

# Wikipedia-based Entity Linking for the Digital Library of Polish and Poland-Related News Pamphlets

Maciej Ogrodniczuk

Linguistic Engineering Group  
Institute of Computer Science  
Polish Academy of Sciences

Włodzimierz Gruszczyński

Institute of Polish Language  
Polish Academy of Sciences

ICADL 2020

30 November 2020

# The Digital Library used in the study

## **CBDU:**

- The Digital Library of Polish and Poland-Related Ephemeral Prints from the 16th, 17th and 18th Centuries
- a thematic digital library of approx. 2,000 Polish and Poland-related pre-press documents
- dated between 1501 and 1729
- available only in image form (PDF files containing scanned originals)
- accompanied with rich metadata taken over from existing bibliography, including item descriptions

<https://cbdu.ijp.pan.pl>

# CBDU: a sample item

Kurzer und wahrhafter  
Bericht, was sich im Herbst  
1598 in dem in Königreich  
Schweden entstandenen  
Kriegswesen zwischen der  
Königlichen Majestät und  
Herzog Karl zugetragen hat

## Full title

[*Tytul, frakt.:*] Kurtzer und warhaff-  
tiger bericht / was sich in nechstverschie-  
nen Herbst / des 1598. Jhars / in dem  
entstande-  
nen Kriegswesen im Königreich Schwe-  
den / zwischen der König: Majest:  
vnd Hertzog Carls F: G:  
zugetragen.  
[*drzeworyt: scena batalistyczna*]  
1598.

## Contents

Relacja o wyprawie króla Zygmunta III we  
wrześniu 1598 do Szwecji oraz o walkach  
prowadzonych z ks. Karolem  
Sudermańskim, m. in. o bitwach pod  
Stegeborg i Linköping.

## Metadata

Identifier:	276
Language:	Deutsch
Place of Publication:	[B. m. dr.]
Date of Publication:	1598
Year:	1598
Format:	4°
Sheets:	Knfb. 4

Kurtzer vnd warhaff-  
tiger bericht/was sich in nechstverschie-  
nen Herbst / des 1598. Jhars / in dem entstande-  
nen Kriegswesen im Königreich Schwe-  
den / zwischen der König: Majest:  
vnd Hertzog Carls F: G:  
zugetragen.



1598.

# Background information needed!

## On two levels:

- related to content (actors, locations, events, facts...)
- related to the item as a whole ('read more')

# Getting this information

## From which source?

- Wikipedia! it's large, universal...
- already contains compensation mechanisms such as redirections

## How to use it?

- getting content-related information by wikization
- getting item-related information by wikisearch

## Wikifier results

### Text

Relacja o wyprawie króla Zygmunta III we wrześniu 1598 do Szwecji oraz o walkach prowadzonych z ks. Karolem Sudermańskim, m.in. o bitwach pod Stegeborg i Linköping.

### Annotations

PR	Annotation	Annotation (pl)	
0.0604	<u>Linköping</u> W D	<u>Linköping</u>	>>
0.0577	<u>Szwecja</u> W D	<u>Szwecja</u>	>>
0.0549	<u>Karol IX</u> <u>Waza</u> W D	<u>Karol IX</u> <u>Waza</u>	>>
0.0361	<u>1598</u> W D	<u>1598</u>	>>
0.0321	<u>Zygmunt III</u> <u>Waza</u> W D	<u>Zygmunt III</u> <u>Waza</u>	>>

# Wikisearch



**WIKIPEDIA**  
Wolna encyklopedia

[Strona główna](#)

[Losuj artykuł](#)

[Kategorie artykułów](#)

[Najlepsze artykuły](#)

[Częste pytania \(FAQ\)](#)

[Dla czytelników](#)

[O Wikipedii](#)

[Kontakt](#)

[Wspomóż Wikipedię](#)

## Bitwa pod Stegeborgiem

**Bitwa pod Stegeborgiem** – bitwa rozegrana na polach pod zamkiem **Stegeborg** w dniu 8 września?/18 września **1598**, w czasie **wyprawy króla** szwedzkiego i polskiego

6 KB (539 słów) - 15:04, 5 maj 2019

## Linköpings blodbad (przekierowanie z [Krwawa łaźnia w Linköping](#))

jest za jedną z największych zbrodni sądowych w dziejach **Szwecji**. **We wrześniu 1598**, po **bitwie pod Stångebro**, na mocy ugody w **Linköping** w ręce księcia

3 KB (374 słowa) - 23:59, 28 lut 2020

## Bitwa pod Linköping

regenta Karola **Sudermańskiego** a wojskami **króla Zygmunta III Wazy** w czasie jego **wyprawy do Szwecji**. Po śmierci Jana III Wazy w 1592 r. tron **Szwecji** objął katolicki

11 KB (1185 słów) - 23:40, 24 sie 2019

# Preliminary results

## A sample print description:

An account of King Sigismund III's expedition to Sweden in September 1598 and the battles with with Prince Charles of Södermanland, including the battles of Stegeborg and Linköping.

Entity in Polish	Entity in English	WF	WS
Zygmunt III Waza	Sigismund III Vasa	✓	
Szwecja	Sweden	✓	
1598	1598	✓	
Karol IX Waza	Charles IX of Sweden	✓	
Linköping	Linköping	✓	
Bitwa pod Stegeborgiem	Battle of Stegeborg		✓
Bitwa pod Linköping	Battle of Stångebro		✓
Unia polsko-szwedzka	Polish–Swedish union		✓



# Improving relevance of results

## Filtering entries:

- 1 **date filter:** links to Wikipedia date pages are obviously non-informative
- 2 **frequency filter:** keep only entries less frequent than 10 000 most frequent lemmatized unigrams in the National Corpus of Polish:
  - × Vilnius
  - ✓ Orsha
- 3 **overlap filter:** in case of overlaps, less-specific entries were discarded:
  - × Orsha
  - ✓ Battle of Orsha

# Final results

## After filtering:

Entity in English	Initially recognized	After date filter	After frequency filter	After overlap removal	Identified manually
Sigismund III Vasa	✓	✓	✓	✓	✓
Sweden	✓	✓			
1598	✓				
Charles IX of Sweden	✓	✓	✓	✓	✓
Linköping	✓	✓	✓		
Battle of Stegeborg	✓	✓	✓	✓	✓
Battle of Stångebro	✓	✓	✓	✓	✓
Polish–Swedish union	✓	✓	✓	✓	

# Conclusions

## Ideas for the future:

- processing textual content of prints
- use different language versions of Wikipedia
- use sources other than Wikipedia
- apply the process to entire library and/or related material

# Thank you!

## **And the funding institution:**

The work was financed by a research grant from the Polish Ministry of Science and Higher Education under the National Programme for the Development of Humanities for the years 2019–2023 (grant 11H 18 0413 86)