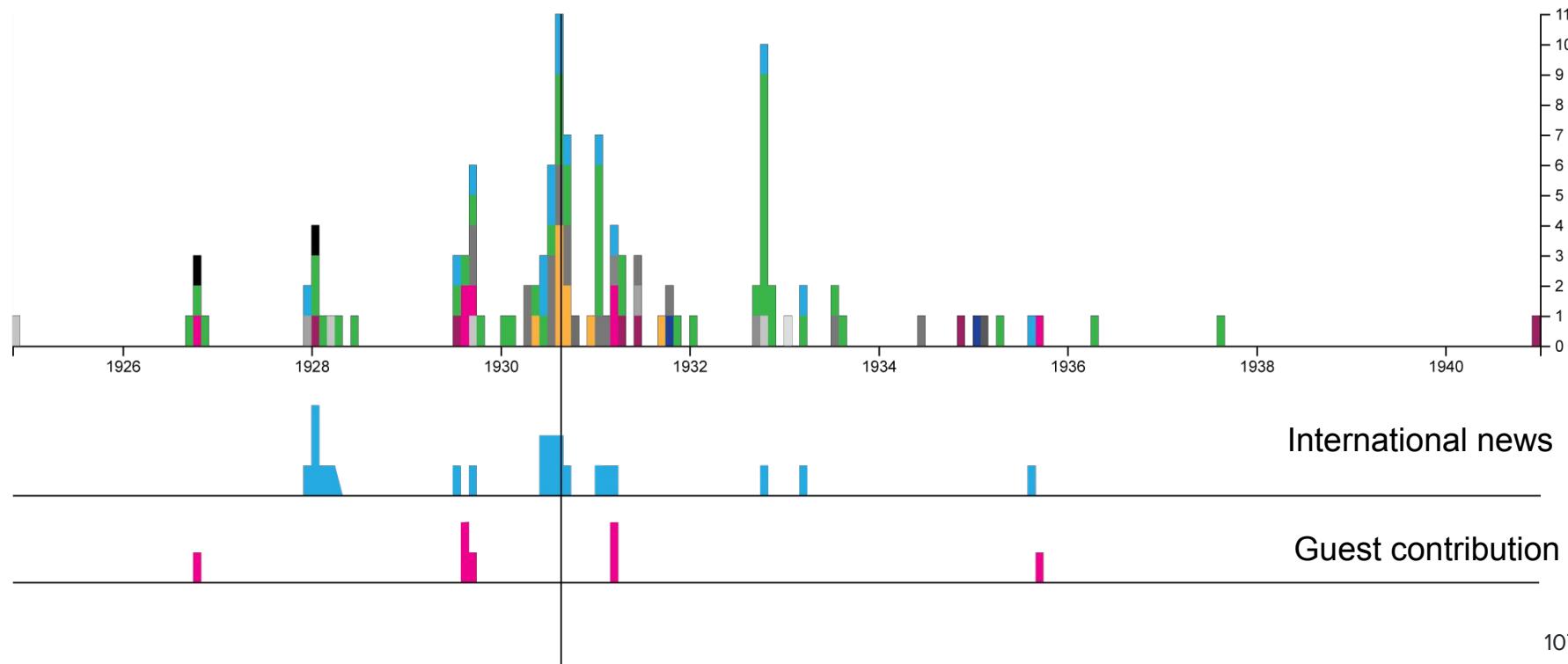
Collecting data from the queries: why

Paneurope use case:

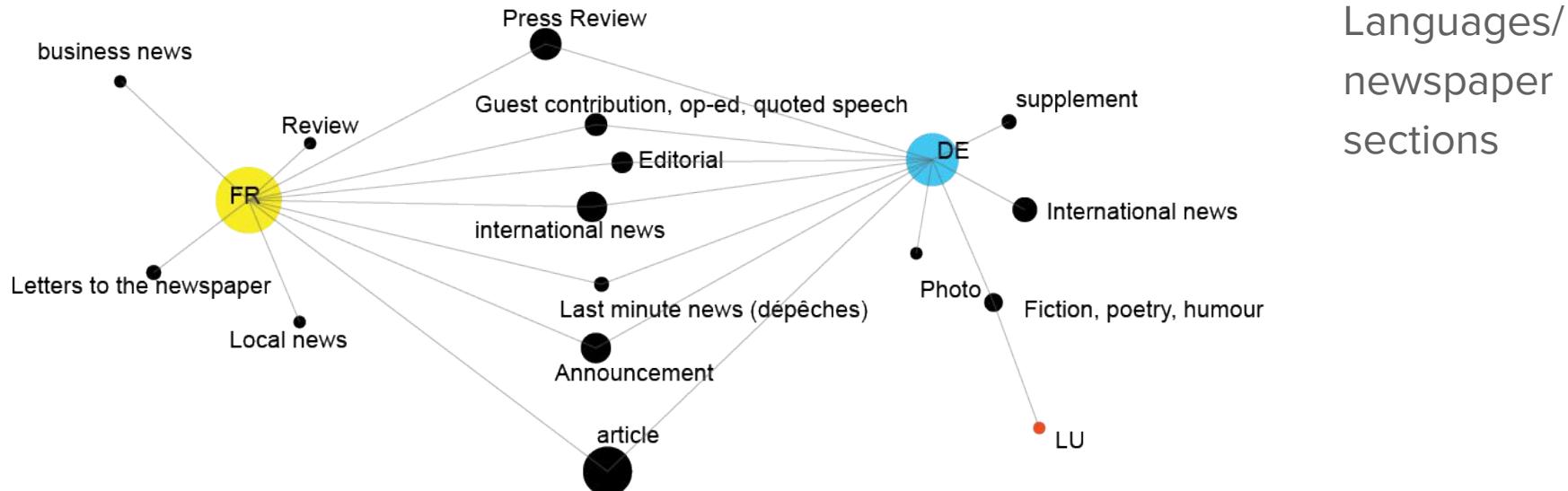
	A	B	D	E	F	G	H	I	J	K	L	N	O	P
1	query	Collection	Metadata/article		Metadata/ Title					Classification /tags				
2	query	Collection	author	title of the article	newspaper title	date-iso	page	URL	nr	section/ typologie des articles	Historical section title	type of newspaper	lang.	KWIC?
6	Paneurope	L'Express/imp a.s.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://		204	Guest contribution, op-ed, quoted sp	Local newspaper	Local newspaper	DE	
9	paneurope	e-luxemburg:s.a.		"Allemagne. M. Coudenhove-Kalergi candidat au pré l' indépendance luxembourgeo		1931-11-23	1 <http://		327	article		Cultural newspaper	FR	
14	Paneurope	Scriptorium (Bc s.a.)		"Après l'Assemblée des délégués de la Fédération La revue		1933-01-19	3		18	Business news	Page Agricole de la General informatio	FR		
15	Paneurope	L'Express/imp a.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 luxembourgeo	1929-08-08	1 <http://		220	Guest contribution, op-ed, quoted sp	Cultural newspaper	FR		
17	paneurope	e-luxemburg:Richard Coudenhove-Kale	"Atlantide"		L'indépendance luxembourgeo	1929-08-09	1 <http://		221	Guest contribution, op-ed, quoted sp	Cultural newspaper	FR		
18	Paneurope	Le temps ari:s.a.		"Au congrès de l'Union Pan-Européenne. M. Briand	Gazette de Lausanne	1930-05-19	6 <http://		137	Announcement	Dernières Dépêch	General informatio	FR	
19	Paneurope	Le temps ari:s.a.		"Au Congrès paneuropéen de Bâle"	Journal de Genève	1932-10-05	10 <http://		272	article		General informatio	FR	
20	Paneurope	L'Express/imp a.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 ari:s.a.		"Au sommaire du dernier numéro de la revue PANE"	Journal de Genève	1927-12-24	2 <http://		352	Announcement	Echos	General informatio	FR	
22	Paneurope	Le temps ari:s.a.		"Aussi longtemps que Paneurope n'aura pas dispar	Gazette de Lausanne	1931-01-21	2/3 <http://		20	Press Review	Dans la presse	General informatio	FR	L'entourage de
23	Paneurope	Le temps ari:s.a.		"BALE-VILLE - Pour l'entente entre les peuples."	Gazette de Lausanne	1933-09-20	2/3 <http://		262	Announcement		General informatio	FR	
24	paneurope	e-luxemburg:Georges Reynaud, ancien	"La menace de l'Anschluss"		L'indépendance luxembourgeo	1931-03-31	1 <http://		90	Guest contributi	Billet extérieur	Cultural newspaper	FR	
25	paneurope	e-luxemburg:Georges Reynaud, ancien	"Malaise européen"		L'indépendance luxembourgeo	1931-03-03	1 <http://		62	Guest contributi	Billet extérieur	Cultural newspaper	FR	
27	paneurope	e-luxemburg:s.a.	"Briards Paneuroparde."		Luxemburger Wort	1930-09-12	1 <http://www.e.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 ari:s.a.		"Ce qui pour M. Briand est un but, pour l'Allemagne	Gazette de Lausanne	1930-07-22	2/3 <http://		200	Press Review	Dans la presse	General informatio	FR	Ces nations sav
31	Paneurope	L'Express/imp a.s.a.		"Conférence pour la trêve douanière. M. Schultess	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.e.article					Local newspaper	DE	

Frequency of the query distributed over NSP sections

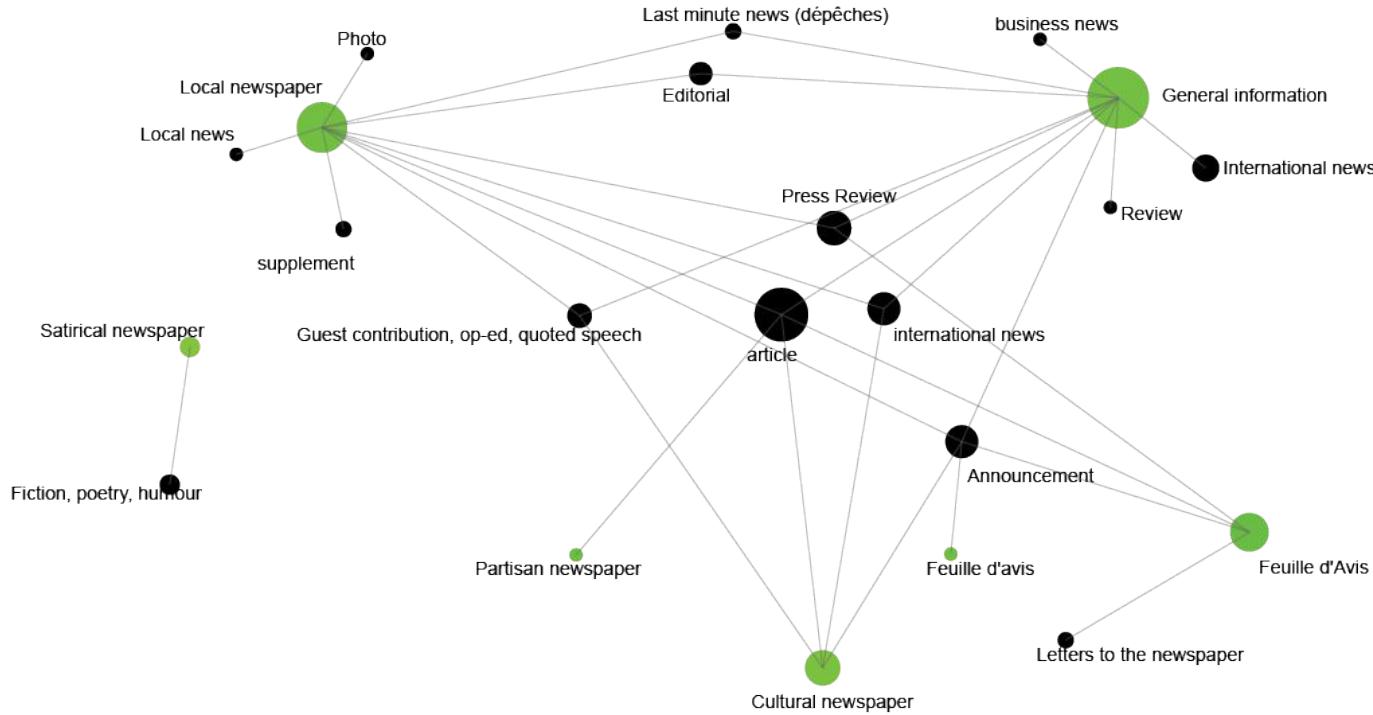
Distribution across all newspaper titles.



Paneurope use case



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:
“**Pan-europe**
Congress-Basel”

NOUVELLES SUISSES

La jeunesse au congrès paneuropéen

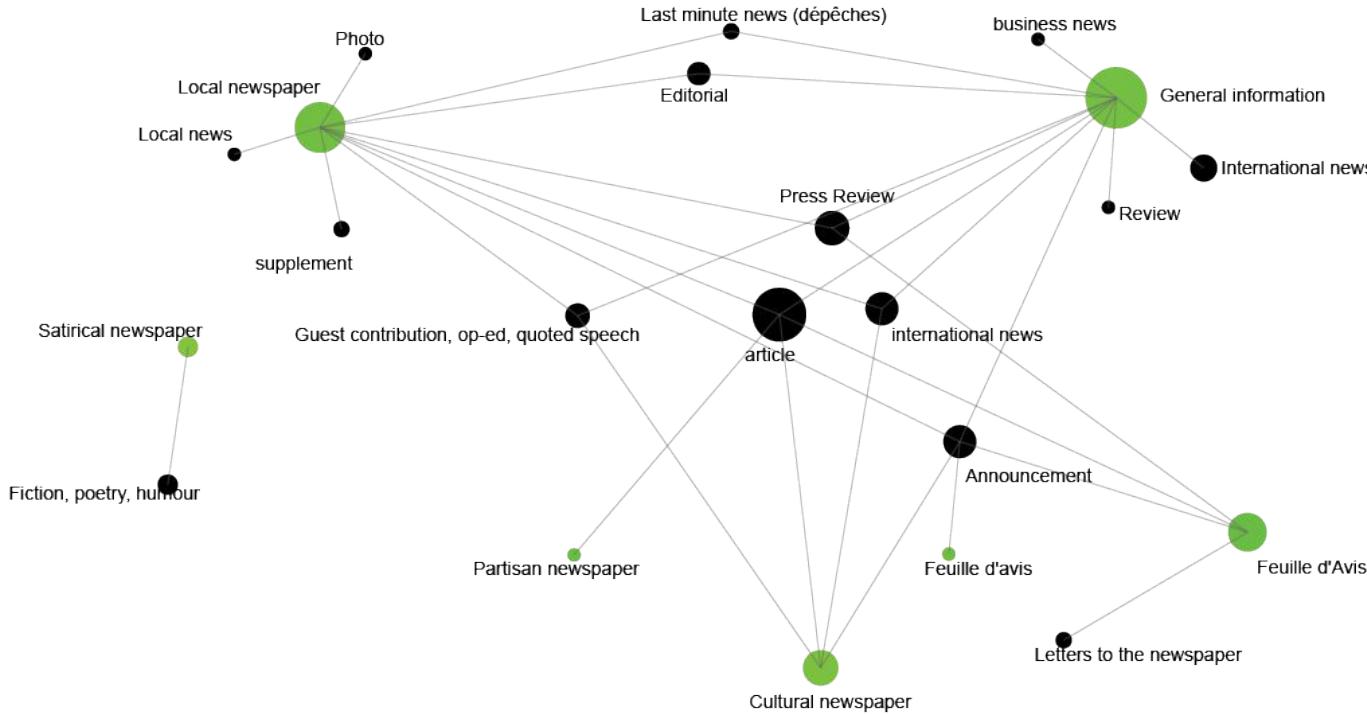
Lundi ,s'est tenue à Bâle, dans le cadre du congrès paneuropéen, une assemblée de la jeunesse européenne, au cours de laquelle de nombreux orateurs, notamment le professeur Recasens-Siches, de l'Université de Madrid, le poète Jacob Wassermann, de Vienne, l'écrivain français Alfred Fabre-Luce, le poète 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 20me siècle, on n'arrivera pas à une Pan-europe. 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
Pan-europe
Liehburg
Moitke
Jacob Wassermann
Recasens-Siches
Francis Delaisi
Caillaux
Gerhard Meyer
Alfred Fabre-Luce
M. Grimm

Paneurope use case

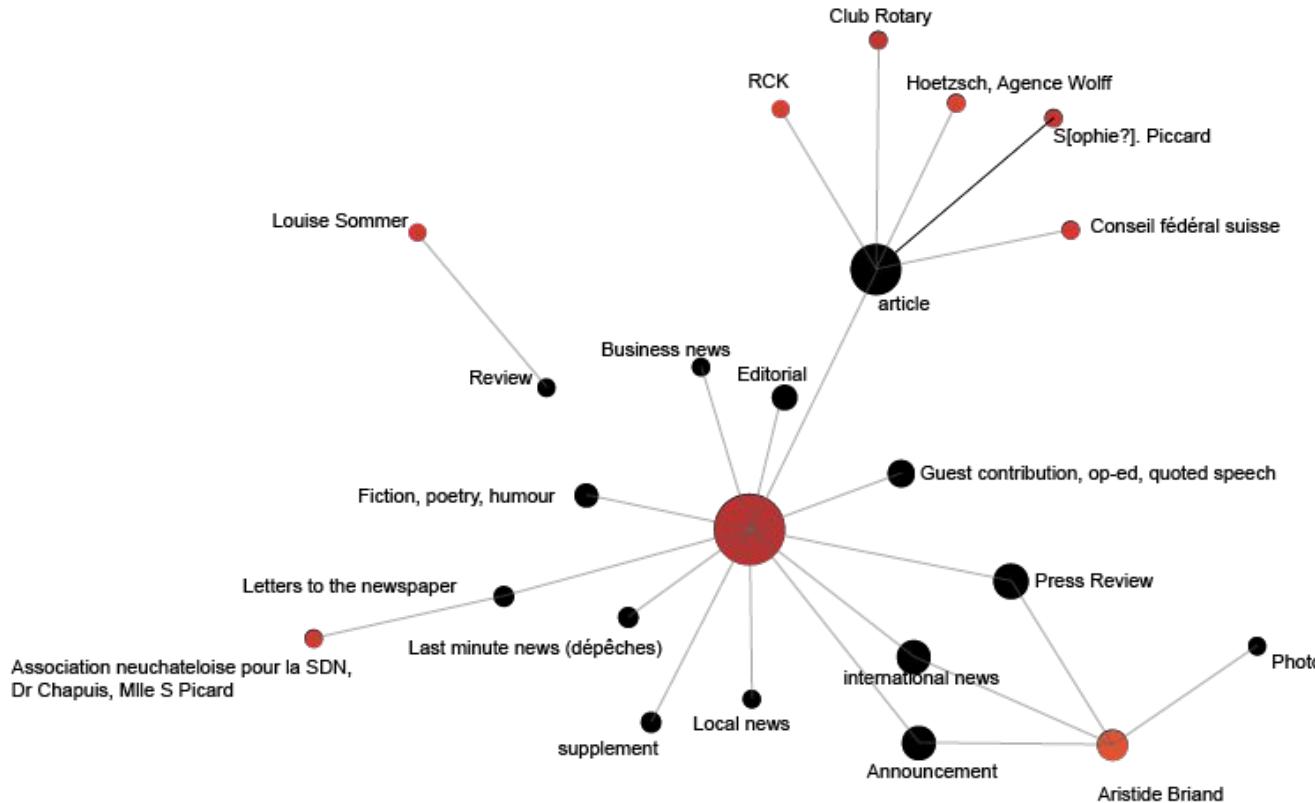


Newspaper type/
newspaper
sections

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, novelLa 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(1)

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(1)

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(1)

[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(1)

[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(1)

[chronique locale]

V / RUE GIRAFFE

pp. 1 - 2

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 Pompei ensevelie sous la cendre. Avec les étrangers, la richesse, le mouvement, la vie, se retiraient. Hôtels et villas se ferment successivement. Les rues semblaient vides. Sur le seuil des magasins fermés, des boutiquiers mélancoliques halètent entre deux échappées, et la plage, démenée 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 retrouva simple ville de province ; et, du même coup, l'esprit de province y revint. 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 Roquescize 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 Wattembrode, une invitation à la campagne.

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

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 plus à mouter une garde empêtrée.

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

Alanguin et nonchalante, sous les plus vaporieux d'un élégant peignoir, ses cheveux bionds 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

phoe

C.1

Suggest
Public Tags
(fixed vocabulary)

C.2

"Format":
newspaper
sections or specific
page regions

C.3

"Contents":
tag mainly relates
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 halètaient 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 héritage, la princesse le départ de son neveu, Mme Guidal l'arrivée de sa fille, la comtesse e Wartembrode, une invitation à la campagne.

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

La santé d'Yvonne commandait l'intérêt, et, enduant une quinzaine, un médecin attentif avait clouée sur une chaise longue, auprès de quelle Hubert s'était plu à mouter 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
Banner	Crime, law and justice news
Box	<input checked="" type="checkbox"/> Fiction, poetry, humour X 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 speech	Reviews (books, entertainment...)
Index, table of content	Science news
Interviews	Sports news
<input checked="" type="checkbox"/> Letters to the newspaper <small>1 user selected this</small>	
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 halataient 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 héritage, la princesse le départ de son neveu, lme Guidal l'arrivée de sa fille, la comtesse e Wattembrode, une invitation à la campagne.

VI

ADD TAG

ADD TAG ...

FORMAT

CONTENTS

no content tags matching "essay"

no content tags matching "essay"

CATEGORY

VALUE

FORMAT

CONTENTS

essay

SHORT DESCRIPTION

SUGGEST NEW

EXPLORE TAGS ...

'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, l'invita à son repas le longage.

who served as the leader of the Republic of China

C.7

"Suggest new"
Add a description for the tag

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)

explore ▾

YOUR SEARCH PAST SEARCHES

FULL TEXT SEARCH REFINE ...

CONTAINS ▾ EXACTLY ▾ **guerre froide**

CONTAINS ▾ ... ▾ **europe**

PUBLISHED IN (DATE RANGE) RESET

IN NEWSPAPER TITLES RESET

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

TYPES OF ARTICLES
112 international news
50 politics

TOP NAMED ENTITIES
103 Napoleon
50 Zurich, Suisse location

LANGUAGE OF ARTICLES
112 French

search

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 ^A.- ^A, ^A H est à l'heure de multiples...

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

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

ADD TO FAVOURITES ADD TO COLLECTION...

ADD TO COLLECTION filter or create collection

favourites
 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 FAVOURITES ADD TO COLLECTION...

GUERRE FROIDE

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

EXPORT CITATIONS ... DOWNLOAD AS ...

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

Or.5

structured overview of your collection,
export citations or metadata

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

personal space / my collections

Alba Rorwacher
Researcher EN ▾

explore ▾

MY COLLECTIONS MY SAVED QUERY

DISPLAY ITEMS AS LIST OF ITEMS VISUALIZE AS NETWORK ORDER BY LAST ADDED ▾

SUMMARY

710 items in the collection **Guerre Froide & Europe**, ordered by latest added

EXPORT CITATIONS ... EXPORT ITEMS METADATA

EDIT TITLE / DESCRIPTION REMOVE COLLECTIONS

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

Page 1

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

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 ...

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 austro-tchécoslovaque Le Kremlin ne paraît pas près de modifier sa politique en Europe

INTERNATIONAL NEWS

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

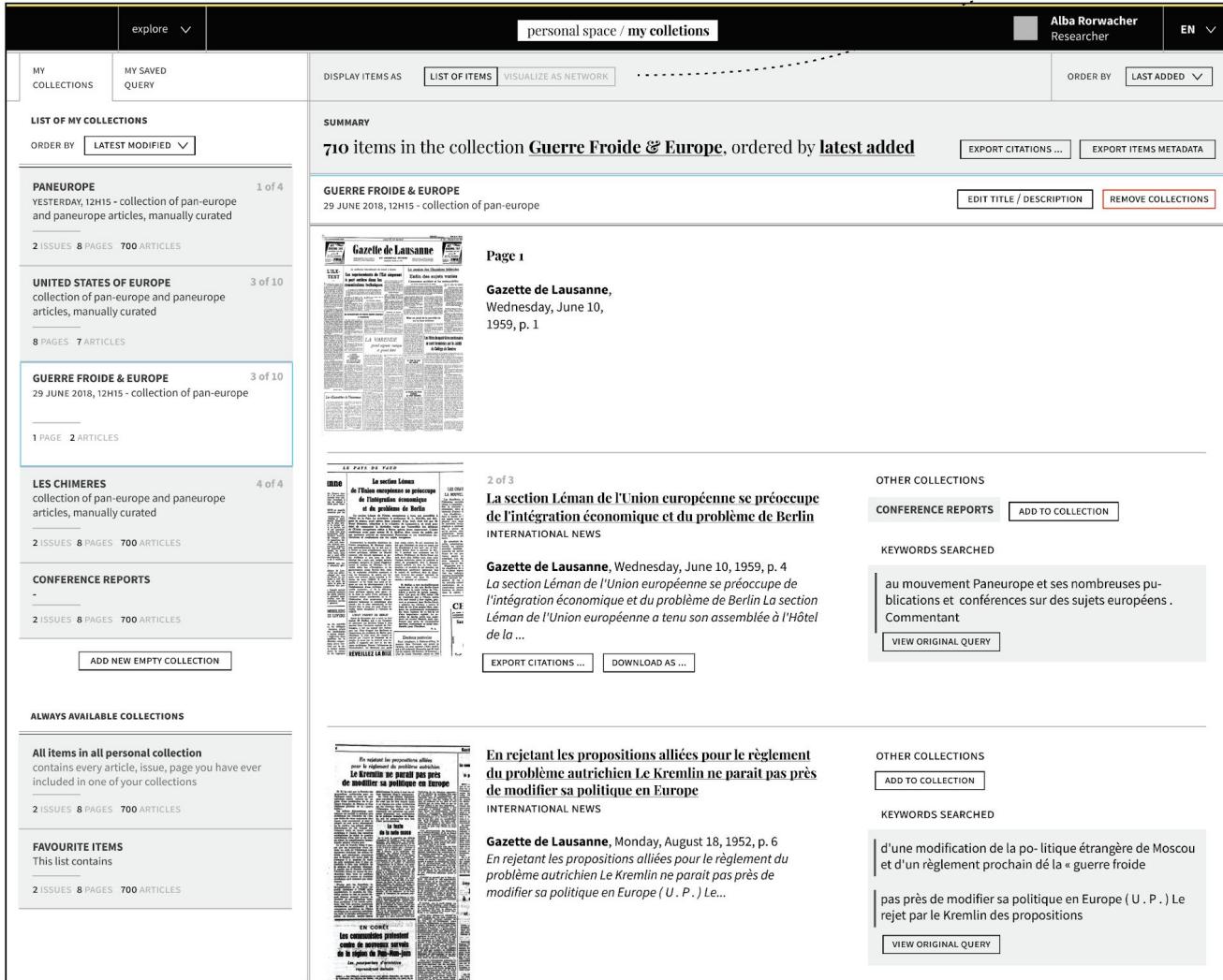
OTHER COLLECTIONS ADD TO COLLECTION

KEYWORDS SEARCHED

d'une modification de la politique étrangère de Moscou et d'un règlement prochain de la « guerre froide

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

VIEW ORIGINAL QUERY



Or.8

“How did you find this item?”
list of search results and link to the original query



LE PAYS DE FAUD

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

La section Léman de l'Union européenne a tenu son assemblée à Genève le 10 juin. Les deux dernières années ont été marquées par une croissance régulière de l'activité de l'Union. Les deux dernières années ont été marquées par une croissance régulière de l'activité de l'Union. Les deux dernières années ont été marquées par une croissance régulière de l'activité de l'Union.

LES CHAÎNES LA HOTEL

CE S...

REVEILLEZ LA BUE

2 of 3

La section Léman de l'Union européenne se préoccupe d'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 d'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

EN COREE Les communistes protestent contre de nouveaux survols de la région de Pan-Mun-jon

Le texte de la note russe

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

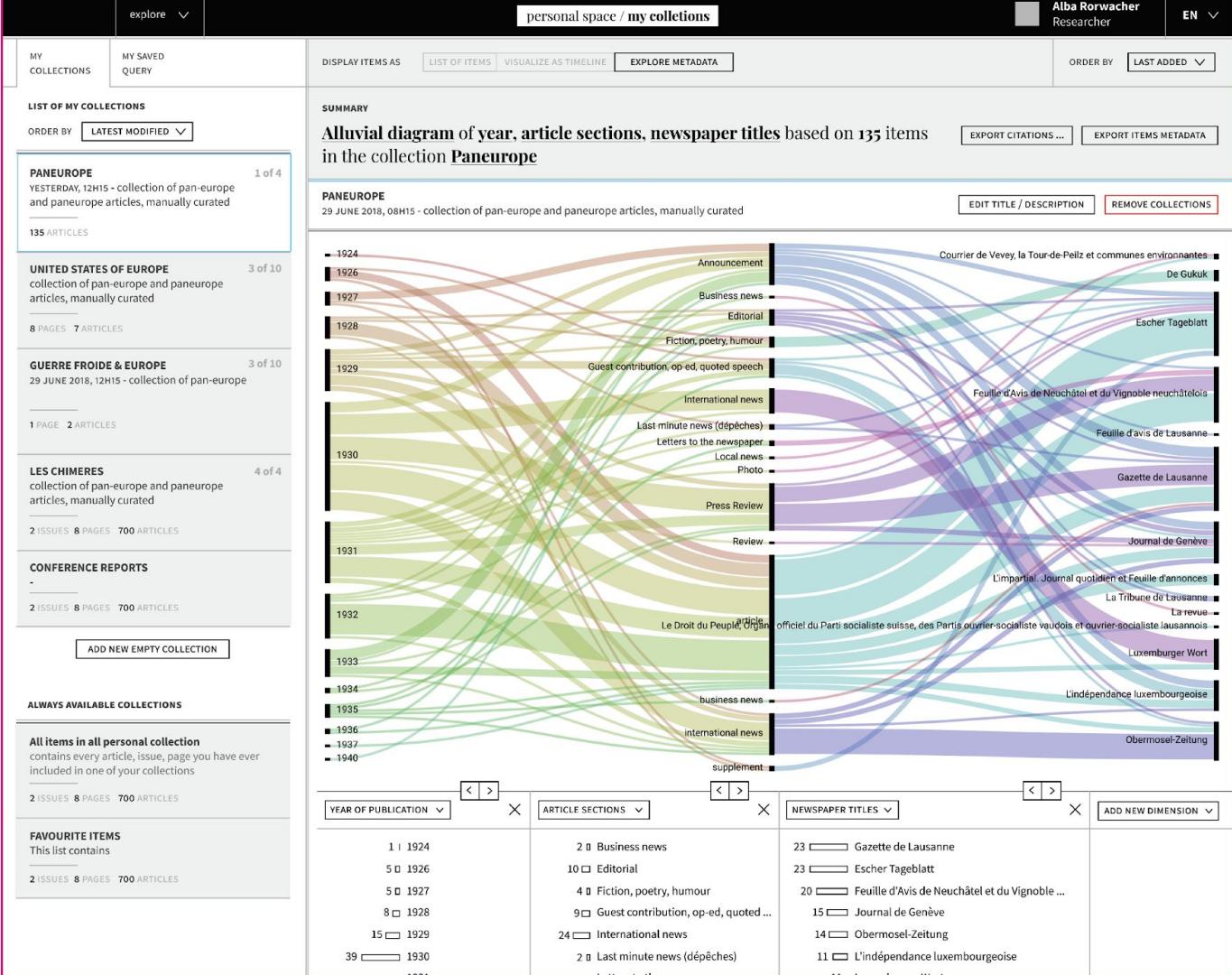
[ADD TO COLLECTION](#)

KEYWORDS SEARCHED

d'une modification de la politique étrangère de Moscou et d'un règlement prochain de la « guerre froide

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

[VIEW ORIGINAL QUERY](#)



switch type of visualisation:
list vs graph, vs alluvial diagrams ...
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

EST MODIFIED

Alluvial diagram of year, article sections, newspaper titles based on 135 items in the collection [PanEurope](#)

[EXPORT CITATIONS ...](#)[EXPORT ITEMS METADATA](#)

i - collection of pan-europe articles, manually curated
1 of 4

PAN-EUROPE 3 of 10
pan-europe and paneurope manually curated

LES

E & EUROPE 3 of 10
e15 - collection of pan-europe

ES

4 of 4
pan-europe and paneurope manually curated

5 700 ARTICLES

REPORTS

5 700 ARTICLES

[NEW EMPTY COLLECTION](#)[.E COLLECTIONS](#)

personal collection
article, issue, page you have ever
of your collections

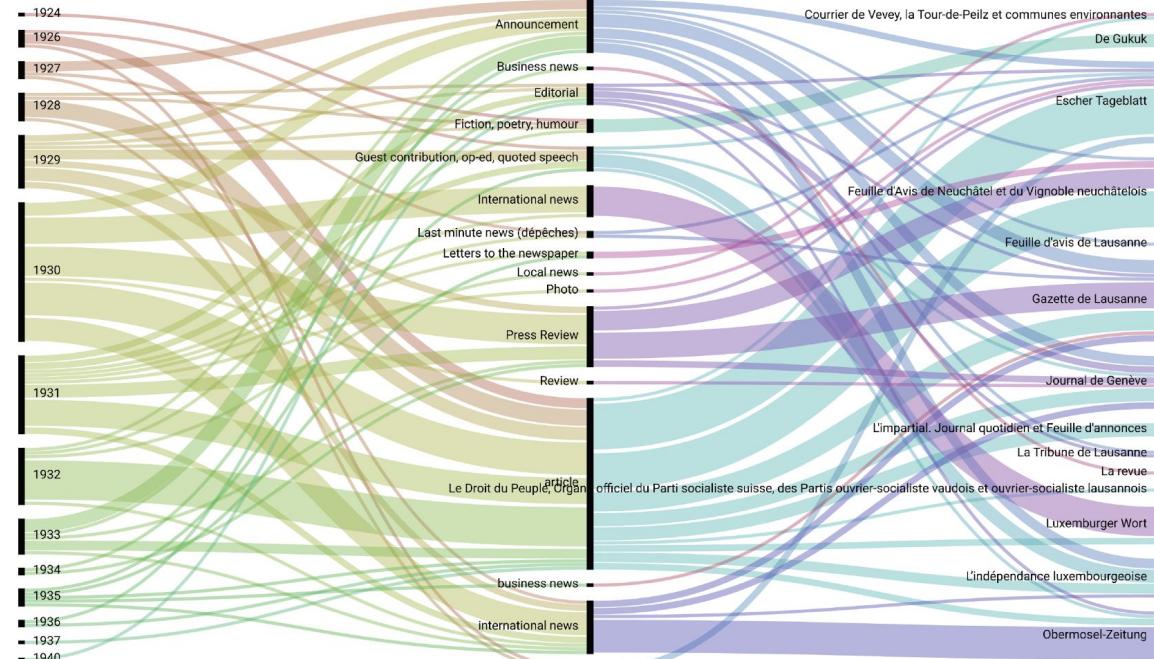
5 700 ARTICLES

MS

5 700 ARTICLES

PANEUROPE

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

[EDIT TITLE / DESCRIPTION](#)[REMOVE COLLECTIONS](#)

YEAR OF PUBLICATION ▾ X ARTICLE SECTIONS ▾ X 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)

Text classification

politics



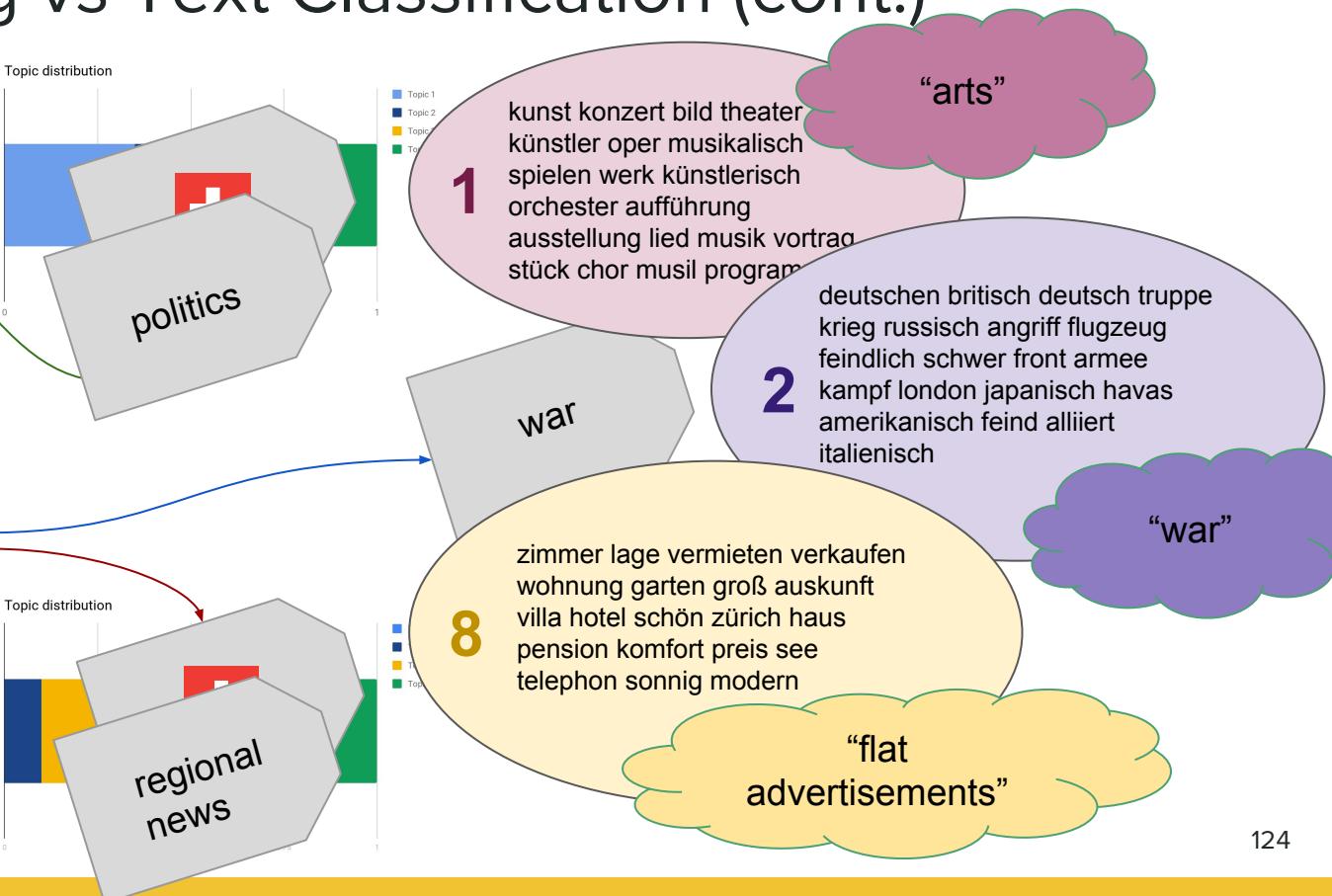
war



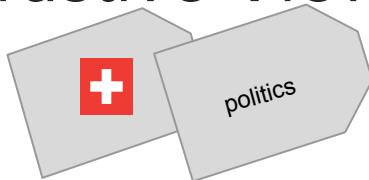
unsupervised

Topic modeling

Topic Modeling vs Text Classification (cont.)

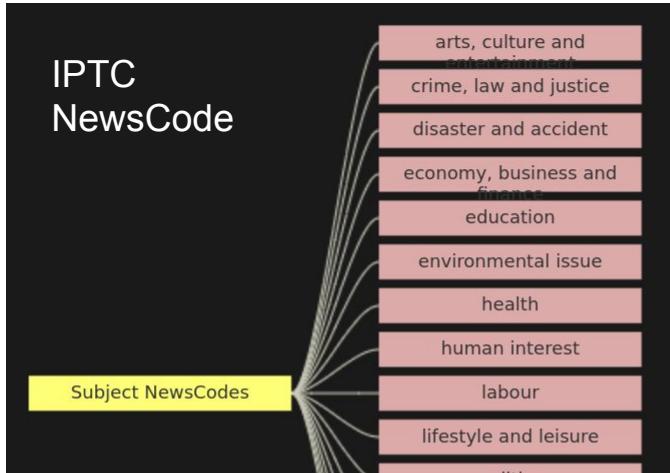


Contrastive View: Classes vs. Topics



classes = defined concepts

- taxonomy, e.g.



<http://show.newscode.org/index.html?newscode=subj&lang=en-GB&startTo>Show>



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



SUSSSE JOURNAL DE GENEVE - 8

Forte diminution de la grippe par rapport à 1972

Renvoi au Conseil fédéral pour réexamen mais va

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

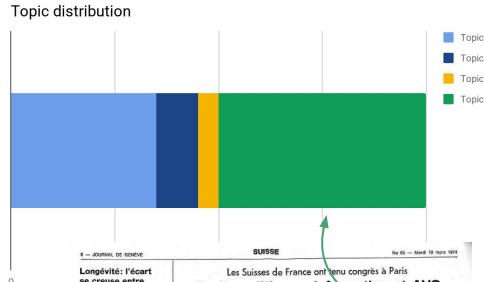
M. Hay ne méritait pas cela

Le résultat de la consultation du DFPJ

Avortement: guère d'unanimité pour l'une ou l'autre des trois variantes du projet

Le résultat final

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



Outlook on the personal space

Exportability of the collections?

- > The collection is a starting point 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 collections...

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?

