

Adapting a topic modelling tool to the task of finding recurring themes in Swedish folk legends

Maria Skeppstedt, Rickard Domeij, and Fredrik Skott

An adapted version of the Topics2Themes tool was applied on a corpus consisting of around 10,000 Swedish folk legends.

This resulted in the automatic extraction of 20 topics.

The screenshot displays the Topics2Themes web application interface. It features three main panels: 'Terms', 'Topics', and 'Texts'. The 'Terms' panel on the left lists various words and phrases, including 'kyrkan / kyrkans / kyrkorna', 'kyrka', 'bygga', 'byggde', 'stannade', 'döda', 'byggdes', 'klockan', 'byggas', 'julotta / julottan', 'fläsk / fläsket / gryta / grytan / gräddade / grädden / häll / jäst / kokt / korv / korven / kött / köttet / mjöl / mjölet / mjölk / mjölken / ost / potatis / ris / smör / smöret / smörgås / socker / spadet / sås / vitlök / örter', 'dräng / drängen / piga / pigan', 'armarna / axlarna / benen / benet / bröstet / fötterna / halsen / huvudet / händerna / knäna / låret / näsan / rumpen / ryggen / skallen / tårna / öronen', 'gubben', 'prästen', 'tre', 'barnet', and 'hästen'. The 'Topics' panel in the center shows a list of topics, with 'kyrkan - kyrka - bygga' highlighted. The 'Texts' panel on the right displays several text snippets, each with a 'Ne i' button. The snippets include: 'Det fanns för mycket länge sedan två jättar. Den ene hette Hund och den andre hette Katt. Den förre ville bygga en kyrka i Brattås, den senare ville att kyrkan skulle ligga vid Högarberg. Det blev Katt som avgick med seger och kyrkan fick namnet Kattunga kyrka. När den väl var färdig, stod Hund en dag uppe på Gallås. Han grep plötsligt en stor sten för att slunga den mot kyrkan. Men stenen klövs itu. Ena hälften hamnade öster om kyrkan i Stens lycka. Den andra föll ned söder om kyrkan på Högarberg. Den kallas Hönsabärsstenen. För mycket länge sedan fanns bara Kattunga kyrka. På den tiden red Stockholms- och Varbergsborna hit. Skjutsar fanns inte då. De släppte sina hästar i en äng som än i dag kallas Kyrkäng. Det finns en sägen om Kattunga kyrka att hur många som kom dit så fick de plats och hur få åhörarna än var så fyllde de hela kyrkan. När kyrkan för hundra år sedan skulle rivas emedan man ansåg sig behöva stenar, var det många som drog sig för arbetet. Den som rubbade den första stenen blev hårt straffad. Han klände ett finger och efter en kort tid var han död. Hästen som drog det första lasset, blev blind. 'De byggde kyrka på en plats i Värö, men varje natt blev det nerrivet. Då släppte de två hopokade tvillingbaggar och där de stannade, skulle kyrkan byggas. När de hade stannat, sa det i luften: "Här vill ja våra, för detta heter Värö. "Där byggde de, och där blev kyrkan ståanes. Mormor trodde på att döingar var i kyrkan om julnatten. Hon brukade tala om att det var så många döingar i kyrkan som det kunde rymmas. Men alla sutto inte vända åt prästen. Det var somliga som satt med ryggen åt prästen. Hon tala om att det kom en käring till kyrkan om natten, innan döingarna hade gått ur kyrkan. Käringen såg att många voro bekanta till henne. Hon skulle springa ut ur kyrkan. Men schalen fick hon inte med sig. Den blev efter i dörren. Och döingarna tog den. Om morgonen när folk kom så låg schalen vid dörren, men det hade blivit så smått som mull av den. På den tiden Nössemarks gamla kyrka brukades, så fanns det inte julmorgon än de såg ljus lysa i fönstren långt innan någon människa hade kommit till kyrkan. Så snart kyrkväktarn kom i närheten av kyrkan, slocknade ljusen. De döda var uppe efter klockan tolv och hade gudstjänst i kyrkan. Jag minns en

The topic modelling tool was configured to use:

- A word space model trained on a Swedish corpus.
- A Swedish stop word list.

The Swedish stop word list consisted of:

- Standard Swedish stop words.
- 380 additional stop words tailored to the content of the corpus. They therefore also included older spelling versions and grammatical forms of Swedish words.

Future versions of the tool will be extended with text summarisation functionality, in order to retain the text overview provided by the tool, also when it is applied on longer folk legends.



Supported by the Swedish Research Council (project number 2017-00626)
Correspondence to: maria.skeppstedt@isof.se
<https://github.com/mariask2/topics2themes>

NATIONELLA
SPRÅKBANKEN

