

impresso Techmeeting - 14.11. 8h30-12h

- output from WS1 (Marten)
- overview of the design process (Daniele and Thijs)
- mockups (Paul)
- scenarios and outlook (Estelle and Paul)
- outlook (all)

Output from WS1 (1/2)

Key word search / Recommender system

- Generating seed keywords (trial and error approach to determine keywords, cumulative construction of search queries, synonyms/antonyms)
- Semantic neighborhood (changing over time, KWIC, see how specific entities & key-words are qualified by newspapers, for a certain period or over time)
- Search across languages

Data / corpus

- OCR quality
- Missing metadata
- Personal storage (Subcorpus creation/comparison : between different periods, regions, languages)
- OLR (for corpus overview, search in titles, newspaper sections)

Output from WS1 (2/2)

Data visualisation

- Corpus overview
- Search in titles, newspaper sections
- Geographical search : search based on location (from metadata, from content)
- Transparency : information about the search scope: in width (coverage of newspapers collection) and in depth (quality of NLP processes, availability of metadata)

NLP

- Text reuse (possibility for suggestion of articles related to the same content as the press release? Using the word distribution, search for similar word mix in other articles)
- Term frequency over time (ngrams)
- Entity search (trace back person in timetracking of references, when texts mention their source)
- Topics as filters

Quick overview of the design process

conversation subjects

lightweight, "challenging" design scenarios with challenges



first set of features (shared ideas)



first drafts of designed components

existing methods & tools



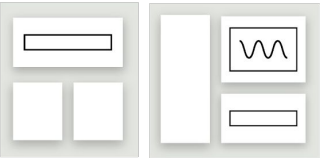
refined set of features (shared ideas)



refined components

use case scenarios

alpha users' use case scenarios recomposing components



introduction of new compositions

conversation subjects, part 2

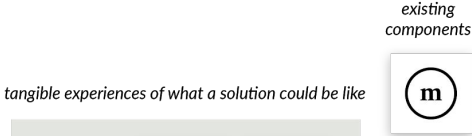
beta users's own uses cases, design scenarios for less simple tasks



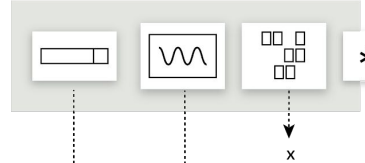
refined set of features (shared ideas)



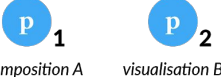
refined/new components



tangible experiences of what a solution could be like



concept solution prototypes



start prototyping!



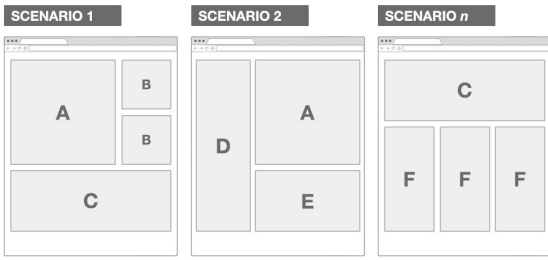
existing methods & tools

breakthrough inspiring new vision

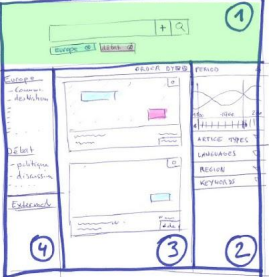
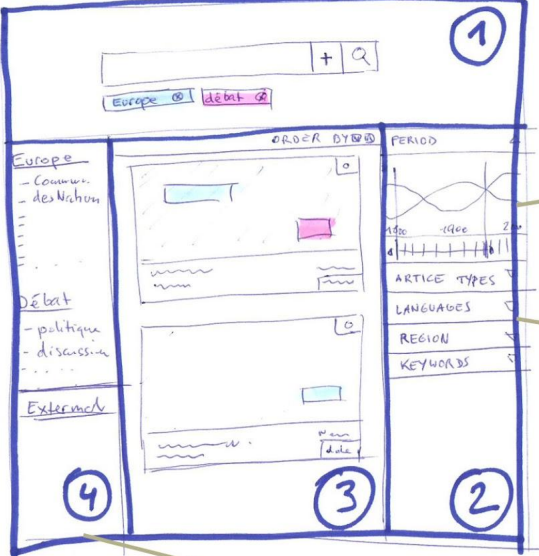
Develop individual components



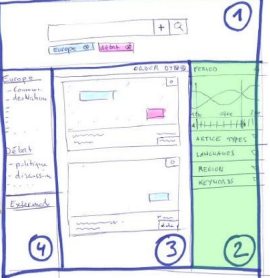
And combine them into layouts



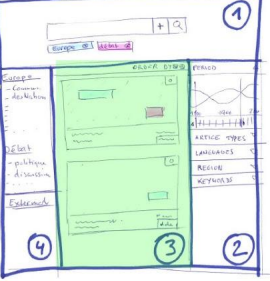
Modules



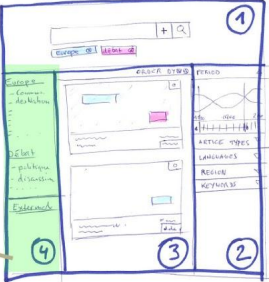
1. Search



2. Filters



3. Results



4. Suggestions

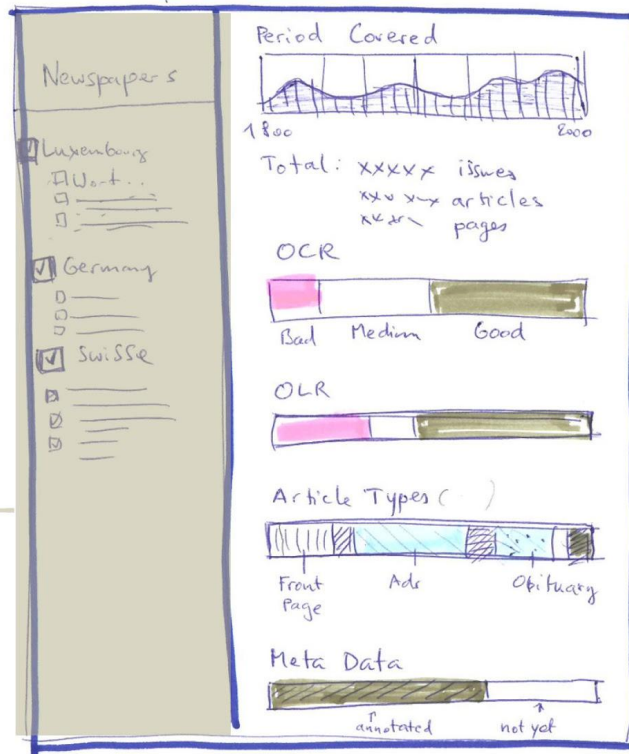
Transparency

TRANSPARENCY PROBLEM:

NEED TO GET AN OVERVIEW OF AVAILABLE DATA

SOLUTION:

PROVIDE STATISTICS ON SUB-CORPUS



Search

sac | Q

sac | + | Q

- Sacramento
country in sacramento
- Swiss Alpine Club (SAC) 1
- Saxony (Sachsen) 2
- more
- containing "sac"

auto-complete +
+ Named Entity Recognition

dedicated suggestion
page

Word Analytics /
Semantic Neighbourhood
... (?)

Vanzetti x sacco

sac | + | Q

add to query
refine search

Overview of the revised scenario

- A. search for a concept: Indogerman/european people
- B. search for a topic within a type of article: Europe in readers letters
- C. search for an event and its memory: Victor Hugo's speech

tested in : <http://www.letempsarchives.ch/>

A. search for a concept: indogerman/european people

First step:

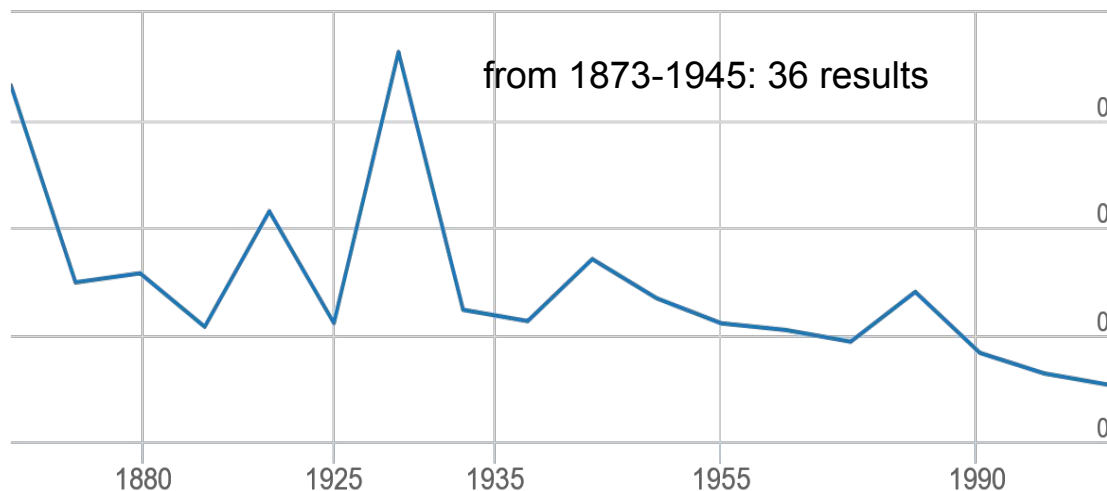
1. check the variation in different languages (via wikipedia for instance)
2. check if there were some prominent defenders of that idea (place, literature)
3. search with the keyword

INDOEUR | + | Q

A. search for a concept: first results/impressions

Zoom : 10 a 100 a Tous

De 1873 à 1994



good recognition of the word even is separated

problèmes relatifs à l'origine des langues indo-européennes». 20 h. 30. Athénée, Classe d'agriculture, avec le

search with indoeur*: 42 results

site, désireuse de s'attacher le jeune maître déjà célèbre, obtint la création d'une chaire d'histoire et comparaison des langues indo-européennes, qu'il vint occuper au début du semestre d'hiver. En vain, l'on e-saya de le rappeler à Paris; en vain, M. Bréal lui offrit sa suppléance au Collège de France où Saussure

racines, étant eux-mêmes dérivés de plus anciennes langues, de souche indoeuropéenne. Les restes de ces langues se re-

missing data?

— indoeuropéen

A. search for a concept

three different conceptions:

- people
- linguistics
- geographic

(manual) selection of key words
surrounding the chosen conception:

“people, nation, famille, ancêtre, race,
souche, tronc...”

option to store the results in different
collection?

no need to refine, not so many results:
need to enlarge? how?

11 issues selected from 1876 to 1944.

indoeuropéen apres:1873 avant:1944

À PROPOS AIDE FR | EN |

... que de ette la- imique énorme ructure s cons-	... et d'Herculanum; rien du lien linguistique qui rattache ensemble les nations indo- européennes; rien du vaste déroulement d'inscriptions sur les tombes et les tem- ples d'Égypte; rien du sens des inscrip-	... Les nouvelles politiques suisses tienn peu de place dans la Gazette de 1824, même d'ailleurs que dans toute la presse
CDL		09.03.1924 — page 1

... ns et Assy- e réformée, délégué de ie politique puis 1893.	— M. BOGATIEWSKI, délégué de l'Univ. Suède Göteborg. M. JOH. VISING, délégué de l'Uni- versité, recteur, professeur de langues indo- européennes. Auteur de travaux sur l'histoire de la langue française et spécialement de français parlé et écrit au moyen âge en Angleterre (anglo-normand).	Pittsburgh. Rev. D. S. Schaff, délégué du Western Theol. Sem. de Pittsburgh et de Presbyt. Church de Philadelphie. D'origine grisonne, professeur d'histoire ecclésiastique, auteur d'une Histoire de l'Eglise chrétienne au moyen âge. Princeton. Rev. F. LANDEY PATTON, dé- légué et président de Theol. Sem. of the Presb. Church. Educateur de haute valeur,
JDC		08.07.1909 — page 5

majorité des niques réu- conserva- l Cox, ob- ns un joli dimbourg: ar, avocat, é sollicitor termes de	En Perse Le correspondant du Daily Chronicle à Tabriz télégraphie le 4 que les royalis- tes ont démoli la ligne télégraphique indo- européenne.	vent vivement l'attitude de Belgrade, que nul, maintenant, ne pourra accuser de troubler la paix. Vienne, 5. Des informations de source privée par- venues de Belgrade à la Zeit mettent en doute la renonciation de la Serbie à des compensations. Elles dépeignent la situa- tion comme très sérieuse.
JDC		06.03.1908 — page 2

A. search for a concept: what's next?

SEMANTIC PROBLEM:

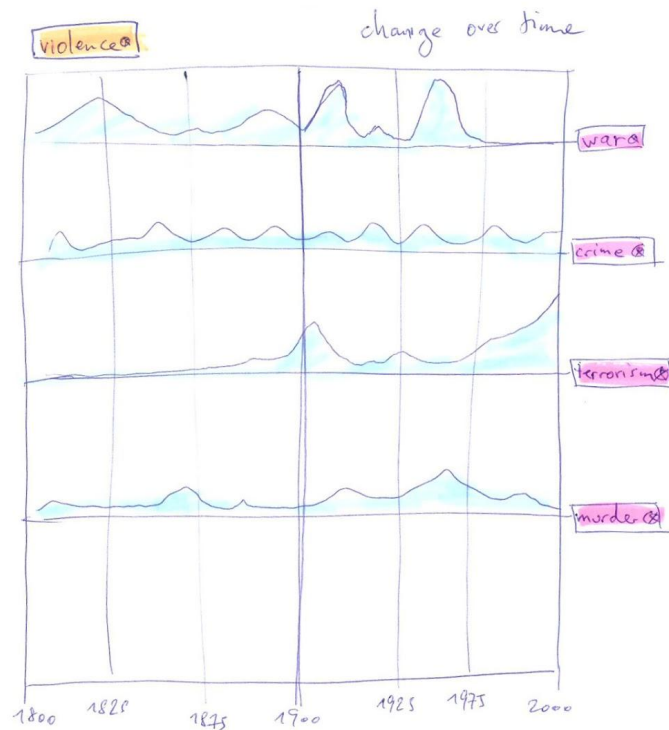
THREE DIFFERENT CONCEPTIONS:

- PEOPLE
- LINGUISTICS
- GEOGRAPHIC

SOLUTION:

SELECTION OF KEY WORDS
SURROUNDING THE CHOSEN
CONCEPTION:

"PEUPLE, NATION, FAMILLE, AN-
CÊTRE, RACE, SOUCHE, TRONC..."



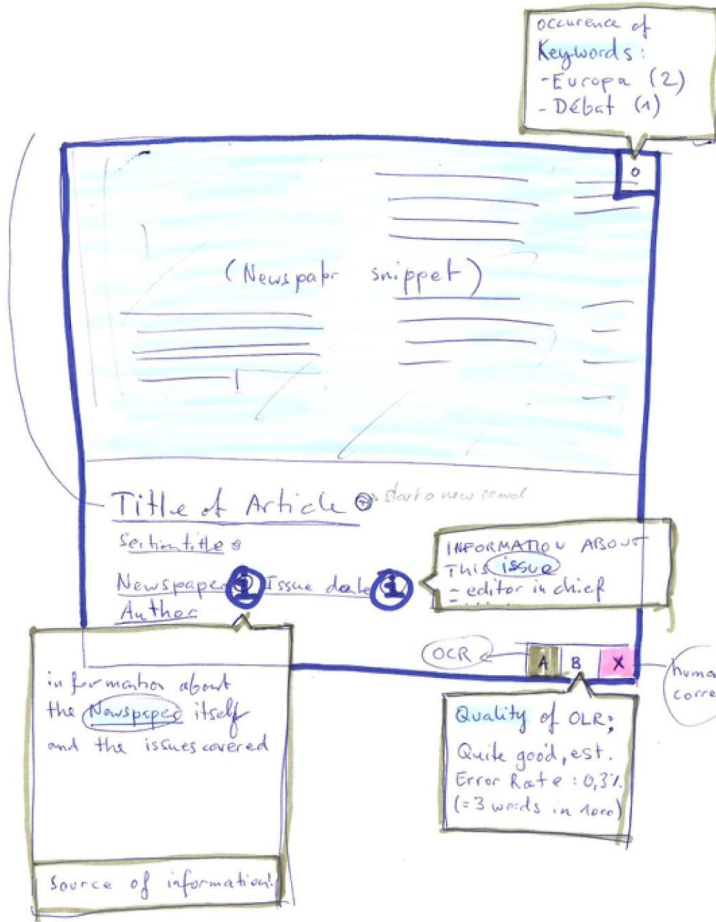
A. search for a concept: what's next?

GENEROSITY

PROVIDE ADDITIONAL
ENTRY POINTS BASED
ON SEARCH RESULTS
THROUGH LINKED
META-DATA

GENEROSITY

PROVIDE INFORMATION
ON META-DATA



TRANSPARENCY

GIVE INSIGHTS ON
RESULT LISTING /
DEMISTIFY SEARCH
ALGORITHM
(DARK BOX)

TRANSPARENCY

GIVE INSIGHTS ON
OCR/OLR QUALITY,
HUMAN CORRECTION

B. search for a type of article: readers letters

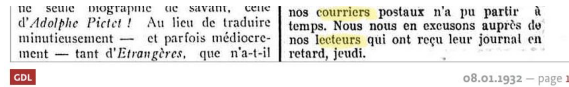
First step:

1. try out several potential titles or keywords connected to that type, taking into account that it may not be a very old practice, or that the wording changed drastically over time
2. refine the search after with a specific topic, for instance: Europe

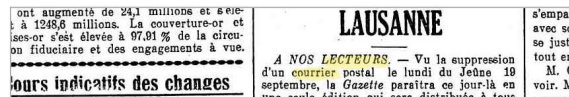
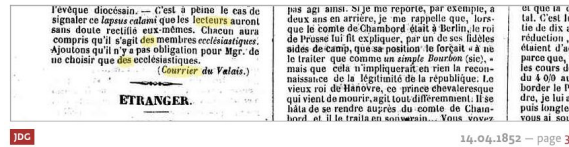
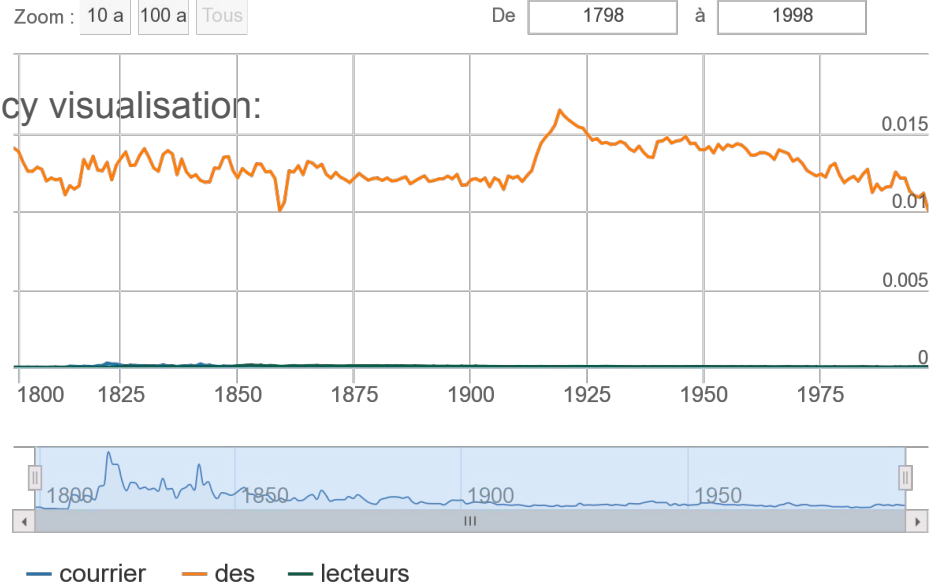
B. search for a type of article: readers letters



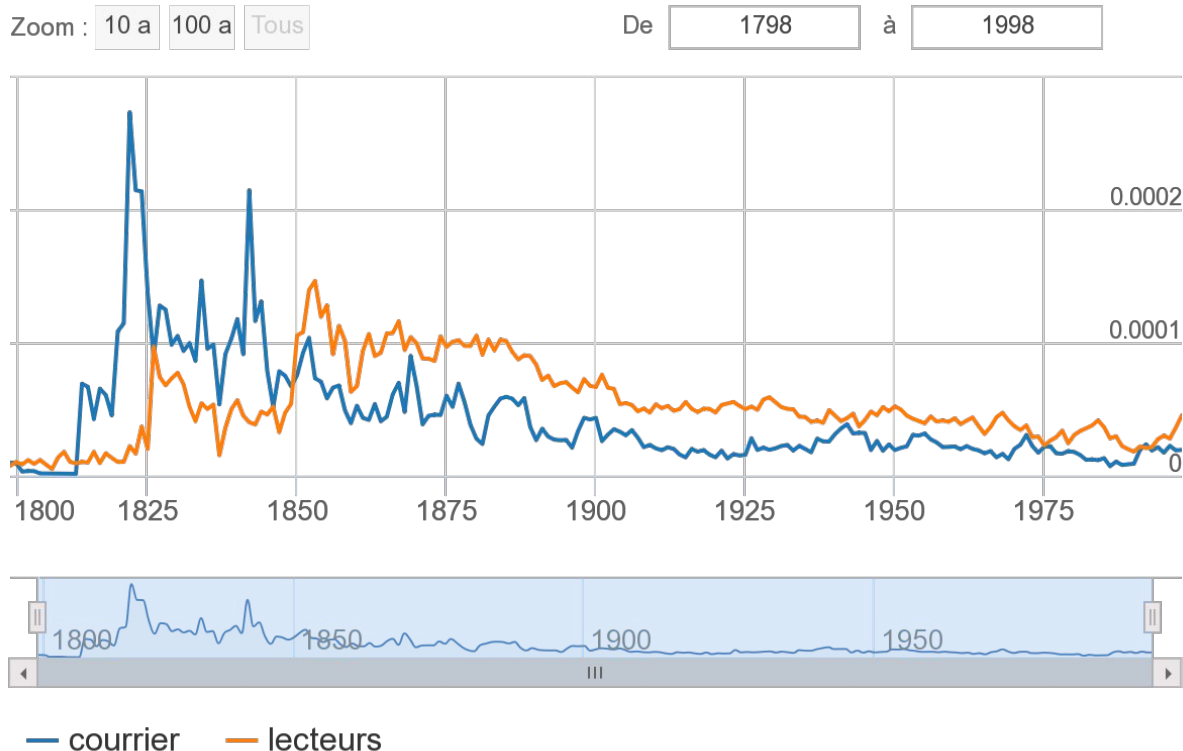
different types of layout but there seems to be not specific section before 1945.



not visible in the frequency visualisation:



B. search for a type of article: readers letters



when searching with two words, more visibility of disconnection of the two words

search for “courrier des lecteurs” before 1945: two irrelevant results.

B. search for a type of article: readers letters

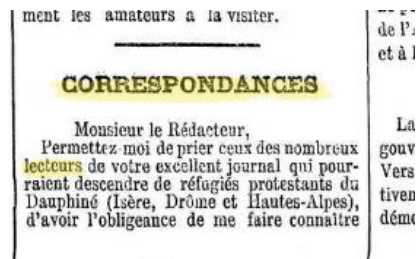
search for “de nos lecteurs” before 1945: a lot of results with “à nos lecteurs”.

then: search with: “correspondance” and “lecteurs”: 178493 results
(sometimes: “correspondance particulière”)

observation: section “correspondances” and letter seem to start often with
“Monsieur le rédacteur”: search with it: 12 461 results.

-> possibility to **combine** or **segregate** several queries according to **time**: eg.
if possible to identify the creation and then the change of a section title,
possibility for one research, collect several queries?

tried to refine with Europe: increase of the results: 187 646, and apparently
no regards for proximity of the terms (even without OLR).



Personal Space

DYNAMIC LISTS

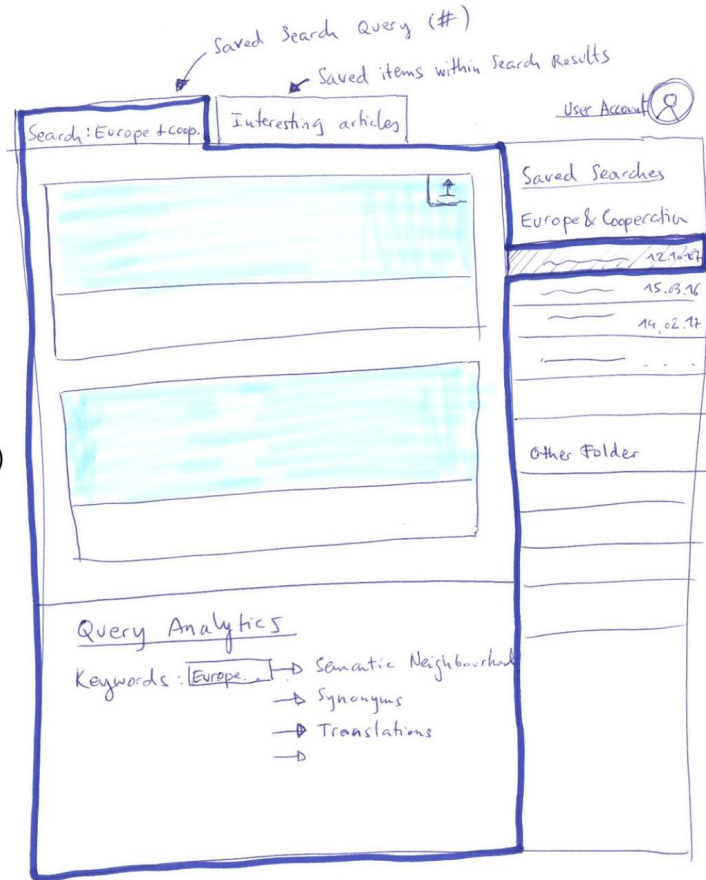
SAVE SEARCH QUERRIES

STATIC LISTS (IDENTIFIED USERS ONLY)

SAVE PERSONNALIZED LISTS

PREFERENCES

SAVE FAVOURITE SETTINGS



LIST TOOLS

ORGANIZE
LISTS INTO
FOLDERS

ORDER LIST
ITEMS

ANALYZE
LISTS

COMPARE
MULTIPLE
LISTS

C. search for an event and its aftermath: Victor Hugo's speech on the United States of Europe

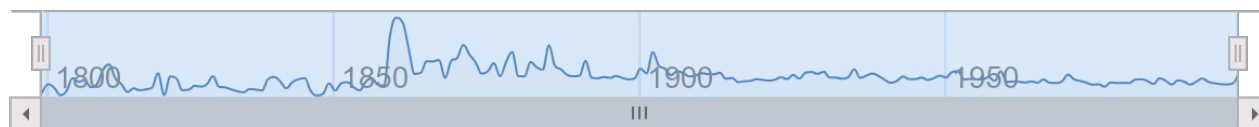
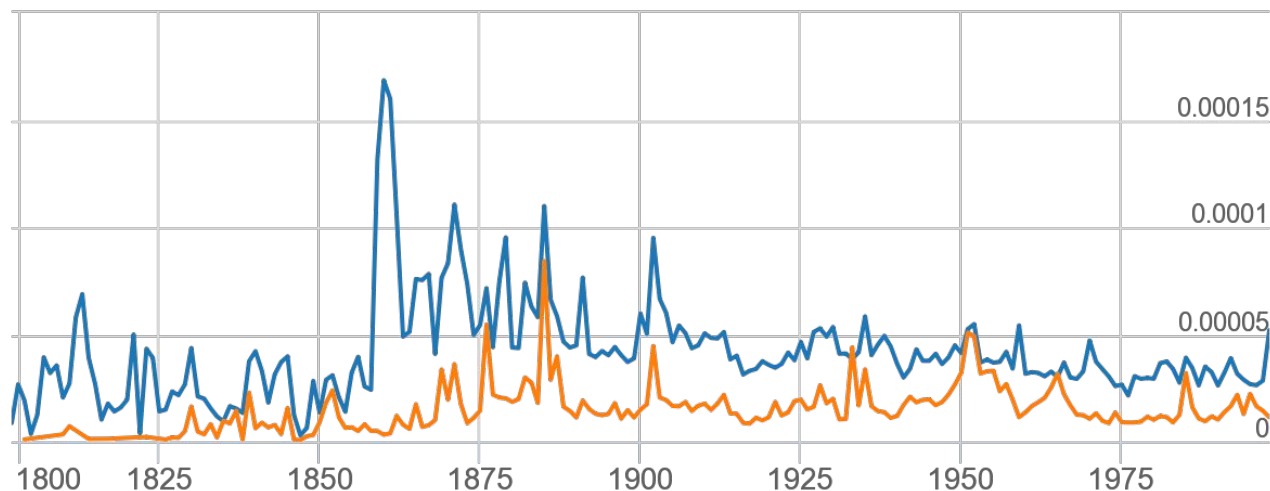
First step:

1. search for the original version
https://fr.wikisource.org/wiki/Discours_sur_les_%C3%89tats-Unis_d%E2%80%99Europe_au_congr%C3%A8s_international_de_la_paix_en_1848_%C3%A0_Paris
2. select salient expressions, names, information about the date and place
3. search for immediate coverage, in the following weeks/months
4. search for the memory of this event/ideas over time, the references to that speech (frequency and context)

C. search for an event and its memory: Victor Hugo's speech

Zoom :

De à



— victor — hugo

C. search for an event and its memory: Victor Hugo's speech

victor hugo



Ordonnance de la justice de
Paris sur Paris.
AGÈMENTS MODERNES
le septennat de Vincent
de notables modifications

LA REINE MARY

La voix d'un peuple entier la ber-
ce en son tombeau.

(Victor Hugo.)

A BONNOC AINE ED I
152, av. Aristide Briand Maternité de Hasenwain
Mulhouse (Haut-Rhin) Mulhouse



02.06.1951

GDL

30.10.1956 — page 4

jours de
table for-

DERNIÈRES DÉPÊCHES

Italie

Genève.

feld
les mem-
sont ren-
sion de

ROME, 28. (S) — En souvenir de Victor
Hugo. — Dans une des prochaines séances
du conseil municipal de Rome, le prince Co-
lonna, syndic de la ville, proposera de don-
ner le nom de Victor-Hugo à l'une des rues

mir-Hissar, où les insurgés se sont montrés
avec beaucoup d'audace. Deux divisions
réduits sont mobilisés, le n° 7 à Andrinople
le n° 20 à Damas, mais elles resteront pro-
soirement dans leurs quartiers de recrui-
ment. Sept bataillons de réduits albanais
garnison à Verissovitz, sur le bruit que pl-
sieurs de leurs compatriotes avaient é-
condamnés à l'exil en Asie Mineure, ont é-
posé les armes, et menacent, si l'on don-

JDC

29.09.1903 — page 3

rains sont
Loetschberg

rondissement des
ous communique

de l'automotrice. Six personnes ont
trouvé la mort dans cet accident.

Le cinquantenaire de Victor Hugo

Dimanche après-midi, à
de 17 h. 30, M. Louis Lavar
jardinier, habitant Paudex,
bain devant Lutry lorsqu'il
frappé de congestion. Il se
noya. MM. Naiss, de Pau

different types of results:

- poetry
- commemorations
- adds of shops on victor hugo streets

FAITS DIVERS.

— Les amis et l'entourage de M. Victor Hugo veulent créer un journal parisien. M. Victor Hugo a donné le titre, assez heureux : le *Rappel*. Cette feuille politique quotidienne aurait l'allure légère et satirique du *Figaro*. On peut déjà faire connaître parmi ses rédacteurs : MM. Paul Meurice, Ch. Hugo, H. Rochefort. M. Vacquerie se chargera de la revue des théâtres, où il compte semer des innovations.

Ce journal paraîtrait le 1^{er} mars. Victor Hugo y enverrait quelquefois de petites poésies.

C. search for an event and its memory: Victor Hugo's speech

same problem as before: no possibility to refine the search results from Victor Hugo and the main reference of that speech:

7438 results for “Victor Hugo” and “États-Unis d'Europe” (no refinement)

search for: “États-Unis d'Europe”: 14 results, including articles about the US and Europe, but mainly about that idea.

search for “etats-unis d'Europe”: 585 results

C. search for an event and its memory: Victor Hugo's speech

Manual collection of surrounding words: allusion, création, constitution, projet, volonté, obstacle, opposition, futurs, conférence, rêve (further located than with "indoeuropean")

La Délégation britannique n'a point mesuré ses applaudissements à M. Briand quand il a parlé ici, la semaine dernière, des **Etats-Unis d'Europe**. Au banquet des Bergues où les Anglais « fortunés convives » figurèrent en bonne place, ils mirraient, tout comme les autres, les bouchées doubles. Interrogeait-on les uns ou les autres sur la « grande idée » mise en circulation, on obtenait des réponses prudentes mais approbatives : « Grande et généreuse **idée**... Il y a certainement

Lorsque M. **Aristide Briand** commença à parler des **Etats-Unis d'Europe**, les journaux se hâtèrent d'utiliser les élogieuses formules d'usage : **plan** ingénieux, noble **intention**, généreuse entreprise... Aujourd'hui le *Matin* croit bon de surenchérisir : il désigne le projet comme : **l'initiative** la plus audacieuse qu'un homme d'Etat ait prise depuis la guerre ». Pourtant le **memorandum** qui vient d'être rendu public me paraît singulièrement pâle et inoffensif.

ooo

People: Mussolini, Briand, Stresemann, Benes...

-> Organisation reusing that concept in the 1930s, supported by A. Briand: how to differentiate and store them separately (both are interesting)?

Les États-Unis d'Europe

L'orateur fait **allusion** aux États-Unis d'Europe et dit qu'entre les peuples géographiquement groupés comme les peuples d'Europe devrait exister une sorte de **lien fédéral**. Il estime que ce lien ne devrait pas seulement exister dans le domaine économique, mais aussi dans le do-

Communiqué

ÉTATS-UNIS D'EUROPE

Samedi et dimanche aura lieu sur le territoire du canton une vente d'insignes autorisée par le Département de justice et police au profit du fonds de propagande du Mouvement européen pour les **Etats-Unis d'Europe** dont le siège est à Genève, et le Conseil provisoire présidé par M. Jean-Jacques Babel, le Comité suisse par M. Auguste Kupferschmid.

Le but poursuivi est passionnant. Aussi longtemps que l'**ordre** européen est inexistant, toute formule de société des nations sur le plan mondial sera une **utopie**.

Où les peuples européens se fédéreront pour reconstruire dans la paix ce que la démeure des hommes a anéanti, ou ils continueront à se haïr et seront irrémédiablement broyés.

Pour sortir du chaos, aidons le Mouvement européen pour les **Etats-Unis d'Europe**, dernière chance de salut.

Samedi et dimanche, que chacun porte l'insigne du M. E. J. symbole de la renaissance européenne.

Prix minimum : 80 centimes.
M. E. J. Case postale Rive 133
Chèques postaux : 110.044

2545

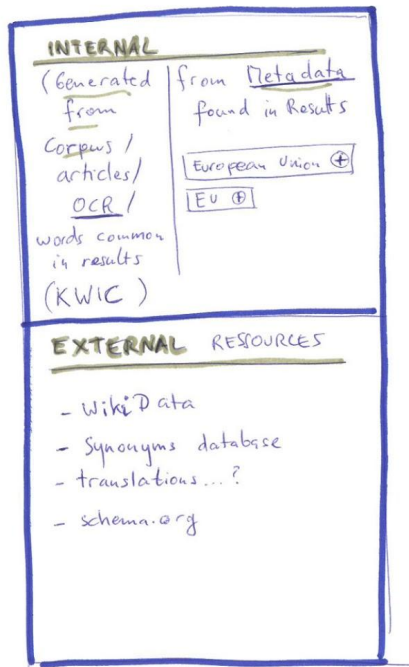
ministrer une douche glaciale aux promoteurs de la **Fédération européenne** : « Une Fédération européenne, disent-ils, non pas, mais une Fédération renforcée de la Grande-Bretagne et des Dominions, oui, voilà le salut. » De sorte que les **promoteurs des États-Unis d'Europe** risquent fort, avec les meilleures **intentions** du monde, de détacher davantage encore de l'Europe la Grande-Bretagne et les Dominions. Ils veulent aggraver, ils dissolvent. Il y a dans cet accueil plus que tête, de quoi faire réfléchir les gouvernements de l'Europe continentale. **Surtout, pas trop de zèle !**

Quant aux États extra-européens, j'ai déjà noté leur **réserve** : L'Amérique du Sud qui fait partie elle-même de l'**Union panaméricaine** n'a aucune raison valable de prendre ombrage des **Etats-Unis d'Europe**. Elle n'en regrette pas moins que ce **projet ait été lancé** à la Société des nations. Quant aux Américains de l'Union nord-américaine, qui s'intéressent vivement à ce qui se fait ici tout en ne figurant pas officiellement dans le *League*, leur opinion est plus difficile à saisir et plus complexe. **M. Briand** a déclaré si haut que les **Etats-Unis d'Europe** n'étaient point dirigés contre leur puissant pays qu'ils auraient pu le croire. Et, de fait, ceux qui fréquentent les couloirs de la S. d. N. ne marquaient à l'égard du projet Briand de **confédération européenne** aucune animosité. Certains d'entre

C. search for an event

GENEROSITY

RECOMMEND BASED ON
CORPUS (KWIC, ...)



RECOMMEND BASED ON
INTERNAL META-DATA

RECOMMEND BASED ON
EXTERNAL RESSOURCES

Transparency on source of information.

Outlook?

- OLR?
- Filters?
- Collection of elements for the constitution of a subcorpus (transversal, vertical) in a personal space?
- text reuse?
- NER?
- temporally differentiated sub-corpora?
- external links? (historical context, contemporary events?)
- > Maybe an option for the Event detection mechanism for Wikipedia content that was applied in CUBRIK (LWI)
- ...

Output from WS1 (1/2)

Key word search / Recommender system

- Generating seed keywords (trial and error approach to determine keywords, cumulative construction of search queries, synonyms/antonyms)
- Semantic neighborhood (changing over time, KWIC, see how specific entities, key-words are qualified by newspapers, for a certain period or over time)
- Search across languages

Data / corpus

- OCR quality
- Missing metadata
- Personal storage (Subcorpus creation/comparison : between different periods, regions, languages)
- OLR (for corpus overview, search in titles, newspaper sections)

Output from WS1 (2/2)

Data visualisation

- Term frequency over time
- Corpus overview
- Search in titles, newspaper sections
- Geographical search : search based on location (from metadata, from content)
- Transparency : information about the search scope: in width (coverage of newspapers collection) and in depth (quality of NLP processes, availability of metadata)

NLP

- Text reuse (possibility for suggestion of articles related to the same content as the press release? Using the word distribution, search for similar word mix in other articles)
- Frequency search
- Entity search (trace back person in timetracking of references, when texts mention their source)
- Topics as filters