The DECRYPT Project

Munich, November 1st 2019

Dr. Nils Kopal

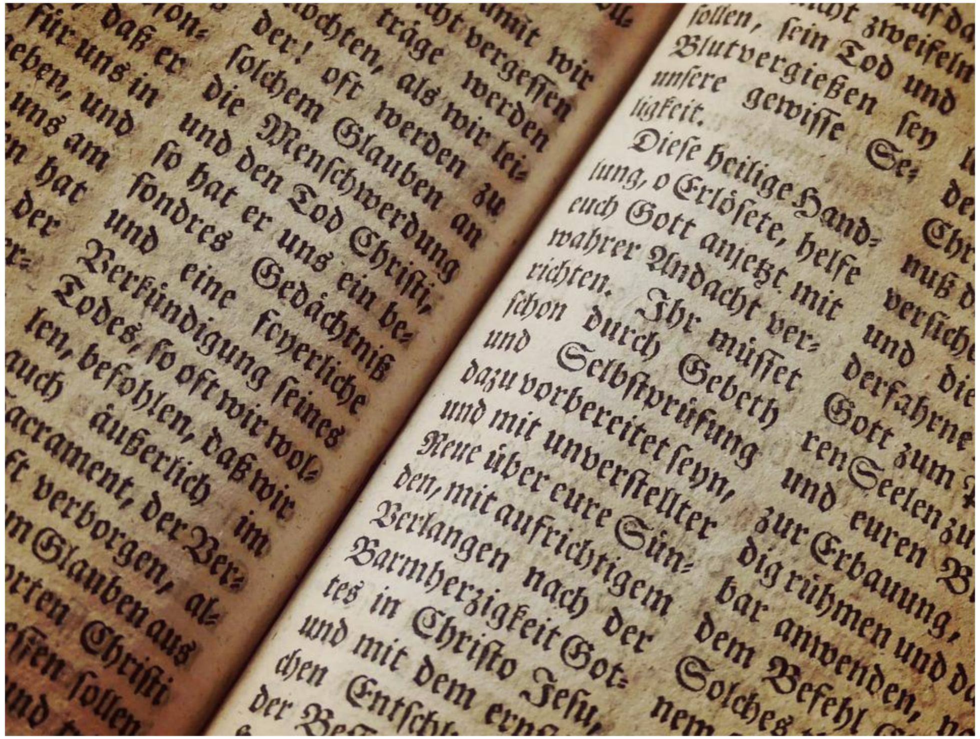
CrypTool 2 Project / DECRYPT Project

kopal@cryptool.org

CrypTool Meeting 20+ Years

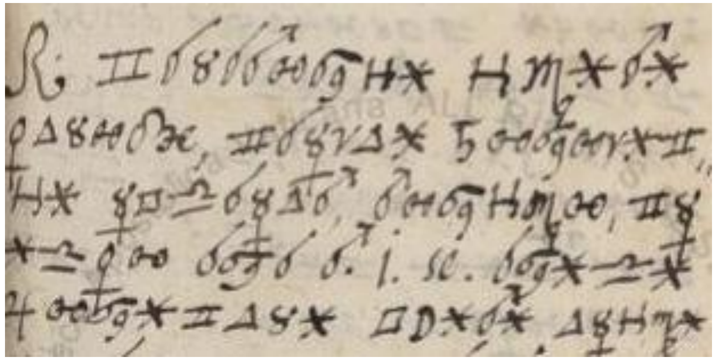
Content

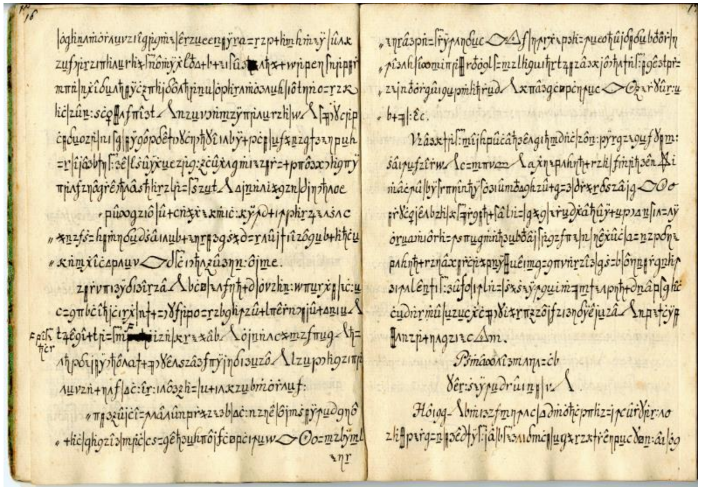
1. Introduction
2. Historical Cryptology
3. The DECRYPT Project
4. Resources & Tools
5. Some Examplary Results of the Project
6. Conclusion



1. Introduction

- **Thousands of encrypted manuscripts can be found in archives**
  - Diplomatic/military correspondence
  - Intelligence reports
  - Scientific writings
  - Private letters and diaries
  - Manuscripts related to secret societies and magic
- **Mostly not yet available for historical research**
- **Many researchers are working on these, but**
  - **uncoordinated**
  - in **various scientific areas**: history, linguistics, philology, computer science, and computational linguistics





1. Introduction

- **Not much is known in detail about encrypted documents throughout the centuries**
- **Large-scale studies were not possible due lack of**
  1. **infrastructural** resources
  2. **tools** for historical cryptology
- **Available tools are mostly unsuitable to these documents, because of**
  1. (old) **hand-writing**
  2. **mistakes/errors**
  3. **non-standardized languages**
  4. **no digitalization** available (**transcriptions**)
  5. often **mixtures** between **cleartext/ciphertext** and **different languages**

1. Introduction

- **Planned solution: DECRYPT project**
- **Project goals:**

1. Build **infrastructural support** for historical cryptology
2. **Collect** various documents throughout the ages, digitize them, **store** them
3. Release **resources and tools** to **digitize**, **process** and **decrypt** historical encrypted sources

2. Historical Cryptology

- „Historical cryptology is the study of encrypted messages from our history aiming at their decryption and contextualization“
- Dimensions

  - **Mathematical/computerized** (cryptanalysis)
  - **Linguistic** and other coding pattern
  - **Historical** context

3. The DECRYPT Project

- **“DECRYPT focuses on the refinement and development of the tools involved in the automatic processing of encrypted historical sources.”**
- **Supported by the Swedish Research Council, grant 2018-06074**
  - Funding for **2+4 years, started in 2019**
  - **Interdisciplinary** (linguists, historians, philologists, computer scientists, and cryptanalysts)
  - Involved universities from Sweden, Hungary, Spain, Germany (see next slide for details)
- **3 steps involved in our project:**

1. **Data collection** and **digitization**
2. **Analysis** needed prior to decryption
3. **Decryption** and **cryptanalysis**

3. The DECRYPT Project

Project Members

- **Uppsala University, Sweden** **(language analysis)**
  - Beata Megyesi (project leader)
- **University of Gothenburg, Sweden** **(language analysis)**
  - Michelle Waldispühl
- **Computer Vision Center, Universitat Autònoma de Barcelona, Spain** **(image processing)**
  - Alicia Fornés
- **Budapesti Müszaki és Gazdaságtudományi Egyetem, Hungary** **(history)**
  - Benedek Láng
- **University of Siegen, Germany** **(cryptanalysis)**
  - Bernhard Esslinger
- **Universität der Bundeswehr München, Germany** **(cryptanalysis)**
  - Arno Wacker

3. The DECRYPT Project

Step 1: Data collection and digitization

- **Archive work**
  - Search for **crypto-related documents**
  - Scan/photograph
  - Cataloge (in **DECODE** database)

- **Researchers: historians**

3. The DECRYPT Project

Step 2: Analysis needed prior to decryption

• **Determine “scenario”**	**(→ attack type):**
1. Found ciphertext	(**Ciphertext-only**)
2. Found ciphertext and plaintext	(**Known-plaintext**)
3. Found ciphertext and key	(**Decryption**)
4. Found plaintext and key	(**Encryption**)

- **Researchers: cryptanalysts**

3. The DECRYPT Project

Step 3: Decryption including cryptanalysis

- **Part 1: (transcription)**
  - Digitization and pre-processing of the historical source resulting in **images**
  - (Semi-)automatic **transcription** of images
  - **Researchers:** image processing experts
- **Part 2a: (analysis)**
  - (Historic) **language models**
  - **Researchers**: computer linguistics, philologists
- **Part 2b: (cryptanalysis)**
  - Breaking of ciphers using e.g. **heuristics**
  - **Researchers:** cryptanalysts
- **Part 3: (historical analysis of plaintexts)**
  - Analysis of new **(historical) findings** concerning plaintexts and methods
  - **Researchers:** historians

4. Resources and Tools

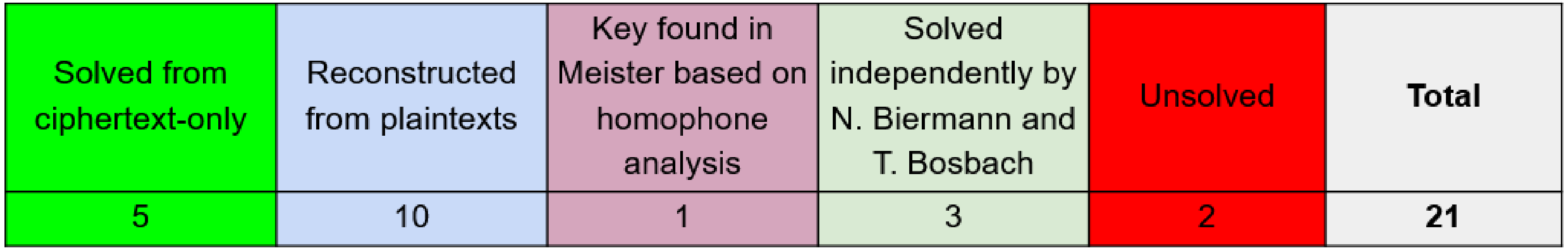
1. **DECODE database**
   - Collection of **ciphertexts, keys, etc.**
2. **Historical corpora (HistCorp) and language models**
   - Collection of **historical original text corpora** in 14 European languages
3. **Tools**
   - **Web service**: for transcription and „easy“ parts of cryptanalysis
   - **CrypTool 2**: tool for supporting „difficult“ parts of cryptanalysis
   - **Console applications** as prototypes → will be migrated with „nice“ UI into CrypTool 2

5. Some Exemplary Results of the Project

- **Huge collection of historic documents in the DECODE database**
  - **~1.000 records**
  - **33% are original keys**
  - **634 cipher texts (205 decrypted, 232 transcribed)**
  - Many “one pagers”; longest document 410 pages

5. Some Exemplary Results of the Project

- Decipherment of collections of historic Vatican ciphers

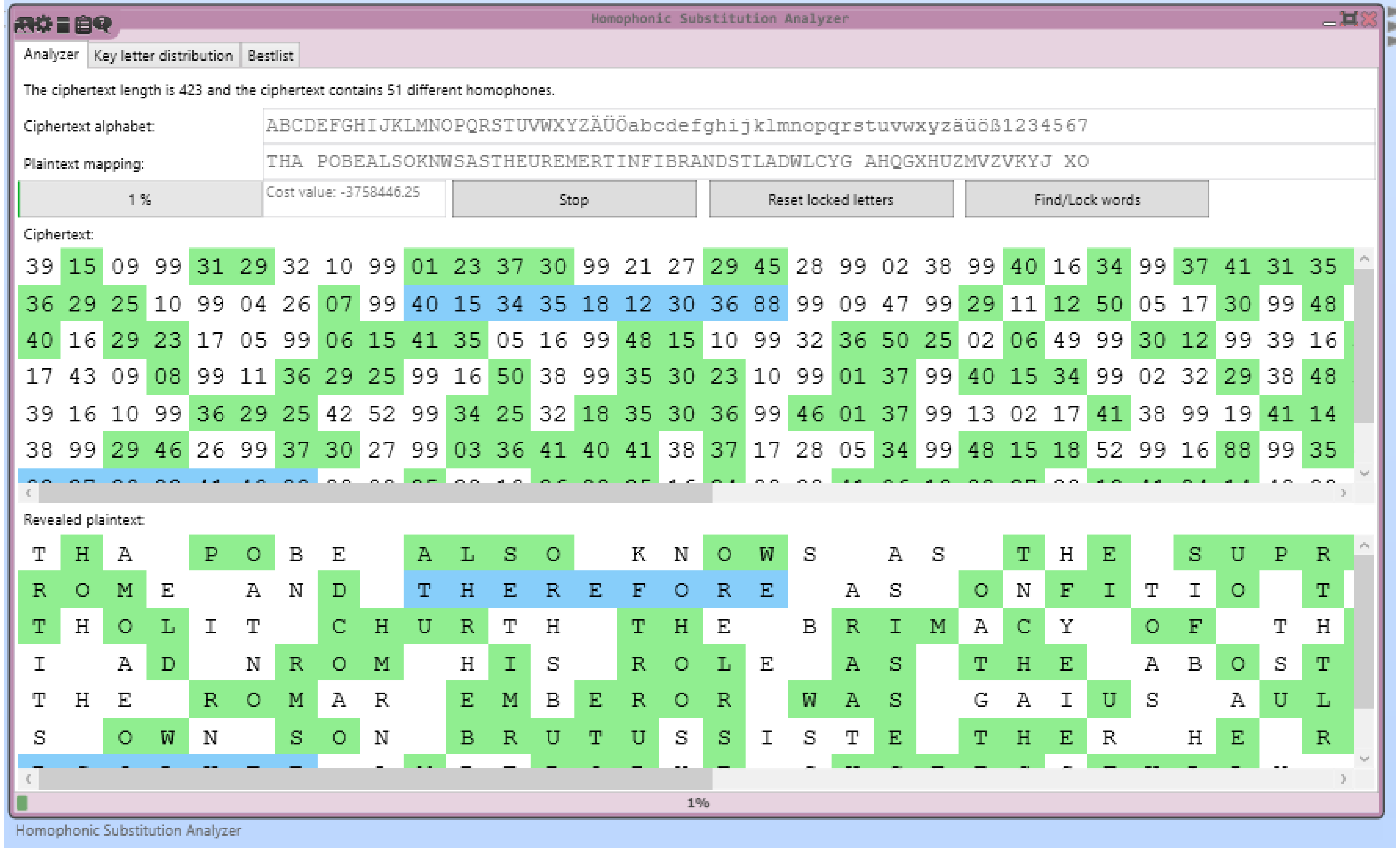


Solved from ciphertext-only	Reconstructed from plaintexts	Key found in Meister based on homophone analysis	Solved independently by N. Biermann and T. Bosbach	Unsolved	**Total**
5	10	1	3	2	**21**

- 19 of 21 collections are solved
- 2 are unsolved
- A huge set of (console) tools developed by Lasry (and currently ported to CT2)

5. Some Exemplary Results of the Project

Tools in CrypTool 2 – Homophonic Substitution Analyzer



The ciphertext length is 423 and the ciphertext contains 51 different homophones.

Ciphertext alphabet: ABCDEFGHIJKLMNOPQRSTUVWXYZÄÜÖabcdefghijklmnopqrstuvwxyzäüöß1234567

Plaintext mapping: THA POBEALSOKNWSASTHEUREMERTINFIBRANDSTLADWLCYG AHQGXHUZMVZVKYJ XO

1 %

Cost value: -3758446.25

Ciphertext:

```
39 15 09 99 31 29 32 10 99 01 23 37 30 99 21 27 29 45 28 99 02 38 99 40 16 34 99 37 41 31 35
36 29 25 10 99 04 26 07 99 40 15 34 35 18 12 30 36 88 99 09 47 99 29 11 12 50 05 17 30 99 48
40 16 29 23 17 05 99 06 15 41 35 05 16 99 48 15 10 99 32 36 50 25 02 06 49 99 30 12 99 39 16
17 43 09 08 99 11 36 29 25 99 16 50 38 99 35 30 23 10 99 01 37 99 40 15 34 99 02 32 29 38 48
39 16 10 99 36 29 25 42 52 99 34 25 32 18 35 30 36 99 46 01 37 99 13 02 17 41 38 99 19 41 14
38 99 29 46 26 99 37 30 27 99 03 36 41 40 41 38 37 17 28 05 34 99 48 15 18 52 99 16 88 99 35
```

Revealed plaintext:

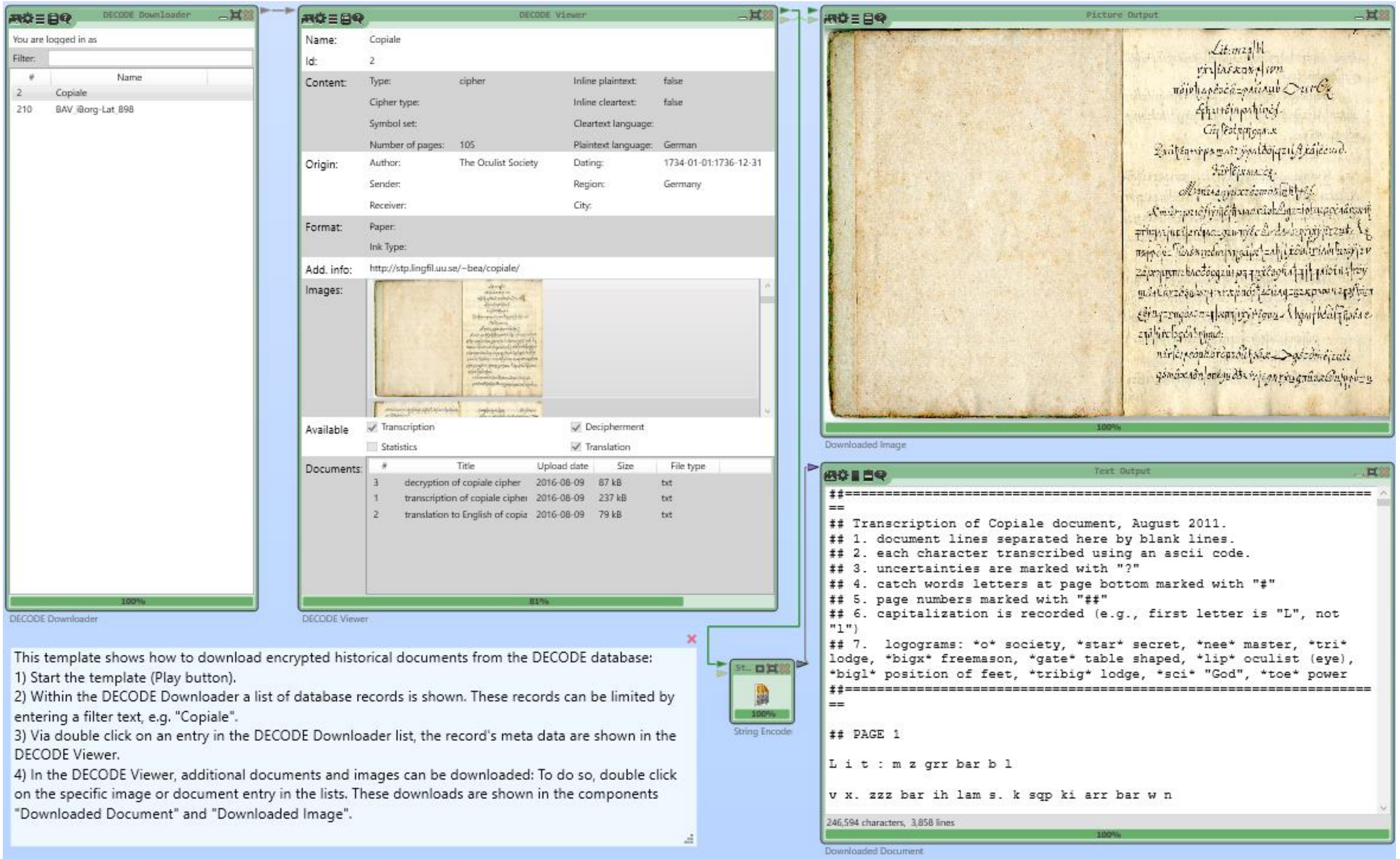
```
T  H  A     P  O  B  E     A  L  S  O     K  N  O  W  S     A  S     T  H  E     S  U  P  R
R  O  M  E     A  N  D     T  H  E  R  E  F  O  R  E     A  S     O  N  F  I  T  I  O     T
T  H  O  L  I  T     C  H  U  R  T  H     T  H  E     B  R  I  M  A  C  Y     O  F     T  H
I     A  D     N  R  O  M     H  I  S     R  O  L  E     A  S     T  H  E     A  B  O  S  T
T  H  E     R  O  M  A  R     E  M  B  E  R  O  R     W  A  S     G  A  I  U  S     A  U  L
S     O  W  N     S  O  N     B  R  U  T  U  S  S  I  S  T  E     T  H  E  R     H  E     R
```

1%

Homophonic Substitution Analyzer

5. Some Exemplary Results of the Project

Tools in CrypTool 2 – DECODE Downloader and DECODE Viewer



You are logged in as

Filter:

#	Name
2	Copiale
210	BAV_Borg-Lat_898

Name:	Copiale			
Id:	2			
Content:	Type:	cipher	Inline plaintext:	false
	Cipher type:		Inline cleartext:	false
	Symbol set:		Cleartext language:	
	Number of pages:	105	Plaintext language:	German
Origin:	Author:	The Oculist Society	Dating:	1734-01-01:1736-12-31
	Sender:		Region:	Germany
	Receiver:		City:	
Format:	Paper:			
	Ink Type:			
Add. info:	http://stp.lingfil.uu.se/~bea/copiale/			
Images:				
Available	☑ Transcription		☑ Decipherment	
	□ Statistics		☑ Translation	

Documents:

#	Title	Upload date	Size	File type
3	decryption of copiale cipher	2016-08-09	87 kB	txt
1	transcription of copiale ciphe	2016-08-09	237 kB	txt
2	translation to English of copia	2016-08-09	79 kB	txt

```
##===================================================================
==
## Transcription of Copiale document, August 2011.
## 1. document lines separated here by blank lines.
## 2. each character transcribed using an ascii code.
## 3. uncertainties are marked with "?"
## 4. catch words letters at page bottom marked with "#"
## 5. page numbers marked with "##"
## 6. capitalization is recorded (e.g., first letter is "L", not
"l")
## 7.  logograms: *o* society, *star* secret, *nee* master, *tri*
lodge, *bigx* freemason, *gate* table shaped, *lip* oculist (eye),
*bigl* position of feet, *tribig* lodge, *sci* "God", *toe* power
##===================================================================
==

## PAGE 1

L i t : m z grr bar b l

v x. zzz bar ih lam s. k sqp ki arr bar w n
```

246,594 characters, 3,858 lines

This template shows how to download encrypted historical documents from the DECODE database:

1) Start the template (Play button).

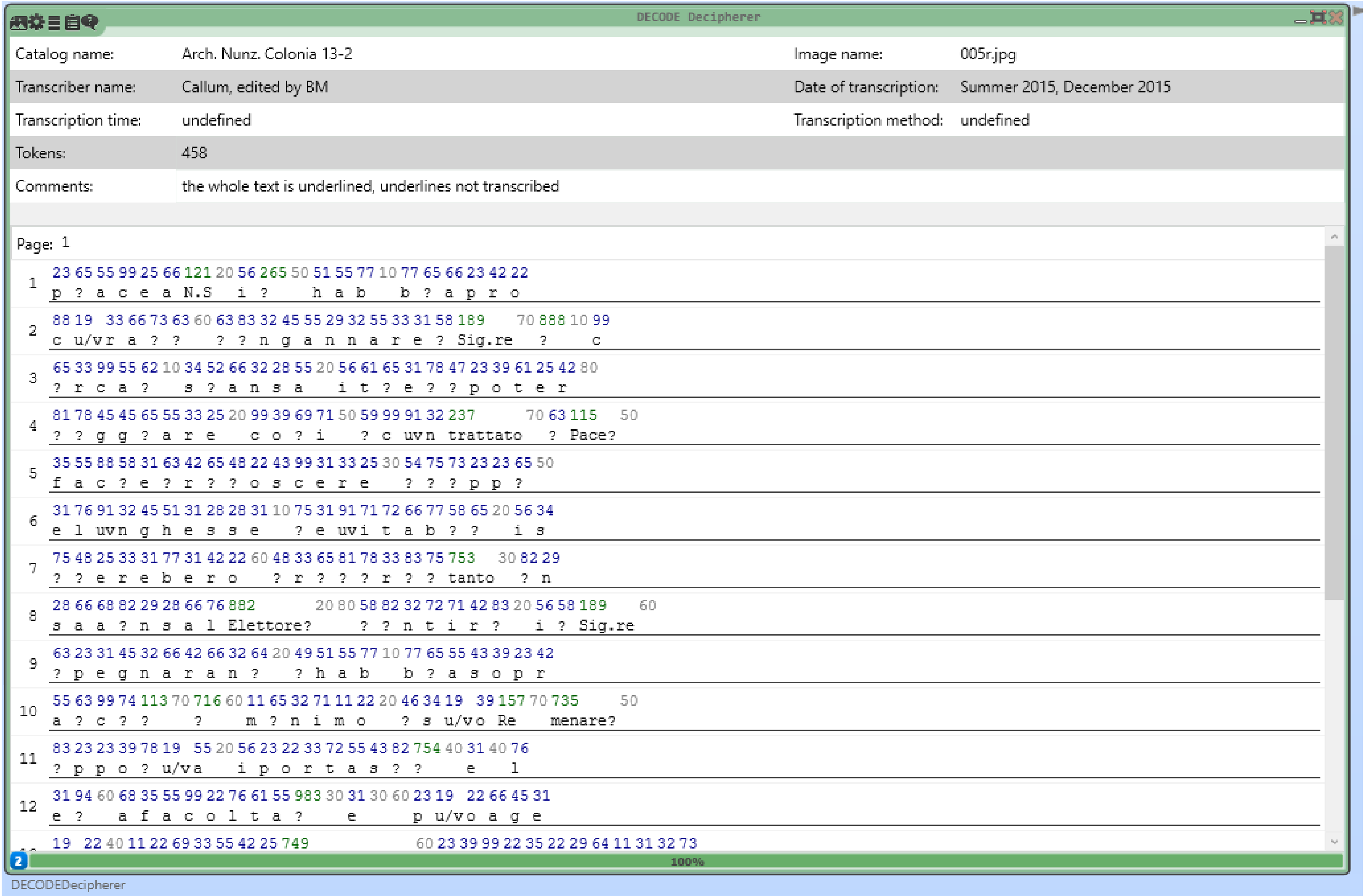
2) Within the DECODE Downloader a list of database records is shown. These records can be limited by entering a filter text, e.g. "Copiale".

3) Via double click on an entry in the DECODE Downloader list, the record's meta data are shown in the DECODE Viewer.

4) In the DECODE Viewer, additional documents and images can be downloaded: To do so, double click on the specific image or document entry in the lists. These downloads are shown in the components "Downloaded Document" and "Downloaded Image".

5. Some Exemplary Results of the Project

Tools in CrypTool 2 – DECODE Decipherer



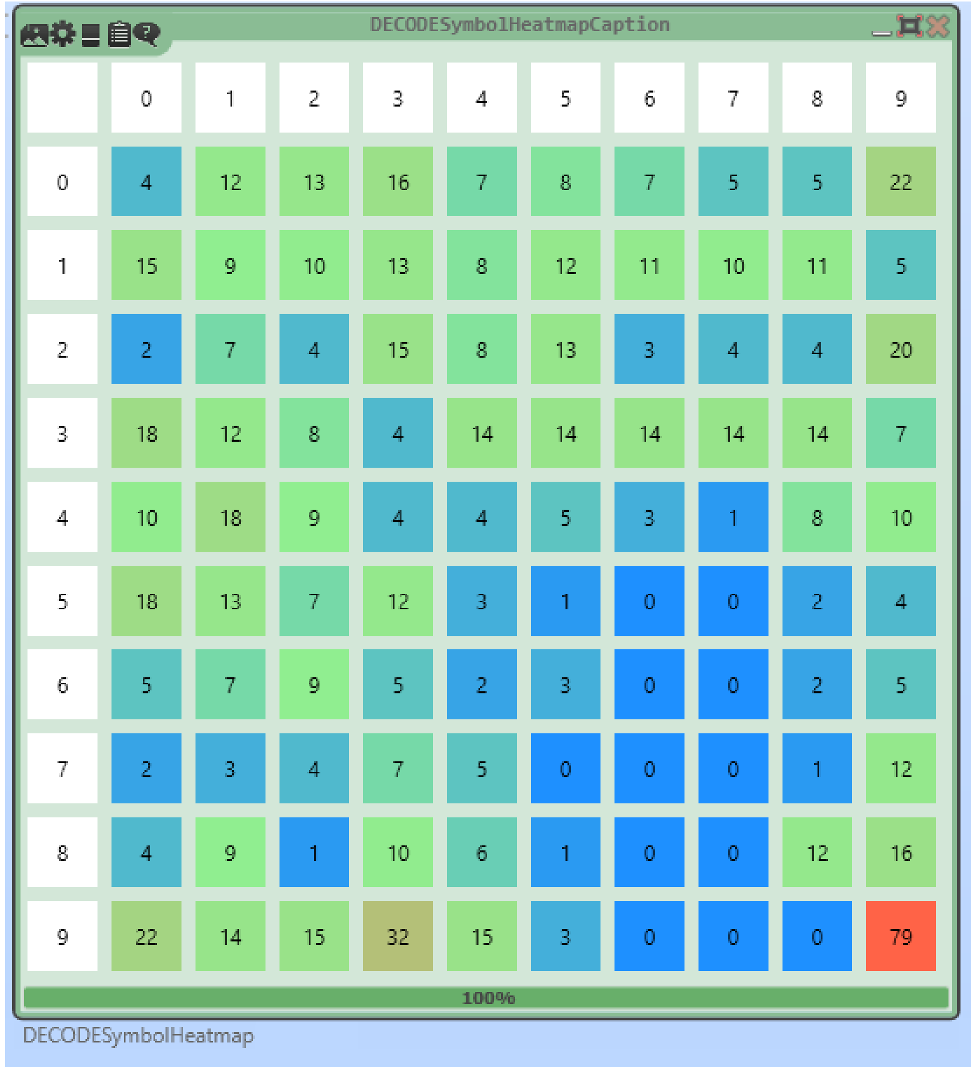
Catalog name:	Arch. Nunz. Colonia 13-2	Image name:	005r.jpg
Transcriber name:	Callum, edited by BM	Date of transcription:	Summer 2015, December 2015
Transcription time:	undefined	Transcription method:	undefined
Tokens:	458		
Comments:	the whole text is underlined, underlines not transcribed		

Page: 1

```
 1  23 65 55 99 25 66 121 20 56 265 50 51 55 77 10 77 65 66 23 42 22
    p ? a c e a N.S  i ?    h a b   b ? a p r o
 2  88 19  33 66 73 63 60 63 83 32 45 55 29 32 55 33 31 58 189    70 888 10 99
    c u/vr a ? ?   ? ? n g a n n a r e ? Sig.re   ?    c
 3  65 33 99 55 62 10 34 52 66 32 28 55 20 56 61 65 31 78 47 23 39 61 25 42 80
    ? r c a ?   s ? a n s a   i t ? e ? ? p o t e r
 4  81 78 45 45 65 55 33 25 20 99 39 69 71 50 59 99 91 32 237      70 63 115   50
    ? ? g g ? a r e   c o ? i   ? c uvn trattato  ? Pace?
 5  35 55 88 58 31 63 42 65 48 22 43 99 31 33 25 30 54 75 73 23 23 65 50
    f a c ? e ? r ? ? o s c e r e   ? ? ? p p ?
 6  31 76 91 32 45 51 31 28 28 31 10 75 31 91 71 72 66 77 58 65 20 56 34
    e l uvn g h e s s e   ? e uvi t a b ? ?   i s
 7  75 48 25 33 31 77 31 42 22 60 48 33 65 81 78 33 83 75 753   30 82 29
    ? ? e r e b e r o   ? r ? ? ? r ? ? tanto   ? n
 8  28 66 68 82 29 28 66 76 882        20 80 58 82 32 72 71 42 83 20 56 58 189    60
    s a a ? n s a l Elettore?      ? ? n t i r ?   i ? Sig.re
 9  63 23 31 45 32 66 42 66 32 64 20 49 51 55 77 10 77 65 55 43 39 23 42
    ? p e g n a r a n ?   ? h a b   b ? a s o p r
10  55 63 99 74 113 70 716 60 11 65 32 71 11 22 20 46 34 19  39 157 70 735     50
    a ? c ? ?     ?     m ? n i m o   ? s u/vo Re    menare?
11  83 23 23 39 78 19  55 20 56 23 22 33 72 55 43 82 754 40 31 40 76
    ? p p o ? u/va   i p o r t a s ? ?     e   l
12  31 94 60 68 35 55 99 22 76 61 55 983 30 31 30 60 23 19  22 66 45 31
    e ?   a f a c o l t a ?     e      p u/vo a g e
    19  22 40 11 22 69 33 55 42 25 749           60 23 39 99 22 35 22 29 64 11 31 32 73
```

5. Some Exemplary Results of the Project

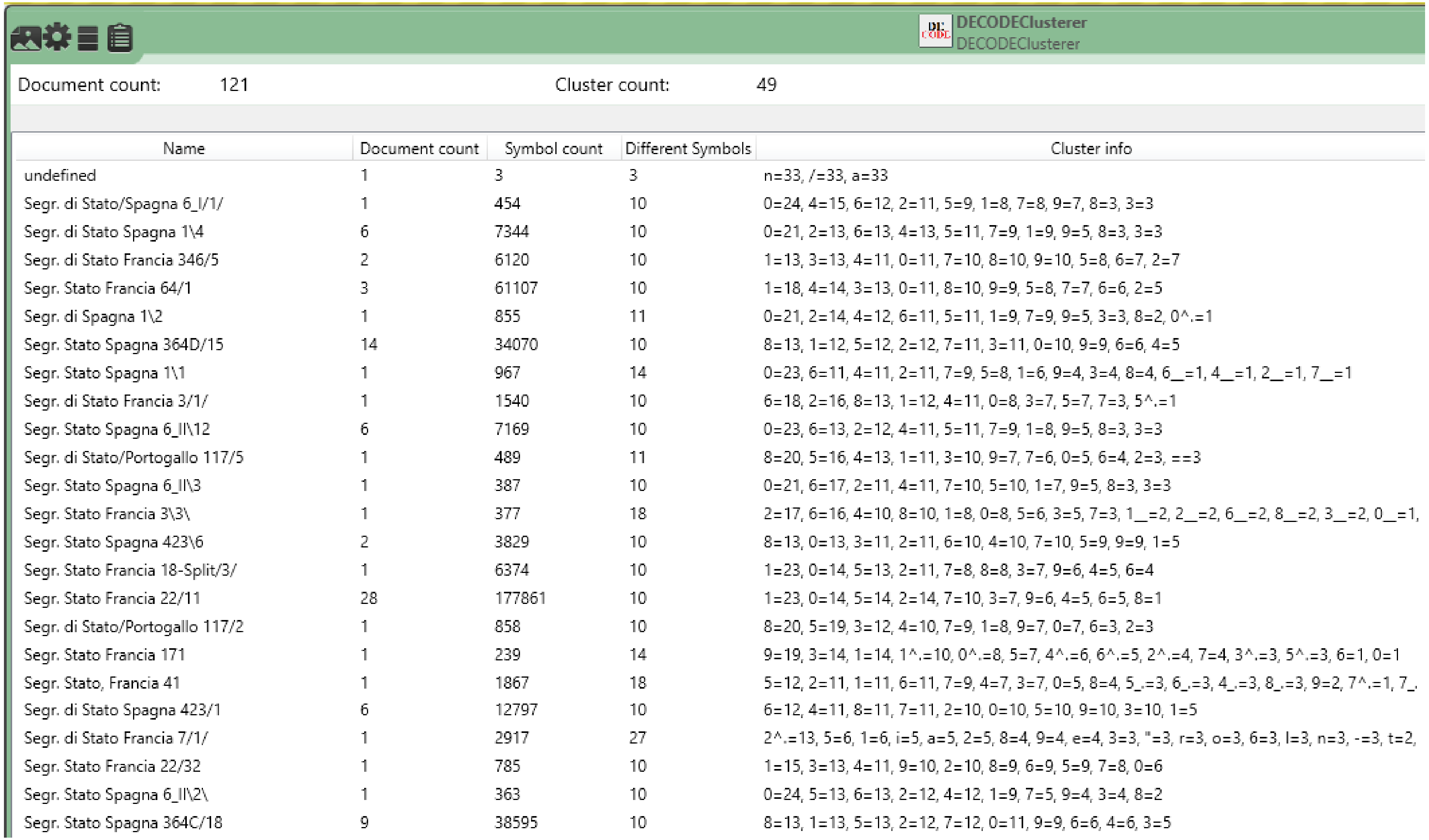
Tools in CrypTool 2 – DECODE Symbol Heatmap



	0	1	2	3	4	5	6	7	8	9
0	4	12	13	16	7	8	7	5	5	22
1	15	9	10	13	8	12	11	10	11	5
2	2	7	4	15	8	13	3	4	4	20
3	18	12	8	4	14	14	14	14	14	7
4	10	18	9	4	4	5	3	1	8	10
5	18	13	7	12	3	1	0	0	2	4
6	5	7	9	5	2	3	0	0	2	5
7	2	3	4	7	5	0	0	0	1	12
8	4	9	1	10	6	1	0	0	12	16
9	22	14	15	32	15	3	0	0	0	79

5. Some Exemplary Results of the Project

Tools in CrypTool 2 – DECODE Key Clusterer

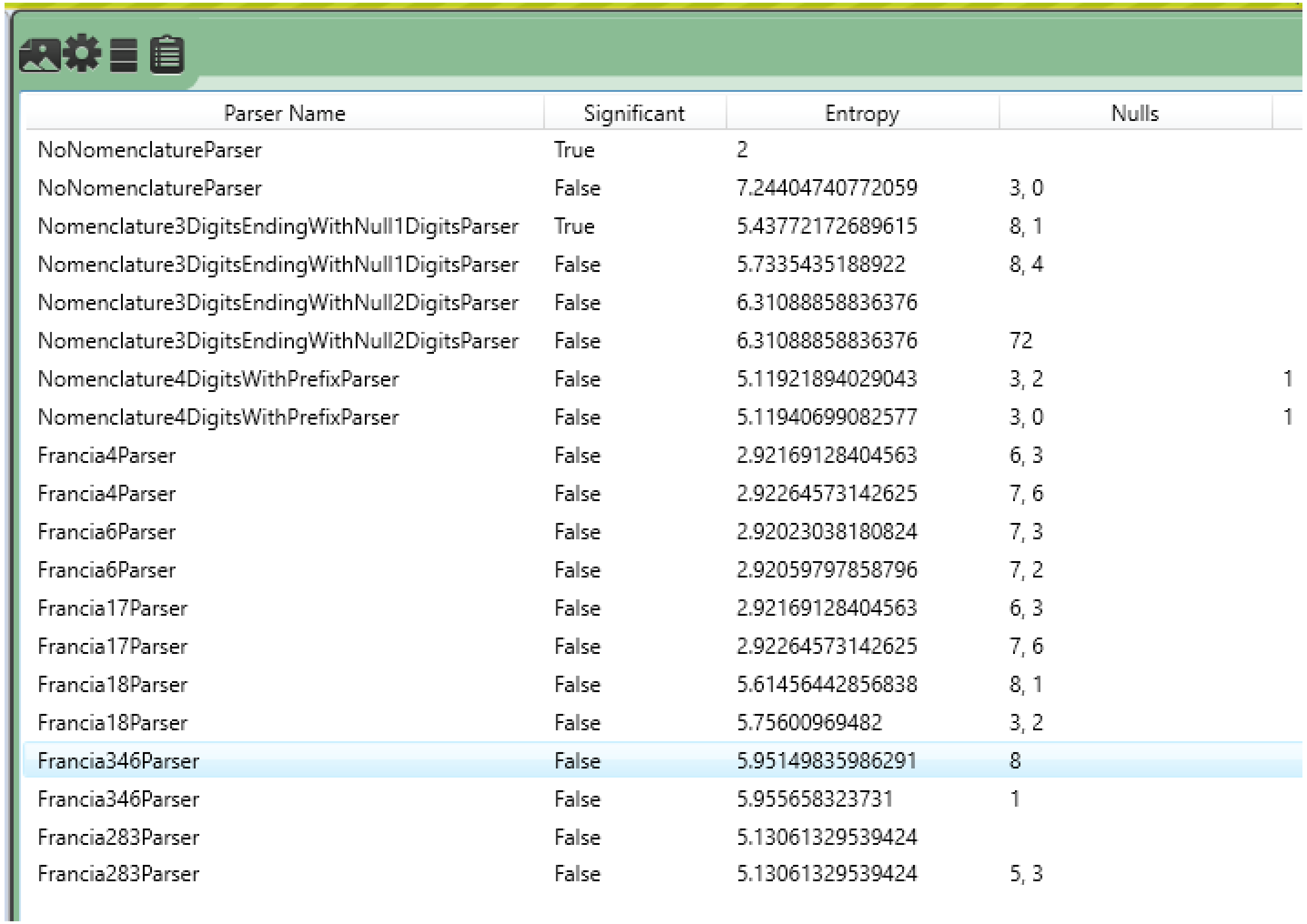


Document count: 121 Cluster count: 49

Name	Document count	Symbol count	Different Symbols	Cluster info
undefined	1	3	3	n=33, /=33, a=33
Segr. di Stato/Spagna 6_I/1/	1	454	10	0=24, 4=15, 6=12, 2=11, 5=9, 1=8, 7=8, 9=7, 8=3, 3=3
Segr. di Stato Spagna 1\4	6	7344	10	0=21, 2=13, 6=13, 4=13, 5=11, 7=9, 1=9, 9=5, 8=3, 3=3
Segr. di Stato Francia 346/5	2	6120	10	1=13, 3=13, 4=11, 0=11, 7=10, 8=10, 9=10, 5=8, 6=7, 2=7
Segr. Stato Francia 64/1	3	61107	10	1=18, 4=14, 3=13, 0=11, 8=10, 9=9, 5=8, 7=7, 6=6, 2=5
Segr. di Spagna 1\2	1	855	11	0=21, 2=14, 4=12, 6=11, 5=11, 1=9, 7=9, 9=5, 3=3, 8=2, 0^.=1
Segr. Stato Spagna 364D/15	14	34070	10	8=13, 1=12, 5=12, 2=12, 7=11, 3=11, 0=10, 9=9, 6=6, 4=5
Segr. Stato Spagna 1\1	1	967	14	0=23, 6=11, 4=11, 2=11, 7=9, 5=8, 1=6, 9=4, 3=4, 8=4, 6__=1, 4__=1, 2__=1, 7__=1
Segr. di Stato Francia 3/1/	1	1540	10	6=18, 2=16, 8=13, 1=12, 4=11, 0=8, 3=7, 5=7, 7=3, 5^.=1
Segr. Stato Spagna 6_II\12	6	7169	10	0=23, 6=13, 2=12, 4=11, 5=11, 7=9, 1=8, 9=5, 8=3, 3=3
Segr. di Stato/Portogallo 117/5	1	489	11	8=20, 5=16, 4=13, 1=11, 3=10, 9=7, 7=6, 0=5, 6=4, 2=3, ==3
Segr. Stato Spagna 6_II\3	1	387	10	0=21, 6=17, 2=11, 4=11, 7=10, 5=10, 1=7, 9=5, 8=3, 3=3
Segr. Stato Francia 3\3\	1	377	18	2=17, 6=16, 4=10, 8=10, 1=8, 0=8, 5=6, 3=5, 7=3, 1__=2, 2__=2, 6__=2, 8__=2, 3__=2, 0__=1,
Segr. Stato Spagna 423\6	2	3829	10	8=13, 0=13, 3=11, 2=11, 6=10, 4=10, 7=10, 5=9, 9=9, 1=5
Segr. Stato Francia 18-Split/3/	1	6374	10	1=23, 0=14, 5=13, 2=11, 7=8, 8=8, 3=7, 9=6, 4=5, 6=4
Segr. Stato Francia 22/11	28	177861	10	1=23, 0=14, 5=14, 2=14, 7=10, 3=7, 9=6, 4=5, 6=5, 8=1
Segr. di Stato/Portogallo 117/2	1	858	10	8=20, 5=19, 3=12, 4=10, 7=9, 1=8, 9=7, 0=7, 6=3, 2=3
Segr. Stato Francia 171	1	239	14	9=19, 3=14, 1=14, 1^.=10, 0^.=8, 5=7, 4^.=6, 6^.=5, 2^.=4, 7=4, 3^.=3, 5^.=3, 6=1, 0=1
Segr. Stato, Francia 41	1	1867	18	5=12, 2=11, 1=11, 6=11, 7=9, 4=7, 3=7, 0=5, 8=4, 5_.=3, 6_.=3, 4_.=3, 8_.=3, 9=2, 7^.=1, 7_.
Segr. di Stato Spagna 423/1	6	12797	10	6=12, 4=11, 8=11, 7=11, 2=10, 0=10, 5=10, 9=10, 3=10, 1=5
Segr. di Stato Francia 7/1/	1	2917	27	2^.=13, 5=6, 1=6, i=5, a=5, 2=5, 8=4, 9=4, e=4, 3=3, "=3, r=3, o=3, 6=3, l=3, n=3, -=3, t=2,
Segr. Stato Francia 22/32	1	785	10	1=15, 3=13, 4=11, 9=10, 2=10, 8=9, 6=9, 5=9, 7=8, 0=6
Segr. Stato Spagna 6_II\2\	1	363	10	0=24, 5=13, 6=13, 2=12, 4=12, 1=9, 7=5, 9=4, 3=4, 8=2
Segr. Stato Spagna 364C/18	9	38595	10	8=13, 1=13, 5=13, 2=12, 7=12, 0=11, 9=9, 6=6, 4=6, 3=5

5. Some Exemplary Results of the Project

Tools in CrypTool 2 – DECODE Parser Tester



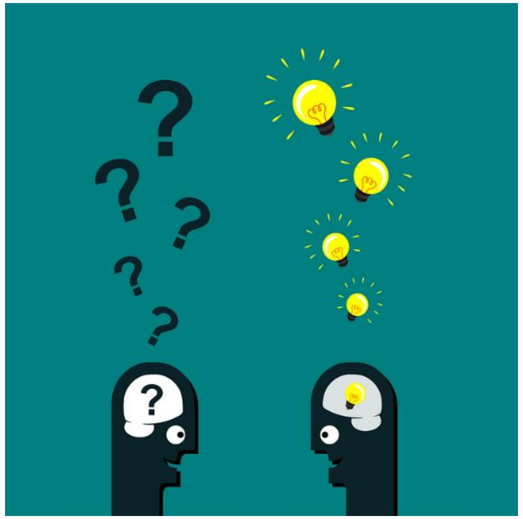
Parser Name	Significant	Entropy	Nulls	
NoNomenclatureParser	True	2		
NoNomenclatureParser	False	7.24404740772059	3, 0	
Nomenclature3DigitsEndingWithNull1DigitsParser	True	5.43772172689615	8, 1	
Nomenclature3DigitsEndingWithNull1DigitsParser	False	5.7335435188922	8, 4	
Nomenclature3DigitsEndingWithNull2DigitsParser	False	6.31088858836376		
Nomenclature3DigitsEndingWithNull2DigitsParser	False	6.31088858836376	72	
Nomenclature4DigitsWithPrefixParser	False	5.11921894029043	3, 2	1
Nomenclature4DigitsWithPrefixParser	False	5.11940699082577	3, 0	1
Francia4Parser	False	2.92169128404563	6, 3	
Francia4Parser	False	2.92264573142625	7, 6	
Francia6Parser	False	2.92023038180824	7, 3	
Francia6Parser	False	2.92059797858796	7, 2	
Francia17Parser	False	2.92169128404563	6, 3	
Francia17Parser	False	2.92264573142625	7, 6	
Francia18Parser	False	5.61456442856838	8, 1	
Francia18Parser	False	5.75600969482	3, 2	
Francia346Parser	False	5.95149835986291	8	
Francia346Parser	False	5.955658323731	1	
Francia283Parser	False	5.13061329539424		
Francia283Parser	False	5.13061329539424	5, 3	

6. Conclusion

- **DECRYPT collects historic ciphers and develops tools and resources for transcription and cryptanalysis**
- **Project limits & challenges**
  - Will **NOT** be able to **solve every cipher automatically**, e.g. non-deterministic ciphers
  - Sometimes, **manual transcriptions are more feasible** than automatic due to errors
  - "**Unknown**" cipher (types) will also **be hard** to be solved automatically
  - Tools often will **support the cryptanalyst** and can not replace him ☺
- **Benefits**
  - Huge set of **original historic crypto material**
  - Many **helpful tools**
  - New insights in **early-modern cryptology**
  - New insights in our **"hidden history"**

Questions and discussion

Thank you very much for your attention!



Do you have questions?