



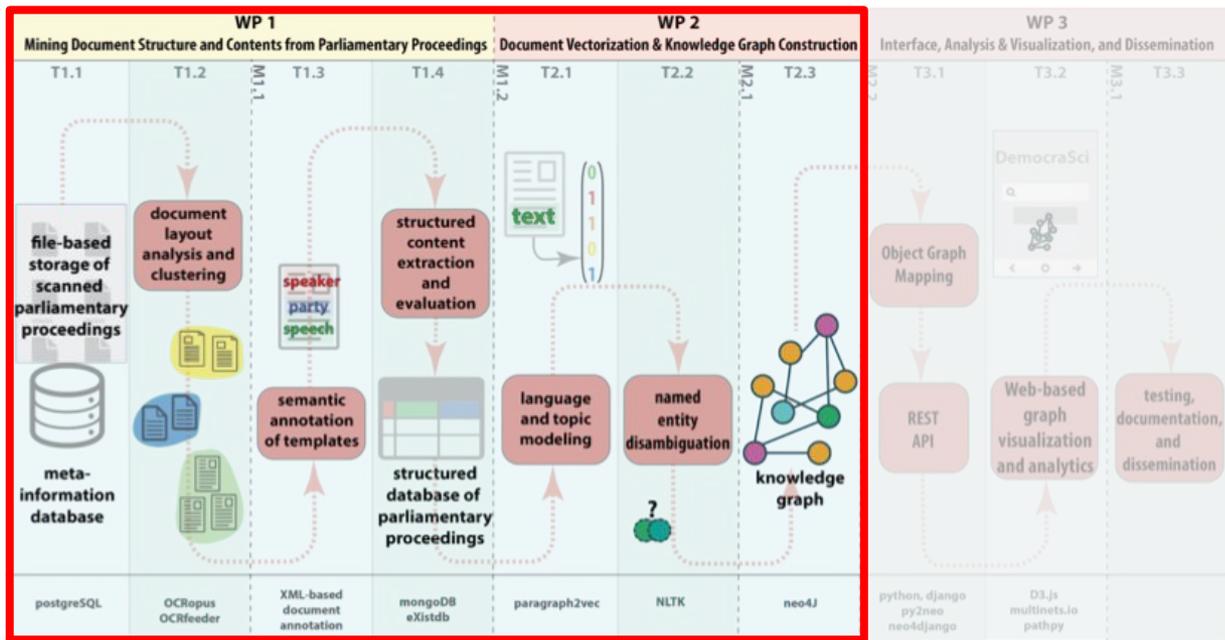
# DemocraSci

Follow-up meeting, October 16th, 2018

# Overview

- The project is nicely structured. We have been focusing on WP1&2 so far.

We are working on both WPs at the same time. This way we can benefit from results or requirements of other stages

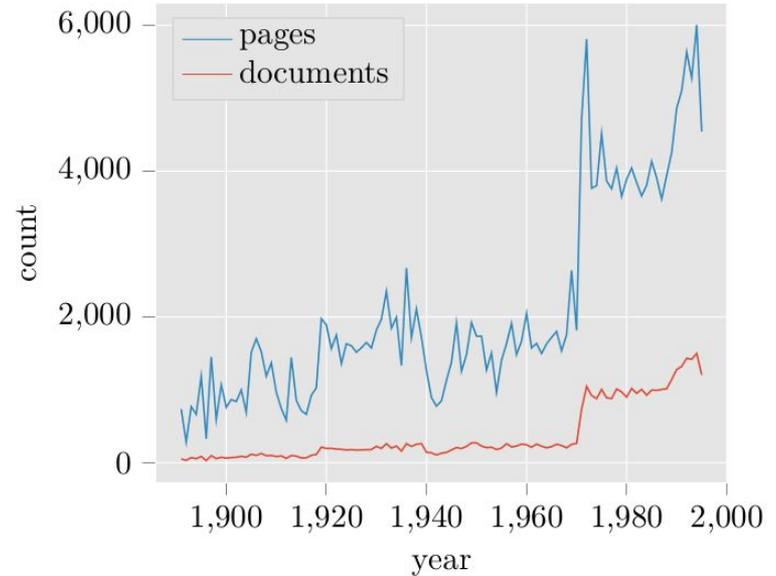


# Data & Storage (T1.1)

- 40'000 OCRed pdf scans:
  - From 1891 to 1995, time period can be extended.
  - Protocols from national council and council of states.
  - Plus metadata on session, date, etc.
- List of politicians.

✓ Data stored on Switch.

↻ Dealing with the large number of files, which complicates the use of Renku, where the database will be set up.



# Preprocessing of extracted XML data (T1.1-T1.2)

- OCR sufficient but problems with the extracted XMLs.

The result is an enriched XML file for each document, with new attributes for the textboxes, corrected bounding boxes, etc.

Implemented a method for:

- Correcting overlaps, mergings textlines, etc.
- Finding the margins, horizontal and vertical lines.
- Identifying page elements (needs to be improved for some).

Rule-based methods perform better than an unsupervised approach.

# Layout analysis and annotation (T1.2.8)

• There are different layouts, but normally comprised in single proceedings.

Bern, Dienstag, 10. April. III. Jahrg. 1893/94 III<sup>e</sup> année Bern, mardi, 10. avril.

Antliches **BULLETIN**  
 stenographisches Bulletin **STENOGRAPHIQUE OFFICIEL**  
 der schweizerischen Bundesversammlung **N° 43** PARENTHÈSE FÉDÉRALE SUISSE

## Verbrechen gegen die öffentliche Sicherheit. Délits contre la sûreté publique.

### Law (German)

#### Bundesgesetz betreffend

Ergänzung des Bundesgesetzes über das Bundesrecht der schweizerischen Eidgenossenschaft vom 6. Februar 1853.

Die Bundesversammlung der schweizerischen Eidgenossenschaft, nach Einsicht einer Entsch. des Bundesrats vom 14. Dezember 1893; in Ausführung des Art. 114 der Bundesverfassung und in Ergänzung des Bundesgesetzes über das Bundesrecht der schweizerischen Eidgenossenschaft vom 4. Februar 1853,

beschliesst:

- Art. 1. Wer Sprengstoffe zu verheerenden Zwecken gebraucht, wird mit Zuchthaus von wenigstens 10 Jahren bestraft.
- Art. 2. Wer Sprengstoffe, von denen er anzunehmen muss, dass sie zu Verbrechen gegen die Sicherheit von Personen oder Sachen gebraucht werden sollen, herstellt oder zu deren Herstellung Anleitung gibt, wird mit Zuchthaus nicht unter 3 Jahren bestraft.
- Art. 3. Wer Sprengstoffe, von denen er anzunehmen muss, dass sie zu Verbrechen gegen die Sicherheit von Personen oder Sachen gebraucht werden sollen, herstellt oder zu deren Herstellung Anleitung gibt, wird mit Zuchthaus nicht unter 3 Jahren bestraft.
- Art. 4. Wer in der Absicht, Schrecken zu verbreiten oder die allgemeine Sicherheit zu erschüttern, zu Verbrechen gegen die Sicherheit von Personen oder Sachen aufzufahren oder Anleitung gibt, wird mit Gefängnis nicht unter 6 Monaten oder mit Zuchthaus bestraft.
- Art. 5. Werden also in Art. 4 bedrohte Verbrechen durch die Druckpresse oder durch ähnliche Mittel begangen, so sind die sämtlichen Teilnehmer (Thäter, Anstifter, Gehilfen und Begünstigter) strafbar und es finden auf dieselben die Vorschriften der Artikel 69 bis 72 des Bundesstrafgesetzes vom 4. Februar 1853 keine Anwendung. Gegen den Begünstigter kann nur blosse Geldbusse erkannt werden.
- Art. 6. Die in den Artikeln 1 bis 4 erwähnten Handlungen unterliegen, auch wenn sie im Ausland begangen werden, den Strafbestimmungen dieses Gesetzes, sofern sie gegen die Eidgenossenschaft oder ihre Angehörigen gerichtet sind.
- Art. 7. Treffen mit den in den Artikeln 1 bis 4 erwähnten Verbrechen Handlungen zusammen, welche unter das kantonale Strafrecht fallen, so sind diese letzteren durch das kantonale Gesetz zu bestrafen und unterstehen dem kantonalen Bundesgesetzrecht.

Beschluss des Nationalrats. 6. April 1894.

Zustimmung, wo nichts anderes bemerkt ist.

Art. 4. Wer in der Absicht, Schrecken zu verbreiten oder die allgemeine Sicherheit zu erschüttern, zu Verbrechen gegen die Sicherheit von Personen oder Sachen aufzufahren oder Anleitung gibt, wird mit Gefängnis nicht unter 6 Monaten oder mit Zuchthaus bestraft.

Art. 5. Werden also in Art. 4 bedrohte Verbrechen durch die Druckpresse oder durch ähnliche Mittel begangen, so sind die sämtlichen Teilnehmer (Thäter, Anstifter, Gehilfen und Begünstigter) strafbar und es finden auf dieselben die Vorschriften der Artikel 69 bis 72 des Bundesstrafgesetzes vom 4. Februar 1853 keine Anwendung. Gegen den Begünstigter kann nur blosse Geldbusse erkannt werden.

Art. 6. Die in den Artikeln 1 bis 4 erwähnten Handlungen unterliegen, auch wenn sie im Ausland begangen werden, den Strafbestimmungen dieses Gesetzes, sofern sie gegen die Eidgenossenschaft oder ihre Angehörigen gerichtet sind.

Art. 7. Treffen mit den in den Artikeln 1 bis 4 erwähnten Verbrechen Handlungen zusammen, welche unter das kantonale Strafrecht fallen, so sind diese letzteren durch das kantonale Gesetz zu bestrafen und unterstehen dem kantonalen Bundesgesetzrecht.

Art. 8. Referendums-Klausel.

Décision du conseil des états. 29 mars 1894.

### Law (French)

Loi fédérale sur le code pénal fédéral du 4 février 1853.

#### L'ASSEMBLÉE FÉDÉRALE DE LA CONFÉDÉRATION SUISSE

vu le message du conseil fédéral du 18 décembre 1893; en application de l'article 114 de la constitution fédérale et en vertu de son pouvoir exécutif fédéral (fédéral) de la Confédération suisse du 4 février 1853,

arrête:

Art. 1<sup>er</sup>. Celui qui fait usage de matières explosibles dans une intention criminelle, sera puni de la mort ou de réclusion à vie.

Art. 2. Celui qui, sachant qu'il les sera destinés à commettre des délits contre des personnes ou des propriétés, fabrique des matières explosives ou des débris destinés à commettre des délits contre des personnes ou des propriétés, sera puni de cinq ans de réclusion au moins.

Art. 1. Celui qui, sachant qu'il les sera destinés à commettre des délits contre des personnes ou des propriétés, fabrique des matières explosives ou des débris destinés à commettre des délits contre des personnes ou des propriétés, sera puni de cinq ans de réclusion au moins.

Art. 2. Celui qui reçoit, détiens, transmet ou transporte des substances explosibles dans une autre intention que celle d'empêcher un crime ou un délit, sera puni de cinq ans de réclusion au moins.

Art. 3. Celui qui, sachant qu'il les sera destinés à commettre des délits contre des personnes ou des propriétés, fabrique des matières explosives ou des débris destinés à commettre des délits contre des personnes ou des propriétés, sera puni de cinq ans de réclusion au moins.

Art. 4. Celui qui, sachant qu'il les sera destinés à commettre des délits contre des personnes ou des propriétés, fabrique des matières explosives ou des débris destinés à commettre des délits contre des personnes ou des propriétés, sera puni de cinq ans de réclusion au moins.

Art. 5. Si les actes mentionnés à l'article 4 sont commis par la voie de la presse ou par des moyens analogues, tous ceux qui ont coopéré au délit (auteurs, médiateurs, complices, fauteurs) sont punissables en vertu des articles 69 à 72 du code pénal fédéral du 4 février 1853 ne leur sont pas applicables. Le fauteur peut s'être puni que de l'amende.

Art. 6. Les délits énumérés aux articles 1 à 4 sont punis conformément aux dispositions de la présente loi s'ils ont été commis même à l'étranger contre la Confédération ou ses ressortissants.

Art. 7. En cas de connexité d'un des délits mentionnés aux articles 1 à 4 avec des actes prévus et punis par le droit pénal cantonal, les tribunaux cantonaux ont à connaître de ces délits et ont le droit de grâces cantonal qui leur est applicable.

Art. 8. Clause référendaire.

Décision du conseil national. 6 avril 1894.

### Discussion

Ständerat. — Conseil des États. Discussion. — 1<sup>re</sup> séance du 18 avril 1894, à 3 heures de relevé.

Vizepräsident: Messager. Président: Messager.

Tagesordnung: — Ordre du jour: Bundesgesetz betreffend Ergänzung des Bundesgesetzes über das Bundesrecht der Schweiz. Eidgenossenschaft vom 4. Februar 1851.

Le loi fédérale complétant le code pénal de 1853.

Differenzen. — Divergences. (Sébe Seite 405 oben. — Voir page 405 ci-dessus.)

Art. 2. Celui qui reçoit, détiens, transmet ou transporte des substances explosibles dans une autre intention que celle d'empêcher un crime ou un délit, sera puni de cinq ans de réclusion au moins.

Art. 3. Celui qui, sachant qu'il les sera destinés à commettre des délits contre des personnes ou des propriétés, fabrique des matières explosives ou des débris destinés à commettre des délits contre des personnes ou des propriétés, sera puni de cinq ans de réclusion au moins.

Diese Druckschrift liegt nun aber einseitig der kantonalen Gesetzgebungen zugrunde und sie ist auch im Bundesrecht nicht acceptiert. Das Bundesrecht enthält nämlich, nach Aufnahme von zwei Artikeln, auch allgemein von Verbrechen) und vertritt darüber das ganze Gebiet des strafbaren Verbrechens, mit Ausnahme des Bundesrechts, der blosse Polizeiverwaltung. Wir beantragen also Fachräten ein Beschliessen des Ständerates, welcher mit der Terminologie des übrigen Bundesrechts übereinstimmt.

Die zweite Änderung, welche vom Nationalrat vorgebracht wird, ist mehr rechtliche Natur. Der Nationalrat glaubt, wenn man blos sagt: «Wer Sprengstoffe, von denen er anzunehmen muss, dass sie zu Verbrechen gegen die Sicherheit von Personen oder Sachen gebraucht werden sollen, herstellt oder zu deren Herstellung Anleitung gibt, wird mit dem zweifachen Tode» zu deren Herstellung Anleitung gibt) nicht genügend ausgesprochen, dass es sich um Sprengstoffe handle, von denen der Betroffene anzunehmen muss, dass sie verheerend zu gebraucht werden sollen. Er möchte darin sagen: «oder zur Herstellung solcher Sprengmittel Anleitung gibt)». Hier ist die Kommission einverstanden, dem Nationalrat zustimmen, obschon sie nicht zu sehen kann, dass die Redaktion dadurch bedeutend verbessert wird.

Was nun den französischen Text betrifft, so schlägt der Nationalrat also etwas andere Fassung vor, als der Ständerat, und ihre Kommission würde dieselbe acceptieren, jedoch mit einer Abänderung. Sie schlägt vor, statt «qui doit admettre» zu sagen «qui doit prouver» und der andere Ausdruck auch anders verstanden werden kann, «convaincu est». Man kann ihn auch verstehen, wie etwa «qui doit avouer», also wenn er zugeben muss, die Sprengstoffe in verheerender Absicht herzustellen zu haben, und mit einem Zweifelsmoment zu vermeiden, schlägt Ihre Kommission vor, den Text des Nationalrats mit der Abänderung zu acceptieren, dass statt «admettre» gesagt werde «prouver».

Die Legation föderale munit offre cette participation. Je reconnais qu'il est difficile de trouver un mot qui à cet égard, nous échapperait les dispositions de la loi fédérale sur le code pénal fédéral de 1853. Mais il n'est pas possible de trouver un mot qui à cet égard, nous échapperait les dispositions de la loi fédérale sur le code pénal fédéral de 1853. Mais il n'est pas possible de trouver un mot qui à cet égard, nous échapperait les dispositions de la loi fédérale sur le code pénal fédéral de 1853.

M. Michard: Je n'ai pas la prétention d'appuyer le texte allemand du projet de loi qui vous est soumis, mais en ce qui concerne la rédaction française, je désire expliquer la différence qui nous sépare de ce qui est proposé par le conseil national. Il s'agit pas question ici de discussion académique, pas plus que de système grammatical; c'est surtout une question de droit proprement dit. L'adjectif que l'on a suggéré ou doit employer les locutions de la session, celles qui sont employées par les auteurs et sont généralement admises. Or en français, le mot «admettre» embrasse l'ensemble des faits; on ne peut pas exclure de la part de celui qui est poursuivi qu'il se livre lui-même, ni quelle que condition à sa condamnation son propre aveu.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

Quelle est l'idée mise des articles 2 et 3? Ces deux articles sont destinés à commettre des délits. Or il est clair que vous ne pouvez pas subordonner la commission des délits à la commission des délits. Mais si vous ne pouvez pas subordonner la commission des délits à la commission des délits, vous ne pouvez pas subordonner la commission des délits à la commission des délits.

# Layout analysis and annotation (T1.2&1.3)

- There are different layouts, but normally comprised in single proceedings.
  - Possible sections: laws, votes, speeches, etc.
- ↻ Analysis of the document at the level of the textboxes (paragraphs, headers, etc.) aiming at identifying different types of texts:
  - Later this will help to cluster documents according to their sections.
  - ✓ Implemented for extracting speeches, as we required them for WP2.

# Structured content extraction (T1.4)

- ✓ We already extracted:
  - All speeches with corresponding speakers.
  - Enriched metadata file.
  - List of politicians and councils presidents.
- This information can be readily integrated into a MongoDB database:
  - One entry per document with previously identified elements as attributes.
  - Associated to a relational database with the list of politicians, their canton, party, etc.

# Preprocessing for WP2

- ✓ For topic modeling on the extracted speeches:
  - Language identification
  - (Thus far only) for German speeches
    - Tokenization
    - POS-tagging
    - Lemmatization
    - All lower case
    - Stopword removal
  - A document are all speeches of one proceeding.
- ✓ For knowledge graph (from extracted speeches):
  - Extract mentioned people (using NER to identify person first)

Der Hund jagt die Katzen

Der - Hund - jagt - die - Katzen

Der A - Hund N - jagt V - die A - Katzen N

Der - Hund - jagen - die - Katze

hund - jagen - katze

Hund - jagen - Katze

# Language and topic modelling (T2.1)

- ✓ Latent Dirichlet allocation (LDA) was applied to:
  - Two legislative periods (39th, 44th).
  - The complete time-span splitted in two sub-periods (1891 - 1970, 1971 - 1995).
- 44th legislative period (1991 - 1995)
  - 2799 documents.
  - ✓ Playing with parameters → good results with 100 topics and neglecting terms that appear in more than 10% of documents and in less than 20 documents.
  - ✓ 49 topics could be named.

# LDA for 44th legislative period (1991 - 1995)

asylum	
$\lambda = 1$	$\lambda = 0$
ausländer	asylbewerber
asylbewerber	asylgesetz
flüchtling	einbürgerung
koller	ausschaffungshaft
ausländerrecht	asylverfahren
zwangsmassnahmen	asylgesuch
einbürgerung	asylsuchender
asylgesetz	anag
standesinitiative	fremdenpolizei
asylpolitik	ausweisung

drugs	
$\lambda = 1$	$\lambda = 0$
droge	drogenpolitik
drogenpolitik	betäubungsmittel
übereinkommen	drogenkonsum
betäubungsmittelgesetz	süchtigen
illegal	heroinabgabe
lette	stupéfiantes
drogenproblem	drogenabhängige
heroin	drogenszen
drogenhandel	süchtig
prävention	drogensucht

$$\text{relevance}(w|t) = \lambda \cdot p(w|t) + (1 - \lambda) \cdot \frac{p(w|t)}{p(w)}$$

given topic name(s)	number
general terms + federal elections	1
French terms	2
parliamentary investigations	3
parliamentary investigations + taxes + federal budget	4
financial center <sup>4</sup> + law	5
federal budget + research	6
parliamentary votes	7
accession to EU	8
unemployment + occupational benefit plan	9
adaption to European law	10
railway	11
world trade	12
social and health insurance for farmers	13
NEAT	14
crime prevention	15
asylum	16
UN blue helmet troops	17
armament	18
agriculture	19
AHV revision	20
genetic engineering	21
public organizations + Italian terms	22
racism criminal norm <sup>5</sup>	23
Armee 95	24

# LDA for 44th legislative period (1991 - 1995)

- Five topics allocated to 1993/20022799
  - Financial center + law (44%)
  - Asylum (32%)
  - Racism criminal norm (13%)
  - Federal budget (10%)
  - Acquisition of fighter aircrafts (1%)

✓ They cover the content well!

- But, “aus der Luft gegriffen”

91.3290

**Postulat Seiler Hanspeter**  
**Unterkunft und Verpflegung**  
**von Asylbewerbern**  
**Demandeurs d'asile.**  
**Gîte et couvert**

*Diskussion – Discussion*

Siehe Jahrgang 1992, Seite 627 – Voir année 1992, page 627

Frau **Goll**: Ich möchte Sie bitten, dieses Postulat abzulehnen, und zwar aus folgenden Gründen:

Ihre Sprache, Herr Seiler, ist äusserst entlarvend. Der in Ihrem Postulat angeschlagene Ton zeugt von tiefster Verachtung gegenüber benachteiligten und diskriminierten Menschengruppen und schürt in erheblichem Masse Fremdenfeindlichkeit und Fremdenhass. Ich möchte mit Ihnen Ihren Text kurz näher anschauen:

Ich bin über die Äusserungen von Frau Goll sehr verwundert, da ich mit ihr noch nie ein Wort über solche Fragen diskutiert habe. **Sie hat irgend etwas aus der Luft gegriffen** und mich als asozialen Typ hingestellt. Ich bitte Sie: Kommen Sie einmal zu uns; dann schauen wir einmal gewisse Sachen an. Was ich

dass gerade Leute wie Sie mit solchen Äusserungen dazu beitragen, diesen Unmut anwachsen zu lassen.

Ihrer Meinung nach sollten sich die Fürsorgeleistungen explizit und ausgerechnet «für Asylbewerber», wie Sie betonen, «in vernünftigen Grenzen halten». Was heisst denn das? Sie sug-

# LDA for 39th legislative period (1971 - 1975)



Mostly similar topics but with a different focus:

- Development aid for Africa and Asia instead of Eastern Europe
- Drugs on alcoholism instead of hard drugs

alcoholism (drugs)	
$\lambda = 1$	$\lambda = 0$
bundesanwaltschaft	alkoholismus
bundesanwalt	droge
alkoholverwaltung	bundesanwalt
lehrer	straflos
strafe	händler
alkoholismus	bundesanwaltschaft
jugendliche	bestrafung
konsum	betäubungsmittel
bestrafen	betäubungsmittelgesetz
droge	strafandrohung

drugs	
$\lambda = 1$	$\lambda = 0$
droge	drogenpolitik
drogenpolitik	betäubungsmittel
übereinkommen	drogenkonsum
betäubungsmittelgesetz	süchtigen
illegal	heroinabgabe
lette	stupéfians
drogenproblem	drogenabhängige
heroin	drogenszen
drogenhandel	süchtig
prävention	drogensucht



Some different topics

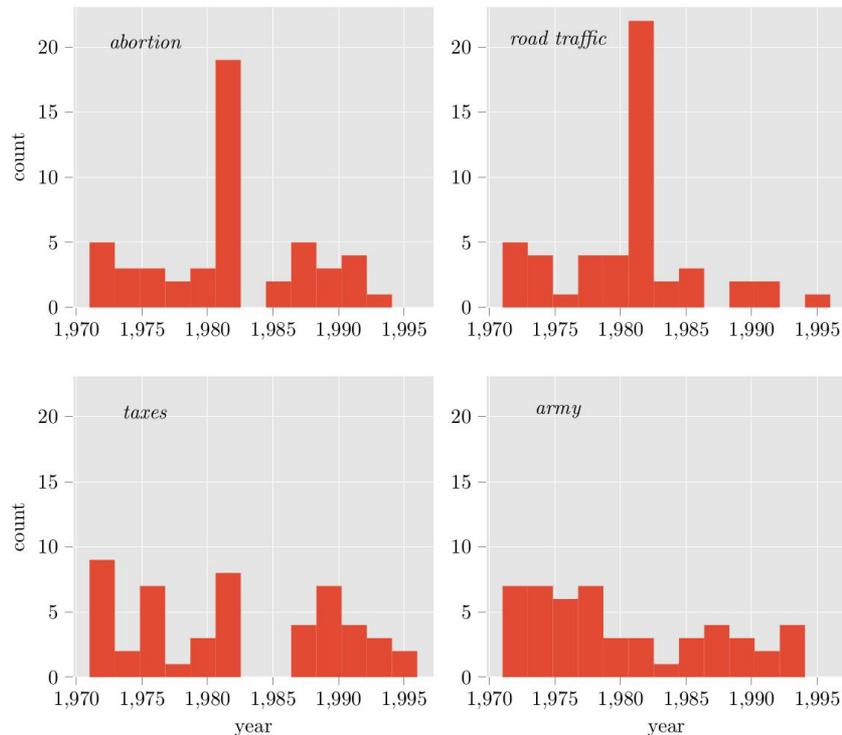
- 39th: electricity from nuclear energy
- 44th: UN blue helmet troops

# LDA for 1971 until 1995



## Dynamic evaluation

- By plotting years of top 50 documents for each topic as a histogram
- Abortion law in 1982
- Introduction of “Autobahn Vignette” in 1985



# Language and topic modelling (T2.1)

- ✓ Summary:
  - Obtained good overview of corpus.
  - LDA proved successful for topic identification.

- Future work:
  - Advanced dynamic topic modeling
  - Approaches to determine hierarchy of topics (general and sub-topics)
  - Embedding techniques for dealing with different languages
  - Revisit after identifying sections and types of documents on WP1
  - Enrich database with extracted topics
  - Other ideas?

# Open issues

- Named entity disambiguation (T2.2)
  - Are the following databases available?
    - Semantically enriched and disambiguated database of delegates of the Swiss Parliament from Prof. Simon Hug.
    - Structured database in PI's group for the disambiguation of committees, members of government, parties and fractions, cities and regions.
- Knowledge graph (T2.3)
  -  We are currently investigating on how to set this up
    - Following example of LegisGraph project
    - Using neo4j and py2neo

