

# User Interface Challenges

for the Study of Folklore and Literature

Timothy R. Tangherlini (UCLA)

# Some key problems

- Design cycle
  - Many Humanities projects have a very short funding cycle, with very limited funds (\$3000 is not an unusual amount)
  - Limits the type of iteration through design that could converge on user centered interfaces
    - Humanities scholar might be an excellent scholar but
    - Not an excellent designer and
    - Not an excellent data scientist
- How do we shift from analog interaction with highly curated data (e.g. reading a novel) to computationally mediated interaction with uncurated data (e.g. trying to find passages from millions of books in HathiTrust)?

# Four Challenges in Folklore and Literature

- Collections and Archives
- Indexing and Classification
  - curation
- Visualization and Navigation
- Analysis

# The Folklore Macroscope

• • •

WitchHunter, TrollFinder, TreasureX, ElfYelp

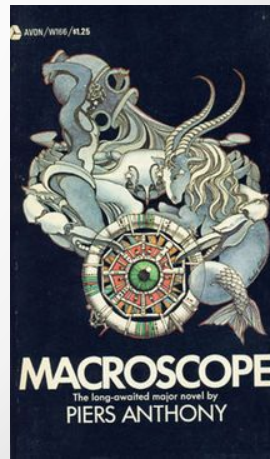
With Peter Broadwell (UCLA)  
& James Abello (Rutgers), and David Mimno (Cornell)



# Macrosopes

- Macrosopes provide a 'vision of the whole,' helping us 'synthesize' the related elements and detect patterns, trends, and outliers while granting access to myriad details. Rather than make things larger or smaller, macrosopes let us observe what is at once too great, slow, or complex for the human eye and mind to notice and comprehend  
-Katy Börner (2011)

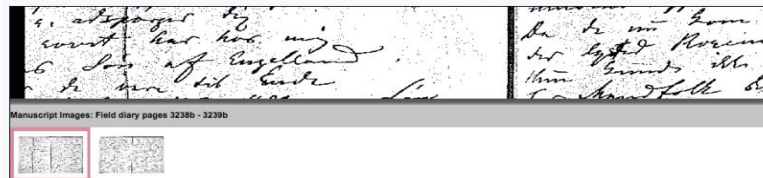
Cf: Joël de Rosnay (1975) Le macroscope. Vers une vision globale. Editions du Seuil.



## Folklore Macroscopic Tools

### The Danish Folklore Nexus (Tangherlini and Broadwell 2013)

A curated study environment, included with *Danish Folktales, Legends, and Other Stories*



### ETKspace (Broadwell and Tangherlini 2014)

A faceted browsing interface to the full Tang Kristensen corpus

Story	Places Mentioned	Keywords	ETK Index
DS_01_0_00001	Aabenraa	Å	About the Devil and being in league with him
DS_01_0_00002	Aabenraa	Å	Churches and their construction: Monasteries, holy spr
DS_01_0_00003	Aabenraa	Å	Concerning earlier agricultural practices
DS_01_0_00004	Aalborg	aalb	Cunning men and women and their activities
DS_01_0_00005	Aarhus	aahr	Death portents
DS_01_0_00006	Aabenraa	aahr	Diverse hauntings and conjuring of revenants
DS_01_0_00007	Aabenraa	aahr	Diverse people
DS_01_0_00008	Aabenraa	aahr	Diverse place legends
DS_01_0_00009	Aabenraa	aahr	Flies
DS_01_0_00010	Aabenraa	aahr	Female revenants
DS_01_0_00011	Aabenraa	aahr	From the time of villainage
DS_01_0_00012	Aabenraa	aahr	Horses and their sport
DS_01_0_00013	Aabenraa	aahr	Household spirits
DS_01_0_00014	Aabenraa	aahr	Houses and life in them
DS_01_0_00015	Aabenraa	aahr	Human and cattle illnesses
DS_01_0_00016	Aabenraa	aahr	Legends about farms and farms

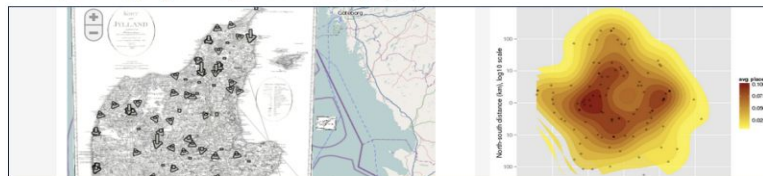
### WitchHunter (Tangherlini and Broadwell 2015)

Devised for geo-semantic mapping and pattern discovery in geo-located folklore



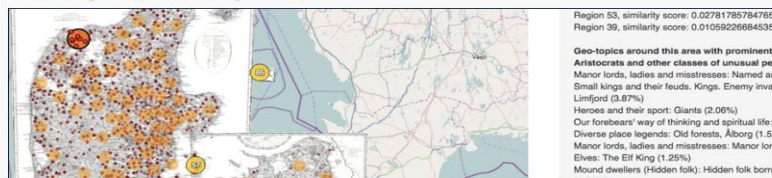
### GhostScope (Broadwell and Tangherlini 2015)

Conceptual mapping of storytellers' mental geographies



### EiYelp (Tangherlini and Broadwell 2015)

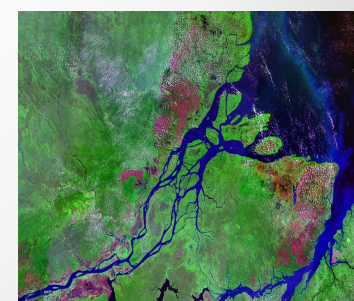
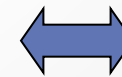
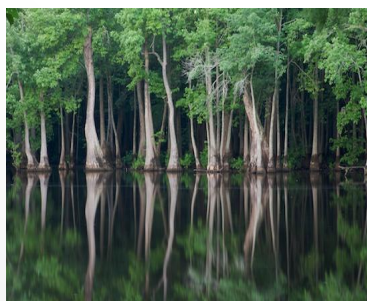
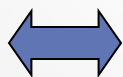
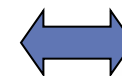
A tool for geo-topic modeling and exploration



# DFL 1.0:

## Classification and Navigation in a Curated Corpus

- Navigation must “add value”
- How do we leverage the huge amount of work that we have done in acquiring, organizing and providing dynamic labels for our research materials?





# Danish Folklore Nexus

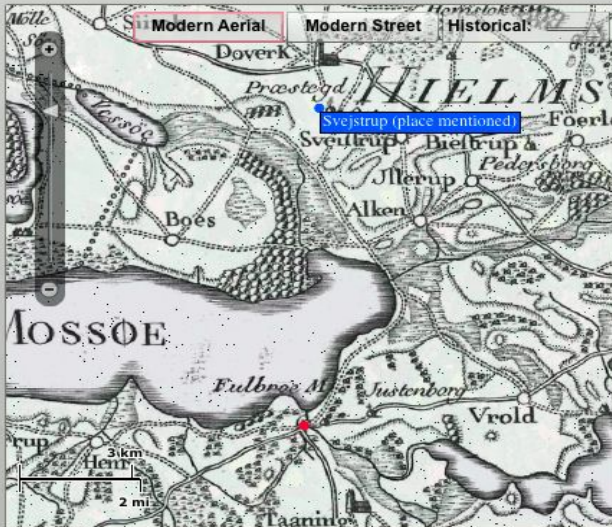


Folklore and 19th Century Denmark   Evald Tang Kristensen   Folktales, Legends and Ballads   Bibliography   Search   About

[HOME](#)   [Informant 123](#)   [Story 380](#)   [Place 615](#)

STORY DETAILS for JAH\_VI\_639 told by Peder Johansen

Close tab



## Story data for JAH\_VI\_639

Order told: 2.06

Recorded during fieldtrip: **Feb 1888**

Fieldtrip dates: 1888-02-04 - 1888-02-17

Place recorded: **Fuldbrø Mølle**

Field diary pages: 3773b - 3774a

Associated keywords: angry, birth, boy, bread, breakfast, chair, cheese, children, clogs, clothing, day, farm, fire, girl, grandmother, house, laughter, Mads, manor lord, Niels, Niels Rebslæer, old, parish clerk, punishment, ringing, school, servant, skin, stove, Søren Jul, tease, throat, trickery, wolf, wood, young

Places mentioned in story: Svejstrup

## JAH\_VI\_639

### INFORMANT BIOGRAPHY

#### PLACES

Fuldbrø Mølle (place recorded)

Svejstrup (place mentioned)

### STORIES TOLD by INFORMANT

## English translation of published version: JAH\_VI\_639 told by Peder Johansen

The oldest parish clerk that I know of in Svejstrup was Niels Ring. He was first a servant for the manor lord at Venge farm and he appointed him to the position. He was so harsh that once when he punished a young girl, he lifted up her clothes and but her bare bottom up on top of a glowing wood-burning stove for so long that her skin was half burned off when she came down. But then he got too old and the children teased him, they filled his chair with pins.

The next parish clerk was Søren Jul. He was a clever man, but strict, and when he got angry he'd throw all the boys together in a heap and then he'd walk all over them with his leather-soled clogs and then he'd yell, "You Goddamn wolf children!" But regardless how strict he was, they still played tricks, they did those boys. One of them was called Mads, he'd been born up at his great grandmother's house. So one day he'd gotten off to a bit of a late start, and when he got to school he wasn't even finished with his breakfast, since he arrived chewing on a cheese rind and a piece of bread. When he gets there, the parish clerk had just started the lessons. Then his desk mate, Niels Rebslæer, sees his chance and while Mads is sitting there chewing on the cheese rind, he reaches out and shoves it down his throat. Now Mads was sitting there and couldn't get the rind either up or down. The other children began to laugh and the parish clerk couldn't control them. "What is going on?" he asks them gruffly. Immediately, Niels Rebslæer jumps up and says, "It's Mads, he's ingenious!" Regardless of how strict he was, the parish clerk couldn't help laughing, but they both had to go and get spanked anyway.

## Story indices for JAH\_VI\_639

**Genre:** Legend

**ETK index:** Our forebears' way of thinking and spiritual life (Vore fædres tankesæt og åndsliv)

**Tangherlini indices**

People: Girl

People: Parish Clerk

Places: Church

Actions or events: Injury

Actions or events: Punishment

Resolution: Ambiguous

Stylistics: Direct Conversation

# Novel Navigation



Danish Folktales, Legends, and Other Stories

http://etkspace.scandinavian.ucla.edu/danishfolklore/#app=6edf&8baf-selectedIndex=1&b0c5-selectedIndex=0

Google

Pandora Radio - Listen to Free In... mac screen shot - Google Search Danish Folktales, Legends, and O...

HOME DS\_IV\_687

MANUSCRIPT/PUBLISHED VERSIONS AND ANNOTATIONS for DS\_IV\_687 told by Jens Peter Pedersen

Close tab

Texts and Annotations Manuscript Images

Danish manuscript English translation

Der var en Herremand her paa Boller i Taars det var en stor Kjæltring og s. slem til at plage hans Bønder med Hoveri og Skatter og Tiender og jog dem fra deres gaarde og turede med dem. S.

6297A

blev h. s. slem til at gaa igjen der paa gaarden s. Manden der var kommen der kunde ikke være der. S. fik h. en klog Præst til at komme og mane ham og h. kom til fastsat Tid og de kom jo ogsaa til at tale sammen og h. gjente ham fra gaarden og forbi Boller [M] Vandmølle. Udenfor Møllen paa [M] Vanddæmningen vilde h. have ham sat ned. Møllekarlen tykte det var skjønt at se hvordan det gik til og h. raammer Hovedet ud og ser paa det. Men s. bliver gj. for stærk og Pr. var ved at skulle have været ned i Mølle-

Danish published version English translation

Der var en herremand her på Boller i Tårs, det var en stor kjæltring, og så slem til at plage hans bønder med hoveri og skatter og tiender, og så jog han dem fra deres gårde og turede med dem. Han blev så slem til at gå igjen der på gården, så den nye mand, som var kommen der, ikke kunde være der. Så fik han en klog præst til at komme og mane ham. Han kom til fastsat tid, og de kom jo også til at tale sammen. Han gjente ham så fra gården og forbi Boller vandmølle. Der vilde han have ham sat ned på vanddæmningen udenfor møllen. Møllekarlen tykte, det var skjønt at se, hvordan det gik til, og han rammer hovedet ud og ser på det. Men så bliver gjengangeren for stærk, og præsten var ved at skulle have været ned i mølledammen. Så bliver karlen forskrækket og trækker sig tilbage. Præsten bliver da herre over manden, og han bliver sat ned.

Related stories based on multi-dimensional classification

Annotations Bibliographic references

The manor lord's haunt is a classic example of a guilty haunt and is widely attested (DS IV 691; Pentikäinen 1969). The manor lord is threatening not only in life, but in death as well.

Jens Peter turns to the wise minister for help, narratively proposing this as an effective strategy against the haunting of the manor lord (Tangherlini 1998). The conjuring does not proceed smoothly, and the farmhand's interruption nearly dooms the minister, a motif that is well known from other stories of conjuring (DS V 728 and 733; PJ 2.05; KMP 1.07). In some stories, the minister actually does lose power over the ghost because of the farmhand's interruption, and these stories often detail the catastrophic results of that interruption (DS IV 744).

Kristian IV's wife, Kristine Munk, was exiled to the manor farm at Boller, and lived there from 1630-1658. Later, the manor farm became part of the larger holdings of the Count of Frijsenborg (PJ 3.18).

Related stories

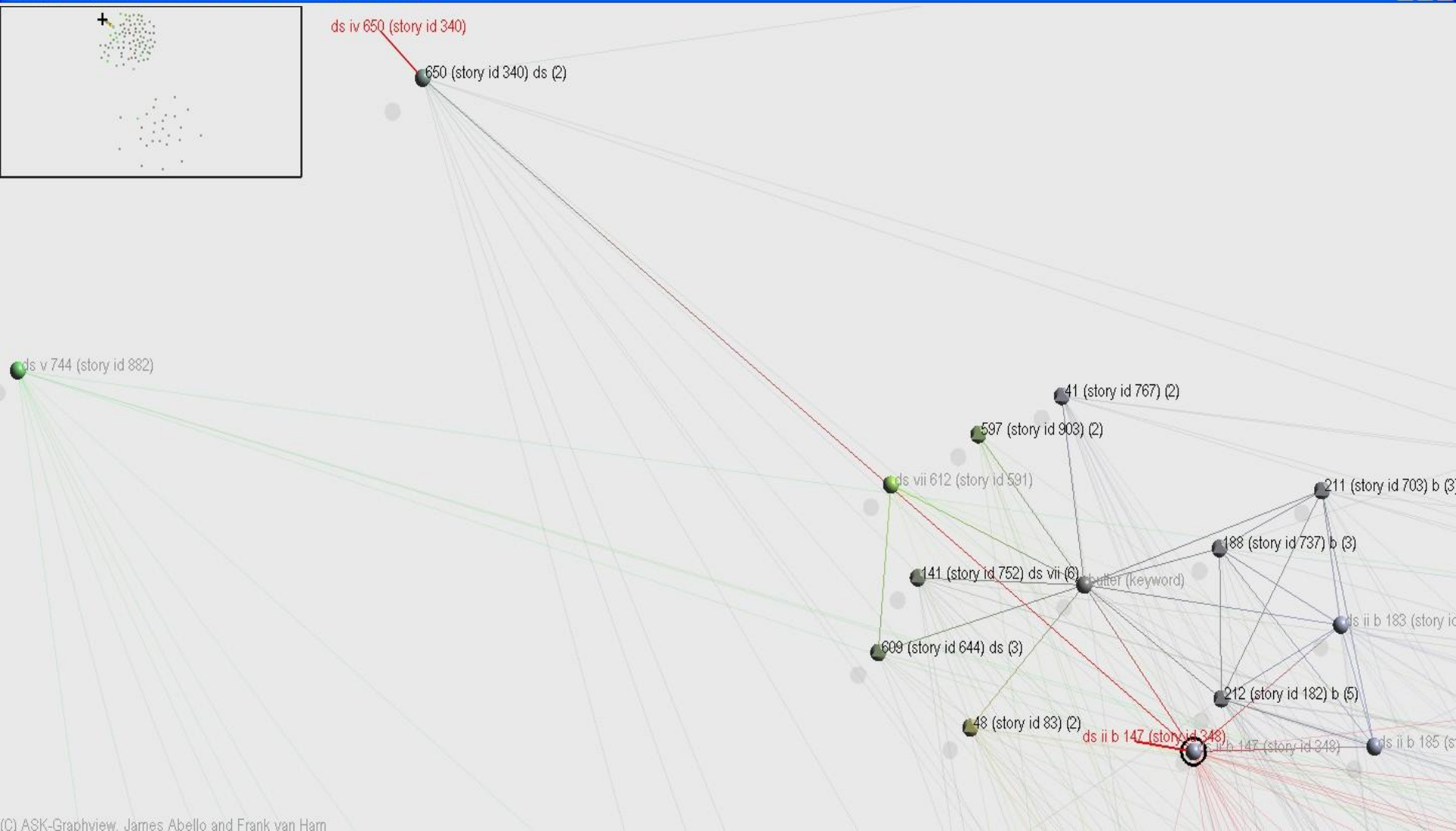
- 241 - 4.25 - DSnr\_IV\_290
- 68 - 0.00 - DS\_V\_513
- 150 - 0.03 - DS\_VII\_1045
- 460 - 0.00 - DS\_V\_728
- 111 - 0.00 - DS\_V\_733
- 123 - 2.05 - DS\_V\_822
- 235 - 1.07 - DS\_V\_810
- 373 - 0.00 - DS\_IV\_744

BACK TO STORY VIEW



# James Abello's *GraphView*

ASK-Graphview : DANISHHYPERGRAPH.WADJ ( DS II V IV VII I VI III E F G B (28))



# Drill-down

Danish Folktales, Legends, and Other Stories

http://etkspace.scandinavian.ucla.edu/danishfolklore/#app=6edf&8baf-selectedIndex=1&b0c5-selectedIndex=6 Google

KQED Twitter Blog: Global pulse DFL Nexus Danish Folklore Interface ETKSpace Topic Models Google Voyant Tools Google Maps YouTube Story Workbench

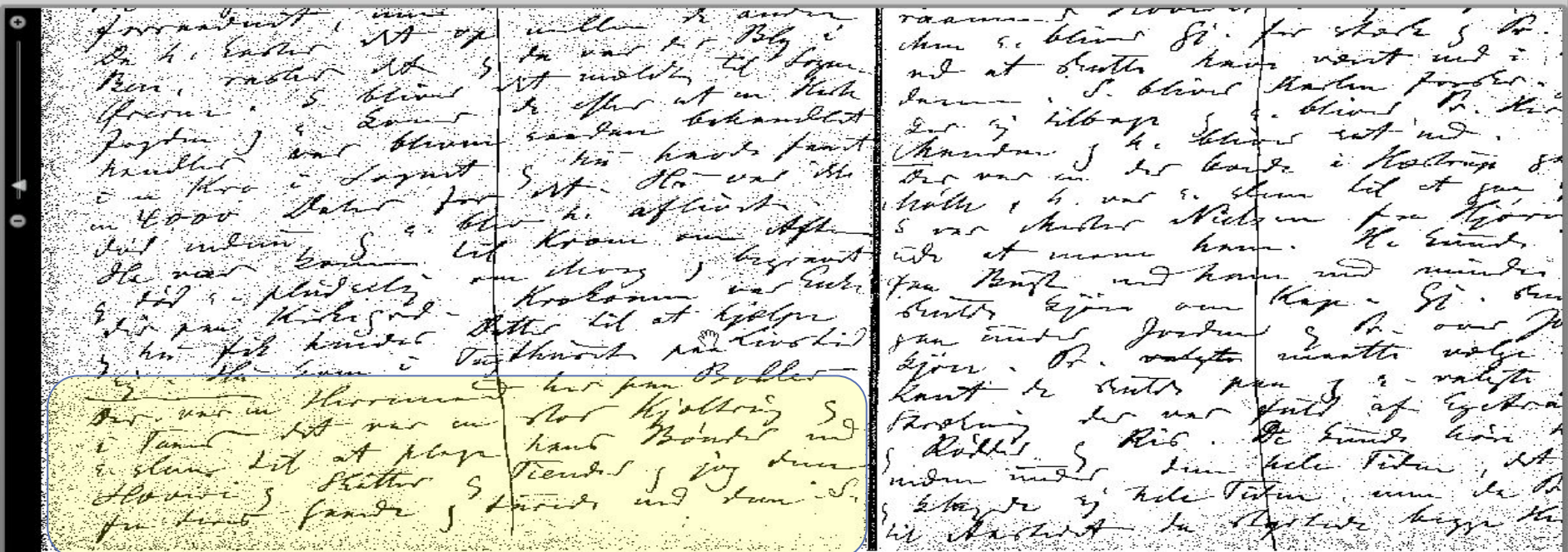
Pandora Radio - Listen to Free In... mac screen shot - Google Search Danish Folktales, Legends, and O...

HOME DS\_IV\_687

MANUSCRIPT/PUBLISHED VERSIONS AND ANNOTATIONS for DS\_IV\_687 told by Jens Peter Pedersen

Close tab

Texts and Annotations Manuscript Images

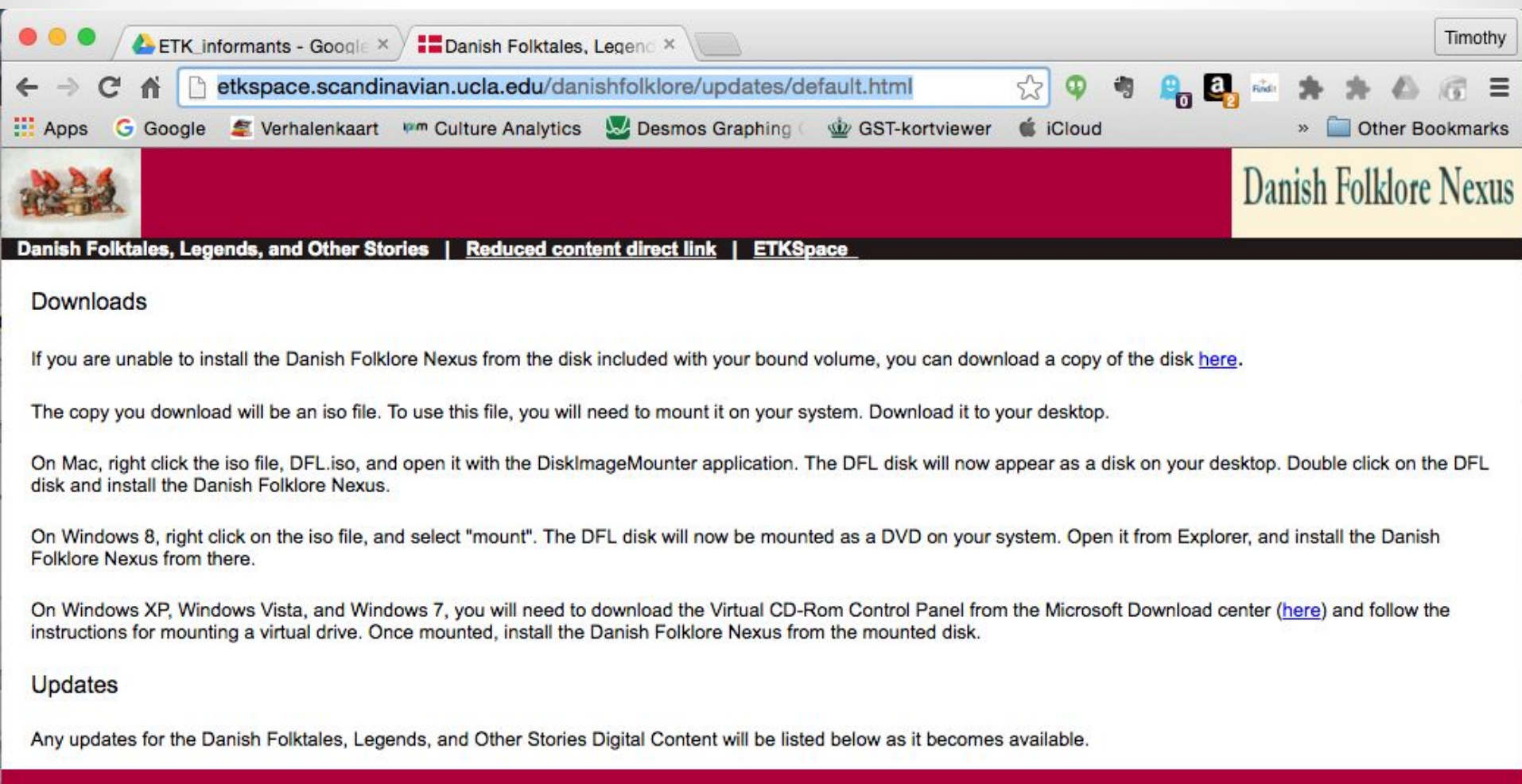


Manuscript Image: Field diary pages 6296b - 6297a

Image 1

BACK TO STORY VIEW

# Benefits



The screenshot shows a web browser window with two tabs: "ETK\_informants - Google" and "Danish Folktales, Legend". The address bar displays the URL `etkspace.scandinavian.ucla.edu/danishfolklore/updates/default.html`. The browser's toolbar includes various icons for search engines (Google, Verhalenkaart), applications (Desmos Graphing, GST-kortviewer), and social media (iCloud). A red banner at the top of the page features the text "Danish Folklore Nexus" on the right and a navigation bar on the left with links: "Danish Folktales, Legends, and Other Stories", "Reduced content direct link", and "ETKSpace".

## Downloads

If you are unable to install the Danish Folklore Nexus from the disk included with your bound volume, you can download a copy of the disk [here](#).

The copy you download will be an iso file. To use this file, you will need to mount it on your system. Download it to your desktop.

On Mac, right click the iso file, DFL.iso, and open it with the DiskImageMounter application. The DFL disk will now appear as a disk on your desktop. Double click on the DFL disk and install the Danish Folklore Nexus.

On Windows 8, right click on the iso file, and select "mount". The DFL disk will now be mounted as a DVD on your system. Open it from Explorer, and install the Danish Folklore Nexus from there.

On Windows XP, Windows Vista, and Windows 7, you will need to download the Virtual CD-Rom Control Panel from the Microsoft Download center ([here](#)) and follow the instructions for mounting a virtual drive. Once mounted, install the Danish Folklore Nexus from the mounted disk.

## Updates

Any updates for the Danish Folktales, Legends, and Other Stories Digital Content will be listed below as it becomes available.



# Disadvantages

- Relatively fixed content
- Predetermined pathways through curated content
- Limited to curated content

# (Un)curated Collections

- Archives (or libraries) are often curated at a fairly high level
  - Metadata
  - Standard classifications
- Often have limited means for discovery
  - Focus on precision
- Goal of scholar often not easily served by these methods to access collections
  - Scholar often wants to generate a (dynamic) research corpus

# DFL 2.0: Faceted Browsing

The screenshot displays a web browser window with the URL <http://scandmacpro.humnet.ucla.edu/~pbroadwell/etkSpace/Documents/index.php#Column.SelectItemHistory%=%21>. The browser's address bar shows the URL, and the search bar contains the text "Google".

The application interface is titled "Danish Folklore: The Evald Tang Kristensen Collection". It features a navigation bar with links for Home, Help, and Feedback. Below the navigation bar, there are tabs for "Generate Topic Models" and "Capture a Screen Shot with Mac OS X".

The main content area displays a faceted browsing interface. The top section shows the current selection: "Kirsten Marie Pedersdatter / January 1894 / JAT\_IV\_212 / (Collection Place)". Below this, there are four columns of data: Person, Fieldtrip, Story, and Collection Place. Each column has a list of items with a vertical scrollbar. The "JAT\_IV\_212" item is highlighted in the "Fieldtrip" column.

Below the data columns, there is a section titled "Available Columns" with a red "X" icon. It lists "Place Mentioned" as an available column.

The main content area is titled "JAT\_IV\_212" in large orange text. Below the title, there are two tabs: "Data" and "Map". The "Data" tab is selected, showing a paragraph of text:

The foods served at this particular banquet would be considered unusual since they constitute everyday fare and not particularly exciting everyday fare at that. In this case, the poor food makes the celebration memorable. Kirsten Marie had previously told a story about another banquet that was memorable, in that case because the hosts should not have been able to afford an elaborate celebration (KMP 2.07). In both cases, she emphasizes the violation of expectations. In the published version, Tang Kristensen has separated these descriptions and, in so doing, erased the continuity of the descriptions and the clear sense of expectations for banquets.

Below the text, there is a section titled "English publication text:" followed by the text: "At one big birth celebration, they weren't served anything but (buckwheat) gruel and fish as far as I can remember."

# Repurposing other software

- Recognition that there are certain interfaces that our users already understand and with which they are comfortable
- Recognition that we do not have support to design ground-up interface
- Recognition that a “one-off” solution does not scale well to other projects
- Challenge: Can we adopt other interfaces for Humanities challenges?

# A sample search

etkSpace: The Evald Tang Kristensen Collection of Danish Folklore

http://etkspace.scandinavian.ucla.edu/etkSpace/index.php#Column.SelectItemHistory%=%26columnclicked%3D

Google

Pandora Radio - Listen to Free In... mac screen shot - Google Search etkSpace: The Evald Tang Kristen...

etkSpace

Home Help Start Over

ScratchPad

/ Churches and their construction. Monasteries, holy springs, etc. / (Places Mentioned) / (Story) / (Keywords)

Add Columns

ETK Index	Places Mentioned	Story	Keywords
About the Devil and being in league with him	Abildgård	DS_03_0_00357	å
Churches and their construction. Monasteries, holy springs, etc.	Åbjaerg	DS_03_0_00358	aa
Concerning earlier agricultural practices	Åbøl	DS_03_0_00359	aab
Cunning men and women and their activities	Åby	DS_03_0_00360	aaben
Death portents	Agerskov	DS_03_0_00361	aabnet
Diverse hauntings and conjuring of revenants	Agger	DS_03_0_00362	aabning
Diverse people	Aggersbøl	DS_03_0_00363	aal
Diverse place legends	Agrie	DS_03_0_00364	aale
Elves	Ajstrup	DS_03_0_00365	aar
Female revenants	Åkirkeby	DS_03_0_00366	aarmundrede
From the time of vilenage	Åkjær	DS_03_0_00367	aarsag
Heroes and their sport	Ål	DS_03_0_00368	aarstal
Household spirits	Ålborg	DS_03_0_00369	åben
and life in them	Åle	DS_03_0_00370	åbenbar
and cattle illnesses	Allersø	DS_03_0_00371	åbenbare
about farms and towns	Alling kloster	DS_03_0_00372	åbne

**Churches and their construction. Monasteries, holy springs, etc. (and) About the Devil and being in league with him (2318 results)**

**Stories Related to Your Selection**

Export metadata as: [XML](#), [CSV](#), [Plain Text](#)

Prev show: 10 | 20 | 50 | 100 | 200 Next

**DS\_03\_0\_00357**

**Publication Info:** DS\_03\_0\_00357  
**Place(s) Mentioned:** Jelling  
**Keywords:** kristi, bekomme, dommer, anno, jylland, herred, syssel, præsteindb  
**ETK Index:** Churches and their construction. Monasteries, holy springs, etc.

**DS\_03\_0\_00358**

**Publication Info:** DS\_03\_0\_00358  
**Place(s) Mentioned:** Romme  
**Keywords:** kirke, afstand, koge, sende, klat, omkreds, danne, fjerding, befolkning, rej, vendelbo, tilkomme, me, gabe, menighed  
**ETK Index:** Churches and their construction. Monasteries, holy springs, etc.

**DS\_03\_0\_00359**

Export



# Drill-down and export

etkSpace: The Evald Tang Kristensen Collection of Danish Folklore

http://etkspace.scandinavian.ucla.edu/etkSpace/index.php#Column.SelectItemHistory%=%26columnclicked%3D Google

Pandora Radio - Listen to Free In... mac screen shot - Google Search etkSpace: The Evald Tang Kristen...

Ch Churches and their construction. Monasteries, holy springs, etc. [1 more] / Jelling / DS\_03\_0\_00357 / (Keywords)

ETK Index Places Mentioned Story Keywords

About the Devil and being in league with him  
Churches and their construction. Monasteries, holy :  
Concerning earlier agricultural practices  
Cunning men and women and their activities  
Death portents  
Diverse hauntings and conjuring of revenants  
Diverse people  
Diverse place legends  
Elves  
Female revenants  
From the time of villeinage  
Heroes and their sport  
Household spirits  
Houses and life in them  
Human and cattle illnesses  
Legends about farms and towns

Jelling  
Jelling høje  
Jennikes Dam  
Jerslev broer  
Jerup  
Jetsmark  
Jordløse  
Jordløse  
Juelslund  
Juelstrup  
Julelund  
Julskov  
Junget  
Kalbygård  
Kallerup vad  
Kalundborg

DS\_03\_0\_00357  
DS\_03\_0\_00358  
DS\_03\_0\_00359  
DS\_03\_0\_00360  
DS\_03\_0\_00361  
DS\_03\_0\_00362  
DS\_03\_0\_00363  
DS\_03\_0\_00364  
DS\_03\_0\_00365  
DS\_03\_0\_00366  
DS\_03\_0\_00367  
DS\_03\_0\_00368  
DS\_03\_0\_00369  
DS\_03\_0\_00370  
DS\_03\_0\_00371  
DS\_03\_0\_00372

anno  
bekomme  
dommer  
herred  
jylland  
kristi  
præsteindb  
syssel

DS\_03\_0\_00357

Story text:

357. Torrilå herred haver sit navn af en dommer over Jylland ved navn Torrilå, som døde for kristi fødsel 1270. Jelling sogn ligger midt i herredet og er herredssognet, haver sit navn af Elling, en dommer, efter Torrilå i Jylland, som døde for kristi fødsel anno 1195. Efter hvilken dommer Jelling syssel også sin navn haver bekommet. Præsteindb. til Ole Worm.

Metadata:

Keywords: herred  
dommer  
bekomme  
præsteindb  
syssel  
jylland  
kristi  
anno

Places Mentioned: Jelling

Story: DS\_03\_0\_00357

ETK Index: Churches and their construction. Monasteries, holy springs, etc.

Export metadata as: XML, CSV, Plain Text

Export

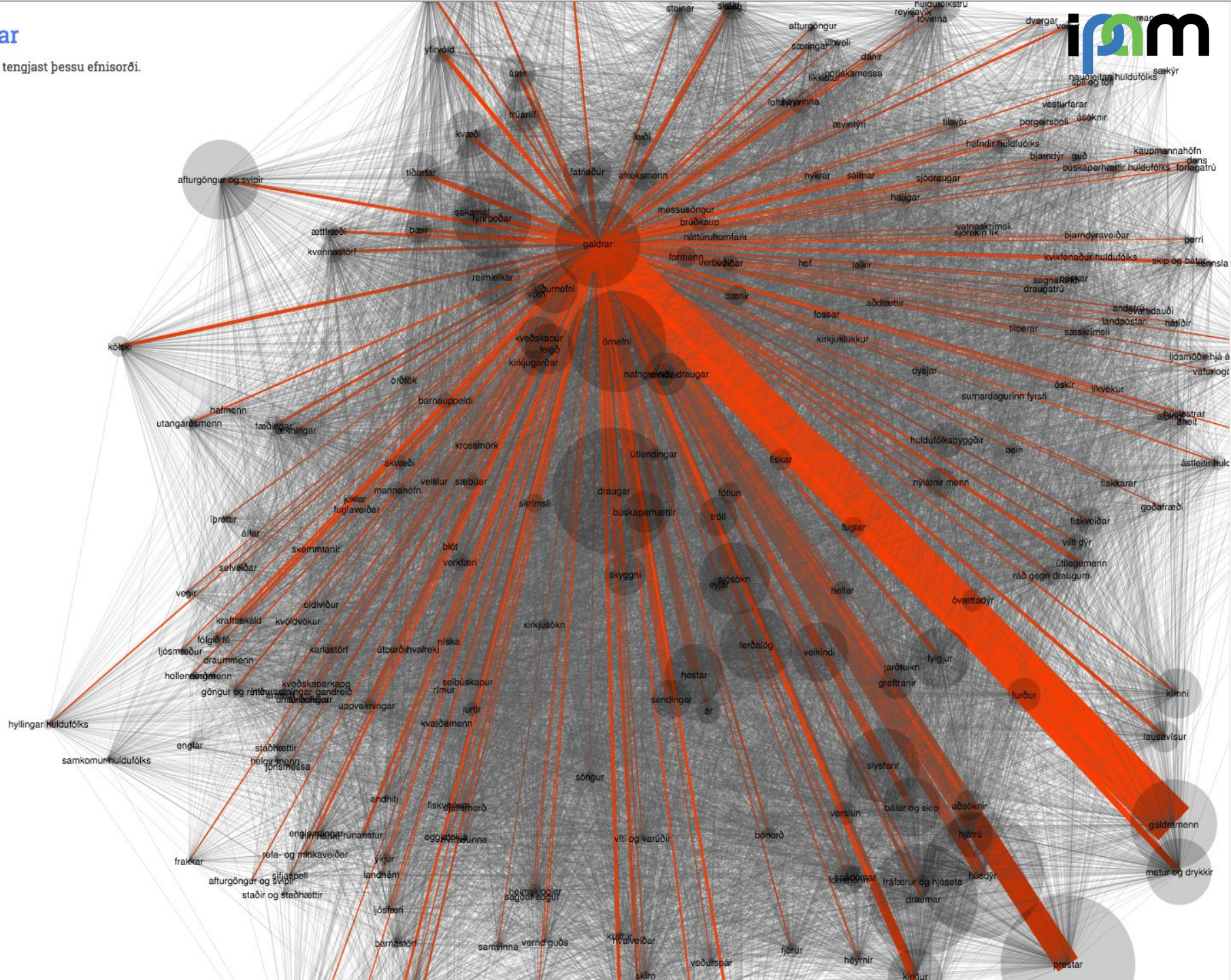
# Advantages of DFL 2.0

- Selection of a subset of the larger corpus
  - Subset can be
    - subjected to other types of analysis (e.g. GIS)
    - packaged for use in DFL 1.0
- User defined path
- Addressable searches
  - Allows for rapid reproduction of searches
- Extensions
  - “features” added to each of the main categories
    - Eg. add gender, occupation, dob to “people” category
    - Information derived from other collections
  - Use map generated bounding boxes to create a search

# Disadvantages

- Fitting a collection to the interface
- Faceted browsing presupposes that there is some fundamental unit (here the story) that is the goal of a search







# Accessing Analytics

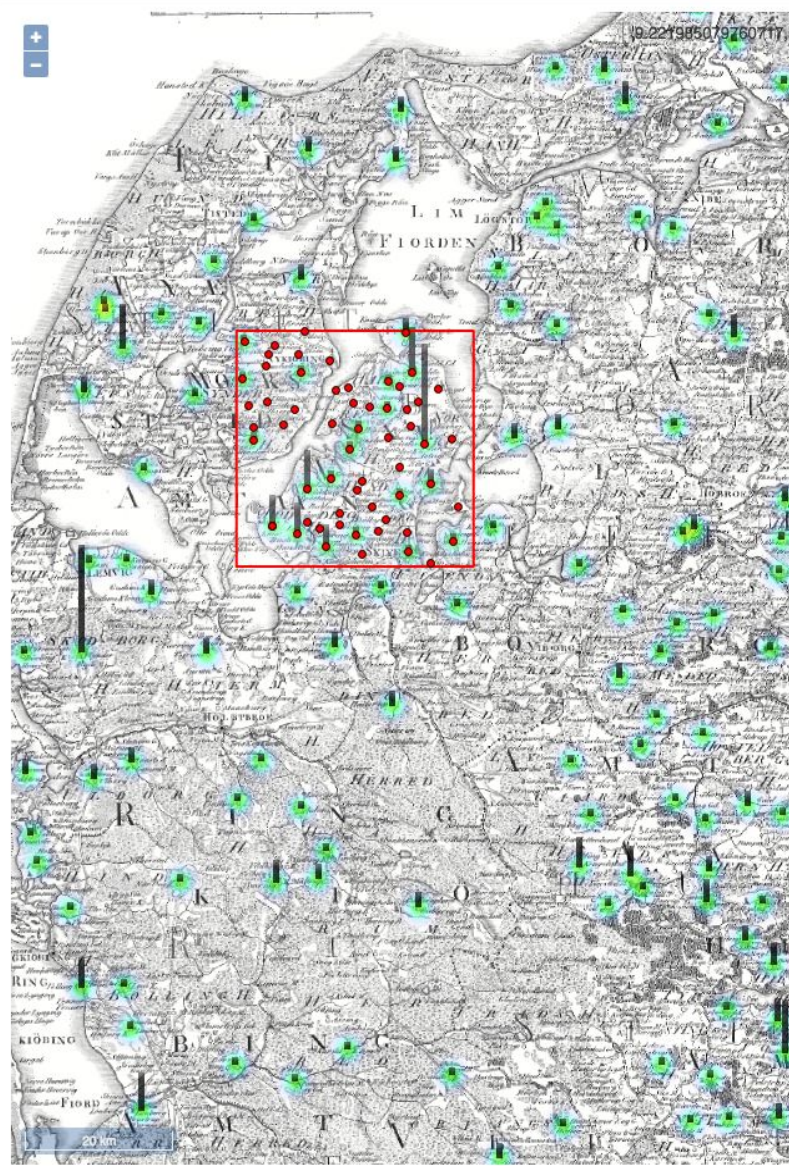
- Many current projects in folklore focus on access to archival materials or visualizations generated on the collection as a whole
- Some of our work has focused on creating a series of tools to access analyses of aspects of the collection in whole or in part
- “Operationalizing Distant Reading”...

# WitchHunter & TrollFinder

WitchHunter & TrollFinder: Mapping the Evald Tang Kristensen Collection

Help

- ☐ Mound dwellers (Hidden folk)
- ☐ Elves
- ☐ Household spirits
- ☐ Traveling monsters
- ☐ Water spirits
- ☐ Wiverns and small creepy-crawlies
- ☐ Werewolves and nightmares
- ☐ Religious legends
- ☐ Death portents
- ☐ Lights and portents
- ☐ Heroes and their sport
- ☐ Churches, monasteries, holy springs
- ☐ Legends about farms and towns
- ☐ Diverse place legends
- ☐ Legends about treasure
- ☐ Small kings and their feuds... Enemy invasions
- ☐ Manor lords, ladies and mistresses
- ☐ Ministers
- ☐ Diverse people
- ☐ Robbers, murderers and thieves
- ☐ Strandings
- ☐ Plague and illnesses
- ☐ Diverse hauntings and conjuring of revenants
- ☒ **Female revenants**
- ☐ Revenants on ill-gotten land
- ☐ Revenants in diverse places
- ☐ About the Devil and being in league with him
- ☐ Cunning men and women and their activities
- ☐ Witches and their sport
- ☐ Human and cattle illnesses
- ☐ Concerning earlier agricultural practices
- ☐ From the time of villeinage
- ☐ Houses and life in them
- ☐ Social gatherings and parties
- ☐ Outdoor life
- ☐ Our forebears' way of thinking and spiritual life



## Region bounding box:

Bottom left corner: 56.55201611368983N, 8.653442599291965E  
Top right corner: 56.8295428192738N, 9.175293185229465E

## Raw keyword counts:

- skrive: 33 times in region
- tigge: 26 times in region
- tage: 26 times in region
- opdage: 25 times in region
- spøge: 24 times in region
- bande: 23 times in region
- kjære: 22 times in region
- tænde: 22 times in region
- øre: 22 times in region
- regne: 21 times in region

## RF-IPF scores:

- skrive: 0.006036 (33 times in region, 374 times globally)
- tage: 0.005765 (26 times in region, 219 times globally)
- tigge: 0.005412 (26 times in region, 264 times globally)
- kjære: 0.004842 (22 times in region, 224 times globally)
- regne: 0.004742 (21 times in region, 207 times globally)
- spøge: 0.004683 (24 times in region, 316 times globally)
- bande: 0.004603 (23 times in region, 295 times globally)
- opdage: 0.004573 (25 times in region, 374 times globally)
- paste: 0.004516 (16 times in region, 95 times globally)
- ellekone: 0.004482 (12 times in region, 27 times globally)

## Normalized by global frequency:

- ålam: 1 (1 times in region, 1 times globally)
- furbo: 0.88 (7 times in region, 8 times globally)
- våre: 0.67 (2 times in region, 3 times globally)
- brask: 0.67 (8 times in region, 12 times globally)
- gie: 0.67 (2 times in region, 3 times globally)
- bønding: 0.6 (3 times in region, 5 times globally)
- strømpebånd: 0.5 (1 times in region, 2 times globally)
- østside: 0.5 (1 times in region, 2 times globally)
- kol: 0.5 (3 times in region, 6 times globally)
- fode: 0.5 (2 times in region, 4 times globally)



- ☐ Mound dwellers (Hidden folk)
- ☐ Elves
- ☐ Household spirits
- ☐ Traveling monsters
- ☐ Water spirits
- ☐ Wiverns and small creepy-crawlies
- ☐ Werewolves and nightmares

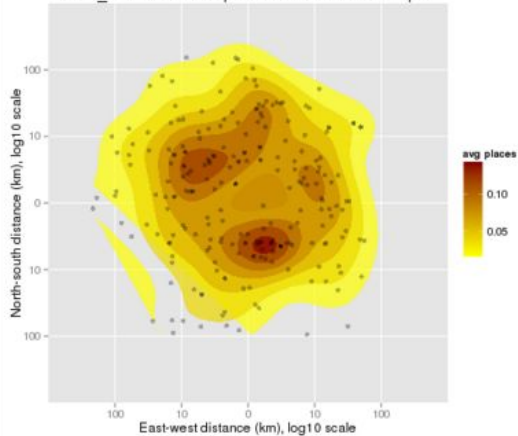
GhostScope & Treas

Storytellers' collective mental geography of this topic:



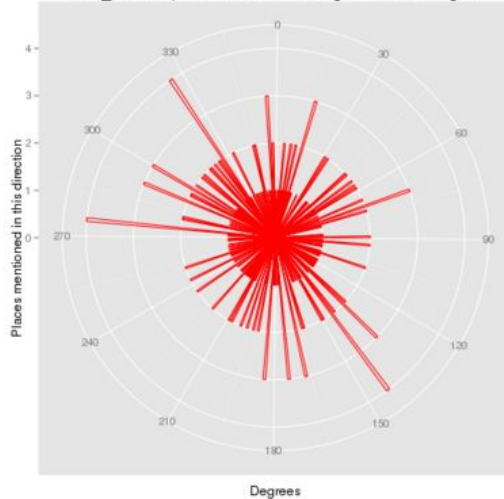
Help

ETK\_treasure relative places mentioned heat map 2



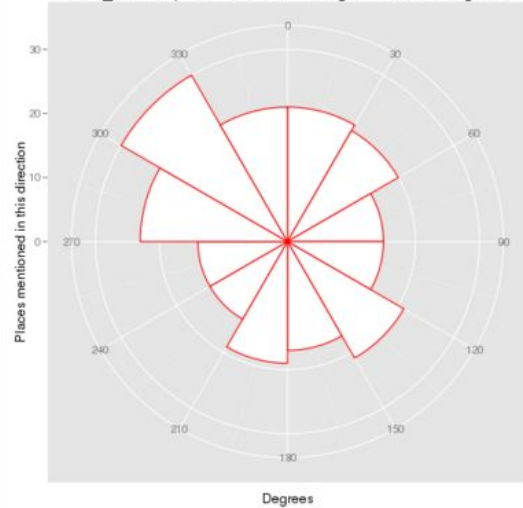
S  
y invasions

ETK\_treasure place mentioned bearings, bin size: 1 degree



revenants

ETK\_treasure place mentioned bearings, bin size: 30 degrees

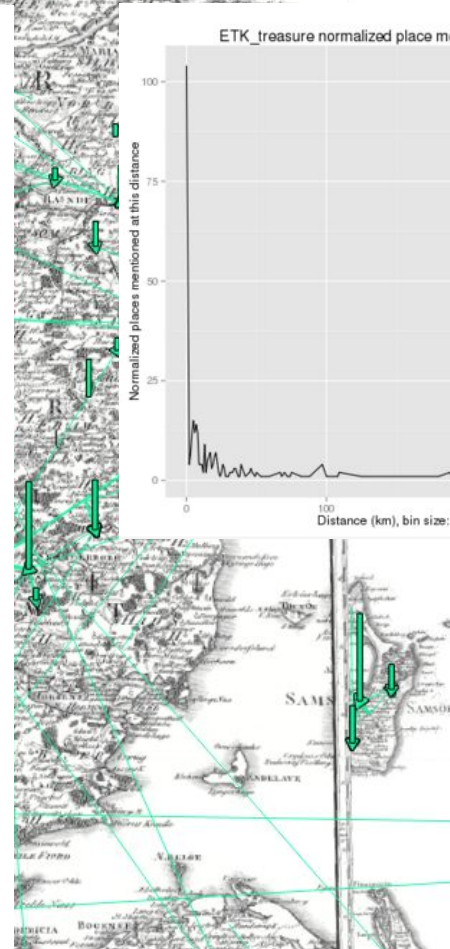
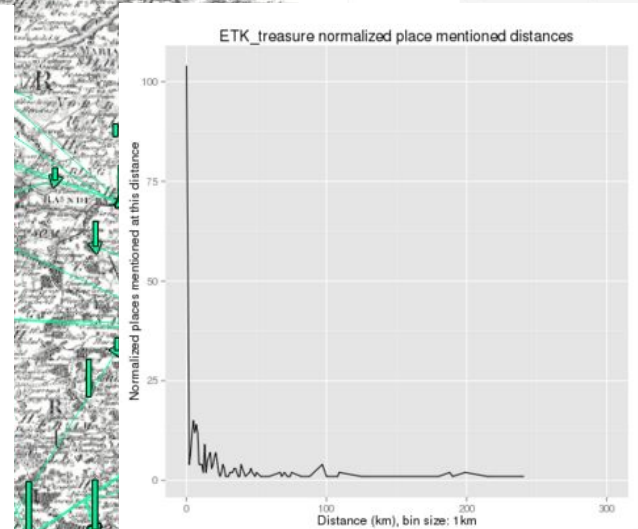
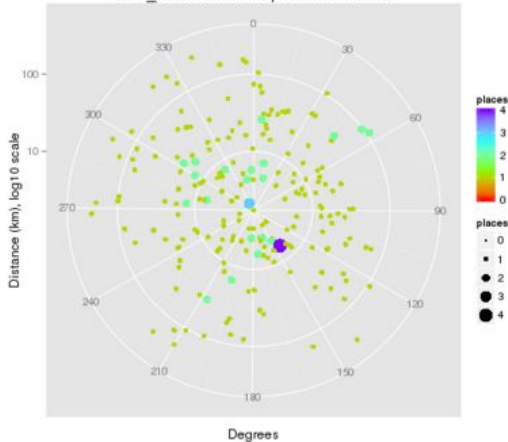


with him  
activities

ices

spiritual life

ETK\_treasure relative places mentioned



# ElfYelp

## Information about the selected region

### Region bounding box:

Bottom left corner: 56.65900196695813N, 8.910586220808362E

Top right corner: 56.806830983861545N, 9.180080806044218E

### Raw keyword counts:

1. bande: 12 times in region
2. kjære: 11 times in region
3. krig: 11 times in region
4. øre: 11 times in region
5. dige: 11 times in region
6. smide: 10 times in region
7. tabe: 10 times in region
8. gård: 10 times in region
9. slæbe: 10 times in region
10. mærke: 10 times in region

### RF-IPF scores:

1. tisse: 0.008658 (8 times in region, 26 times globally)
2. borggård: 0.008079 (9 times in region, 63 times globally)
3. brask: 0.007461 (6 times in region, 12 times globally)
4. kjære: 0.006964 (11 times in region, 224 times globally)
5. bande: 0.006908 (12 times in region, 295 times globally)
6. lavning: 0.00669 (9 times in region, 132 times globally)
7. satan: 0.006553 (9 times in region, 142 times globally)
8. grobe: 0.006542 (6 times in region, 25 times globally)
9. tage: 0.006378 (10 times in region, 219 times globally)
10. sølv: 0.006294 (9 times in region, 163 times globally)

### Normalized by global frequency:

1. brask: 0.5 (6 times in region, 12 times globally)
2. furbo: 0.5 (4 times in region, 8 times globally)
3. langkniv: 0.4 (2 times in region, 5 times globally)
4. lire: 0.4 (2 times in region, 5 times globally)
5. sile: 0.36 (4 times in region, 11 times globally)
6. gie: 0.33 (1 times in region, 3 times globally)
7. sok: 0.31 (5 times in region, 16 times globally)
8. tisse: 0.31 (8 times in region, 26 times globally)
9. snabe: 0.3 (3 times in region, 10 times globally)
10. kuske: 0.29 (2 times in region, 7 times globally)

### Mound dwellers, ghosts and Satan's influence (2.22%):

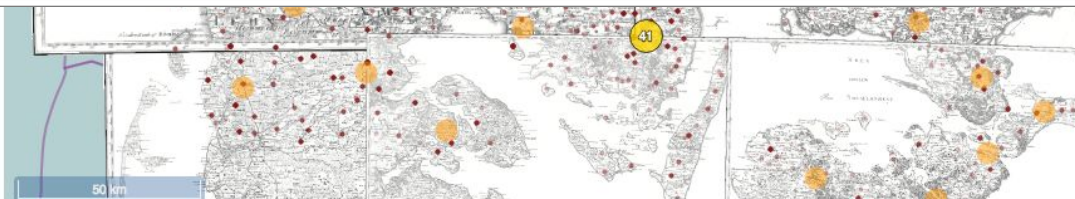
- Mound dwellers (Hidden folk): Cessation of the destruction of a mound (2.09%)
- Legends about farms and towns: Old named villages (2.02%)
- Mound dwellers (Hidden folk): Changelings, the old child (1.82%)
- Our forebears' way of thinking and spiritual life: Bad ministers (1.8%)
- Heroes and their sport: Giants build churches (1.73%)
- Heroes and their sport: Giants throw stones at churches (1.69%)
- Lights and portents: Lights for buildings, etc. Moving (1.6%)
- Outdoor life: Anecdotes from named nackers' lives (1.57%)
- Manor lords, ladies and mistresses: Blood spots (1.53%)
- Manor lords, ladies and mistresses: Walled in (1.36%)
- Female revenants: The murdered child, mother with the knife and washing it or the clothes (1.28%)
- Heroes and their sport: Dams (1.21%)
- About the Devil and being in league with him: Black dogs and the like show themselves (1.2%)
- Mound dwellers (Hidden folk): Smiths in mounds (1.17%)
- Mound dwellers (Hidden folk): Fording of the hidden folk (1.15%)

### Things that happen in fields (1.24%):

- Heroes and their sport: Stones and earth thrown at churches, builders disagree (4.76%)
- From the time of villeinage: Tenancy and purchase of farms (4.21%)
- Small kings and their feuds. Kings. Enemy invasions: Swedes north of Skanderborg and Silkeborg (3.59%)
- Diverse hauntings and conjuring of revenants: Black shapes (3.51%)
- Diverse hauntings and conjuring of revenants: Sheriffs, bailiffs, etc (3.17%)
- Our forebears' way of thinking and spiritual life: Confirmation preparations (3.07%)
- Female revenants: In with maidens (2.88%)
- Diverse place legends: Sounds, fjords, and sailways (2.78%)
- Religious legends: The Sybil's prophecy (2.75%)
- Cunning men and women and their activities: Wise Norwegians (2.58%)
- Traveling monsters: Odin's hunters go through a house or farm (2.47%)
- Outdoor life: Floods and sand movement (2.25%)
- Female revenants: Old hags (2.16%)
- Mound dwellers (Hidden folk): Hidden folks' silver and gold (2.11%)
- Diverse place legends: Sand movements (2.05%)

### Liminal areas and people or boundaries (0.57%):

- Churches and their construction. Monasteries, holy springs, etc.: Virgins build churches (5.99%)
- Outdoor life: Road- and postal service (2.73%)
- Diverse place legends: Old forests, Viborg (2.23%)
- Elves: The dance of the elves (2.19%)
- Diverse place legends: Sounds, fjords, and sailways (2.18%)
- Social gatherings and parties: Engagement and wedding (2.17%)
- Lights and portents: Lights in churches (2.77%)
- From the time of villeinage: Evil manor lords (2.21%)
- Social gatherings and parties: Birth and churching women, going to the altar, etc (2.19%)
- Witches and their sport: Witch burned on the pyre (2.04%)
- Social gatherings and parties: Lent, Easter, Pentecost, etc (1.88%)
- Diverse place legends: Sand movements (1.64%)
- Outdoor life: Various livelihoods (1.61%)
- Our forebears' way of thinking and spiritual life: Landscape descriptions (1.55%)



# Teaching Folklore



Getting Students to Think Like a Folklorist  
through Interface Design



## Folklore Archive

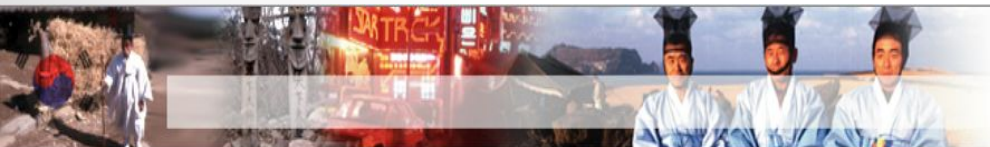
- Conceptualized as a student driven project
- Intends to teach about
  - Fieldwork
  - Archiving
  - Human Subjects Protection
- While generating an anonymized, usable dynamic dataset for the study of Korean American traditional expressive forms
- Two Goals: Pedagogy and Research

# Design drivers

- “Folklore is traditional expressive forms circulating in and across social networks embedded in time and place”
  - Implications:
    - Collectors
    - Informants (Consultants)
    - Contexts
    - Field data
- Study of folklore is based on the analysis of many variants
  - Implications
    - Classification
    - Historic-geographic search



# Search



KOREAN FOLKLORE  
*Online Archive*

[Home](#)  
[Search](#)  
[Admin](#)  
[Change Passcode](#)  
[Dashboard](#)  
[Add Consultant](#)  
[Add Context](#)  
[Add Field Data](#)  
[Archive](#)  
[Logout](#)

**Results** (Crawl by date of collection, most recent list)

**Collector**

Male, 28, College, student, english, korean

**Consultant**

Female, 25, College, student, English, Korean, Korean, LA, CA, USA

**Context**

January 1, 2012, afternoon, Sunny  
Business, Los angeles, ca, usa

**Data**

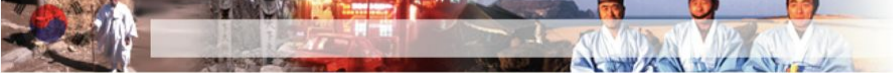
Fieldnotes

She did not know where the story originated, however, her version of the story, was more based on science rather than myth or superstition. During the deathly hot summer of her younger years in Korea, her mother had informed her about not leaving the fan on directly towards the face. If she had, she would die. The oxygen would be repelled away from the nose, thus the person would suffocate and die. When asked about leaving it on in another area of the room, she also stated that the door or window needed to be open to circulate oxygen.

Science

# The Folklore Collector

## Dashboard



**KOREAN** **FOLKLORE**  
*Online Archive*

[Home](#)  
[Search](#)  
[Admin](#)  
[Change Passcode](#)  
[Dashboard](#)  
[Add Consultant](#)  
[Add Context](#)  
[Add Field Data](#)  
[Archive](#)  
[Logout](#)

### DASHBOARD

<b>Collector Profile</b> Timothy Tangherlini tango@humnet.ucla.edu Roebling Ave, Los Angeles, CA 90024 USA DOB: 1593-10-02 Age: 417 Gender: Male Marital Status: married Occupation: used car salesman Educational Level: No Schooling Heritage: European Languages Spoken: English, Danish, Korean, and all other languages	<b>Contexts</b> <b>TT Birthday Party 1/16/2014 *</b> <a href="#">Delete</a> Date: 2014-01-16 Time: afternoon Description: Tommy's birthday party took place in abandoned wer... <b>Test *</b> <a href="#">Delete</a> Date: 0000-00-00 Time: afternoon Description: one two three...
<b>Consultants</b> <b>Tommy Tran *</b> <a href="#">Delete</a> <b>Jihoon Kim *</b> <a href="#">Delete</a> <b>Donald Trump</b> <a href="#">Delete</a>	<b>Data</b> <b>Hardcore Gang Birthday Parties</b> <a href="#">Delete</a> Description: TT: What did you think of the party? JH: Pretty rad. TT: I was scared, really, most of the time.... Filename: Screen Shot 2014-01-13 at 8.12.43 PM.png <b>Hardcore Gang Birthday Parties</b> <a href="#">Delete</a> Description: Photograph of rickety stairs leading to party hall.... Filename: Screen Shot 2014-01-13 at 8.25.56 PM.png <b>Hardcore Gang Birthday Parties</b> <a href="#">Delete</a> Description: blah Filename: <b>Test</b> <a href="#">Delete</a> Description: hello... Filename: 02_toc_final.pdf

# Collectors

[Home](#)  
[Search](#)  
[Admin](#)  
[Change Passcode](#)  
[Dashboard](#)  
[Add Consultant](#)  
[Add Context](#)  
[Add Field Data](#)  
[Archive](#)  
[Logout](#)

## COLLECTOR PROFILE

<b>Last Name: *</b>	<b>First Name: *</b>	<b>Initial:</b>
<input type="text" value="Tangherlini"/>	<input type="text" value="Timothy"/>	<input type="text" value="R"/>
<b>Email Address: *</b>		
<input type="text" value="tango@humnet.ucla.edu"/>		
<b>Street Address:</b>	<b>City:</b>	<b>State/Province:</b>
<input type="text" value="Roebling Ave"/>	<input type="text" value="Los Angeles"/>	<input type="text" value="CA"/>
<b>Zip/Postal Code:</b>	<b>Country:</b>	
<input type="text" value="90024"/>	<input type="text" value="USA"/>	
<b>DOB (yyyy/mm/dd): *</b>	<b>Age:</b>	
<input type="text" value="1593-10-02"/>	<input type="text" value="417"/>	
<b>Gender: *</b>	<b>Marital status: *</b>	<b>Occupation: *</b>
<input type="text" value="Male"/>	<input type="text" value="married"/>	<input type="text" value="used car salesman"/>
<b>Educational Level :</b>		
<input checked="" type="radio"/> No Schooling <input type="radio"/> Elementary School <input type="radio"/> Middle School		
<input type="radio"/> High School <input type="radio"/> College <input type="radio"/> Professional Degree		
<input type="radio"/> Graduate Degree		
<b>Heritage: *</b>		
<input type="checkbox"/> Korean <input type="checkbox"/> Japanese <input type="checkbox"/> Chinese <input type="checkbox"/> Vietnamese		
<input type="checkbox"/> Khmer <input type="checkbox"/> Lao <input type="checkbox"/> Thai <input type="checkbox"/> Philippino		
<input type="checkbox"/> Taiwanese <input type="checkbox"/> Burmese <input type="checkbox"/> Malay <input type="checkbox"/> Indonesian		
<input type="checkbox"/> Native American <input type="checkbox"/> Latino <input checked="" type="checkbox"/> European <input type="checkbox"/> African American		
<input type="checkbox"/> South Asian <input type="checkbox"/> Arabic <input type="checkbox"/> North African <input type="checkbox"/> Central African		
<input type="checkbox"/> South African <input type="checkbox"/> Pacific Islander <input type="checkbox"/> Central Asian <input type="checkbox"/> Central American		
<input type="checkbox"/> South American		
<b>Languages spoken: *</b>		
<input type="text" value="English, Danish, Korean, and all other langua"/>		
<input type="button" value="Submit"/> <input type="button" value="Cancel"/>		

# Consultants

## (not informants)

[Home](#)
[Search](#)
[Admin](#)
[Change Passcode](#)
[Dashboard](#)
[Add Consultant](#)
[Add Context](#)
[Add Field Data](#)
[Archive](#)
[Logout](#)

### CONSULTANT PROFILE

<b>Last Name: *</b>	<b>First Name: *</b>	<b>Initial:</b>
Tran	Tommy	
<b>Show Level: *</b>	None	

<b>Street Address:</b>	<b>City:</b>	<b>State/Province:</b>
1220 Levering Ave	Los Angeles	CA
<b>Zip/Postal Code:</b>	<b>Country:</b>	
90024	USA	

<b>Date of Birth (yyyy/mm/dd):</b>	<b>Age:</b>
1997-02-04	17

**Place of Birth:**

<b>City:</b>	<b>State/Province:</b>	<b>Country: *</b>
Cheju City		Korea

<b>Gender: *</b>	<b>Marital status:</b>	<b>Occupation:</b>
Male	its complicated	seahorse wrangler

**Educational Level :**

☐ No Schooling
 ☐ Elementary School
 ☐ Middle School
 ☐ High School
 ☐ College
 ☐ Professional Degree
 ☒ Graduate Degree

**Income Level:**

☒ 0 - 10,000
 ☐ 10,001 - 35,000
 ☐ 35,001 - 50,000
 ☐ 50,001 - 100,000
 ☐ 100,000+
 ☐ Decline to State

**Heritage: \***

☐ Korean
 ☐ Japanese
 ☒ Chinese
 ☒ Vietnamese
 ☐ Khmer
 ☐ Lao
 ☐ Thai
 ☐ Philippino
 ☐ Taiwanese
 ☐ Burmese
 ☐ Malay
 ☐ Indonesian
 ☐ Native American
 ☐ Latino
 ☐ European
 ☐ African American
 ☐ South Asian
 ☐ Arabic
 ☐ North African
 ☐ Central African
 ☐ South African
 ☐ Pacific Islander
 ☐ Central Asian
 ☐ Central American
 ☐ South American

<b>Date of Birth (yyyy/mm/dd):</b>	<b>Age:</b>
1993-12-09	22

**Place of Birth:**

<b>City:</b>	<b>State/Province:</b>	<b>Country: *</b>
Northridge	CA	United States

<b>Gender: *</b>	<b>Marital status:</b>	<b>Occupation:</b>
Female	Single	Student

**Educational Level :**

☐ No Schooling
 ☐ Elementary School
 ☐ Middle School
 ☐ High School
 ☒ College
 ☐ Professional Degree
 ☐ Graduate Degree

**Income Level:**

☒ 0 - 10,000
 ☐ 10,001 - 35,000
 ☐ 35,001 - 50,000
 ☐ 50,001 - 100,000
 ☐ 100,000+
 ☐ Decline to State

**Heritage: \***

☒ Korean
 ☐ Japanese
 ☐ Chinese
 ☐ Vietnamese
 ☐ Khmer
 ☐ Lao
 ☐ Thai
 ☐ Philippino
 ☐ Taiwanese
 ☐ Burmese
 ☐ Malay
 ☐ Indonesian
 ☐ Native American
 ☐ Latino
 ☒ European
 ☐ African American
 ☐ South Asian
 ☐ Arabic
 ☐ North African
 ☐ Central African
 ☐ South African
 ☐ Pacific Islander
 ☐ Central Asian
 ☐ Central American
 ☐ South American

**Languages spoken: \***

English

**Did you immigrate to the United States? \***

☐ Yes
 ☒ No

**If yes, Date of Immigration (yyyy/mm/dd):**

0000-00-00

**Uploaded Consent Form:**

File Name: 20160302\_090332.jpeg      File Type: image/jpeg      File Size: 1092405 Bytes

**Upload a new consent form to replace the existing one:**

No file chosen



# Context(s)

## CONTEXT

### Event or Interview Title: \*

Fan Death Interview- 01/18/16

### Type of event or expression: (Check all that apply) \*

#### ☐ a. Celebration

☐ Birthday ☐ Seasonal/Holiday ☐ Wedding ☐ Funeral ☐ Graduation ☐ Other Celebration

#### ☐ b. Performance

☐ Oral History ☒ Storytelling ☐ Folk Speech/Gesture ☐ Drama ☐ Song ☐ Dance ☐ Other Performance

#### ☐ c. Material Culture

☐ Architecture ☐ Costume/Clothing ☐ Body /

#### ☒ d. General Observation

### Time of collection: \*

☐ Morning ☒ Afternoon ☐ Evening ☐ Night

### Weather on the day of collection: \*

☐ Sunny ☒ Overcast

### Language(s) of collection: \*

☐ Korean ☐ Japanese ☐ Portuguese ☐ Other

### If other, please specify:

### Place of collection: \*

☐ Business ☒ Residence ☐ Public Place

### Address of Collection:

#### Street Address:

Poner Street

#### City:

La Mirada

#### State/Province:

CA

#### Zipcode/Postal Code:

90638

#### Country:

United States

### If the address of collection is unknown, please enter GPS Info of Collection:

#### Latitude:

Direction:

Please Select

Degrees [0, 180]:

0

Minutes Seconds:

0

#### Longitude:

Direction:

Please Select

Degrees [0, 180]:

0

Minutes Seconds:

0

This interview took place in a residence in the afternoon on an overcast day which smelled of imminent rain.

The residence was enveloped in a soft blanket of silence. Considering that it was a holiday, the streets were fairly silent, except for the appearance of a random ice cream truck blaring out promises of delicious sweets to lure children into buying mediocre ice cream. The occasional greetings of people visiting the neighbors' houses could be heard from the distance.

Lin Lee

☐ Helen Kim

Van Jeong

☐ Ashley Feldman

Helle Chang

☐ Grace Lee

me-age-as

☐ Older than main consultant

☐ Still Camera

☒ Notes

view took place in a residence in the afternoon on st day which smelled of imminent rain.

The residence was enveloped in a soft blanket of silence. Considering that it was a holiday, the streets were fairly silent, except for the appearance of a random ice cream truck blaring out promises of delicious sweets to lure children into buying mediocre ice cream. The occasional greetings of people visiting the neighbors' houses could be heard from the distance.

Submit

Cancel

# Field Data

Home  
Search  
Admin  
Change Passcode  
Dashboard  
Add Consultant  
Add Context  
Add Field Data  
Archive  
Logout

## FIELD DATA

### Title of project: \*

Fan Death Interview: Short Fieldwork Exercise

### Context: \*

- ☐ Feb 8, 2016- Comfort Food Interview 1  
☐ Fan Death Interview- 01/18/16  
☐ Holiday (Chu Seok) Interview- 01/15/16  
☒ Fan Death Interview - 01/16/16  
☐ Feb. 8, 2016- Comfort Food Interview 2  
☐ Feb. 10, 2016- Comfort Food Interview 3  
☐ Feb. 13, 2016- Comfort Food Interview 4  
☐ Feb. 14, 2016- Comfort Food Interview 5  
☐ March 1, 2016- Comfort Food Interview 6

### Collection Media:

- ☒ Fieldnotes  
☐ Images  
☐ Video  
☐ Audio

### Data Description: \*

An interview about the the

Collector: Can you explain ...  
M: If you keep the fan on while you sleep, you can die because the fan lowers your body temperature.

Collector: As a person who studied science, do you think this theory has any scientific basis?  
M: I'm not sure. I've heard stuff on the news but I personally don't know anyone who has been hurt from it.

### Upload Media File (<100MB):

Choose File No file chosen

Submit Cancel

Collector: Can you explain what fan death is?  
M: If you keep the fan on while you sleep, you can die because the fan lowers your body temperature.  
  
Collector: As a person who studied science, do you think this theory has any scientific basis?  
M: I'm not sure. I've heard stuff on the news but I personally don't know anyone who has been hurt from it.

# Summary Challenges

- Folklore collections are often organized according to limiting indexing systems or idiosyncratic organizations
- Goal is to be able to (a) navigate large collections based on researcher determined criteria and (b) discover patterns at scale that can lead to (c) the selection of resources for further examination
- Intuitive interfaces that operate at human-interactive timescales
- Making apparent the underlying organizational schema to help users infer limitations

# Sub-Corpus Topic Modeling

• • •

Challenges for Search and Discovery in the Study of Literature

With Peter Leonard (Yale)



# General flow

## Sub-Corpus Topic Modeling (STM)

1 Sub-corpus  
Selection



2 Chunking



3 Sub-corpus  
Modeling (LDA)



4 STM curation



6 Expert input /  
refinement



5 STM based search



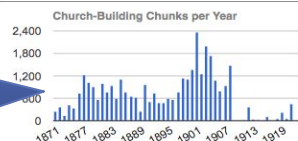
# Topic curation



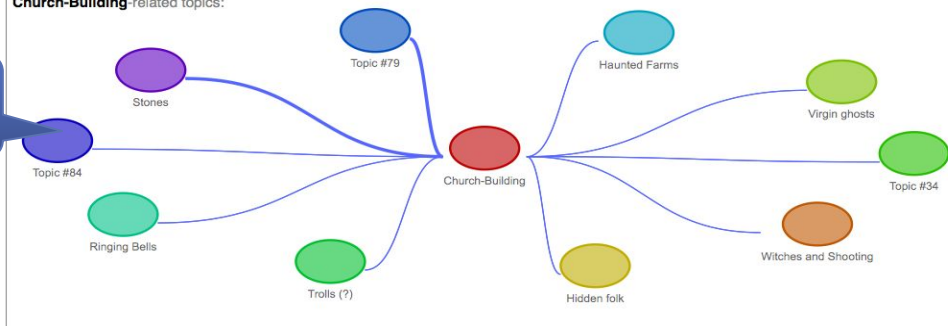
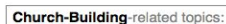
**Church-Building** words:

A word cloud of Danish words related to church building. The most prominent words are 'kirken' (the church) and 'kirke' (church), both in large red font. Other words include 'byggede' (built), 'bygget' (built), 'bygges' (being built), 'tårnet' (the tower), 'første' (first), 'alteret' (the altar), 'natten' (the night), 'færdig' (finished), 'søndag' (Sunday), 'sognet' (the parish), 'kirket' (church), 'navn' (name), 'et' (a), 'står' (stands), 'kirkegården' (the churchyard), and 'bygger' (builder).

Chunks  
per year



## Related topics



**Church-Building** phrases:

A word cloud of Norwegian church-related terms. The most prominent words are 'kirke bygget' and 'kirken bygget', both in large blue font. Other visible words include 'byggede kirken', 'kirkes tårn', 'byggede kirke', 'kirken alteret', 'kirke henger', 'bygget kirke', 'kirken natten', 'kirken færdig', and 'byggede kirke'. The words are arranged in a somewhat circular pattern, with 'byggede kirken' at the top and 'kirken bygget' at the bottom.

Book Chunks 1871-1922:

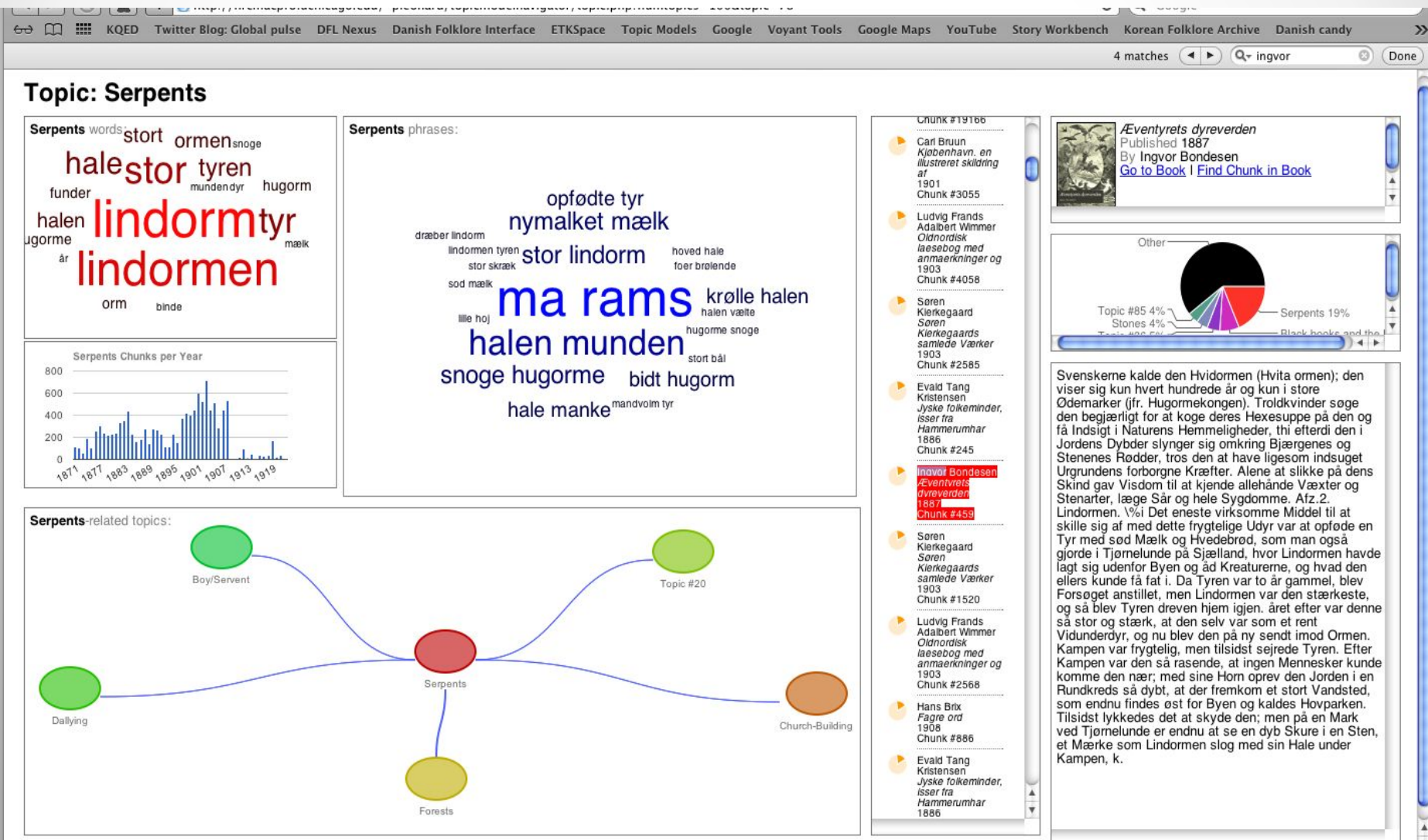
-  Ewald Tang  
Kristensen  
Jyske folkeminder,  
isser fra  
Hammerumhø  
1986  
Chuk # 292
-  Ewald Tang  
Kristensen:  
Samfundet til den  
jyske folkeminder,  
isser fra  
Hammerum-Hø  
1886  
Chuk # 282
-  Henrik Carl Bering  
Lisberg  
Kortlægning i gamle  
dage og livet i  
Købe  
1901  
Chuk # 1542
-  Ewald Tang  
Kristensen;  
Samfundet til den  
jyske folkeminder,  
isser fra  
Hammerum-Hø  
1886  
Chuk # 85
-  Vendtsyssel folk og  
land  
1908  
Chuk # 662
-  Frederik Reinholdt  
Frils  
Bibliotek til Dansk  
kusthistorie  
1901  
Chuk # 3068
-  Ewald Tang  
Kristensen  
Jyske folkeminder,  
isser fra  
Hammerumhø  
1886  
Chuk # 110
-  Ewald Tang  
Kristensen  
Jyske folkeminder,  
isser fra  
Hammerumhø  
1886  
Chuk # 86

*Jyske folkeminder, isser fra  
Hammerumhæred: Sagn og overtro fra  
Jylland, 2.-3. samling*  
Published 1886  
By Evald Tang Kristensen  
[Go to Book](#) | [Find Chunk in Book](#)



149. For skjonhed og for syns skyld- To jomfruer byggede Sal og Ejsing kirker. Den, der byggede den sidstnævnte, byggede for skjOnheden, og den har da ogsaa forhen været meget skjOn; men Sal kirke byggedes for at være synet, og tårnet har ogsaa været uhyre n. Kæmper, kiinger, kirker m. m. 77

# Forgotten Works: Ingvor Bondesen





# Extensions: Finding Thor

## Topic Modeling Old Norse

① Source: Edda Snorra Sturlusonar Topics: 50 20 Stoplisted ☒ ③ Target: Heimskringla Go

Zero words:

þórs  
stóð þór gekk  
segja  
varð  
hefir þórr hér  
þau  
útgarða-loki of aftr  
engi skal

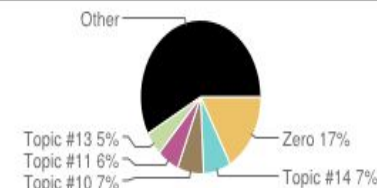
Zero phrases:

þórr jattu  
þórr hafrá  
þórr rerir  
komit hefða hefir verit sannara segja  
nókkut munr  
segja stóð þórr  
rð þórr reiðr

Zero chunks:

Óláfs konungs  
kyrra Chunk 8  
Haralds saga  
hins hárfagra  
Chunk 25  
Haralds saga  
hins hárfagra  
Chunk 73  
Haraldi konungi  
gráteld  
Chunk 17  
Haralds saga  
hins hárfagra  
Chunk 6  
Magnús konungs  
húðr

Heimskringla



Jartegnir Óláfs konungs. Sá atburðr gerðisk í Niðarósi, at skrin Óláfs konungs var borit um stræti, at hofugt varð skrinít svá at eigi fengu menn borit fram ór stað; en síðan var skrinít niðr sett ok brotit upp strætít ok leitát, hvat þar var undir, ok fanzk þar barns lík, er myrt hafði verit ok fólgt þar; var þat þá á brot borit, en bætt apt'r strætít, svá sem áðr hafði verit, en borit skrin at vanda.

Topic #1 words:

mjóð hér  
grundar  
taka jarl of  
lætr  
einarr gramr  
brynju gramr  
hiakkar  
skálaglamm  
gefr  
gram  
her

Topic #1 phrases:

jarl lætr  
gram svinnum guilsendir  
seggjum dáðrókkum veidr dauða  
of lög  
vísir æfr  
of ítran hlími  
tyggja dýrum  
skerr hvatt  
glymvinði lætr  
kristr skóp

Topic #1 chunks:

Magnús konungs  
berfœttis  
Chunk 30  
Magnús blinda  
Chunk 22  
Haralds saga  
hins hárfagra  
Chunk 18  
Haraldi konungi  
gráteld  
Chunk 35

# STM 2.0

- Our original STM machine was a proof of concept machine
  - Finely tuned to the target corpus
  - Preprocessing and modeling done offline and refined multiple times
  - It provided excellent and unexpected results:
    - Discovered discourse in the social science literature that reflected the growing importance of Social Darwinism in late 19<sup>th</sup> and early 20<sup>th</sup> century Denmark
    - Provided an interesting method to confirm the findings of Pil Dahlerup in her mid 1980s “Det moderne gennembruds kvinder”, a rejoinder to the Georg Brandes’s classic, “Det moderne gennembruds mænd”



# Motivating question(s):

- How can we make this
- **Broadly available**,
- **Flexible** to the heterogeneous textual data with which Humanities researchers work,
- Provide **results interpretable** by people new to the field of computational Humanities

# Multi-step process

- Upload
  - Corpus
  - Sub-corpus
- Preprocessing
- Modeling
- Curation
  - Labeling
  - Hiding topics
  - Multi-level modeling (differing levels of granularity)
- Exploration
  - visualization

# Upload

- Devise a corpus or subcorpus
- Upload it to the site

# Corpus upload

STM, Scandinavian Section, UCLA

Curation

Topics

Manage Corpus

Topic Modeling

## Metadata

Corpus  
name

Corpus

Subcorpus

niels\_lyhne



Add files

## Add Files (929 files in corpus)

Drop files here to upload

☐ Already Preprocessed

## Delete Corpus

Delete existing  
corpus

Delete



# Topic browsing og curation

STM, Scandinavian Section, UCLA

Curation Topics

Browse Graph

🏠 > Topics > niels\_lyhne\_stopped (corpus)

Topics Articles

**Inferencer** niels\_lyhne\_stopped (corpus) ✓

Number of Topics

50

100

75

25

Valg af modellets granularitet

# Preprocess

- Chunking (currently off-line)
  - By paragraph
    - Our original STM approach
    - See also Algee-Hewitt, Heuser and Moretti (2015)
  - By page
  - By chapter
  - User defined chunking
- Stop-words
- Tokenizing and/or stemming

# Preprocess

STM, Scandinavian Section, UCLA

Curation

Topics

## Inference

Infer topics on self

Infer topics from ...

## Preprocessing

### Regex Token

Latin

Classical Chinese

Custom

`\p{L}[\p{L}\p{P}]*\p{L}`

### Stopwords

English

Custom

Choice of regex  
tokenization

Choice of  
stopword files

## Process

### Save as

niels\_lyhne (corpus)



Process

# Model the corpus

- Model the corpus
  - Choose level of granularity



# Modeling

(currently LDA, plan to add other topic models)

I, Scandinavian Section, UCLA

Curation

Topics

Language Corpus

Topic Modeling

Process New Ingested Corpus

Select Existing Ingested Corpus

## Ingested Corpus

niels\_lyhne\_stopped (corpus) ✓

Ingested from corpus **niels\_lyhne**. Inferring topics on self.

Choose  
corpus

Model type

## Topic Modeling

Number of Topics

25 ✓

Granularity

Building Inferencer for **niels\_lyhne\_stopped (corpus)** ✓

Inferring Topic Saturation in **niels\_lyhne\_stopped (corpus)** ✓

# Infer topics

- On the main corpus based on the subcorpus
- Standard topic modeling—
  - Make the subcorpus = main corpus

# Build model

n Section, UCLA

Cu

Topic Modeling

Process New Ingested Corpus

Select Existing Ingested Corpus

Choose  
corpus

Select Corpus

Corpus

Subcorpus

niels\_lyhne



Inference

Infer topics on self

Infer topics from ...

Choose  
modeling

# Curation

- Labeling and hiding topics



# Multi-level topic curation

M, Scandinavian Section, UCLA

Curation

Topics

rowse

Graph

Topic granularity

> Topics > niels\_lyhne\_stopped (corpus) > 25 topics

Topics in niels\_lyhne\_stopped (corpus) (25 topics)



Show / hide  
topics

Håb, liv og selvet



vil, skal, maa, ...

end, selv, kun, ...



vilde, skulde, saadan, ...



saae, lille, samme, ...



niels, meget, aar, ...



fordi, end, ganske, ...



efter, dag, ligesom, ...



blev, kom, igjen, ...



Kærlighed



# Topic browsing and search

## Top Phrases

### falskhed forstillelse

højt loftet  
 sveget ideal haanlig  
 elskov kærlighed  
 mandens ideal  
 ædlere end elskede  
 ydmyge bøje knæ  
 elskede hele  
 siden busten  
 hvem mødt  
 hele sjæl igjennem  
 al glans al renheds  
 muligt end  
 forglemmelse selv stolthed ydmyghed  
 egnes skjønhed  
 derunder selv  
 mod dens  
 fremmed grund  
 sorg langt  
 siden gardinet

recommender

graph

## Related Topics

alle, selv, samme, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



ret, ganske, gamle, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



laa, meget, graa, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



blev, saae, efter, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



haand, ord, hen, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



blev, mellem, tale, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



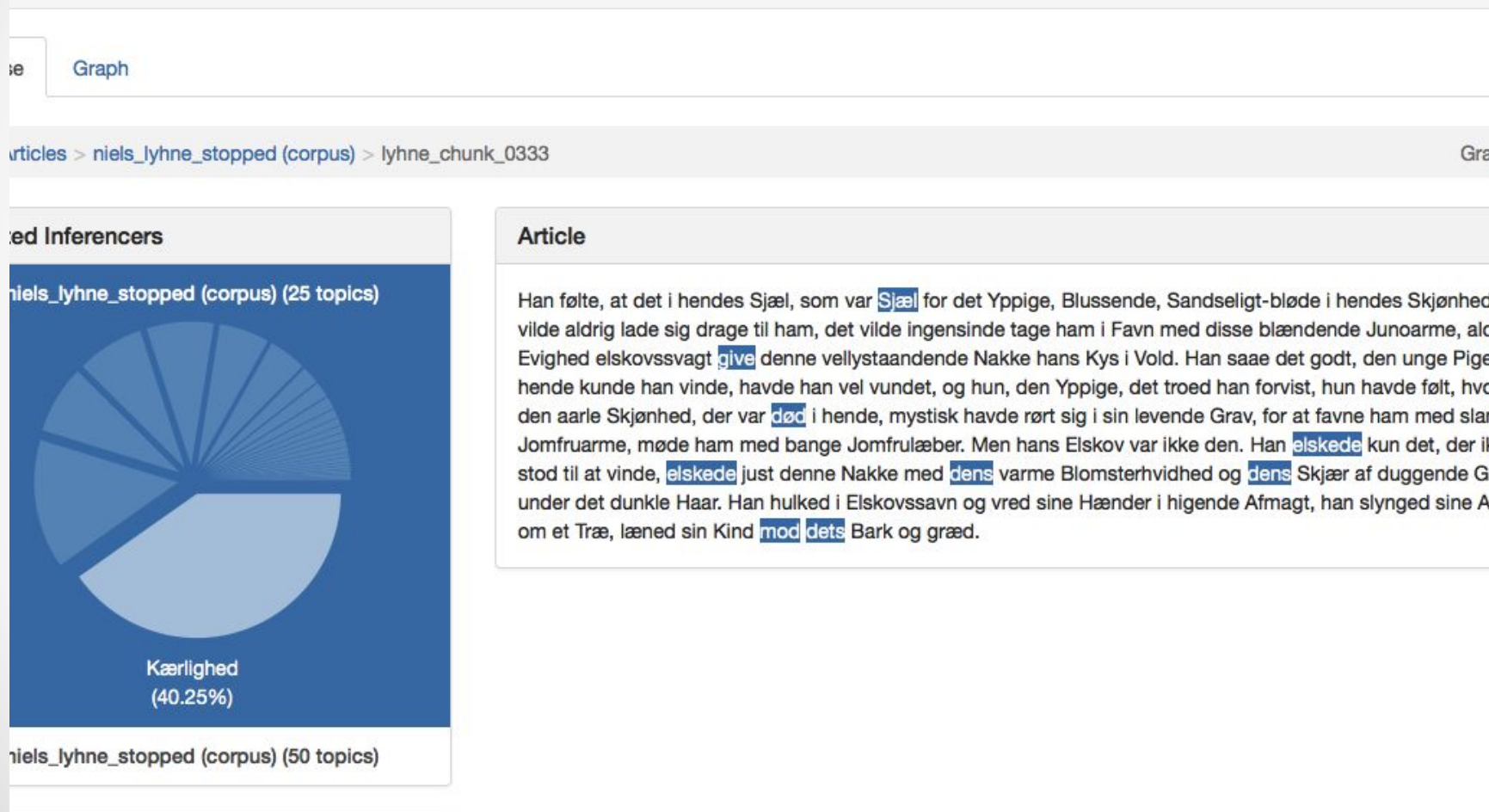
sige, vilde, maa, ...  
of niels\_lyhne\_stopped (corpus) (50 topics)



# Navigation and Discovery

- Topic saturation per chunk
- Discovering similar chunks
- Network representation

# Topic explorer





# Networks of topics

STM, Scandinavian Section, UCLA

Curation

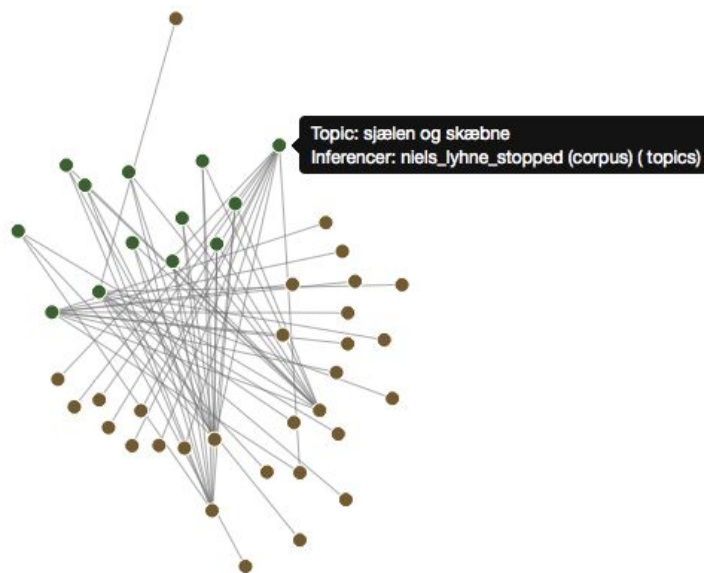
Topics

Browse

Graph

Number of nodes to expand

10



# Ongoing challenges

- Add user login and profiles and sharing so that STM can be a private / collaborative space
- Move the computation into the cloud, so that creating models is not as time intensive as it currently is
- Make large collections freely available for:
  - Subcorpus selection and creation
  - Corpus search

# Beta version of Search/Selection

## Danish Books 1860-1870

73 Books filtered from 1605 originally ([Reset All Filters](#))

sorted by: [labels](#); [then by...](#) • ☐ grouped as sorted

1 • 2 • 3 • 4 [Next »](#)

### *Breve fra og til Johan Ludvig Heiberg*

By Johan L. Heiberg  
Reitzel, 1862. 240 pages

1. 



*Breve fra og til Johan Ludvig Heiberg*  
By Johan Ludvig Heiberg and Thomasine Gyllembourg  
C.A. Reitzels forlag, 1862. 240 pages



2.



### *Johan Ludvig Heibergs poetiske skrifter*

By Johan Ludvig Heiberg  
C. A. Reitzel, 1862. 496 pages



3.



### *Johan Ludvig Heibergs poetiske skrifter*

By Johan Ludvig Heiberg  
C. A. Reitzel, 1862. 872 pages



4.



### *Johan Ludvig Heibergs poetiske skrifter*

By Johan Ludvig Heiberg  
C. A. Reitzel, 1862. 500 pages



Type an Author's name...

heiberg

...or select from a list

- 60 [Johan Ludvig Heiberg](#)
- 12 [Johan L. Heiberg](#)
- 1 [Fiske Icelandic Collection](#)
- 1 [Peder Andreas Christian Heiberg](#)
- 1 [Thomasine Gyllembourg](#)

Type a Publisher...

...or select from a list

- 24 [C. A. Reitzel](#)
- 18 [Reitzel](#)
- 12 [C.A. Reitzel](#)
- 5 [C. A. Reitzel.](#)
- 1 [C.A. Reitzels forlag](#)
- 1 [P. G. Philipsen](#)
- 1 [Reitzels Forlag](#)
- 1 [W. Prior](#)

Publication Date

1856 - 1877

Full Title Search

Type a Subject...

...or select from a list

- 17 [History / General](#)
- 15 [Danish literature](#)
- 15 [Drama](#)
- 15 [Literature](#)
- 11 [Drama / English, Irish, Scottish, Welsh](#)
- 11 [Drama / General](#)
- 1 [Diatoms](#)
- 1 [Mythology, Norse](#)
- 1 [Performing Arts / Theater / General](#)
- 1 [Performing Arts / Theater / Stagecraft](#)
- 1 [Poets, Danish](#)
- 1 [Theaters](#)



# Challenges in the representation of Old Norse-Icelandic

...

IceMorph

<http://tango.bol.ucla.edu/andreas09/index.html>

Work with Kryztof Urban (DataCrush), Aurelijus Vijunas (Natl Kaohsiung Univ), Peter M Broadwell



### Launch Tools

MORPHOLOGICAL  
ANALYZER

INFLECTED DICTIONARY  
BROWSER

## ICEMorph: A Morphological Analysis Tool for Old Icelandic



ICEMorph is a morphological analysis and look-up tool for the study of Old Norse-Icelandic, among the most morphologically complex of the Germanic languages.

The analysis tool uses a functional programming language, FM/Haskell, to tackle the problem of this complexity. Machine

## ICEMorph: Download



ICEMorph is a morphological analysis and look-up tool for the study of Old Norse-Icelandic, among the most morphologically complex of the Germanic languages.

The analysis tool uses a functional programming language, FM/Haskell, to tackle the problem of this complexity. Machine learning algorithms boost the performance of the system.

The look-up tool is based on two of the most important Old Icelandic-English dictionaries: Cleasby-Vigfusson *An Icelandic-English Dictionary* (1874), and Zoëga *Old Icelandic Dictionary* (1910).

## Download

You can download the latest version of the tagged corpus below.

ICEMorph and the tagged corpus is covered by the **Creative Commons Open Archives non-commercial "by" license**. The current citation for ICEMorph and the tagged corpus is:

Tangherlini, Timothy R., Kryztof Urban, Aurelijus Vijunas. 2013. *ICEMorph: An Automated Morphological Analysis System and Dictionary Look-up Tool for Old Icelandic*. <http://icemorph.scandinavian.ucla.edu>

The tagged texts are based on:

Guðni Jónsson and Bjarni Vilhjálmsson. 1943-1944. *Fornaldarsögur Norðurlanda*. Reykjavík : Bókaútgáfan Forni.

Click on the link below to download the current tagged corpus.

**Fornaldarsögur Norðurlanda**



# Why is this hard?

- Old Icelandic is *highly inflected*
  - Making search quite difficult
  - Making learning it quite difficult
- Old Icelandic corpus is normalized
  - To different normalization standards
  - Scholars often want to work with variant readings
- Challenge:
  - Can we devise tools to make accurate/sophisticate research a bit easier
  - Can we help students learn and give non-experts access

Challenges

- NOUNS
- VERBS

Vowel Changes: Umlaut and Breaking

A large part of the complexity of OIc. morphology is due to activity of the so-called "umlauts" and "breakings". Whereas the previous term is merely a synonym of "vowel assimilation", used in a more narrow sense, the interpretation of the latter term, "breaking", has been a matter of debate, and the final solution has not been proposed yet (although it seems to show some similarities to umlauts).

Dissimilation

Dissimilation is fairly rare in Icelandic, yet it has contributed to the development of several irregular paradigms, cf. below:

maðr 'man' (sg.)	teðr/tennr 'teeth' (pl.)
Nom.maðr -- *mann-r	teðr/tennr: teðr -- *tenn-r, whereas the by-form tennr is analogical
Accmann	teðr/tennr
Dat. manni	to , nnum
Gen.man(n)s	tanna

Degemination

Degemination occurs commonly in consonant clusters of the type -C1C1C2(C3)-, cf. sökva 'sink' (transitive v.) vs. sökpa 'id.' (1. p. sg. preterite), or spenna 'to make tense' vs. spenta 'id.' (1. p. sg. pret.), etc. In numerous instances geminates are restored contrary to what seems to be a phonological reduction rule, cf. sökta, spennta, etc. The reason for the restoration of geminates is primarily analogy, to be discussed in 2. below. In certain instances of a different type, cf. leggja 'lay' (inf.) or 'legs' (gen. pl.), bekkja 'benches' (gen. pl.), etc., the spelling would seem to be calling for degemination. However, it does not happen, because the consonant j in such verbs only denotes the frontness of the preceding stop /g' :/, and does not represent an individual phoneme /j/.

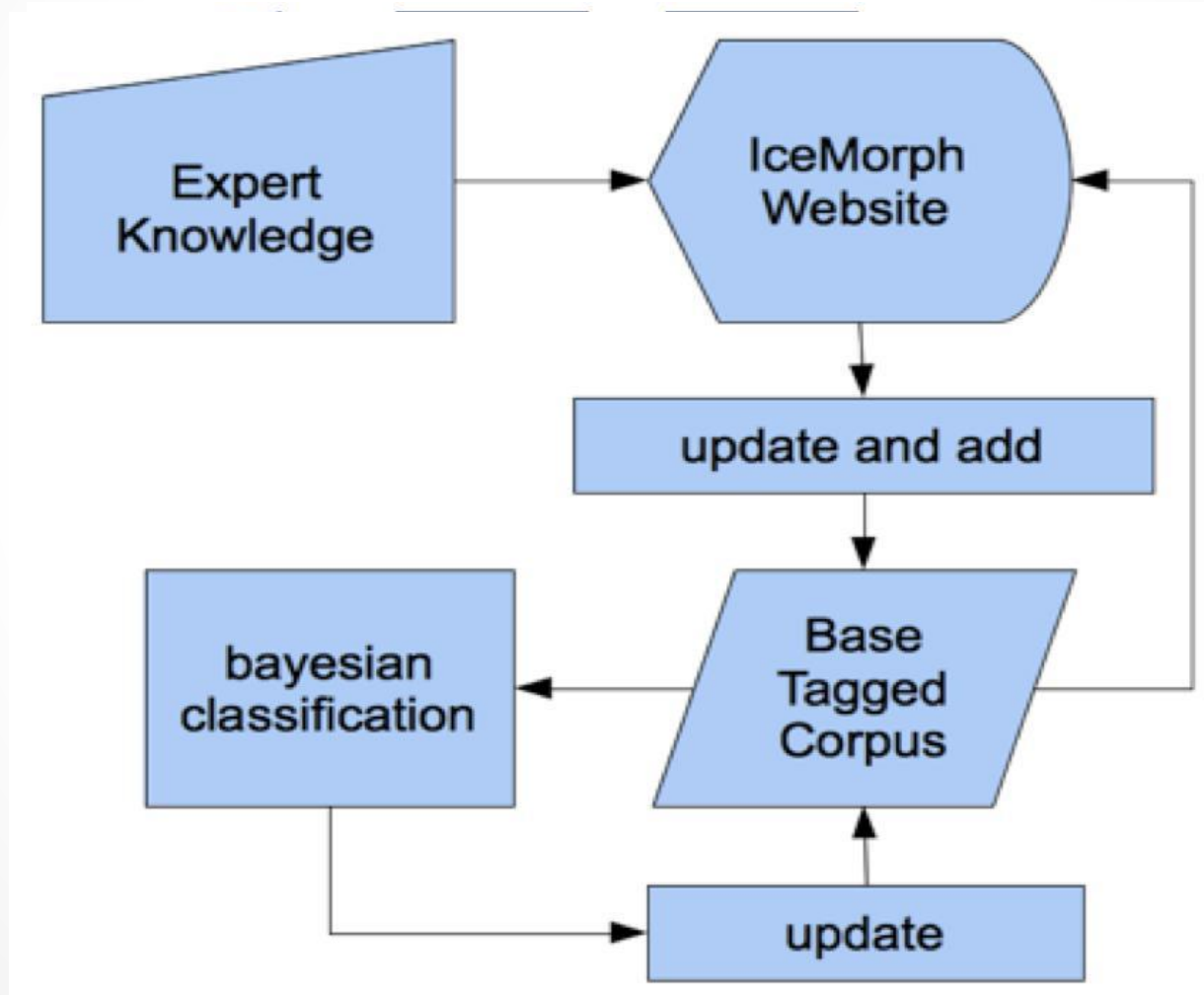
Devoicing in word final position

Devoicing in word-final position has led to certain irregularities as well. The consonants that were affected more frequently than others were d, g and γ <g>. The consonant γ in absolute final position would become /h/ and disappear, engendering irregular development of the preceding mono- and diphthongs, cf. the development of the 1./3. p. sg. preterite of the verb stiga 'step' (-- \*stiγan): Proto-Scandinavian \*staiγ -- \*staih -- \*stē -- OIc. sté. Otherwise, the diphthong \*ai usually developed into OIc. ei, cf. \*baid 'waited' -- OIc. beið, \*skain 'shined' -- OIc. skein, etc. It also developed into ei in those word-forms, where γ was not final, e.g. 2. p. sg. steig-t -- \*staiγ-t. However, due to system pressure ("Systemzwang") a new phonologically irregular by-form steig was formed both for the 1. and the 3. persons.

The disappearance of the word-final /h/ (originally, γ) in a similar way affected the diphthong au, irregularly changing it to ö (OIc. ó), cf. OIc. fló 'flew' -- \*flauh -- \*flauγ instead of the expected au, cf. OIc. skaut 'shot' -- \*skaut, etc. In a similar fashion, the monophthong a was lengthened to ā, cf. vá 'hit (with a weapon)' -- \*wah -- \*way.

As a result of devoicing in word-final position, the consonant /d/ often devoiced and subsequently assimilated preceding sonorants, cf. OIc. batt 'bound' (1./3. p. sg. pret.) -- \*bant -- \*bant -- \*band, etc. The consonant /g/ often behaved in the same way, devoicing and assimilating a preceding nasal, cf. sprakk 'exploded' (1./3. p. sg. pret.) -- \*sprank -- \*sprank -- \*sprang. Since in those word-forms where the conditions for word-final devoicing were not met, devoicing did not occur, paradigms eventually became quite irregular, cf. singular sprakk vs. plural sprungum, or singular batt vs. plural bundum.

# Sketching the process



# Sketching the interface

ON Saga - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.icemorph.ucla.edu/saga/landnama

Most Visited Merriam-Webster Online Wikipedia Welcome to Aeries Br... Latest Headlines Danish Folklore Data ... Danish Folklore Datab...

ON Saga

Saga Edit Dictionary

Select prototype

	Singular	Plural
Nominative	maðr	menn
Accusative	mann	menn
Genitive	manns	manna
Dative	manni	mönnum

Submit Done

Comments/discussion

Corpus Instance: manni,

Lexeme: maðr

POS and CLASS: noun m\_r

Declension: dat\_sg

Translation: (1) man (irrespective of sex), human being (gu skapaði siðarst menn tvá, er ættir eru frá kom sýndi m manni, one showed it to another, it w from from hand to hand; fjöldi manns, a great number of people; múgrmanns, crowd of peo

It might make sense to have the prototype inflection appear here for nouns (I'm not sure what to do about verbs) so you can swap back and forth. The prototype could also be labeled "ja-stem" or whatever depending on example

only appear for staff + admin.

shouldn't the assigned prototype appear here? The "Select prototype" text could be to the side of the pull down

what about doing that as tabs or a second tab that could be discussion?

Done

start

zotero

11:17 AM

# Three Tools

- Morphological analyzer

<http://icemorph.scandinavian.ucla.edu/saga/landnama>

- Inflected dictionary browser

<http://etkspace.scandinavian.ucla.edu/~monit/search.php>

- Reading environment

<http://etkspace.scandinavian.ucla.edu/~monit/saga/>



## Overview

## Lexeme

## Inflection

## Dictionary Entries

(1) man (irrespective of sex), human being (guð skapaði síðarst menn tvá, er ættir eru frá komnar); sýndi m manni, one showed it to another, it went from from hand to hand; fjöldi manns, a great number of people; múgrmanns, crowd of people; (2) degree in kinship; vera at þriðja, frórðax

## Example

## Notes

Source Zoega

## Edit

Lemma maðr

Normalized maðr

Compound maðr

Normalized maðr

## Translation

m., *nnr* und *ranni*, Vellekla, (for the change of *nn* before *r* into *ð* see the introduction to letter N); gen. manns, dat. manni, acc. mann, plur. menn, qs. menn-r; with the article, menninir, so always in old writers, but in mod. mennirnir erroneously, as if from mennir: the plur. meðr, answering to the sing. maðr, occurs in old poets—meðr vituð öðling æðra, Fms. vii. 87 (in a verse); Norð-meðr róa naðri, vi. 309 (in a verse); meðr fengu mikit veðr, Edda 102; hirð-meðr, veðja, Rekst., all verses of the 11th and 12th centuries; er meðr Myrkvið kalla, Akv. 5: meðr hlutu sár. Fbr. 75 new Ed. (in a verse): gen. pl. manna. dat. mönnum. acc. menn. In Ballads

# Inflected diction

## Nearby Words

leistabrækr  
leistabrækr  
leistalauss  
leistr  
leit  
leita  
leiti  
leiti.i.  
leizla  
leizludrykkja  
leið  
leiða

## leita

[Edit Lexeme and Class](#)

POSC: *verb weak (1)*

[View Inflections](#)

### Translations

[Edit Translation](#)

### Cleasby-Vigfusson:

að, prop. a causal from líta, [Ulf. *wlaiton* = περιβλέπεσθαι; North. E. *to l* at e-u, or absol., héraðsmenn leituðu hennar ok fundu hana eigi, Nj. 14; le Landn. 27, 32, Fms. i. 27, 71; leita e-m kvánfangs, Eg. 22, passim: metapl þótt hans væri eigi vel leitað, Fb. ii. 73:—hón leitaði í eina hirzlu, Fms. iv manna-forráð, Nj. 149: leita at e-m, *to seek for*, Fms. ix. 218; ok leiti þér a Fms. i. 69: the phrase, leita sér staðar, *to go on one's business, cacare*, Hm annarra en mín ef þú þarft nokkurs við, Nj. 74; leita sér heilla, Landn. 33; leita ráða, *to seek for advice*, Nj. 75. **2.** *to enquire, examine*, Sks. 638, Hor aggressive, Nj. 16, Ó. H. 222, (á-leitinn); leita á við e-n, *to contest, call in leita?* Nj. 3; en þat ræð ek, at þú leitar eigi optarr á hreysti mína, *that thou after, seek for*, passim; leita eptir máli, *to follow a case, take it up*, Nj. 75; enquire into, Eg. 536: leita upp, *to seek out*, Germ. *ansuchen*, Fms. x. 71: 382; leita við

[Edit Translation](#)

### Zoega:

	Active Indicative			
	Present		Past	
	Singular	Plural	Singular	Plural
1	leita	leitum	leitaða	leituðum
2	leitar	leitið	leitaðir	leituðuð
3	leitar	leita	leitaði	leituðu

	Active Optative			
	Present		Past	
	Singular	Plural	Singular	Plural
1	leita	leitim	leitaða	leitaðim
2	leitir	leitið	leitaðir	leitaðið
3	leiti	leiti	leitaði	leitaði

	Middle Indicative			
	Present		Past	
	Singular	Plural	Singular	Plural
1	leitumk	leitumsk	leituðumk	leituðumsk
2	leitask	leitizk	leitaðisk	leituðuzk
3	leitask	leitask	leitaðisk	leituðusk



# Saga reader

close

close

maðr maðr

(**noun m\_r**) m., nnnr und ranni, Vellekla, (for the change of nn before r into ð see the introduction to letter N); gen. manns, dat. manni, acc. mann, plur. menn, qs. menn-r; with the article, menninir, so always in old writers, but in mod. mennirnir erroneously, as if from mennir: the plur. meðr, answering to the sing. maðr, occurs in old poets—meðr vituð öðling æðra, Fms. vii. 87 (in a verse); Norð-meðr róa naðri, vi. 309 (in a verse); meðr fengu mikit veðr, Edda 102; hirð-meðr, veðja, Rekst., all verses of the 11th and 12th centuries; er meðr Myrkvið kalla, Akv. 5: meðr hlutu sár, Fbr. 75 new Ed. (in a verse): gen. pl. manna, dat. mönnum, acc. menn. In Ballads and Rímur after the 15th century, and hence in eccl. writers of later times, a nom. mann is now and then used, esp. in compds influenced by Germ. and Engl., e. g. hreysti-mann, Skíða R. 58; or for the sake of rhyme, ætla þú ekki, aumr mann | af komast muni strafflaust hann, Pass. 14. 17: [Ulf. manna = ἄνθρωπος; in other Teut. languages spelt man, or better mann.] (translation from: Cleasby-Vigfusson)

(**noun m\_r**) (1) man (irrespective of sex), human being (guð skapaði síðarst menn tvá, er ættir eru frá komnar); sýndi m manni, one showed it to another, it went from from hand to hand; fjöldi manns, a great number of people; múgrmanns, crowd of people; (2) degree in kinship; vera at þrjú, frórðax (translation from: Zoega)

Chapter 19  
Chapter 20  
Chapter 21  
Chapter 22  
Chapter 23

## Is-Morph Saga Browser

### I, Saga 1: Völsunga saga, Chapter 1

þeim manni, er Sigi er nefndr ok kallaðr, at hétu nefndr til sögunnar, er Skaði hét. Hann var ríkr ok Sigi þeira enn ríkari ok ættstærri, at því er Skaði átti þræl þann, er nokkut verðr at geta við hann er fróðr við þat, er hann skyldi at hafast. því jafnframt hinum, er meira þóttu verðir, eða segja eitthvert sinn, at Sigi ferr á dýraveiði, ok síða dýr um daginn allt til aptans. En er þeir bera hann, þá hafði Breði veitt miklu fleira ok meira en þá ok segir, at sik undri, at einn þræll skuli sik því at honum ok drepr hann; dysjar síðan líkit í hann um kveldit ok segir, at Breði hafi riðit frá hann senn ór augliti mér, ok veit ek ekki til hans." getr, at vera munu svik hans ok mun Sigi hafa leita hans, ok lýkr svá leitinni, at þeir fundu þeli Skaði, at þann skafl skyldi kalla Breðafönn heðan af, ok hafa menn nú þat eptir síðan ok kalla svá hverja fönn, er mikil er. Þá kemr upp, at Sigi hefir drepit þrælinn ok myrðan. Þá kalla þeir hann varg í véum, ok má hann nú eigi heima vera með feðr sínum. Óðinn fylgir honum nú af landi brott, svá langa leið, at stóru bar, ok eigi létti hann fyrr en

#### manni

Inflection: Singular Dative

POSC: noun, masculine gender, r-stem

	Singular	Plural
Nominative	maðr	menn
Accusative	mann	menn
Genitive	manns	manna
Dative	manni	mönnum

# Thank you

- UCLA Council on Research
- John Simon Guggenheim Foundation
- American Council of Learned Societies
- The Mellon Foundation
- Google Books
- NSF
- Institute for Pure and Applied Mathematics
- Dansk Folkemindesamling
  
- Collaborators: Peter M. Broadwell, Peter Leonard, James Abello, David Mimno, Mads Rosendahl Thomsen, Kryztof Urban, Aurelijus Vijunas