

MiMoText in 6 Stationen

Station 2: Romankorpus

Collection de romans français du dix-huitième siècle (1750-1800)

MiMoText / roman18

<> Code Issues 1 Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 1 tag

Go to file

Add file

Code

roettger Update		82ba438 7 days ago	631 commits
Archiv	Add README.md for Archiv folder	4 months ago	
NER-with-SpaCy	Merge branch 'master' of https://github.com/MiMoText/roman18	5 months ago	
OCR4all_output	Corrected OCR output	2 months ago	
Python-Scripts	metadata update	20 days ago	
Schemas	added eltec schema folder	11 months ago	
XML-TEI	added new file	16 days ago	
obvil-romans	manual test conversion	5 months ago	
plain	renewed some txt-files	22 days ago	
sentiments	Update	7 days ago	
work-in-progress	Endversion for the hyphen-cases. Input for workxmls	12 months ago	
README.md	Updated script for plain text extraction: tei2txt.py instead of get_t...	19 days ago	
desktop.ini	Metadaten der 70 Romane	7 months ago	
xml-tei_metadata.tsv	metadata update	20 days ago	

README.md

DOI10.5281/zenodo.4061903

roman18

About

Collection de romans français du dix-huitième siècle (1750-1800) / Collection of Eighteenth-Century French Novels (1750-1800)

enlightenment novels french literature trier 1750-1800

Readme

Releases

1 tags

Create a new release

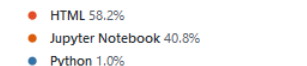
Packages

No packages published
Publish your first package

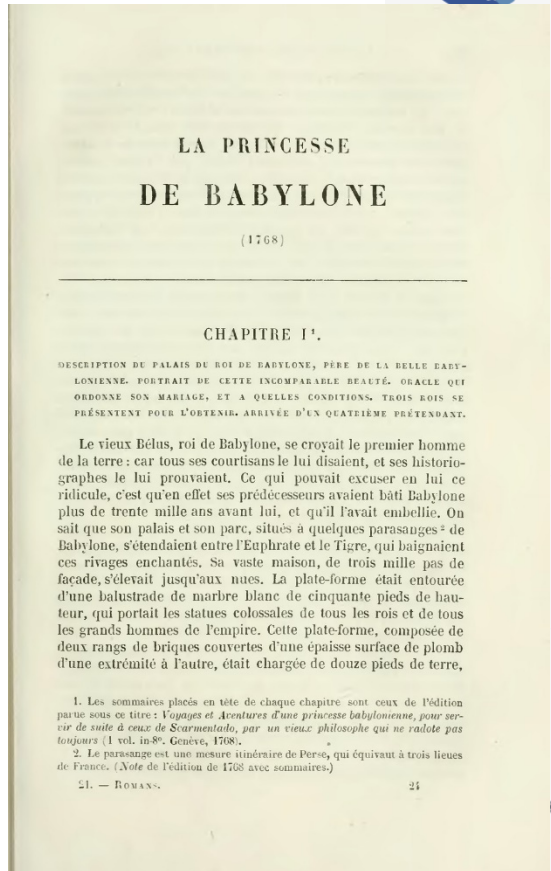
Contributors 7



Languages



TEI / XML



< > Page Discussion Fichier ^

Page:Voltaire - Œuvres complètes Garnier tome2

Cette page a été validée par deux contributeurs.

LA PRINCESSE DE BABYLONE

(1768)

CHAPITRE I^[1].

DESCRIPTION DU PALAIS DU ROI DE BABYLONE, PÈRE DE LA BELLE BABYLONIENNE. PORTRAIT DE CETTE INCOMPARABLE BEAUTÉ. ORACLE QUI ORDONNE SON MARIAGE, ET À QUELLES CONDITIONS. TROIS ROIS SE PRÉSENTENT POUR L'OBTENIR. ARRIVÉE D'UN QUATRIÈME PRÉTENDANT.

Le vieux Bélus, roi de Babylone, se croyait le premier homme de la terre : car tous ses courtisans le lui disaient, et ses historographes le lui prouvaient. Ce qui pouvait excuser en lui ce ridicule, c'est qu'en effet ses prédécesseurs avaient bâti Babylone plus de trente mille ans avant lui, et qu'il l'avait embellie. On sait que son palais et son parc, situés à quelques parasanges^[2] de Babylone, s'étendaient entre l'Euphrate et le Tigre, qui baignaient ces rivages enchantés. Sa vaste maison, de trois mille pas de façade, s'élevait jusqu'aux nues. La plate-forme était entourée d'une balustrade de marbre blanc de cinquante pieds de hauteur, qui portait les statues colossales de tous les rois et de tous les grands hommes de l'empire. Cette plate-forme, composée de deux rangs de briques couvertes d'une épaisse surface de plomb d'une extrémité à l'autre, était chargée de douze pieds de terre,

```
8 <title ref="bgrf:68.55">La Princesse de Babylone: MiMoText edition</title>
9 <author ref="viaf:36925746;wikidata:Q9068">François-Marie Arouet Voltaire(1694-1778)</author>
10 <respStmt>
11 <resp>data capture</resp>
12 <name>wikisource contributors</name>
13 </respStmt>
14 <respStmt>
15 <resp>encoding</resp>
16 <name>
17 Julia Dudar
18 </name>
19 </respStmt>
20 <respStmt>
21 <resp>editor</resp>
22 <name>
23 Julia Röttgermann
24 </name>
25 </respStmt>
26 </titleStmt>
27 <extent>
28 <measure unit="words">27224</measure>
29 </extent>
30 <publicationStmt>
31 <publisher ref="https://mimotext.uni-trier.de">
32 Mining and Modeling Text
33 </publisher>
34 <distributor ref="https://github.com/mimotext/roman-dixhuit">
35 Github
36 </distributor>
37 <date>
38 2020
39 </date>
40 <availability>
41 <licence target="https://creativecommons.org/publicdomain/zero/1.0/deed.en"/>
42 </availability>
43 </publicationStmt>
44 <sourceDesc>
45 <bibl type="digitalSource">
46 <ref target="https://fr.wikisource.org/wiki/La_Princesse_de_Babylone"/>
47 </bibl>
48 <bibl type="printSource">
49 <date>1877</date>
50 </bibl>
```

Topic Modeling - Workflow



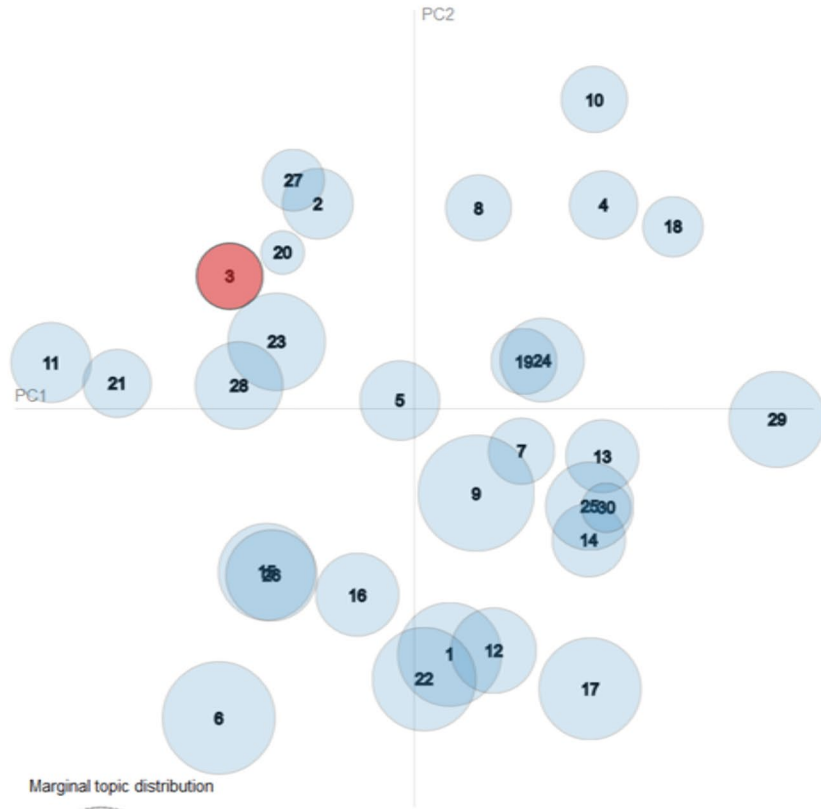
Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾

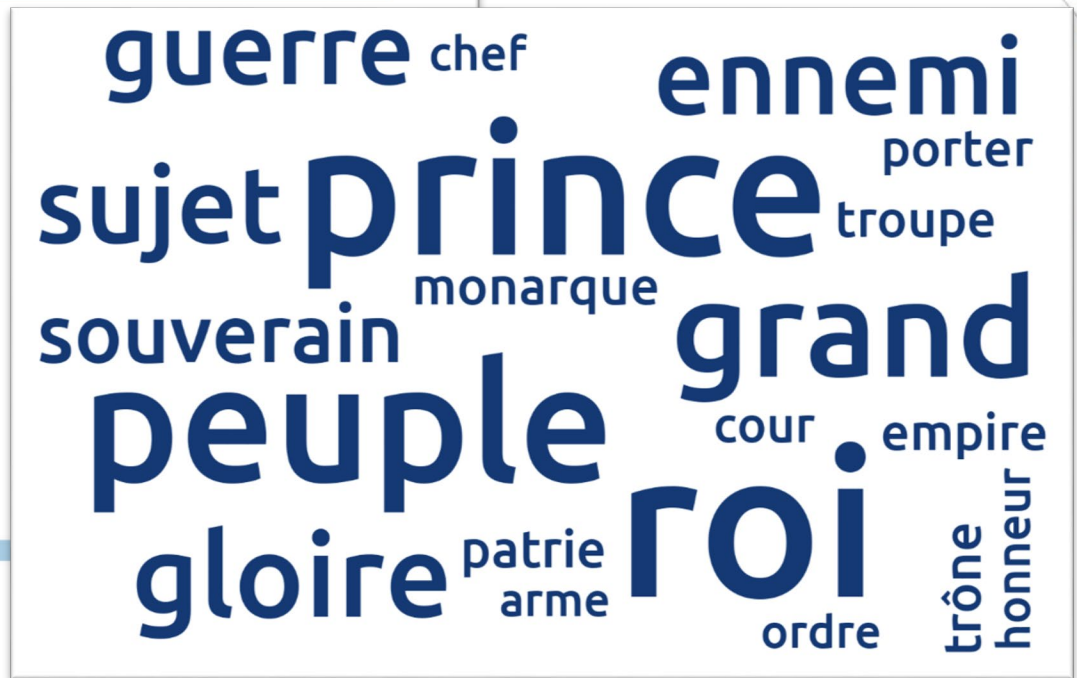
$\lambda = 1$

0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 3 (2.2% of tokens)



Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t)/p(w)$; see Sievert & Shirley (2014)



Beispiel „Voltaire: La Princesse de Babylone“

- Voltaire_Princesse IS_ABOUT "monarchie"
- Voltaire_Princesse IS_ABOUT "voyage"
- Voltaire_Princesse IS_ABOUT "sociabilité"
- Voltaire_Princesse IS_ABOUT "nature"
- Voltaire_Princesse IS_ABOUT "nuit"