



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## ISEBEL an Intelligent Search Engine for Belief Legends

Ding, QiQing; Meder, Theo; Windhouwer, Menzo

2019

### **document version**

Peer reviewed version

### **document license**

CC BY

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

Ding, Q., Meder, T., & Windhouwer, M. (2019). *ISEBEL an Intelligent Search Engine for Belief Legends*. Poster session presented at Digital Humanities Conference 2019, Utrecht, Netherlands.  
<https://dev.clariah.nl/files/dh2019/boa/0439.html>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[pure@knaw.nl](mailto:pure@knaw.nl)

# ISEBEL

## an Intelligent Search Engine for Belief Legends

Distributed around the globe more databases of folktales, including belief legends, have come into existence. Combining them might open up new and exciting research possibilities. ISEBEL (Meertens Institute, 2019a) is a project aiming to create a search engine that makes exactly this possible by providing unified search over the participant's database, while dealing intelligently with the various languages. The following databases are currently providing their stories to the project:

1. the Dutch Folktale Database (Meertens Institute, 2018),
2. the Evald Tang Kristensen Collection (UCLA, 2018), and
3. the WossiDia archive (Wossidlo Institute, 2018) (Rostock).

We are looking out to and working on including other interesting archives and languages throughout Europe - especially a few more Nordic countries already have shown interest, i.e., Iceland, Norway, Sweden and Estonia.

In the project a set of metadata elements about the stories in the databases have been identified - apart from the text, there is information about narrator, time and place of narrating (including geocodes), keywords/tags and the URL leading to the story and metadata in the original database. These elements are part of a core metadata schema, which already provides hooks to be easily extended with more (database specific) information. The databases use this to provide metadata about their stories. The central catalogue of ISEBEL (Meertens Institute, 2019b) runs a harvester periodically, which collects this metadata. The technical basis of this communication is provided by the OAI-PMH protocol.

The collected metadata is loaded and indexed by CKAN (CKAN Association, 2018), an open source data portal. The harvested data from each 'data provider' are stored as individual XML files by the harvester. The XML files, which contain single stories in each one of them, are imported using local API calls to CKAN. Researchers can then search for the data through CKAN and its underlying indexer, namely Solr (Apache Software Foundation, 2018).

The project is still under heavy development to reach its design goals, e.g., to provide researchers with cross-language search results rather than in only one language. In order to achieve this goal, the translation and keywords extraction must work together with keywords mapping to interconnect the stories and keywords in different languages. The domain specific keywords are the keywords specific to one language, which are either manually attached to the stories by the data provider, difficult to translate using machine translation or even not available in other language(s). Those keywords will be manually maintained and mapped to the computer generated keywords, discussed next, to make the mapping complete.

The computer generated keywords are those keywords extracted from the English story translated from the local language of the data provider. According to the test runs, keywords quality extracted from translated stories are much better than translated keywords. The main reason is that the entire story provides more context for an adequate translation.

**Organizations**

Meertens Institute 26608

UCLA 21675

Wossidia 1559

---

**Keywords**

*There are no Keywords that match this search*

---

**Danish Keywords**

smide 377

begrave 371

forsvinde 367

time 365

troid 362

øre 357

følge 356

opdage 351

spøge 347

ende 342

**Show More Danish Keywords**

---

**Dutch Keywords**

dood 4049

man 3464

sterven 2899

Q

**49,842 datasets found** Order by: Relevance

---

**jah\_01\_0\_00265**  
Kvægpesten kom drivende i luften akkurat ligesom en tåge. Hvor den tåge kom over, døde kreaturerne næsten helt ud. Osten gården har vi et par alen nede i jorden fundet mange...

---

**jah\_01\_0\_00264**  
Jes Outsens i Risjarup fortæller, at en dag han gik til Brumle, da så han noget som et stort spindelvæv ellerr forklæde af bla farve komme flyvende fra Brande og fare ned i Nis...

---

**jah\_01\_0\_00263**  
Den sommer, te fæsygenen var her, da var der sådan varme og tanning. te alting døde hen, og alt græsset svejtest, og søen (Mos sø) var overtrukken med et grønt lag, formedelst...

---

**jah\_01\_0\_00262**  
Der var kvægsyge her. Der var sådan sær stank i luften; en mand sagde, at det var fjorden, der skad over. Kreaturerne stritt som fluer, og de stank, forend de blev rigtig...

---

**jah\_01\_0\_00261**  
Der har været kvægpest her to gange. Norden for Vejle Dal var der ingen pest, og så skulde der jo passes på, at den ikke kom derover, og der blev sendt soldater derud for at...

---

**jah\_01\_0\_00260**  
Der har været kvægpest her tre gange. Æ høvder var da ilge så dyre som æ garde. Peder Kristensen. Rottesgård.

---

**jah\_01\_0\_00259**  
I den store kvægsyge døde alle staldstude på Kjeldgård, og de blev kastede ned i en stor kule uden for gården, hvor haven senere blev anlagt. Træerne star ualmindelig frodig,...

---

**jah\_01\_0\_00258**

At the moment ISEBEL is focussing on the specific genre of the traditional belief legend, mainly because all three databases have this folktale genre in common (almost 60.000 legends in total). These legends mainly deal with traditional folk belief in the supernatural, like ghosts, hauntings, devils, witches, wizards, spells, werewolves, nightmares, giants, trolls, goblins etc., as well as stories about hidden treasures, famous robbers, underground passages and sunken castles. The information on folktales collected from the initial three databases make it possible to analyze folktales and traditional folk belief in a large coastal region of the North Sea and the Baltic Sea, where especially Migratory Legends are interesting. Research can include geographic dispersion. For instance, the dispersion of legends about mermaids will most likely show narratives in the direct coastal areas. Another possible research could be gender related: what is the difference between male and female repertoires? How close do the legends play around the own home, and is there a difference here between male and female narrators?

**Apache Software Foundation** (2018). *Apache Solr*. [lucene.apache.org/solr/](http://lucene.apache.org/solr/) (accessed on 26 November 2018).

**CKAN Association** (2018). *CKAN, the world's leading Open Source data portal platform*. [ckan.org](http://ckan.org) (accessed on 21 November 2018).

**Meertens Institute** (2018). *Nederlands VolksverhalenBank*. [verhalenbank.nl](http://verhalenbank.nl) (accessed on 21 November 2018).

**Meertens Institute** (2019a). *ISEBEL*. [isebel.eu](http://isebel.eu) (accessed on 25 April 2019).

**Meertens Institute** (2019b). *ISEBEL central catalogue*. [search.isebel.eu](http://search.isebel.eu) (accessed on 25 April 2019).

**UCLA** (2018). *The Evald Tang Kristensen Collection*. [etkspace.scandinavian.ucla.edu](http://etkspace.scandinavian.ucla.edu) (accessed on 21 November 2018).

**Wossidlo Institute** (2018). *WossiDiA: The Digital Wossidlo Archive*. [wossidia.de](http://wossidia.de) (accessed on 21 November 2018).