Folktale similarity based on ontological abstraction

Marijn Schraagen

Digital Humanities Lab
Utrecht University, The Netherlands

Global WordNet Conference
January 30, 2016

Universiteit Utrecht

Research task

- Compute pair-wise similarity of folktale texts using WordNet

Research task

- Compute pair-wise similarity of folktale texts using WordNet
- Capture common elements in actors and events at an abstract level

Research task

- Compute pair-wise similarity of folktale texts using WordNet
- Capture common elements in actors and events at an abstract level
- Complement existing folktale classification standards

Data

- Data taken from Dutch Folktale Database
  - `http://www.verhalenbank.nl/`, in Dutch

Data

- Data taken from Dutch Folktale Database
  - http://www.verhalenbank.nl/, in Dutch
- Subcorpus for proof of concept

Data

- Data taken from Dutch Folktale Database
  - `http://www.verhalenbank.nl/`, in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words

Data

- Data taken from Dutch Folktale Database
  - `http://www.verhalenbank.nl/`, in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words
- Grammatically correct, modern Dutch

Data

- Data taken from Dutch Folktale Database
  - `http://www.verhalenbank.nl/`, in Dutch
- Subcorpus for proof of concept
- 16 folktales, 33,022 words
- Grammatically correct, modern Dutch
- "Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje."
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn't it? She was not actually called Little Red Riding Hood.*

Preprocessing

- Preprocessing using Frog

Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging

Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs

Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs
- “Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje.”
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.*

Preprocessing

- Preprocessing using Frog
- Tokenization, lemmatization, POS-tagging
- Keep nouns (proper names), adjectives, (non-function) verbs
- “Er was eens een klein meisje, dat Roodkapje heette. Wat een gekke naam, hè? Ze heette ook niet echt Roodkapje.”
- *Once upon a time there lived a little girl, called Little Red Riding Hood. What a strange name, isn’t it? She was not actually called Little Red Riding Hood.*
- klein meisje Roodkapje heten. gek naam hè. heten echt Roodkapje
- *little girl Little_Red call. strange name eh. call really Little_Red.*

Similarity computation

- Count number of matching terms

Similarity computation

- Count number of matching terms
- Sentence level comparison

Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet

Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto

Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction

Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets

Similarity computation

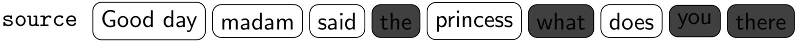
- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets
- Directed similarity score for each sentence

Similarity computation

- Count number of matching terms
- Sentence level comparison
- Check for exact match or abstract match using WordNet
  - Dutch WordNet: Cornetto
- Match similarity score based on level of abstraction
- Sentence similarity score based on match similarity relative to size of lemma sets
- Directed similarity score for each sentence
- First synset used

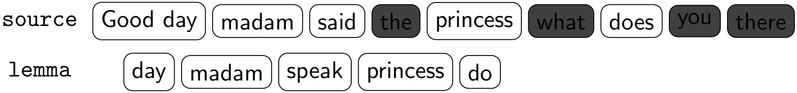
Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



source Good day madam said the princess what does you there

lemma day madam speak princess do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

source: Good day | madam | said | the | princess | what | does | you | there

lemma: day | madam | speak | princess | do

synset: day | lady | speak | royal daughter | do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

Good afternoon | basket maker | said | the | gnome

day | lady | speak | royal daughter | do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

Good afternoon | basket maker | said | the | gnome

good afternoon | basket maker | speak | gnome

day | lady | speak | royal daughter | do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

Good afternoon | basket maker | said | the | gnome

good afternoon | basket maker | speak | gnome

good afternoon | basket maker | speak | gnome

day | lady | speak | royal daughter | do

Similarity computation example

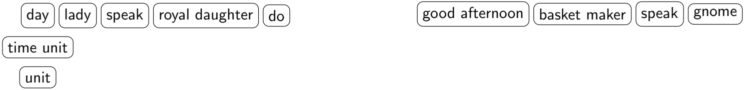
Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

day lady speak royal daughter do
time unit

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



day lady speak royal daughter do

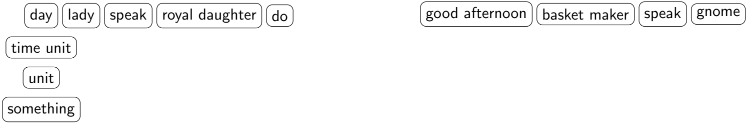
time unit

unit

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



day lady speak royal daughter do

time unit

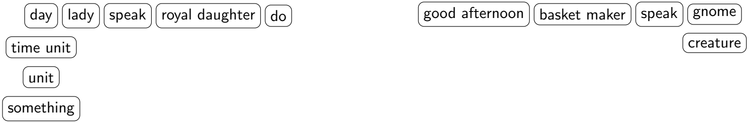
unit

something

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

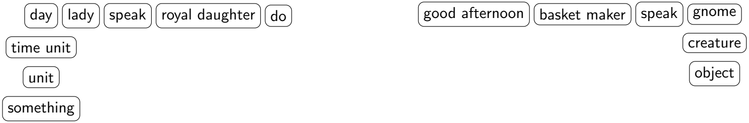


day
lady
speak
royal daughter
do
time unit
unit
something

good afternoon
basket maker
speak
gnome
creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



day lady speak royal daughter do

time unit

unit

something

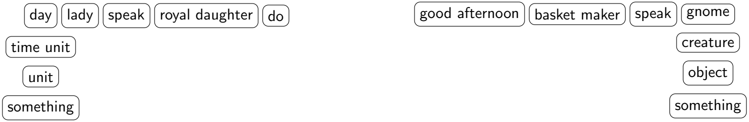
good afternoon basket maker speak gnome

creature

object

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



day lady speak royal daughter do

time unit

unit

something

good afternoon basket maker speak gnome

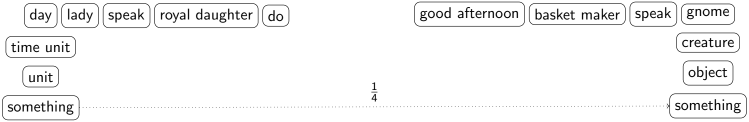
creature

object

something

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}$

day | lady | speak | royal daughter | do | figure

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}$

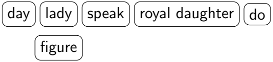
day lady speak royal daughter do
figure

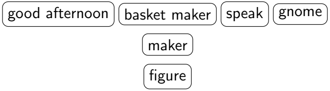
good afternoon basket maker speak gnome
maker

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}$

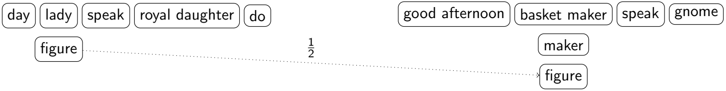




Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}$



day lady speak royal daughter do

figure

good afternoon basket maker speak gnome

maker

figure

$\frac{1}{2}$

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}$

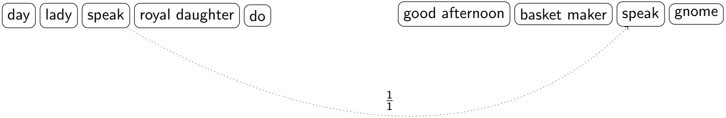
day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2}$



day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

$\frac{1}{1}$

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1$

day lady speak royal daughter do

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1$

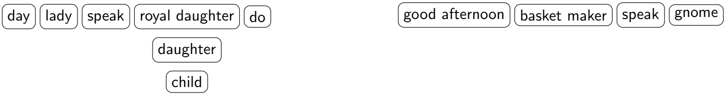
day | lady | speak | royal daughter | do
daughter

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1$



day lady speak royal daughter do

daughter

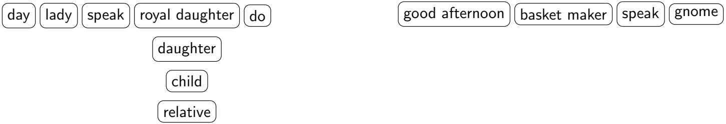
child

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1$



day | lady | speak | royal daughter | do

daughter

child

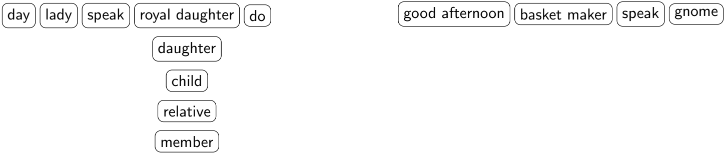
relative

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

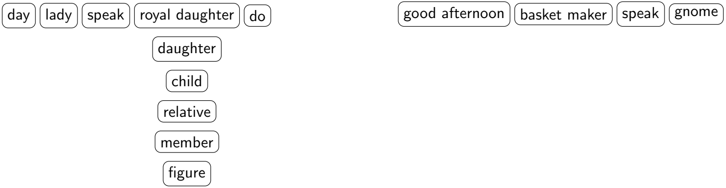
$((\frac{1}{4} + \frac{1}{2} + 1$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1$



day lady speak royal daughter do

daughter

child

relative

member

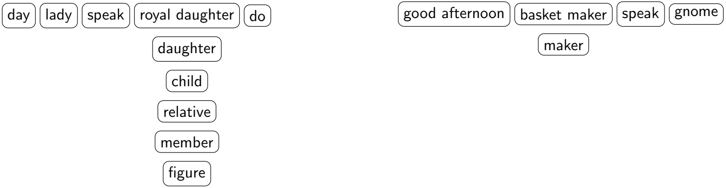
figure

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

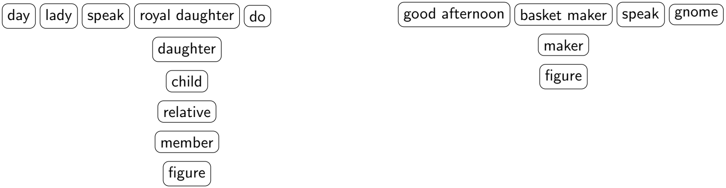
$((\frac{1}{4}+\frac{1}{2}+1$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1$



day lady speak royal daughter do

daughter

child

relative

member

figure

good afternoon basket maker speak gnome

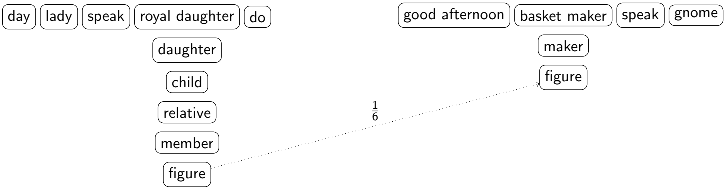
maker

figure

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$((\frac{1}{4}+\frac{1}{2}+1$$



day lady speak royal daughter do

daughter

child

relative

member

figure

good afternoon basket maker speak gnome

maker

figure

$\frac{1}{6}$

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}$

day lady speak royal daughter do

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?

Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}$

day lady speak royal daughter do

act

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}$

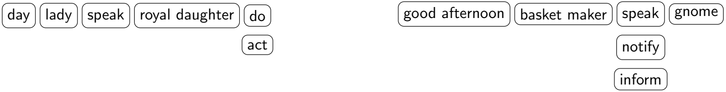
day lady speak royal daughter do
act

good afternoon basket maker speak gnome
notify

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

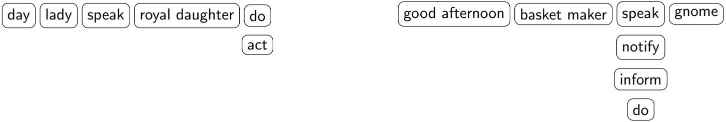
$$((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}$$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}$



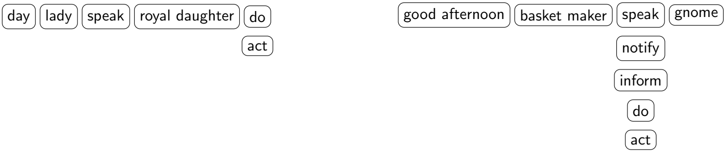
day lady speak royal daughter do
act

good afternoon basket maker speak gnome
notify
inform
do

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}$



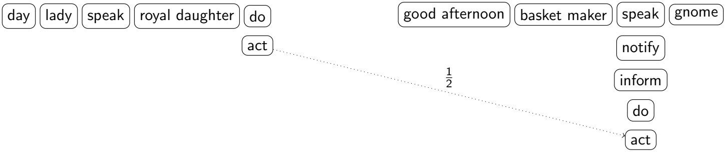
day lady speak royal daughter do
act

good afternoon basket maker speak gnome
notify
inform
do
act

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6}$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

no match

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0$

day
lady
speak
royal daughter
do

good afternoon
basket maker
speak
gnome
maker

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0$

day lady speak royal daughter do

good afternoon basket maker speak gnome

maker

figure

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0$

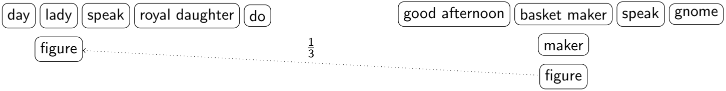
day lady speak royal daughter do
figure

good afternoon basket maker speak gnome
maker
figure

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0$



day lady speak royal daughter do

figure

$\frac{1}{3}$

good afternoon basket maker speak gnome

maker

figure

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}$

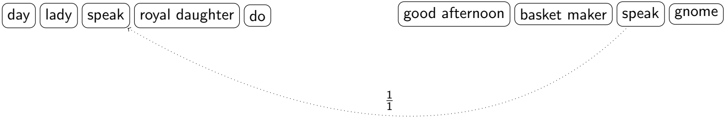
day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$

day lady speak royal daughter do

good afternoon basket maker speak gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4} + \frac{1}{2} + 1 + \frac{1}{6} + \frac{1}{2}) + (0 + \frac{1}{3} + 1$

day | lady | speak | royal daughter | do

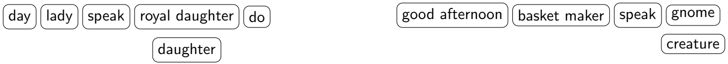
good afternoon | basket maker | speak | gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

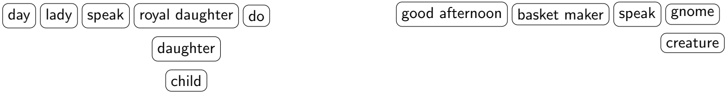
$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$

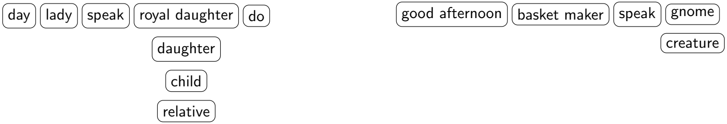

day lady speak royal daughter do
daughter
child

good afternoon basket maker speak gnome
creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



day lady speak royal daughter do

daughter

child

relative

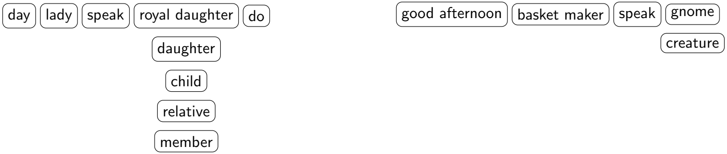
good afternoon basket maker speak gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



day lady speak royal daughter do

daughter

child

relative

member

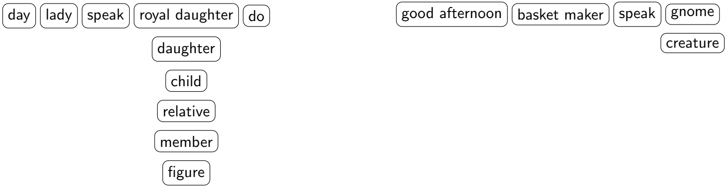
good afternoon basket maker speak gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



day lady speak royal daughter do

daughter

child

relative

member

figure

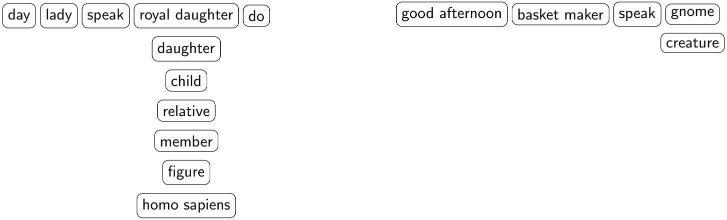
good afternoon basket maker speak gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

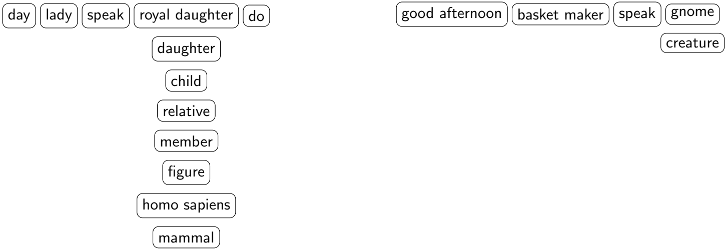
$$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



day lady speak royal daughter do

daughter

child

relative

member

figure

homo sapiens

mammal

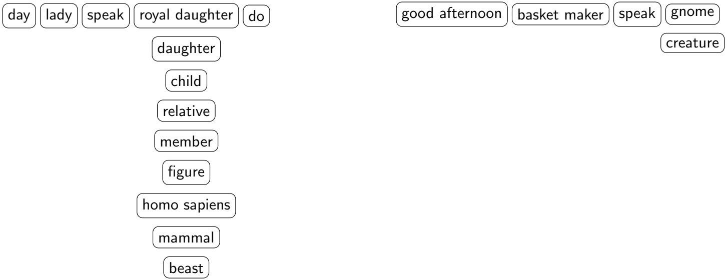
good afternoon basket maker speak gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



day lady speak royal daughter do

daughter
child
relative
member
figure
homo sapiens
mammal
beast

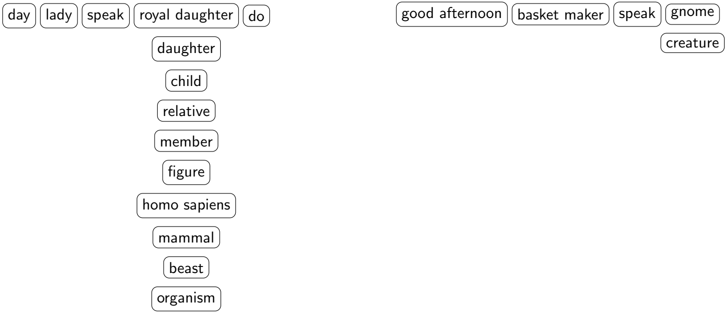
good afternoon basket maker speak gnome

creature

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

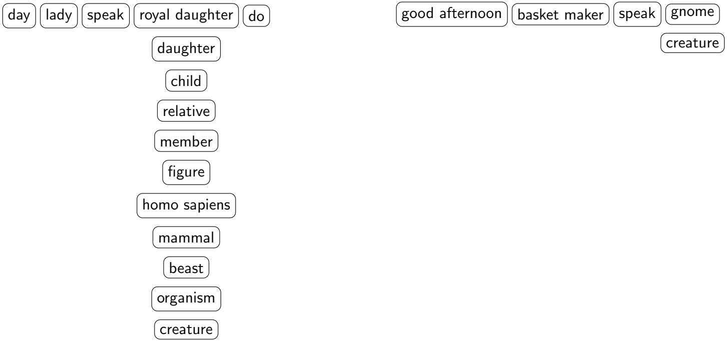
$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

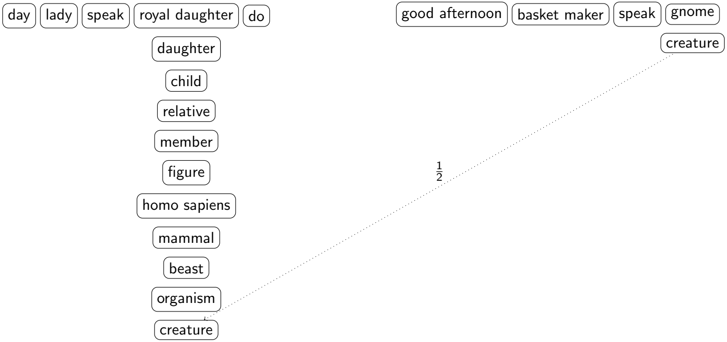
$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$



Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1$$



Universiteit Utrecht

Similarity computation example

Good day madam, said the princess, what are you doing?

Good afternoon basket maker, said the gnome.

$$\left(\left(\tfrac{1}{4}+\tfrac{1}{2}+1+\tfrac{1}{6}+\tfrac{1}{2}\right)+\left(0+\tfrac{1}{3}+1+\tfrac{1}{2}\right)\right)$$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\frac{1}{4}+\frac{1}{2}+1+\frac{1}{6}+\frac{1}{2})+(0+\frac{1}{3}+1+\frac{1}{2}))\,/(5$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$((\tfrac{1}{4}+\tfrac{1}{2}+1+\tfrac{1}{6}+\tfrac{1}{2})+(0+\tfrac{1}{3}+1+\tfrac{1}{2}))\,/(5+4)$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Similarity computation example

Good day madam, said the princess, what are you doing?
Good afternoon basket maker, said the gnome.

$$((\tfrac{1}{4}+\tfrac{1}{2}+1+\tfrac{1}{6}+\tfrac{1}{2})+(0+\tfrac{1}{3}+1+\tfrac{1}{2}))\,/(5+4)=0.47$$

day | lady | speak | royal daughter | do

good afternoon | basket maker | speak | gnome

Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus

Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B

Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
- Ranking-based method

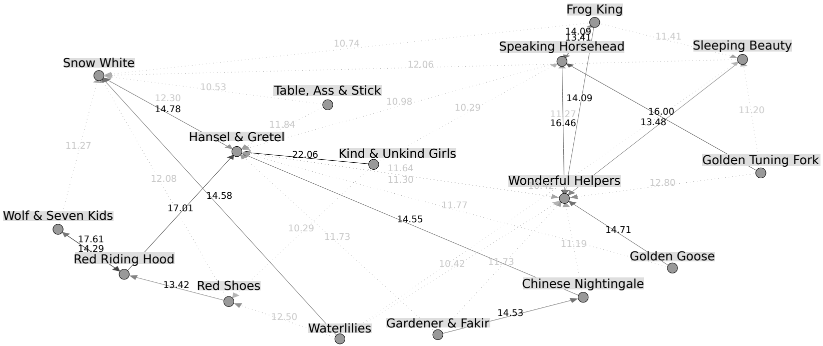
Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
- Ranking-based method
- Non-symmetrical

Document similarity and clustering

- For each sentence in a folktale, find most similar sentence from all sentences in the corpus
- Score for each document pair (A,B) the (relative) amount of sentences from A for which the most similar sentence was found in B
- Ranking-based method
- Non-symmetrical
- Clusters based on similarity thresholds

Similarity computation results



- Central nodes and clusters visible
- Royal protagonists, moral values vs. civilian protagonists, dangerous circumstances

Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings

Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora

Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora
- Folktales: approaches for story variants
  - Evaluation using variant-tagged folktale corpora

Differences with other approaches

- Approaches for pairs of concepts
  - Evaluation using human concept similarity ratings
- Approaches for document categorization
  - Evaluation using gold standard categorized corpora
- Folktales: approaches for story variants
  - Evaluation using variant-tagged folktale corpora
- Current approach: pair-wise document similarity
  - Evaluation less straightforward

Differences with other approaches

- WordNet graph measures

Differences with other approaches

- WordNet graph measures
    - Wu-Palmer: length from shared node to root node

Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth

Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge

Differences with other approaches

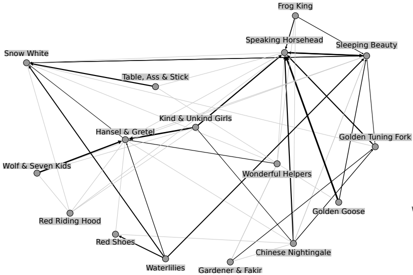
- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge
- Current measure: length from current node to first shared node

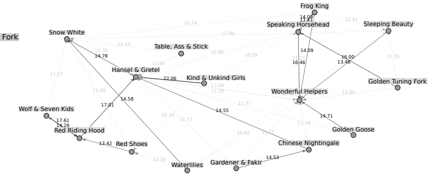
Differences with other approaches

- WordNet graph measures
  - Wu-Palmer: length from shared node to root node
  - Leacock-Chodorow: Shortest path, scaled for local hierarchy depth
  - PageRank, path length weighting, domain knowledge
- Current measure: length from current node to first shared node
- Intended as measure of actor/event relatedness at some level of abstraction, instead of similarity

Evaluation

- Same method, no WordNet, lemma's only





- Clustering and central nodes less apparent

Universiteit Utrecht

Evaluation

- Similarity measure vs. human ratings

scored term	McNo	McRel	McSim	RgNo	RgRel	RgSim
source	**0.64**	**0.60**	0.64	0.54	0.48	0.55
target	0.44	0.39	0.49	0.53	0.53	0.54
lowest	0.59	0.54	0.63	0.53	0.52	0.55
average	0.62	0.56	**0.65**	**0.58**	**0.55**	**0.59**
highest	0.58	0.53	0.61	**0.58**	0.54	**0.59**

- Miller & Charles, Rubenstein & Goodenough word pairs
- No instruction, report similarity, report relatedness

Evaluation

- Similarity measure vs. human ratings

scored term	McNo	McRel	McSim	RgNo	RgRel	RgSim
source	**0.64**	**0.60**	0.64	0.54	0.48	0.55
target	0.44	0.39	0.49	0.53	0.53	0.54
lowest	0.59	0.54	0.63	0.53	0.52	0.55
average	0.62	0.56	**0.65**	**0.58**	**0.55**	**0.59**
highest	0.58	0.53	0.61	**0.58**	0.54	**0.59**

- Miller & Charles, Rubenstein & Goodenough word pairs
- No instruction, report similarity, report relatedness
- Correlations lower than Postma & Vossen (2014), around 0.8

Evaluation

- Similarity measure vs. human ratings

scored term	McNo	McRel	McSim	RgNo	RgRel	RgSim
source	**0.64**	**0.60**	0.64	0.54	0.48	0.55
target	0.44	0.39	0.49	0.53	0.53	0.54
lowest	0.59	0.54	0.63	0.53	0.52	0.55
average	0.62	0.56	**0.65**	**0.58**	**0.55**	**0.59**
highest	0.58	0.53	0.61	**0.58**	0.54	**0.59**

- Miller & Charles, Rubenstein & Goodenough word pairs
- No instruction, report similarity, report relatedness
- Correlations lower than Postma & Vossen (2014), around 0.8
- Different type of similarity measure

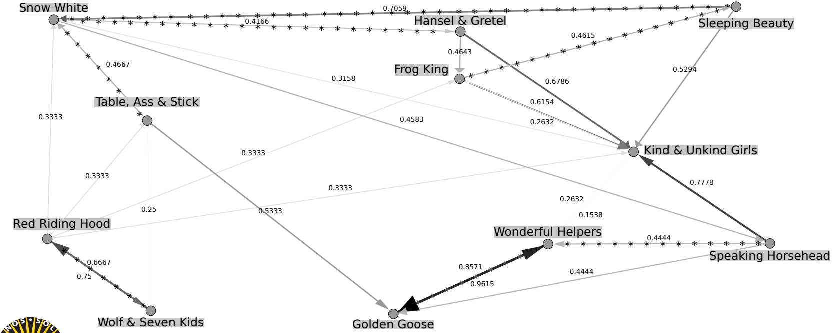
Evaluation

- Comparison with Thompson Motif Index
- Limited number of (semi-)abstract story elements for many (but not all) folktales

ATU	*Title*	*Motif description*	*Motif code*	*match level*
123	The Wolf & the Seven Kids	Disguise by changing voice	K1832	
333	Little Red Riding Hood	Wolf puts flour on his paw to disguise himself	K1839.1	4
533	The Speaking Horsehead	Disguise as goose-girl (turkey-girl)	K1816.5	3
533	The Speaking Horsehead	Imposter forces oath of secrecy	K1933	2
709	Snow White	Compassionate executioner: substituted heart	K0512.2	1

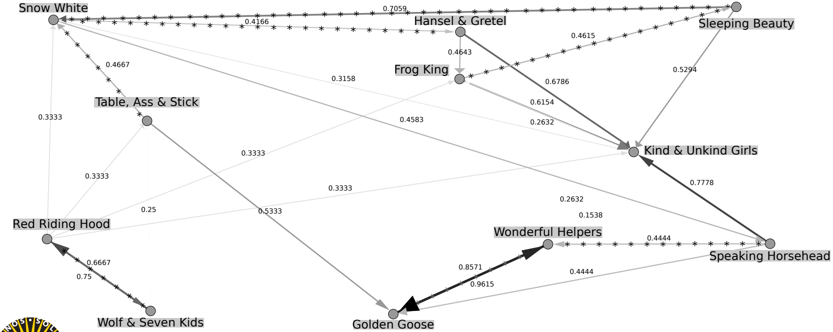
Evaluation

- Directed motif overlap, 2 most similar documents per node
- Asterisk (*) indicates relation also found by WordNet method



Evaluation

- Directed motif overlap, 2 most similar documents per node
- Asterisk (*) indicates relation also found by WordNet method
- TMI contains different type of relation



Discussion

- Sensible clusters and central nodes found

Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward

Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise

Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise
- Use larger and/or more heterogeneous corpus

Discussion

- Sensible clusters and central nodes found
- Evaluation not straightforward
- Many options for similarity computation using WordNet or otherwise
- Use larger and/or more heterogeneous corpus
- Address computational efficiency and scalability

Discussion

Questions?