Applications of language technology to Digital Humanities

Roman Yangarber

Department of Computer Science
University of Helsinki

HELDIG 18 October 2017

Areas of research

- **PULS**
  → newsweb.cs.helsinki.fi
  Web-scale surveillance of news media
- **Revita**
  → revita.cs.helsinki.fi
  Support for language learning, and for endangered languages
- **Etymon**
  → etymon.cs.helsinki.fi
  Models of language evolution
- ...

PULS: Web-scale surveillance of news media

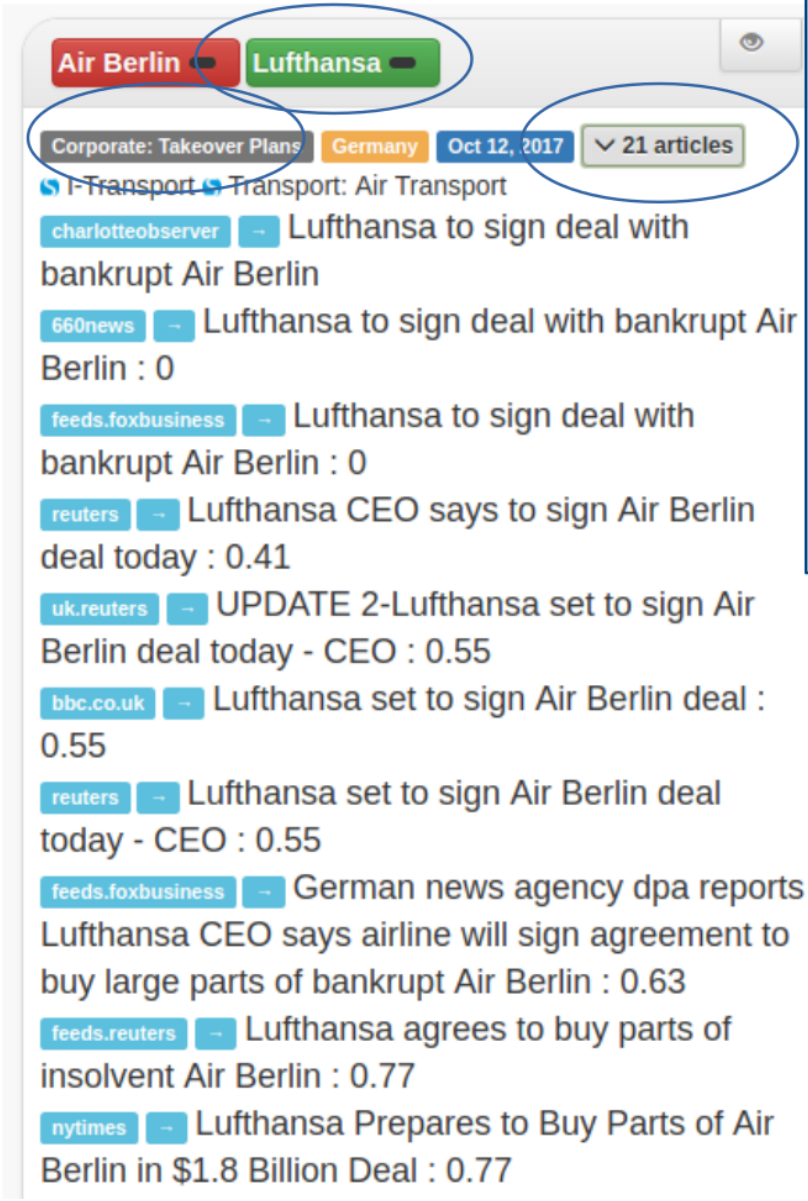
TEKES, companies, EU agencies

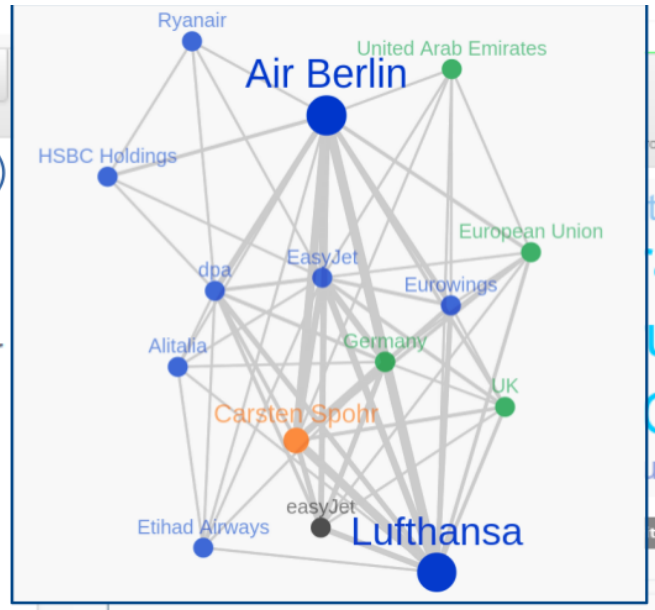
- business news. security, epidemics, ...

Challenges: find interesting patterns in data:
(millions of articles, across time and sources)

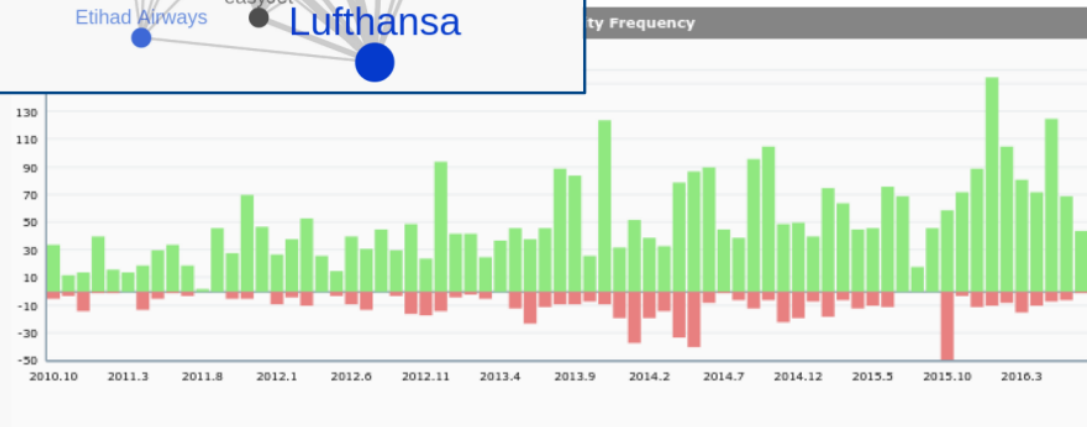
- is an entity receiving attention—from many sources?
- positive or negative attention?
- what entities are similar to given entity, and how?
- ...

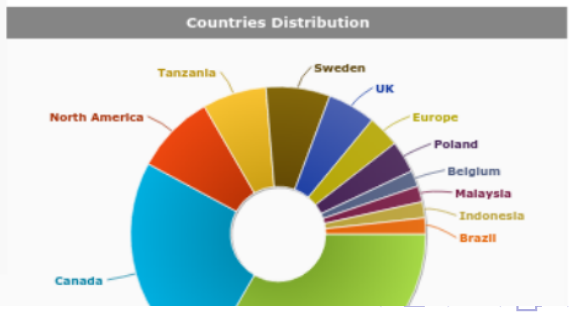
Challenges

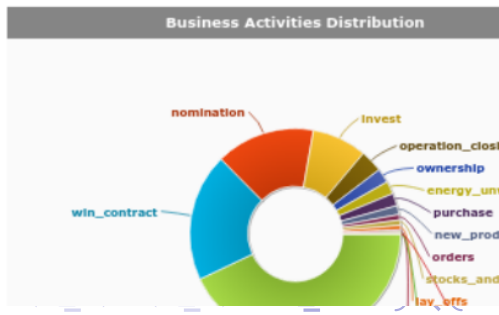

Air Berlin
Lufthansa
Corporate: Takeover Plans
Germany
Oct 12, 2017
21 articles
I-Transport
Transport: Air Transport
charlotteobserver → Lufthansa to sign deal with bankrupt Air Berlin
660news → Lufthansa to sign deal with bankrupt Air Berlin : 0
feeds.foxbusiness → Lufthansa to sign deal with bankrupt Air Berlin : 0
reuters → Lufthansa CEO says to sign Air Berlin deal today : 0.41
uk.reuters → UPDATE 2-Lufthansa set to sign Air Berlin deal today - CEO : 0.55
bbc.co.uk → Lufthansa set to sign Air Berlin deal : 0.55
reuters → Lufthansa set to sign Air Berlin deal today - CEO : 0.55
feeds.foxbusiness → German news agency dpa reports Lufthansa CEO says airline will sign agreement to buy large parts of bankrupt Air Berlin : 0.63
feeds.reuters → Lufthansa agrees to buy parts of insolvent Air Berlin : 0.77
nytimes → Lufthansa Prepares to Buy Parts of Air Berlin in $1.8 Billion Deal : 0.77


Ryanair
United Arab Emirates
Air Berlin
HSBC Holdings
European Union
EasyJet
dpa
Eurowings
Alitalia
Germany
UK
Carsten Spohr
easyJet
Etihad Airways
Lufthansa

To
9/14/2016
Get profile
energy group(29)
PetroChina
Sinopec
Lukoil
BP
Pemex


Frequency
130
110
90
70
50
30
10
-10
-30
-50
2010.10
2011.3
2011.8
2012.1
2012.6
2012.11
2013.4
2013.9
2014.2
2014.7
2014.12
2015.5
2015.10
2016.3


Countries Distribution
Tanzania
Sweden
UK
North America
Europe
Poland
Belgium
Malaysia
Indonesia
Brazil
Canada


Business Activities Distribution
nomination
Invest
win_contract
ownership
purchase
orders

Revita: language learning

Directions:

- → Support for endangered languages
- → Support for language learning: Finnish, Swedish, Russian, German

Key innovations:

- Intermediate-advanced learners
- Work with any text → **dynamic** generation of exercises
- Provide **best** exercises for learner
  → **model** learner competency: what does s/he know ?

Challenges: simulate a good teacher: **model learner's competency**

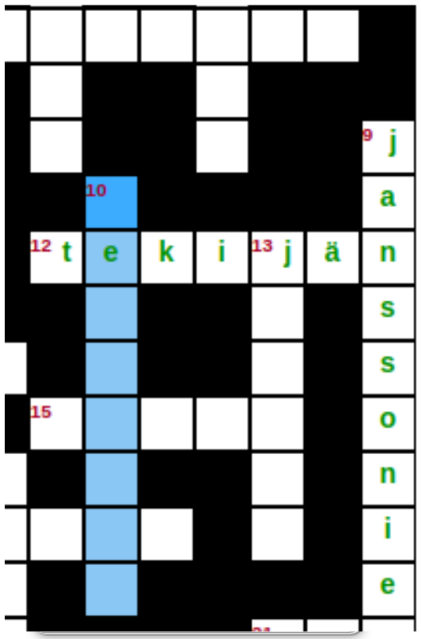
- if exercises too difficult → user becomes discouraged
- if exercises too simple → user becomes bored

Revita: Features

Muumit

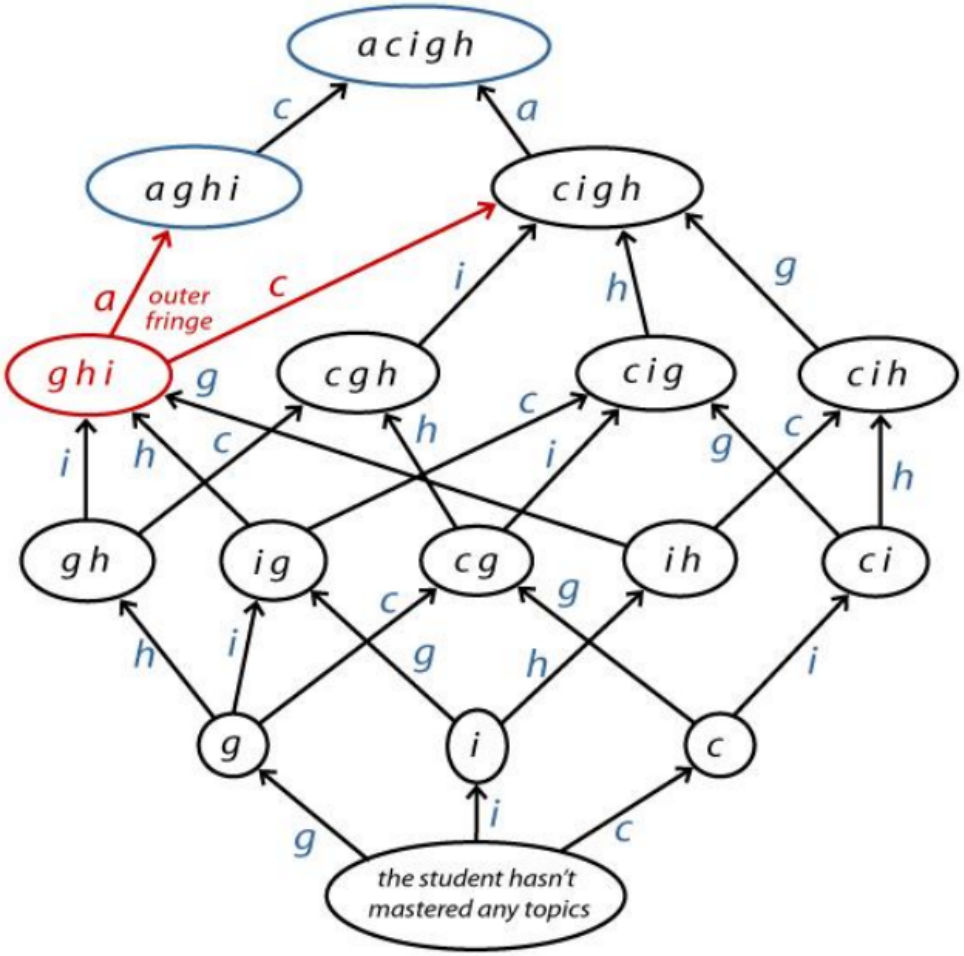
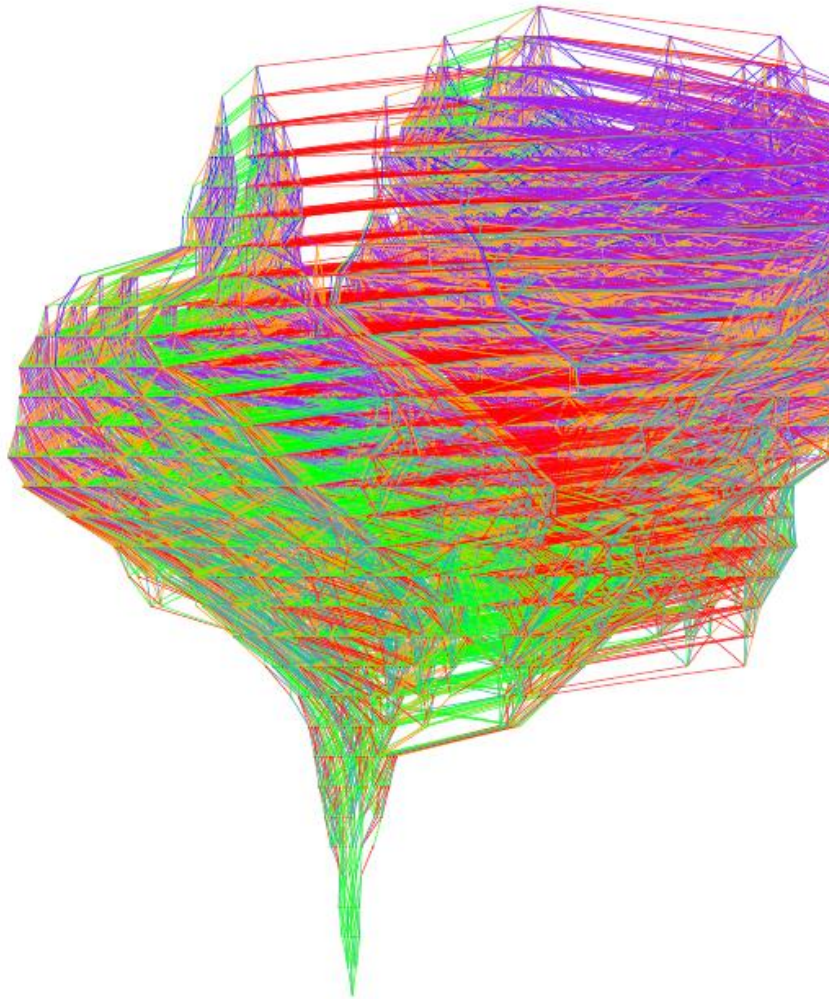
muumi on suomenruotsalainen kirjailijan ___ taiteilijan Tove Janssonin luoma satuhahmolaji. Yksi tunnetuimmista muumihahmoista on Muumipeikko, joka asuu huoleton ja seikkailunhaluinen muumiperheensä kanssa Muumitalossa Muumilaaksossa. Jansson kirjoittaa ruot lastenkirjallisuus Muumi-kirjoja, jotka olla suosituinta ja käännetyintä suomalaista lastenkirjallisuutta .

0 / 0 Check Answers 4 / 5



Muumit

Muumien synty

Muumipeikon hahmo syntyi Tove Jan mielikuvituksessa jo 1930-luvulla. En piirros muumista syntyi **Janssonien** ulkohuussin 21 ↓ . 1930-luvun hänen akvarelleissaan esiintyy musti kaltaisia olentoja, ja 39 ↓ Jansson hahmoa Garmin pilapiirroksissaan, jo **tekijän** tunnusmerkkinä yhdessä sig

Learner model: learning knowledge space from data