

Research

# WAHSP

**Towards a flexible and stable  
CLARIN-supported *open-source*  
web-application for historical  
sentiment mining in public media**



KB

Huygens Instituut



Universiteit Utrecht



DESCARTES CENTRE

UNIVERSITEIT VAN AMSTERDAM

## PROBLEMS IN HISTORICAL RESEARCH

History as science is based on a critical investigation of sources.

The validity of conclusions is strengthened by increased representativeness of the data.



## **PROBLEMS IN HISTORICAL RESEARCH**

Available data too extended and/or too scattered for comprehensive study and analysis.

If indices on data are available, the researcher is dependent on the author of these indices.

They are almost always incomplete and, over larger periods of times, inconsistent.



Research

## KB Krantendatabase

### **Databank Digitale Dagbladen**

Journals in the collection of the Royal Library (KB) in The Hague (Krantenmagazijn Koninklijke Bibliotheek)

The project Databank Digitale Dagbladen digitalizes on a large scale Dutch national, regional, local and colonial newspapers and make these online and free available.

### **Online available**

Since 28 May 2010 first results on the webservice

*Historische kranten*

<http://kranten.kb.nl>

*Eight* million newspaper pages 1618-1945.

Will be augmented in stages, up until 1995.



Universiteit Utrecht

August 30, 2011

Research

## **Problem:**

Context and background of Dutch drug politics in time

## **Aim**

Understanding and evaluation of public sentiments around drugs, drug use and addiction in the Netherlands, 1900-1945

## **Research questions**

Dynamics of public opinion, related to pillarization

Sentiments and attitudes around addiction, patients, doctors and treatment

Sentiments and attitudes around illegal and legal trade, public policies

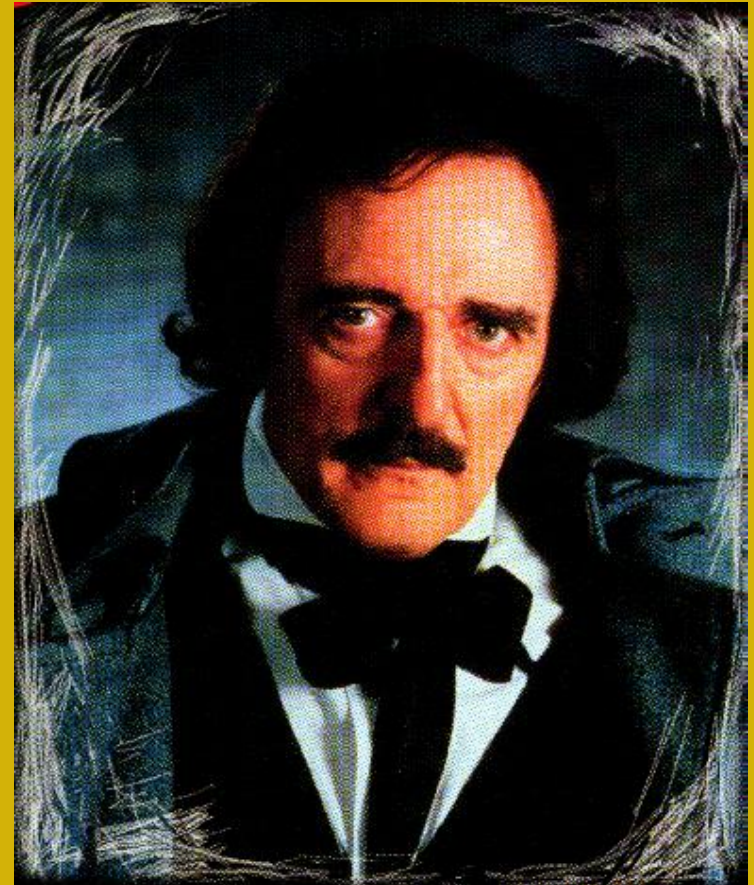


Research

## Edgar Allan Poe's detective

'Experience has shown,  
and a true philosophy  
will always show, that a  
vast, perhaps the larger  
portion of truth arises  
from the seemingly  
irrelevant.'

*(The Mysterie of Marie  
Roget, 1842)*

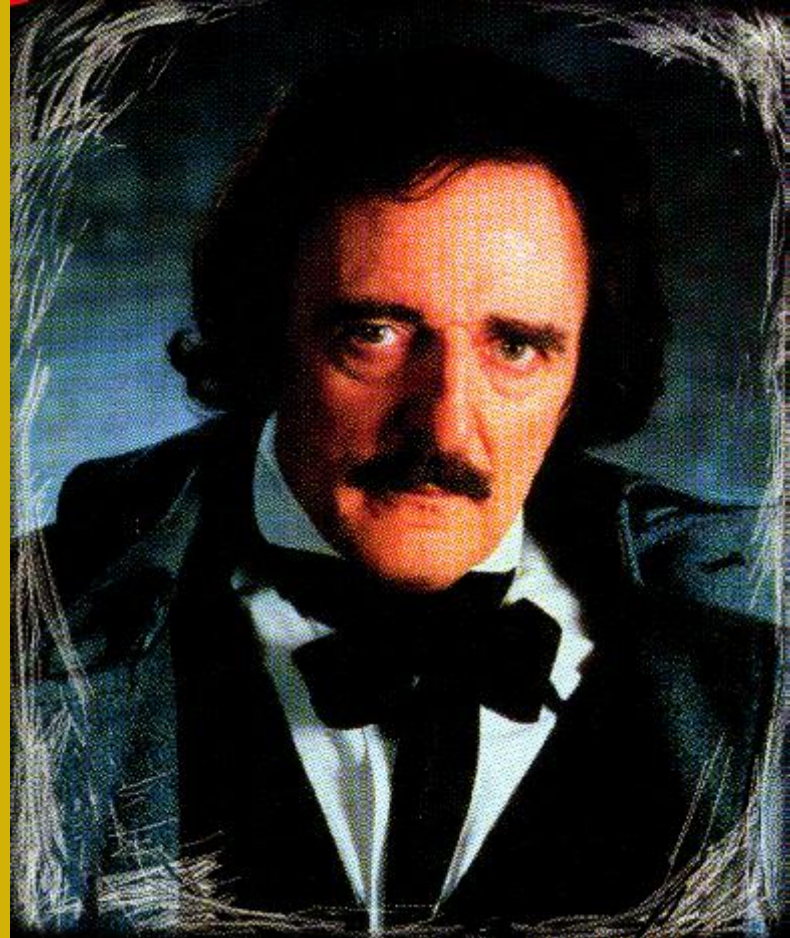


Universiteit Utrecht

## Research

Poe's detective finds the truth by using data in those newspaper articles that *do not* concern the murder.

In a similar way we will find sentiments in those newspaper articles that are at first sight irrelevant.



## What does the user of WAHSP want?

- Auxiliary tool for observing and analyzing trends and patterns
- Interactive tool with the possibility to adapt the original lexicon
- Possibility of production and analysis of subsets of data
- Identification of individual key documents





## **WE NEED:**

A semi-automatic and interactive *open-source* application that extracts relevant data from a mass of seeming irrelevance.

An application that does not replace, but supports the intuition and insights of the researcher.

An application that is user-friendly.





**Zoekopdracht**  
t met  
**Zoekwoorden**  
n en  
**Synoniemen**



**Woordenwo**  
**lken**  
**Sentiment**

## WAHSP



## Zoeken

Zoek

Filter

Found 9276 records.

[volgende](#)1. [INTERNAT. BENDE VAN OPIUMSMOKKELAARS ONTMASK...](#)

Het Vaderland : staat- en letterkundig nieuwsblad  
1934/11/03, Avond.

2. [Het Opium-vraagstuk en de Internationale Op...](#)

Nieuwe Rotterdamsche Courant  
1912/01/30, Ochtend.

3. [OPIUM-HISTORIES](#)

Het Vaderland : staat- en letterkundig nieuwsblad  
1924/08/19, Avond.

4. [De-Opium in haar fabricatie en verbruik.](#)

Het Centrum  
1930/02/22, Dag.

5. [Oeconomische en Financieels Mededeelingen.](#)

Nieuwe Rotterdamsche Courant  
1912/08/01, Ochtend.

6. [Het Opium-vraagstuk en de Interantionale Opi...](#)

Nieuwe Rotterdamsche Courant  
1912/01/27, Avond.

7. [Heropening der vergadering te Genève. Engels...](#)

Het Vaderland : staat- en letterkundig nieuwsblad  
1925/01/20, Ochtend.

8. [DE BEDRIEGER BEDROGEN... Een ambtenaar bij d...](#)

Het Vaderland : staat- en letterkundig nieuwsblad

Selecteer een artikel.

Tekst

Lexicon

## WAHSP



## Zoeken

((cql.serverChoice ex

Zoek

Filter

1900 1905 1910 1915 1920 1925 1930 1935 1940 1945

opium

term

+

Found 9276 records.

[volgende](#)**1. [INTERNAT. BENDE VAN OPIUMSMOKKELAARS ONTMASK...](#)**

Het Vaderland : staat- en letterkundig nieuwsblad  
1934/11/03, Avond.

**2. [Het Opium-vraagstuk en de Internationale Op...](#)**

Nieuwe Rotterdamsche Courant  
1912/01/30, Ochtend.

**3. [OPIUM-HISTORIES](#)**

Het Vaderland : staat- en letterkundig nieuwsblad  
1924/08/19, Avond.

**4. [De-Opium in haar fabricatie en verbruik,](#)**

Het Centrum  
1930/02/22, Dag.

**5. [Oeconomische en Financieels Mededeelingen,](#)**

Nieuwe Rotterdamsche Courant  
1912/08/01, Ochtend.

**6. [Het Opium-vraagstuk en de Interantionale Opi...](#)**

Nieuwe Rotterdamsche Courant  
1912/01/27, Avond.

**7. [Heropening der vergadering te Genève. Engels...](#)**

Het Vaderland : staat- en letterkundig nieuwshlad

Selecteer een artikel.

Tekst

Lexicon

## WAHSP



Zoeken

Lexicon

Hier kun je de lexicon opbouwen. Ieder onderdeel in het lexicon heeft een titel en kan worden gemaakt op basis van een zoekopdracht.

synthetica, 2011-07-20 15:39:25



Morphinisme, 2011-07-18 17:19:05



cocaine, 2011-07-13 15:00:50



Chinezen, 2011-06-30 16:11:03



narcotica2, 2011-05-18 21:31:48



Opium, 2011-05-18 13:58:09



Test Complex, 2011-05-05 10:46:42



Test, 2011-05-05 10:44:39



Selecteer een artikel.

Tekst

## Zoeken

((cql.serverChoice ex

Zoek

Filter

1900 1905 1910 1915 1920 1925 1930 1935 1940 1945

Found 1632 records.

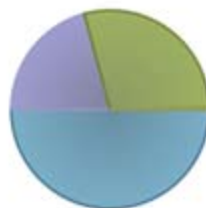
[volgende](#)

1. [Een vervangmiddel voor Cocaine.](#)  
Het nieuws van den dag voor Nederlandsch-Indië  
1925/11/30, Dag. Nederlands-Indië / Indonesië
2. [Cocaine en haar vervangmiddeleu.](#)  
Nieuwe Rotterdamsche Courant  
1929/05/09, Ochtend.
3. [Cocaine Uit de geschiedenis der Locaalanesth...](#)  
Het Vaderland : staat- en letterkundig nieuwsblad  
1927/11/24, Avond.
4. [COCAINE-SMOKKEL](#)  
Limburger koerier : provinciaal dagblad  
1937/02/16, Dag. Regionaal/mokaal
5. [Het moderne geneesmiddel.](#)  
De Sumatra post  
1932/08/18, Dag. Nederlands-Indië / Indonesië
6. [COCAINE-SMOKKEL](#)  
Limburger koerier : provinciaal dagblad  
1937/02/16, Dag. Regionaal/mokaal
7. [Het moderne geneesmiddel.](#)  
De Sumatra post  
1932/08/18, Dag. Nederlands-Indië / Indonesië
8. [Zeker en Waarschijnlijk.](#)  
Het nieuws van den dag voor Nederlandsch-Indië

Lexicon

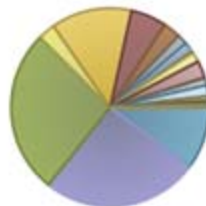
### edition

Dag: 793  
Ochtend: 331  
Avond: 461  
Middag: 3



### papertitle

Het nieuws van den dag voor Nederlandsch-Indië: 162  
Nieuwe Rotterdamsche Courant: 405  
Het Vaderland : staat- en letterkundig nieuwsblad: 426  
Limburger koerier : provinciaal dagblad: 44  
De Sumatra post: 209  
Het Centrum: 93  
Nieuwe Tilburgsche Courant: 45  
Tilburgsche courant: 27



Tekst

tijd alleen grote zullen goed politie kwam werden opium plaats komen onder all  
hadden cocaine geheel weer vrouw zoo komt ver jaar eerste slechts in twee echte  
maken heel heer de bet land liet waar eenige regeering gaan dien welke moeten weg leven wor  
De

Generated by [Pestis](#) on 2011-08-19 15:31:26.245555 | completed | Processed 1413 documents out of 1413 (errors: 0) | [permalink](#)

## Zoeken

Zoek Filter

1900 1905 1910 1915 1920 1925 1930 1935 1940 1945

Found 1632 records. [volgende](#)

- [Een vervangmiddel voor Cocaine.](#)  
Het nieuws van den dag voor Nederlandsch-Indië  
1925/11/30, Dag. Nederlands-Indië / Indonesië
- [Cocaine en haar vervangmiddeleu.](#)  
Nieuwe Rotterdamsche Courant  
1929/05/09, Ochtend.
- [Cocaine Uit de geschiedenis der Locaalanesth...](#)  
Het Vaderland : staat- en letterkundig nieuwsblad  
1927/11/24, Avond.
- [COCAINE-SMOKKEL](#)  
Limburger koerier : provinciaal dagblad  
1937/02/16, Dag. Regionaal/mokaal
- [Het moderne geneesmiddel.](#)  
De Sumatra post  
1932/08/18, Dag. Nederlands-Indië / Indonesië
- [COCAINE-SMOKKEL](#)  
Limburger koerier : provinciaal dagblad  
1937/02/16, Dag. Regionaal/mokaal
- [Het moderne geneesmiddel.](#)  
De Sumatra post  
1932/08/18, Dag. Nederlands-Indië / Indonesië
- [Zeker en Waarschijnlijk.](#)  
Het nieuws van den dag voor Nederlandsch-Indië

Lexicon

### Een vervangmiddel voor Cocaine.

Een van de meest schadelijke stoffen, die naast alcohol voor verdooving gebruikt worden, is wel de cocaine en men heeft dan ook den handel in cocaine aan banden gelegd zonder er echter met veel succes in te slagen om de cocaine als verslavingsmiddel uit onze samenleving te verdryven. In Nederland wordt nog weinig cocaine „gesnoven“, in Parys en Berlijn echter des te meer.

Gemakkelijker zou de stryd tegen doze stof zyn, als men de fabricage ervan geheel kon verbieden, maar tot heden bleek dit nog niet mogelyk, omdat de cocaine als pynstillend middel nog heel wat in de geneeskunde wordt aangewend. De medici gebruiken het middel op twee manieren.

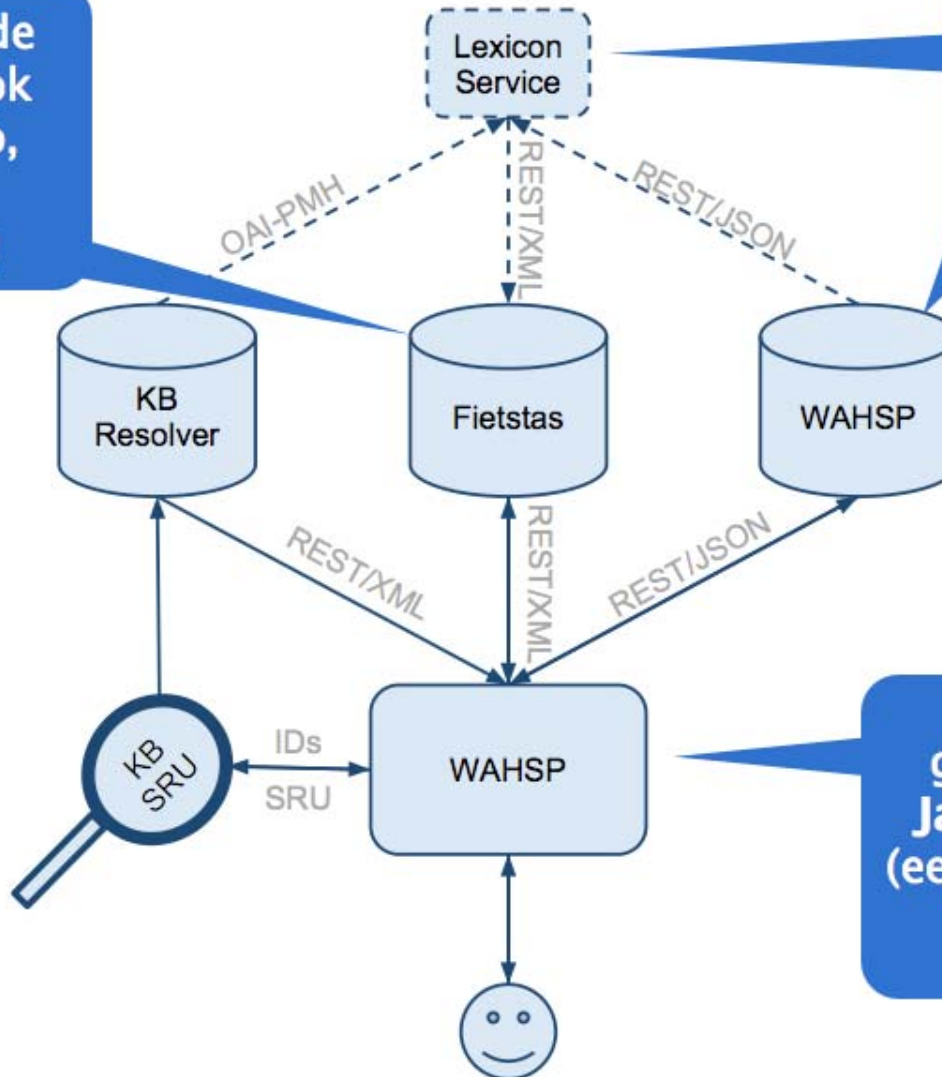
Zy spuiten de stof onderhuids in en bereiken zodoende de zenuwen direct, of wel zy besmeren er slymvliezen van oog, keel, neus, strottenhoofd of blaas mede, om deze ongevoelig te maken. Voor inspuitingen onder de huid heeft men sinds geruimen tyd vervangpreparatengevonden, die by evengoede anaesthetische werking minder giftig zijn en waaraan men zich bovendien niet kan verslaven; zij geven niet als cocaine aanleiding tot aangename gevoelens.

Een der voornaamste is het novocaine; in ons land bv. wordt voor onderhuidsche gevoelloosmaking byna uitsluitend deze stof gebruikt.

Tekst  
Origineel  
Bekijk bij KB

# Infrastructuur

Webservice van de UvA, gebruikt ook **Python, Django, Celery** en voor opslag



**Django**, aangevuld met **Celery**, waarmee taken over meerdere servers verdeeld kunnen

Web applicatie gemaakt met **HTML, JavaScript en Django** (een raamwerk voor web applicaties in de