

---

---

# Identifying Motifs in Folktales

---

---

# What is a folktale?

- Stories passed down through generations, mainly by telling.
- Fairy tales, myths ,legends etc.



# Why are they important?

- Part of our collective cultural heritage.
- Teach society's past, values and lifestyles.
- Significant role in educating children.

# Motivation

- Folktales are being digitized (need to be catalogued to be browsed and easily found).
- Allow researchers to investigate, for example, how folktales have changed through time.
- Motifs play a key role in the classification (into folktale types).



# What is a motif?

- **“What makes your ears so big?”**

[Redacted]

- **Glass slipper**

[Redacted]

- **Cruel stepmother who makes the heroine's life miserable**

[Redacted]



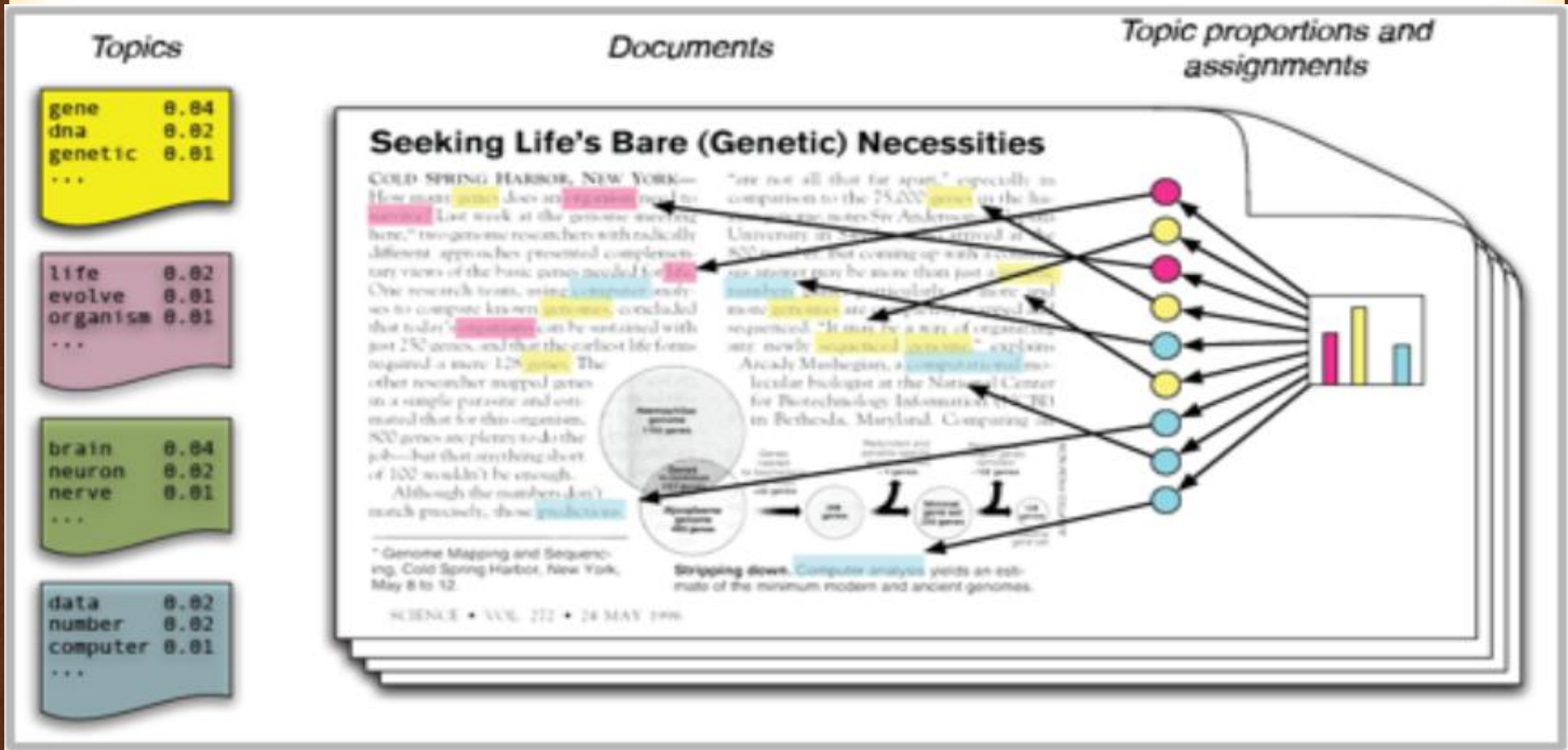
# “Identifying motifs with Topic models”

- A research conducted by Folgert Karsdrop and Antal Van Den Bosch.
- Automatically identify motifs in folktales.
- Multi-label classification task in which we attempt to assign a set of motifs to unseen, unlabeled folktales.

# What is a Topic model?

- A Topic model is a statistical model for discovering the "topics" that occur in a collection of documents.
  - "fetch" and "bone" - documents about dogs
  - "scratch" and "purr" - documents about cats
  - "is", "the" – don't know

- The "topics" produced by topic modeling techniques are clusters of similar words. A topic model captures this intuition in a mathematical framework.





So where do we get Topics?



# Motif-Index (Thompson, 1955 1958)

- over 45,000 motifs, hierarchically ordered in a tree structure.
- There are 23 top-level categories ranging from mythological motifs to motifs concerning traits of character.  
E.g. - *"Magic"*, *"Ogres"*, *"The wise and the foolish"*
- Leaf examples: *"Troll as ogre"*, *"Transformation: pumpkin to carriage"*, *"The shepherd who cried "Wolf!" too often"*



# Thompson's Motif definition

- "Worthy of note because of something out of the ordinary, something of sufficiently striking character to become a part of tradition, oral or literary.

Commonplace experiences, such as eating and sleeping, are not traditional in this sense. But they may become so by having attached to them something remarkable or worthy of remembering"



# TMI - Examples

- “troll under a bridge” → “troll as ogre”, “ogre attacks intruders on bridge”.
- “Wolf cut open and filled with stones as punishment”
- “Glass shoes”
- ‘Transformation: pumpkin to carriage’ →  
‘Transformation: object to object’ →  
‘Transformation’ →  
‘Magic’ (top level node)



So where do we get bags of  
words?



# Aarne-Thompson-Uther (ATU) catalog

- “The Shepherd Boy. A child who herds animals **finds three objects** (of glass) which he **gives back** to their owners. They promise to **reward him [Q42]**. With the help of the last owner, a giant, the boy fulfills three tasks. He acquires a castle in which a princess is confined. **He rescues her and marries her [L161].**”
- This tale contains two motifs, **Q42 ‘Generosity rewarded’** and **L161 ‘Lowly hero marries princess’**.



## Multilingual Folk Tale Database

[Home](#) • [Browse Stories](#) • [Classification](#)

### Aarne-Thompson-Uther Classification of Folk Tales

There are many different folk tales in the world, but many tales are variations on a theme. The classification system of Aarne-Thompson and later by Uther, is intended to bring out the similarities between tales.

Below is the full tree of the ATU classification. Click on a title to see all the stories within that category.

- **ANIMAL TALES** 1-299
  - Wild Animals 1-99
    - The Clever Fox (Other Animal) 1-69
    - Other Wild Animals 70-99
  - Wild Animals and Domestic Animals 100-149
  - Wild Animals and Humans 150-199
  - Domestic Animals 200-219
  - Other Animals and Objects 220-299
- **TALES OF MAGIC** 300-749
  - Supernatural Adversaries 300-399
  - Supernatural or Enchanted Wife (Husband) or Other Relative 400-424

#### Aarne-Thompson-Uther Classification

##### 1-69: The Clever Fox (Other Animal)

- 1-299 ANIMAL TALES
- 1-99 Wild Animals
- 1-69 The Clever Fox (Other Animal)
  - 1 The theft of fish
  - 2 How the bear lost his tail
  - 3 Sham Blood and Brains
  - 4 Carrying the Sham-Sick Trickster
  - 5 Biting the foot
  - 7 The calling of three tree names
  - 8 "Painting" on the Haycock
  - 9 The unjust partner
  - 15 Stealing the Partner's Butter

# Datasets

- Dutch folktales - 1,098 tales. Average number of words per story is 468
- Frisian tales - 1,373 tales. Average number of words per story is 194
- Preprocessing - Remove all punctuations and exclude one letter words and numbers



Now lets get to work!



# What are the challenges?

- Folktales can contains a lot of words and motifs.
- The set of potential labels is large
- Certain motifs will be more strongly tied to the particular folktale.
- So we actually have a **ranking problem.**



# Suggested methods

- Supervised topic model **Labeled LDA** (L-LDA) (Ramage et al., 2009)
- A 'simple' **retrieval model that uses Okapi BM25** as its ranking function.



# Models - Labeled LDA

- Latent Dirichlet Allocation (LDA) (2003) - popular method for extracting topics from texts
- Good in highly skewed datasets
- LDA is a probabilistic model: documents as distributions over topics → distributions over words.
- Each word in a document belongs to a single topic.

## LLDA ( K topics) – Algorithm steps:

1. For each document randomly assign each word to one of the K topics.
2. For each document d, Go through each word w in d:
3. For each topic t compute :  
$$\Pr(\text{topic } t \mid \text{document } d) * \Pr(\text{word } w \mid \text{topic } t)$$
4. The word will be reassign to the topic which maximize the function above.
5. Repeating steps 2-4 a large number of times. Will reach a steady state.



# Models – Big Document Model

- For each motif in the collection merge all documents into one big document.
- For each big document, we compute the TF·IDF (Okapi BM25) weighting function for all words.
- This provides a ranked list of how strongly a word is associated with a big document, i.e. a motif.



# Done!

- Now we have 2 tools to cluster words into topics.
- Which means we have 2 models to classify unseen folktales to a list of motifs.



# Ranking experiment

1. How well did both models ranked motifs of unseen folktales?
2. How well did both models cluster words to motifs?





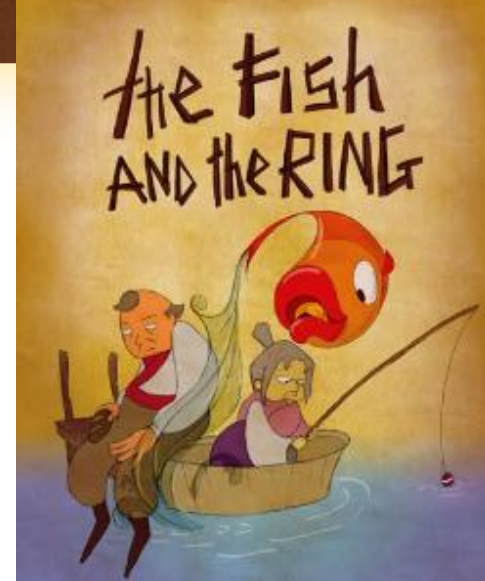
# Ranking motifs - evaluation

- Average Precision – Are most of the target motifs high up in the ranking?
- One Error – For what fraction of documents is the highest-ranked motif incorrect?
- Is Error – What fraction of rankings is not perfect?
- Margin – Average of differences between the highest ranked irrelevant motif and the lowest ranked relevant motif.

	Model	AP	One Error	Is Error	Margin
Dutch	BM25	0.78	0.26	0.27	10.69
	L-LDA	0.72	0.30	0.39	26.48
Frisian	BM25	0.88	0.15	0.15	4.46
	L-LDA	0.88	0.16	0.16	7.0

# Clustering

- Word distributions - LLDA vs. BDM



TD-IDF

L-LDA

## Q426: Wolf cut open and filled with stones as punishment.

wolf, Little Red Riding Hood, grandmother, her, children, little kids, your, Oud-Bovetje, big, granny, belly, goat, mother, angry

wolf, children, mother, door, said, open, her, little kids, so, entire, still, belly, surely, Oud-Bovetje, went

## N211.1: Lost ring found in fish.

Stavoren, teeth, cod, her, denture, ring, sea, wheat, ships, fish, shipper, harbor, she, the Heerhugowaard

the, her, and, she, the, of, in, was, a, lady, Stavoren, she, ring, sea, denture

## K343.2.1: The stingy parson and the slaughtered pig.

clerk, pastor, pig, stolen, will, slaughter, farmers, tonight, everyone, belief, sexton's house, fattened, excellent, slaughter time, pig meat, insignificant

clerk, pastor, pig, will, said, asked, everyone, stolen, mine, yes, so, must, against

## J2321.1: Parson made to believe that he will bear a calf.

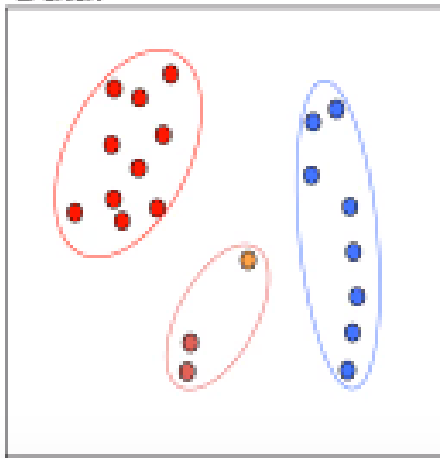
student, pastor, little bottle, cork, uroscopy, monkey, John, clerk, pregnant, rubber band, quack, butt, give birth, your, spins

the (de), a, pastor, John, student, the (het), to be, must, water, says, comes, to (te), to (om), and, surely

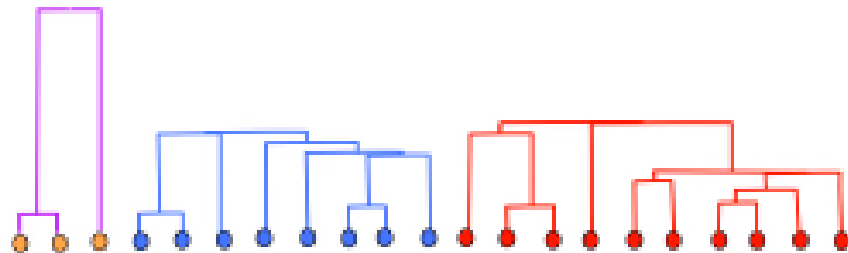
# What is Cluster analysis?

- Method that divides clusters into groups, based on the distance between clusters.
- Hierarchical cluster analysis for L-LDA and BDM.
- “Distances” by cosine similarity metric.

Data:



Dendrogram:



# The clusters test

- The founded groups of clusters were evaluated against the top 23 categories in Thompson's Motif Index.



## Evaluate the cluster solution:

- Homogeneity – Only members of the same class?
- Completeness – All members of the same class?
- V-measure – An entropy-based measure that expresses the harmonic mean of homogeneity and completeness.

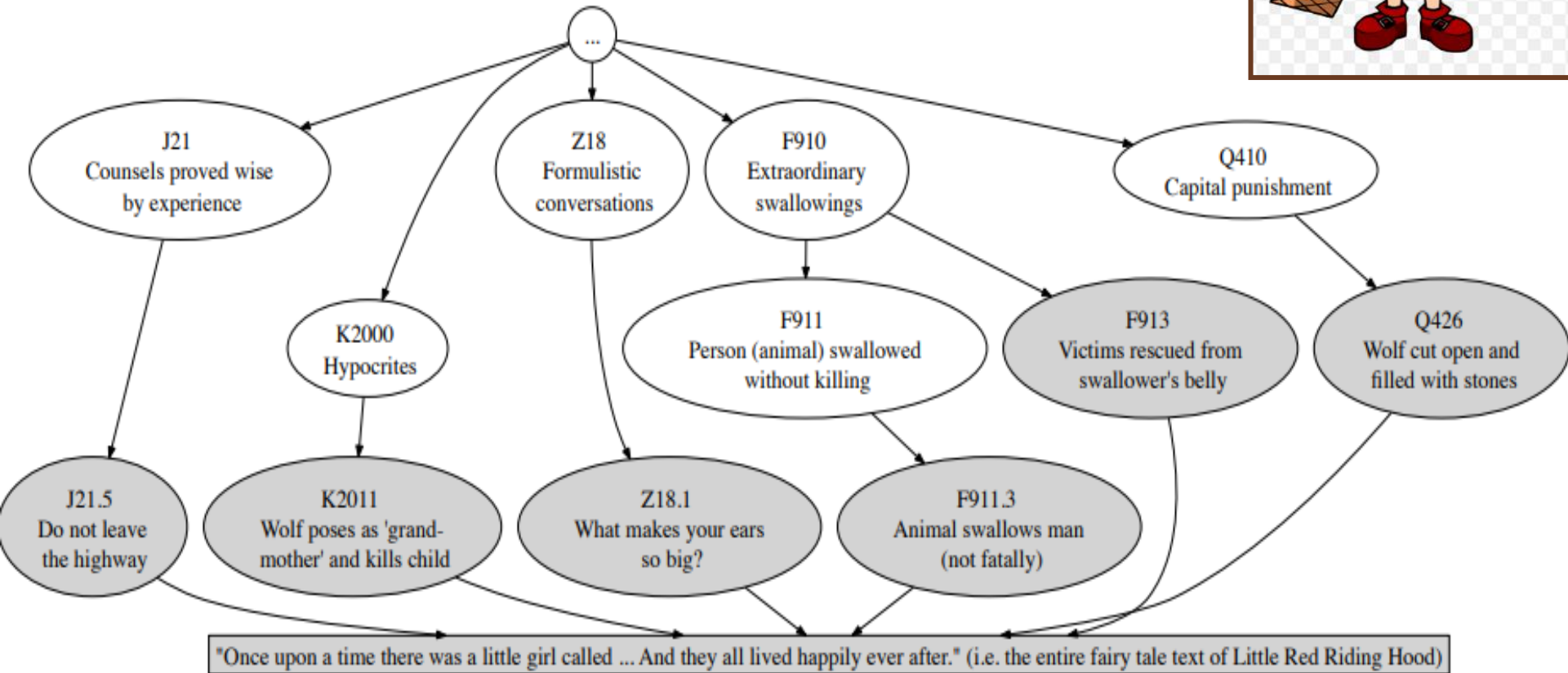
	Model	homogeneity	completeness	V-measure
Dutch	BDM	0.365	0.330	0.347
	L-LDA	0.344	0.281	0.310
Frisian	BDM	0.354	0.299	0.324
	L-LDA	0.358	0.270	0.308

# Exploiting the hierarchical structure of the Motif Index

- In the model described above the set of possible motifs was restricted those in the training data.
- Can we exploit the hierarchical structure of the motif index?

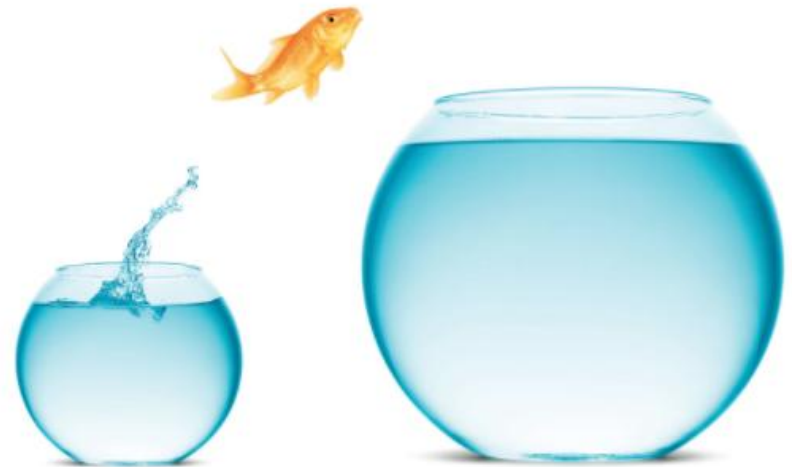


# Little Red Riding Hood



# Expanding the motifs set

- Adding all ancestral motifs in Thompson's Motif Index.
- Non-terminal nodes with at least two children.
- 410 motifs in the Dutch dataset and 293 motifs in the Frisian dataset





# Results of expanded model

	Model	homogeneity	completeness	V-measure
Dutch	BDM	0.339	0.315	0.327
	L-LDA	0.159	0.177	0.168
Frisian	BDM	0.414	0.377	0.394
	L-LDA	0.197	0.199	0.198



# What happened to LLDA?

- L-LDA - many content-free
- L-LDA assigns each word in a document to a single motif. i.e. specific motifs will get more lexically specific words.
- L-LDA is capable of finding good representations of motifs, the knowledge from higher level motifs is not inherited by their children.

J: 'The wise and the foolish'

wolf, Little Red Riding Hood, grandmother, blacksmith, cow, fish, emperor, Hodja, money, Hannes, farmer, little kids, children, table, Seppe, boy, devil, donkey, Oudbovetje

a, done, the, people, hear, maybe, old, soon, entire, tell, normal, we, in, someone, again, to look, real

a conduct'

key, sir, angry, naart, people, xden shoes, me, all, sit, does, i, walk, Nasreddin, ge, winter, eyes, iss

J1700.-J2749: 'Fools (and other unwise persons)'

blacksmith, cow, emperor, Hannes, table, devil, money, Pieterke, donkey, farmer, boy, bag

money, village, farmer, came, donkey, happened, could, stand, John, hour, was, city, Bakel, must, he, all, sat

J2328: 'The moving church tower.'

coats, church, Bakel, movement, how far, push, exceeded, disinterested, savvy, concord, helpful, cooperation, united

church, coats, the, power, possible, Bakel, push, building, perceive, chariot, disinterested, offer, annoying, agree, exceeded

J1750.-J1849: 'Absurd misunderstandings'

vivat bakelorum, bear, woodcutter, Chelm, ships, moon, oranges, 'nmosquito, pigeon, village

the, and, in, those, she, there, you, not, then, moon, whole, is, way, out, went, surely, said, became, already

## Future Research

- Improving the quality of motif representations as discovered by L-LDA
- Or, developing a system that incorporates the motif representations found by BDM, by finding those parts of a text that support a detected motif best.

If we can automatically find motifs in a story, what else can we do?



Can a machine write or “understand”  
folktales?

TOM SIMONITE BUSINESS 10.13.17 07:00 AM

# GOOGLE'S LEARNING SOFTWARE LEARNS TO WRITE LEARNING SOFTWARE

## Now AI Machines Are Learning to Understand Stories

Face and speech recognition is now child's play for the most advanced AI machines. But understanding stories is much harder. That looks set to change.

December 14, 2015

# AI can write surprisingly scary and creative horror stories

She's called Shelley, and you can find her on Twitter.



**Shelley**  
@shelley\_ai

**NEW** I then saw a shadow in the shadows. It was tall, but I could tell that it was a little older than my own age, so I could make out a 1/3

1:45 PM - Oct 31, 2017



**Shelley**  
@shelley\_ai

stars, with no eyes. 3/3 #yourturn

1:45 PM - Oct 31, 2017



**Shelley** @shelley\_ai

31 Oct

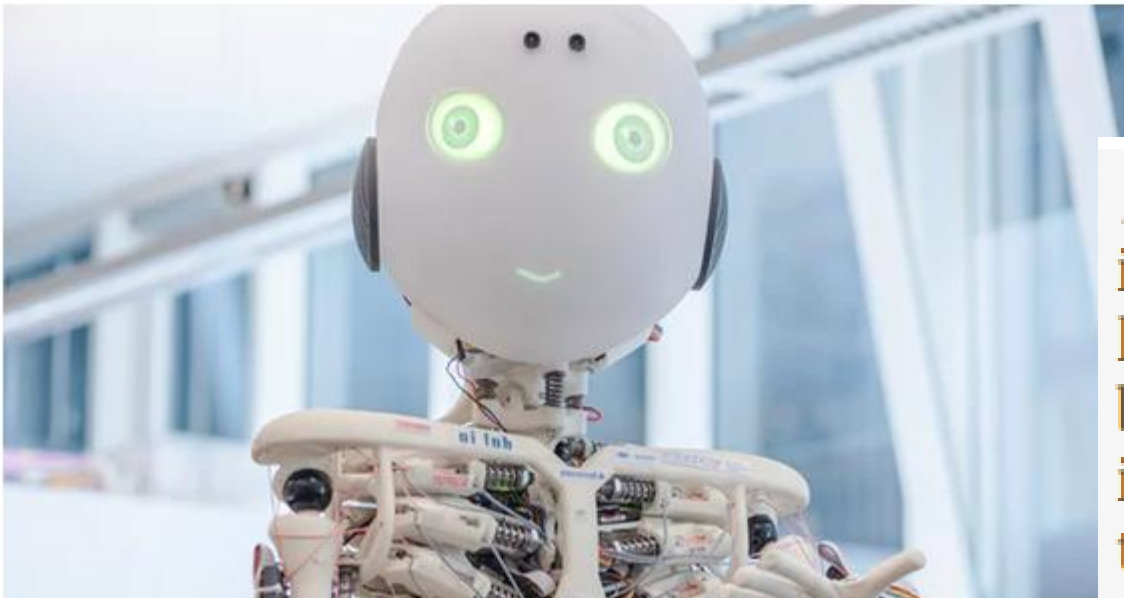
Replying to @shelley\_ai

face. I started to panic, then I looked at it. And it was staring directly at me when I looked right at it. It was black with black 2/3

# Can AI learn from stories?

## Robots could learn human values by reading stories, research suggests

Scientists have been running tests where artificial intelligences cultivate appropriate social behaviour by responding to simple narratives



“ We think that an intelligent entity can learn what it means to be human by immersing itself in the stories it produces  
Associate professor  
Mark Riedl



# Computational reasoning

- A culture - shared beliefs, customs, and products
- no user manual for being human
- Not all humans act morally all the time, but they tend to follow social norms.

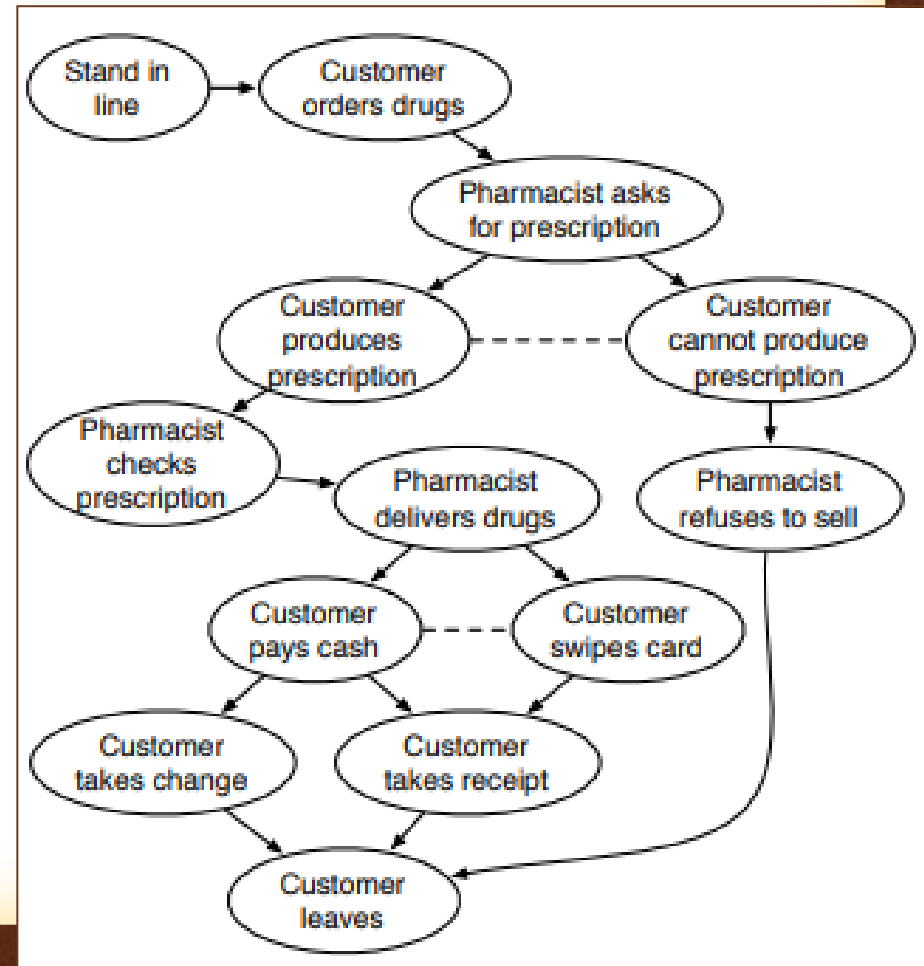
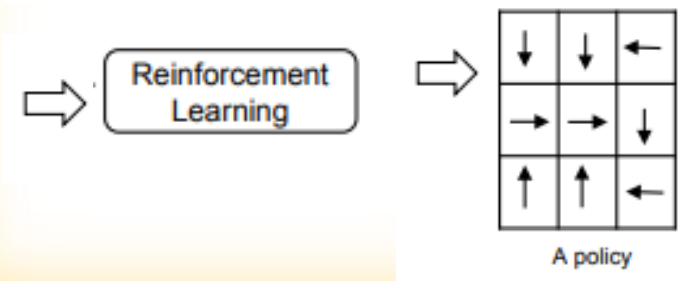
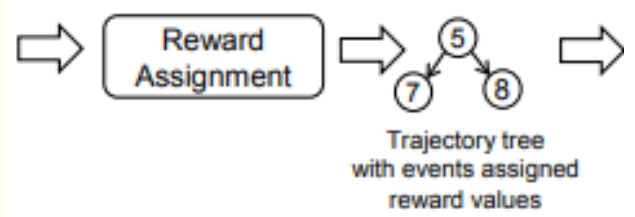
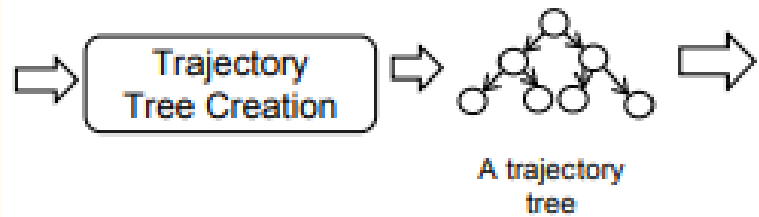
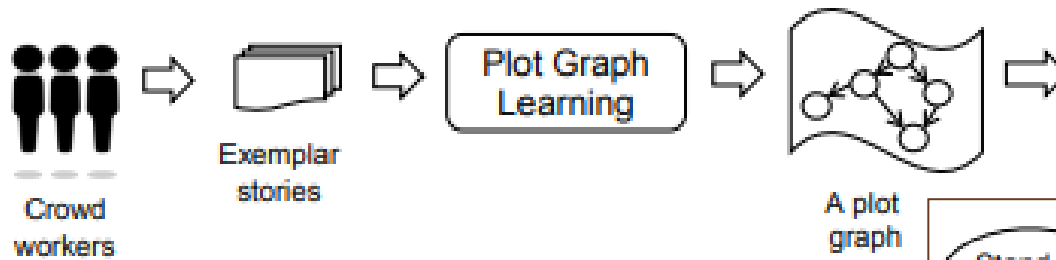
# No User Manual?

- “Childhood” - To raise an entity in a human environment
- Observe human behavior
- Storytelling - communicating tacit knowledge

# Reinforcement learning (RL)

- Learning how to act in a world so as to maximize a reward signal.
- The problem is defined as  $\langle S, A, P, R \rangle$ :
  - $S$  is a set of states
  - $A$  is a set of actions/effectors the agent can perform
  - $P : \{S \times A \times S\} \rightarrow [0, 1]$  is a transition function
  - $R : S \rightarrow R$  is a reward function.
- The solution to a RL problem is a policy  $\pi : S \rightarrow A$ .
- An optimal policy ensures that the agent receives maximal long-term expected reward.

# The process for generating value-aligned behavior from crowdsourced stories.



# Problems..

- Stories written in natural language can contain events and actions that are not executable by an artificial agent.
- Stories are written by humans for humans and thus make use of commonly shared knowledge, leaving many things unstated.
- Stories frequently skip over events that do not directly impact the telling of the story, and sometimes also employ flashbacks which may confuse an artificial learner.

# Computational Narrative Intelligence

- Narrative intelligence is the ability to craft, tell, and understand stories.
- Winston (2011) argues that narrative intelligence is one of the abilities that sets humans apart from other animals and non-human-like artificial intelligences.

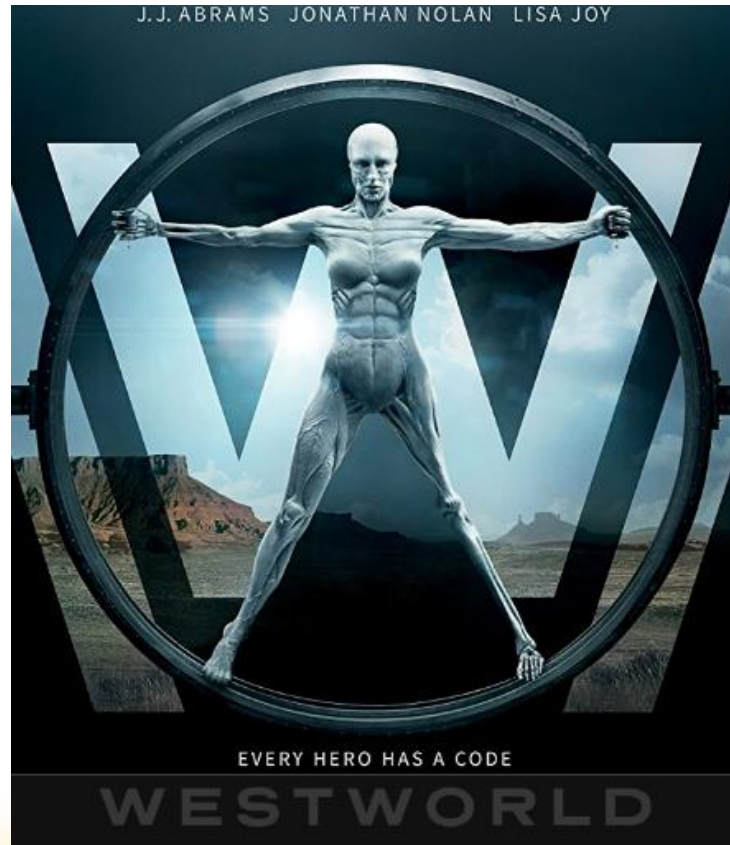
# To conclude

- Finding Motifs automatically from folktales can be achieved in a rather good accuracy. (which will get better)
- Computational reasoning is a not that far from reach.



# Westworld trailer

- <https://www.youtube.com/watch?v=IuS5huqOND4>
- (till 00:51)





# Resources

- **“Identifying Motifs in Folktales using Topic Models”** by Folgert Karsdorp and Antal van den Bosch
- **“Learning about World Cultures through Folktales”** -  
<http://www.socialstudies.org/sites/default/files/publications/yl/1101/110104.html>
- <https://en.wikipedia.org/>
- <https://www.wired.com/story/googles-learning-software-learns-to-write-learning-software/>

# Resources

- <https://www.technologyreview.com/s/544506/now-ai-machines-are-learning-to-understand-stories/>
- <https://www.engadget.com/2017/10/31/shelley-ai-writes-horror-stories-on-twitter/>
- **“Using Stories to Teach Human Values to Artificial Agents”** by Mark O. Riedl and Brent Harrison: <https://www.cc.gatech.edu/~riedl/pubs/aaai-ethics16.pdf>
- **“Learning a Better Motif Index: Toward Automated Motif Extraction”** by W. Victor H. Yarlott and Mark A. Finlayson: <http://drops.dagstuhl.de/opus/volltexte/2016/6708/pdf/OASICS-CMN-2016-7.pdf>