CLOSED DOMAIN QUESTION ANSWERING, EMBEDDING SPACES

AN INTELLIGENT SYSTEM FOR ENTERPRISE-WIDE OPEN DOMAIN QUESTION ANSWERING

MUNTIS RUDZĪTIS

MUNTIS.RUDZITIS@GMAIL.COM

DARBA VADĪTĀJS: DR. DAT., PROF. GUNTIS BĀRZDIŅŠ

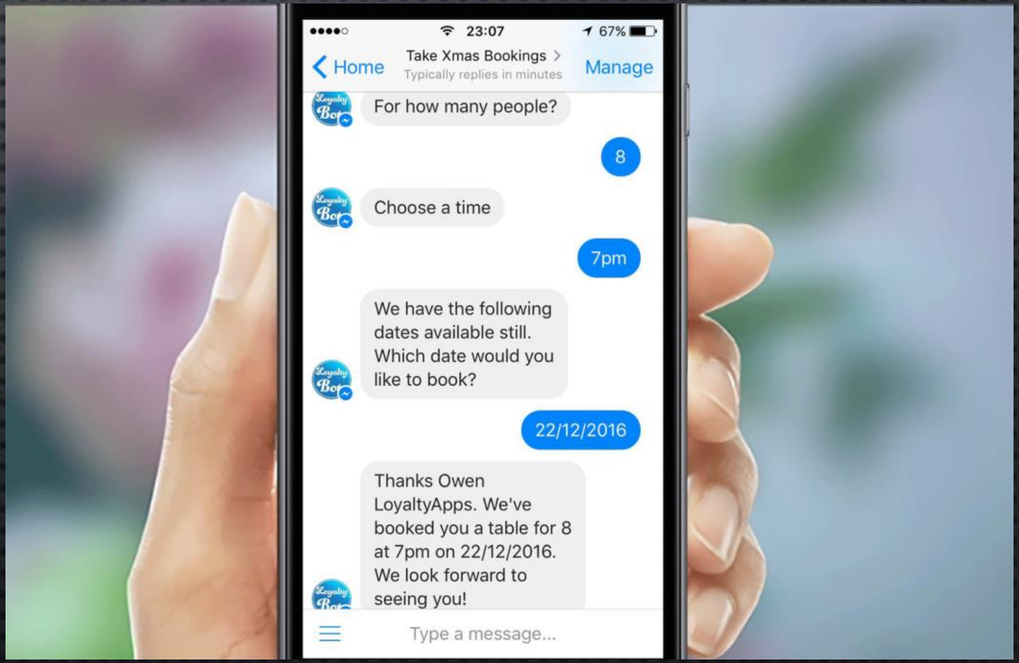
Virtual Assistant

- Amazon echo / Alexa
- Siri
- Google Assistant

- Messenger / chat bots
- Image / AR bots
- Programmable API / Hooks to devices

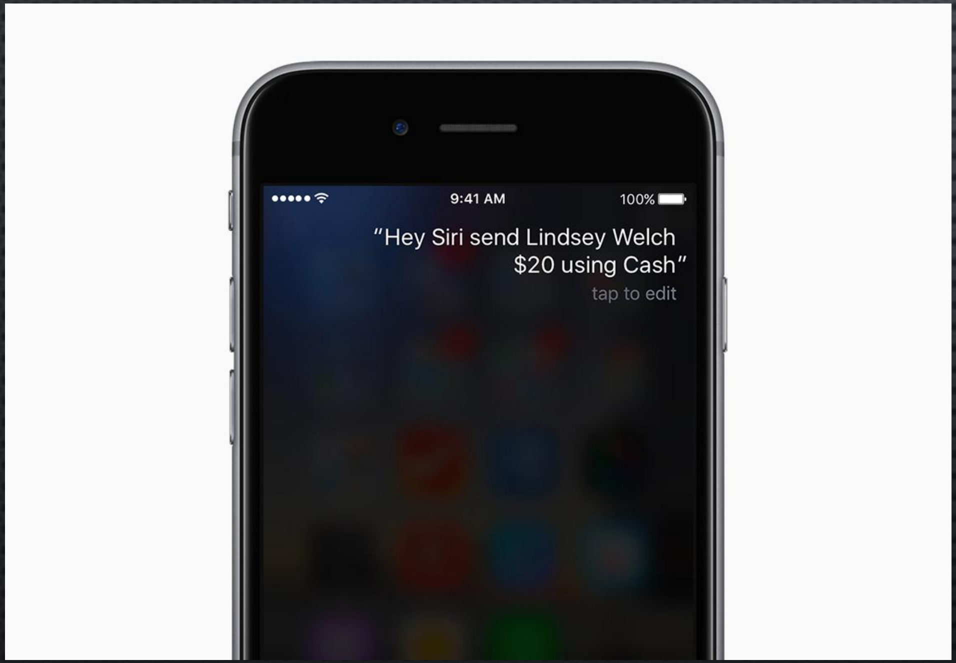
ASSISTANTS & CHATBOTS

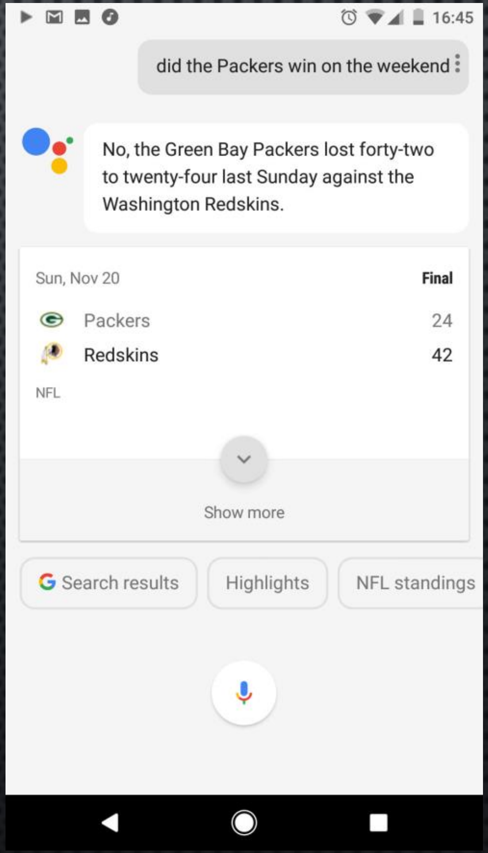
- Text / voice interface
- Simple questions
- Integrated in device (phone) or whole platform (Google)
- Intent recognition
- Not context aware (usually)



https://chatbotslife.com/5-benefits-of-using-a-chatbot-937b8b793826

USE CASES

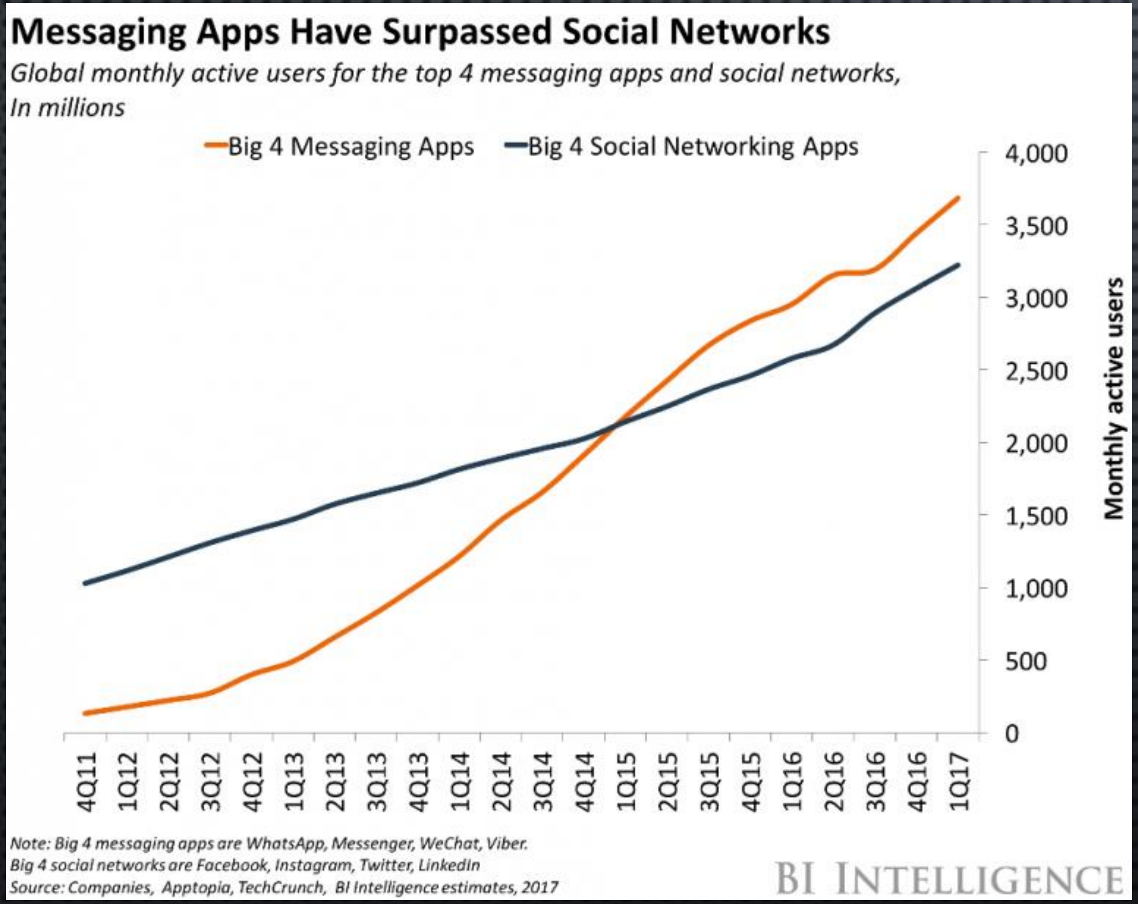


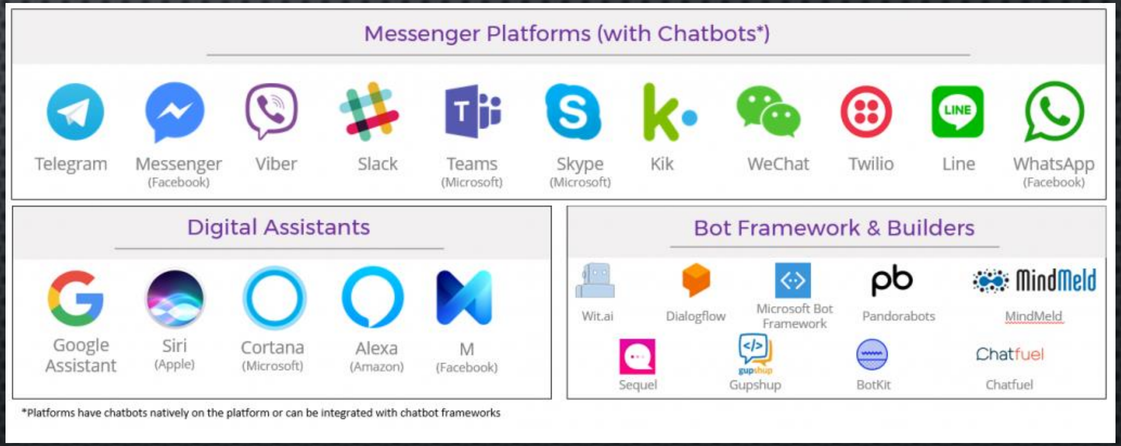


https://media.idownloadblog.com/wp-content/uploads/2016/09/iOS-10-Siri-Cash-payment-teaser-001.jpg

https://www.androidcentral.com/heres-what-google-assistant-can-do-your-pixel

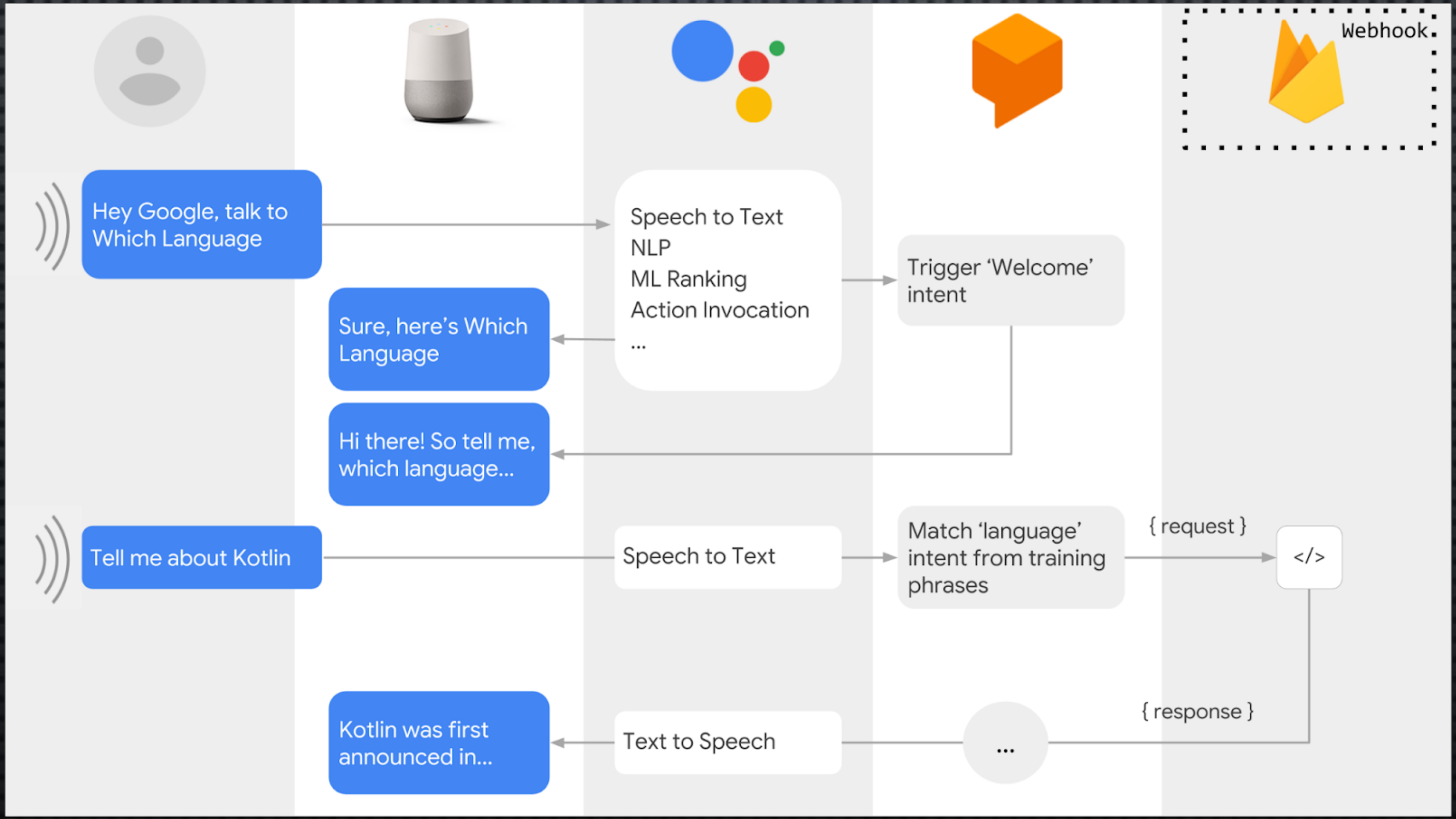
CHATBOTS & CHATBOT PLATFORMS



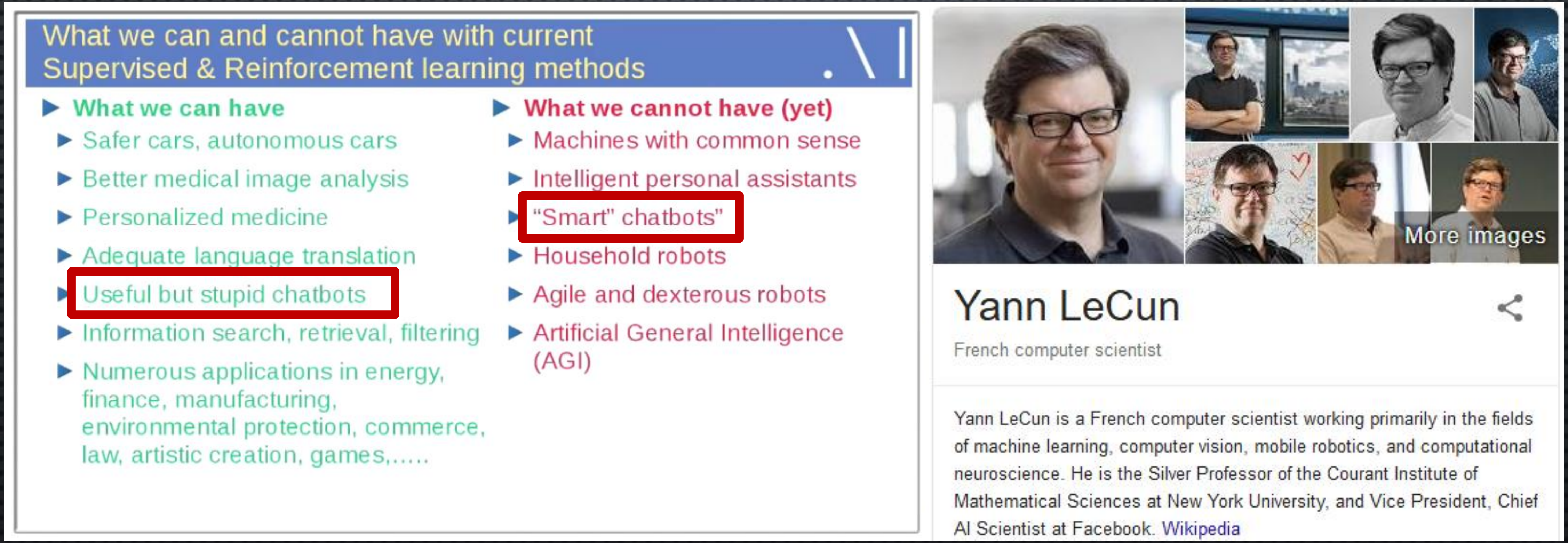


https://lenoxhill.com.au/blog/chatbots-fad-future/

HOW IT WORKS


Webhook
Hey Google, talk to Which Language
Speech to Text
NLP
ML Ranking
Action Invocation
...
Trigger 'Welcome' intent
Sure, here's Which Language
Hi there! So tell me, which language...
Tell me about Kotlin
Speech to Text
Match 'language' intent from training phrases
{ request }
</>
{ response }
Kotlin was first announced in...
Text to Speech
...

YANN LECUN - THE NEXT STEP TOWARDS ARTIFICIAL INTELLIGENCE

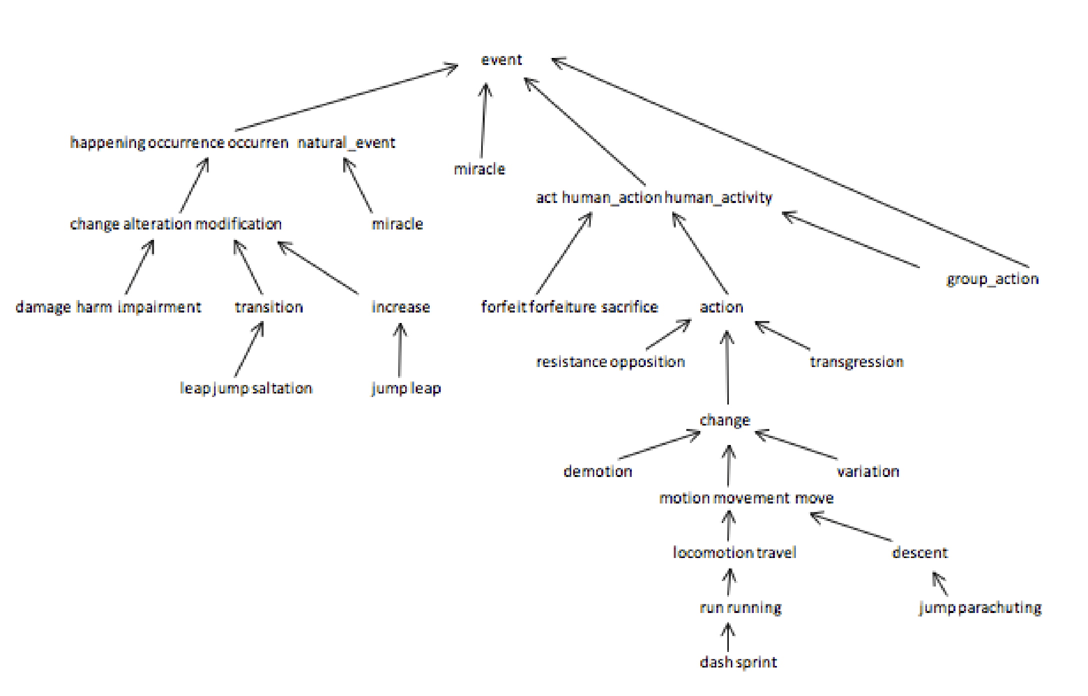


https://www.youtube.com/watch?v=U2mhZ9E8Fk8

INFORMATION & KNOWLEDGE EXTRACTION

- Structured data
  - Requires structured knowledge
  - WordNet

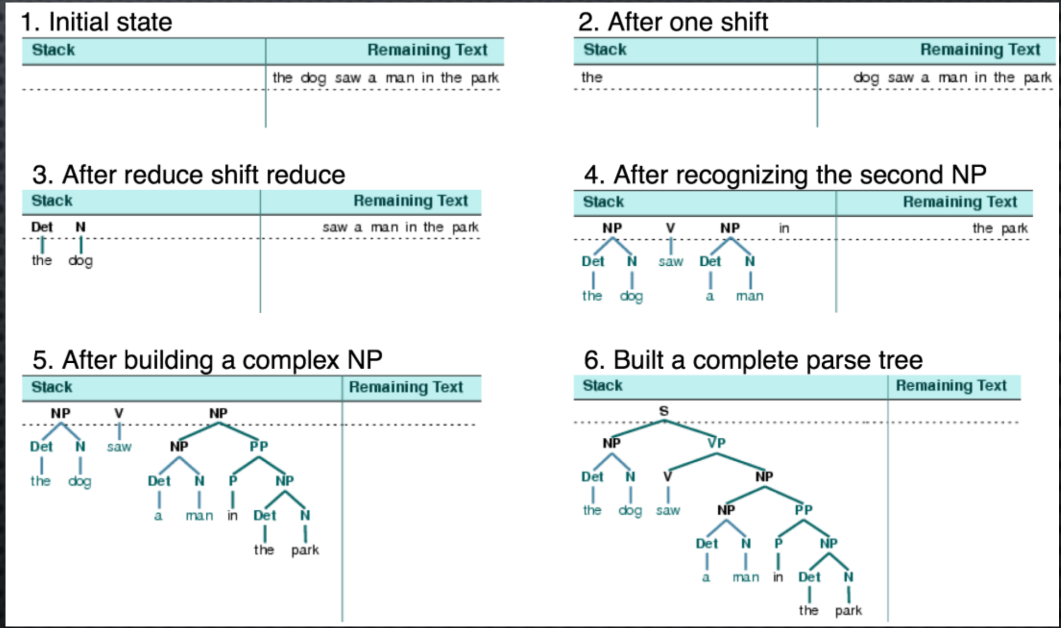
- Not enough structured data for production use



http://www.cs.princeton.edu/courses/archive/fall11/cos226/assignments/wordnet-fig1.png

INFORMATION & KNOWLEDGE EXTRACTION

- Unstructured data:
  - Can work with (mostly) no structured knowledge
- Natural Language Processing (NLP)
- Enough data for production use, requires preprocessing
- Embeddings



https://www.nltk.org/book/ch08.html

WORD EMBEDDINGS

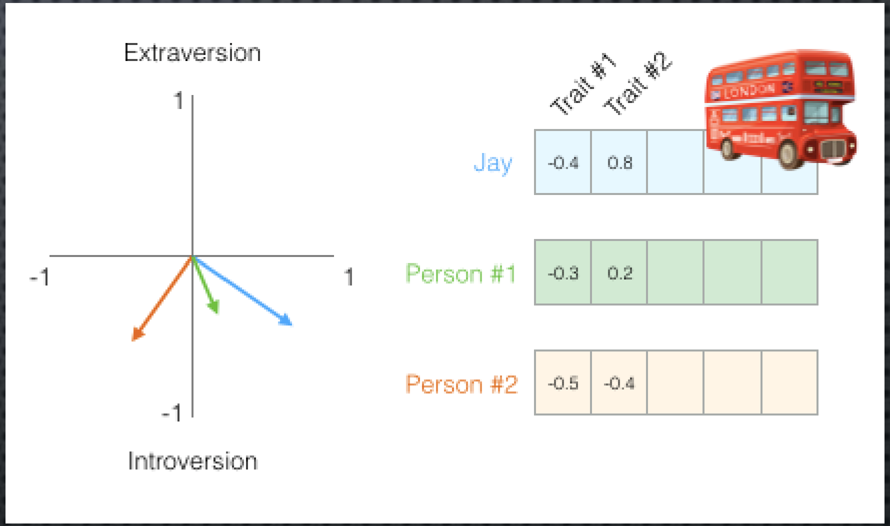
"You shall know a word by the company it keeps"
John Rupert Firth, 1957

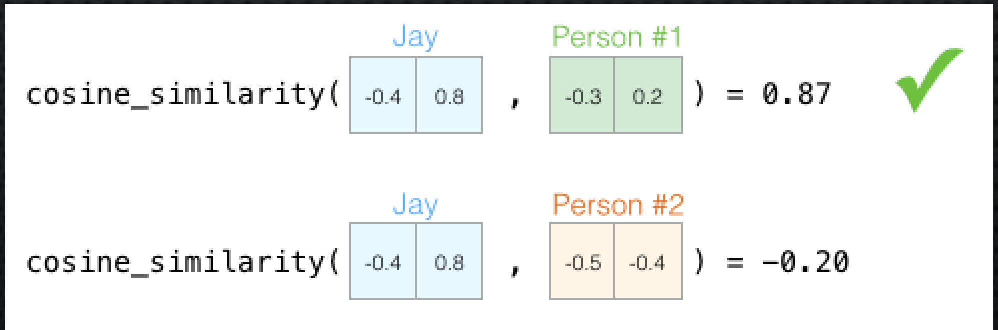
WORD EMBEDDINGS

- **Efficient Estimation of Word Representations in Vector Space (Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean, 2013)** https://arxiv.org/abs/1301.3781
- Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation

WORD EMBEDDINGS

Openness to experience	79 out of 100
Agreeableness	75 out of 100
Conscientiousness	42 out of 100
Negative emotionality	50 out of 100
Extraversion	58 out of 100





WORD EMBEDDINGS

- I like deep learning.
- I like NLP.
- I enjoy flying.

counts	I	like	enjoy	deep	learning	NLP	flying	.
I	0	2	1	0	0	0	0	0
like	2	0	0	1	0	1	0	0
enjoy	1	0	0	0	0	0	1	0
deep	0	1	0	0	1	0	0	0
learning	0	0	0	1	0	0	0	1
NLP	0	1	0	0	0	0	0	1
flying	0	0	1	0	0	0	0	1
.	0	0	0	0	1	1	1	0

Co-correlation matrix:

- Not preserving word order!
- Gets HUGE fast (n^2)

http://www.erogol.com/wp-content/uploads/2016/08/cooccurance_matrix.png

WORD EMBEDDINGS

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make
not	make	a
make	a	machine
a	machine	in

https://jalammar.github.io/illustrated-word2vec/

WORD EMBEDDINGS

Continuous Bag of Words (CBOW)

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make
not	make	a
make	a	machine
a	machine	in

Skipgram

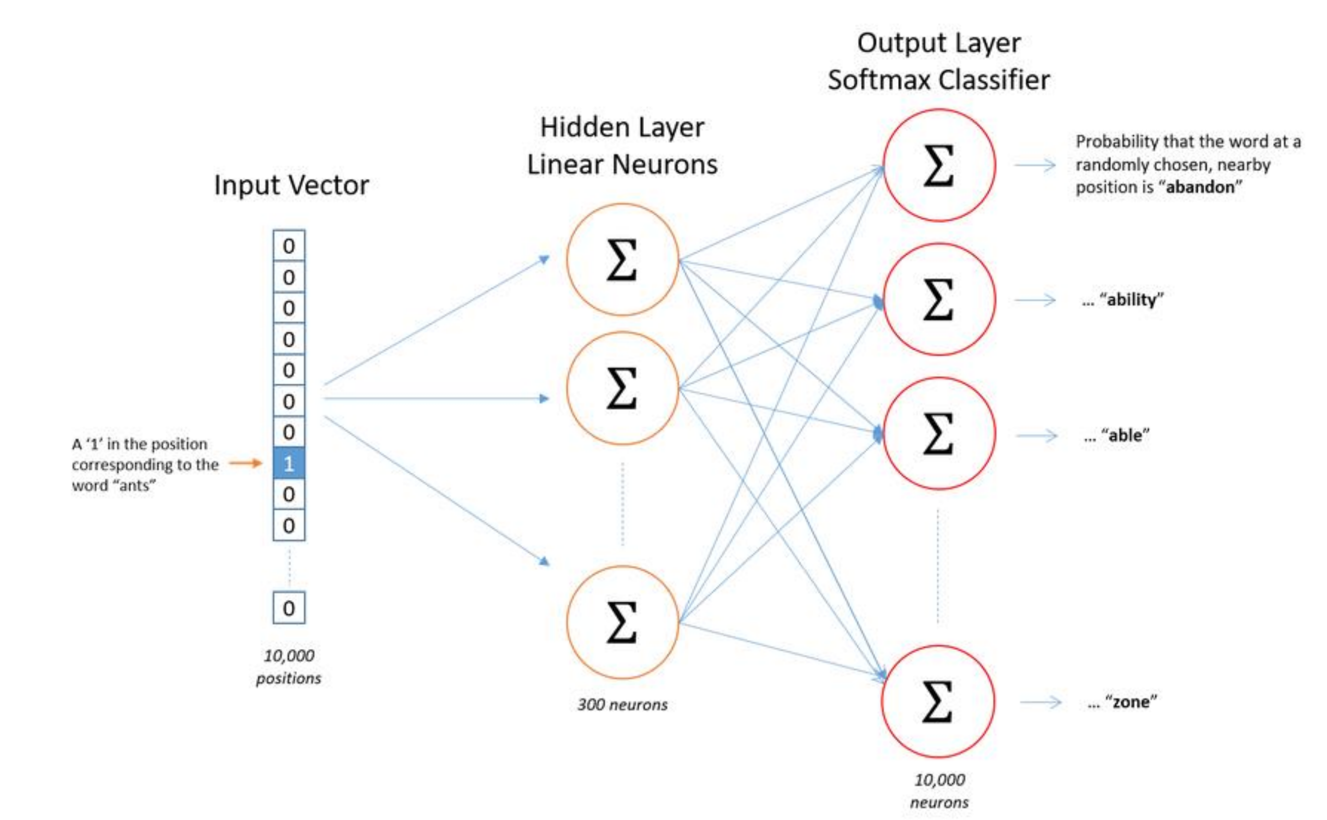
Thou shalt not make a machine in the likeness of a human mind

thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	…

input word	target word
not	thou
not	shalt
not	make
not	a
make	shalt
make	not
make	a
make	machine
a	not
a	make
a	machine
a	in
machine	make
machine	a
machine	in
machine	the
in	a
in	machine
in	the
in	likeness

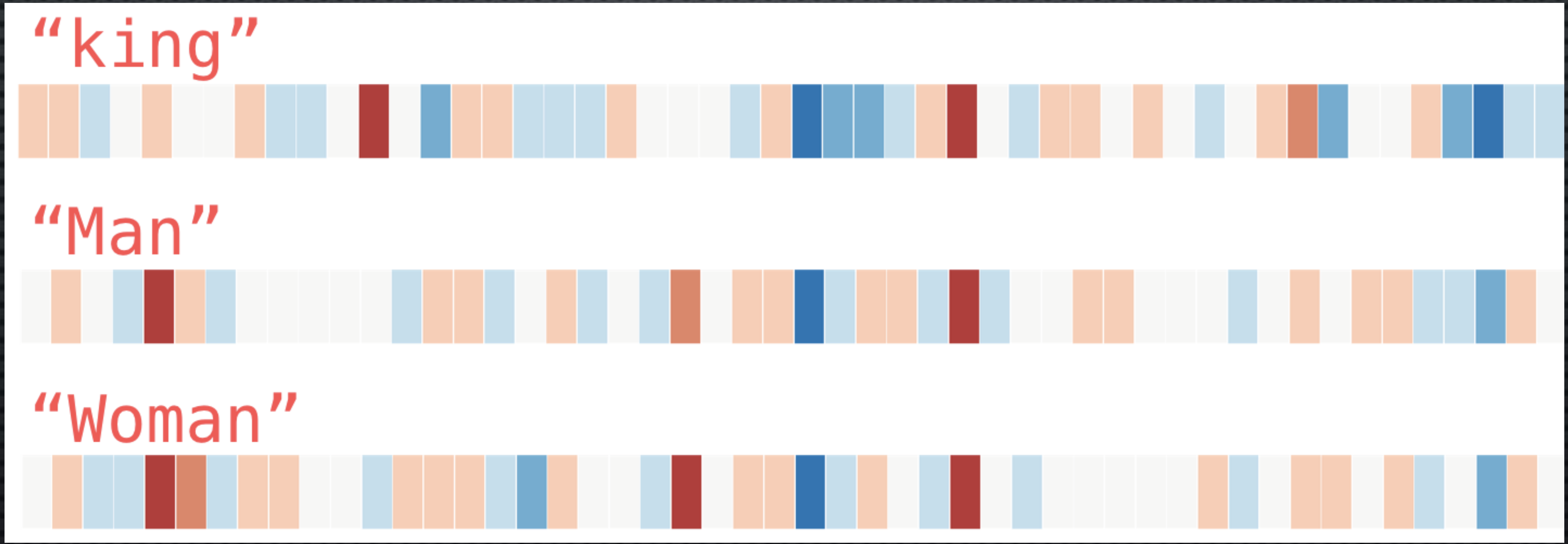
https://jalammar.github.io/illustrated-word2vec/

WORD EMBEDDINGS



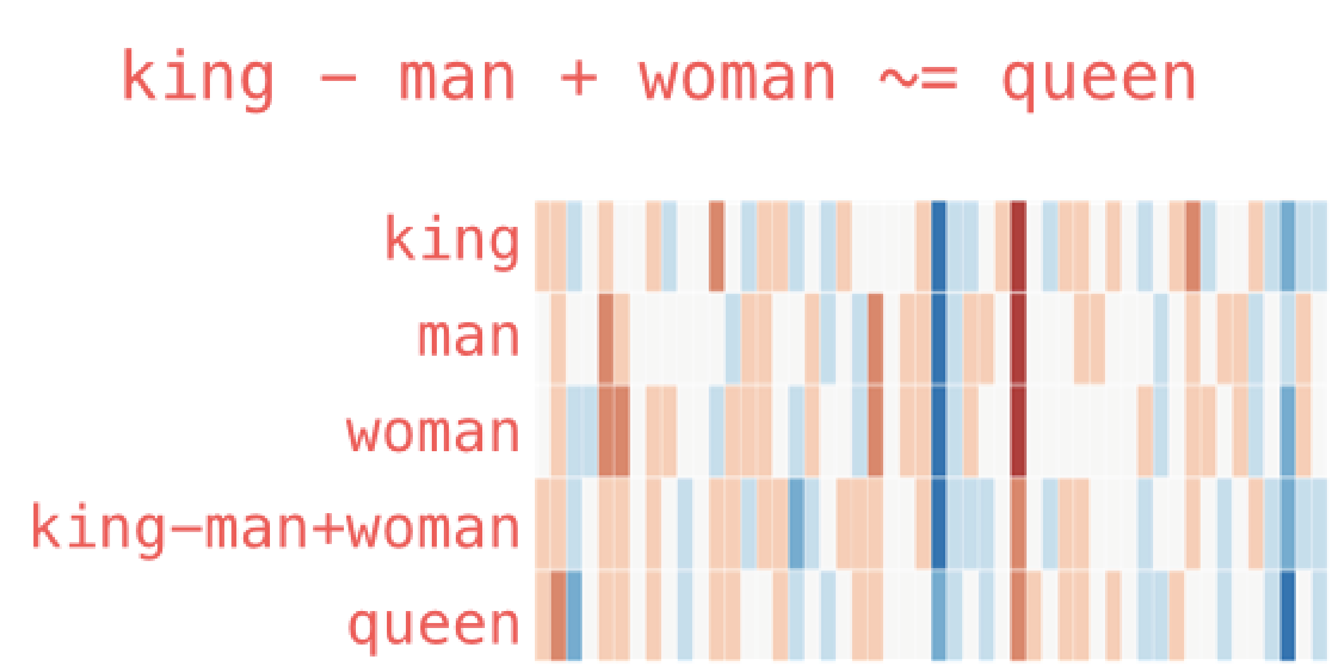
http://mccormickml.com/2016/04/19/word2vec-tutorial-the-skip-gram-model/

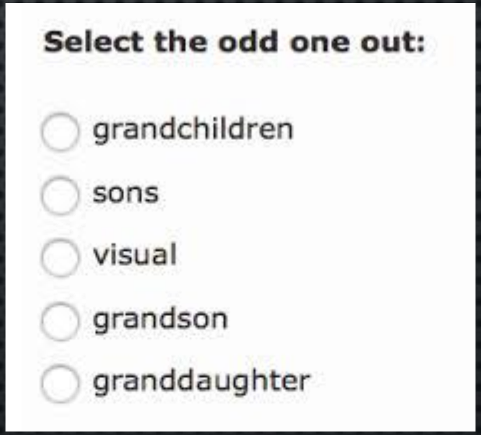
WORD EMBEDDINGS



https://jalammar.github.io/illustrated-word2vec/

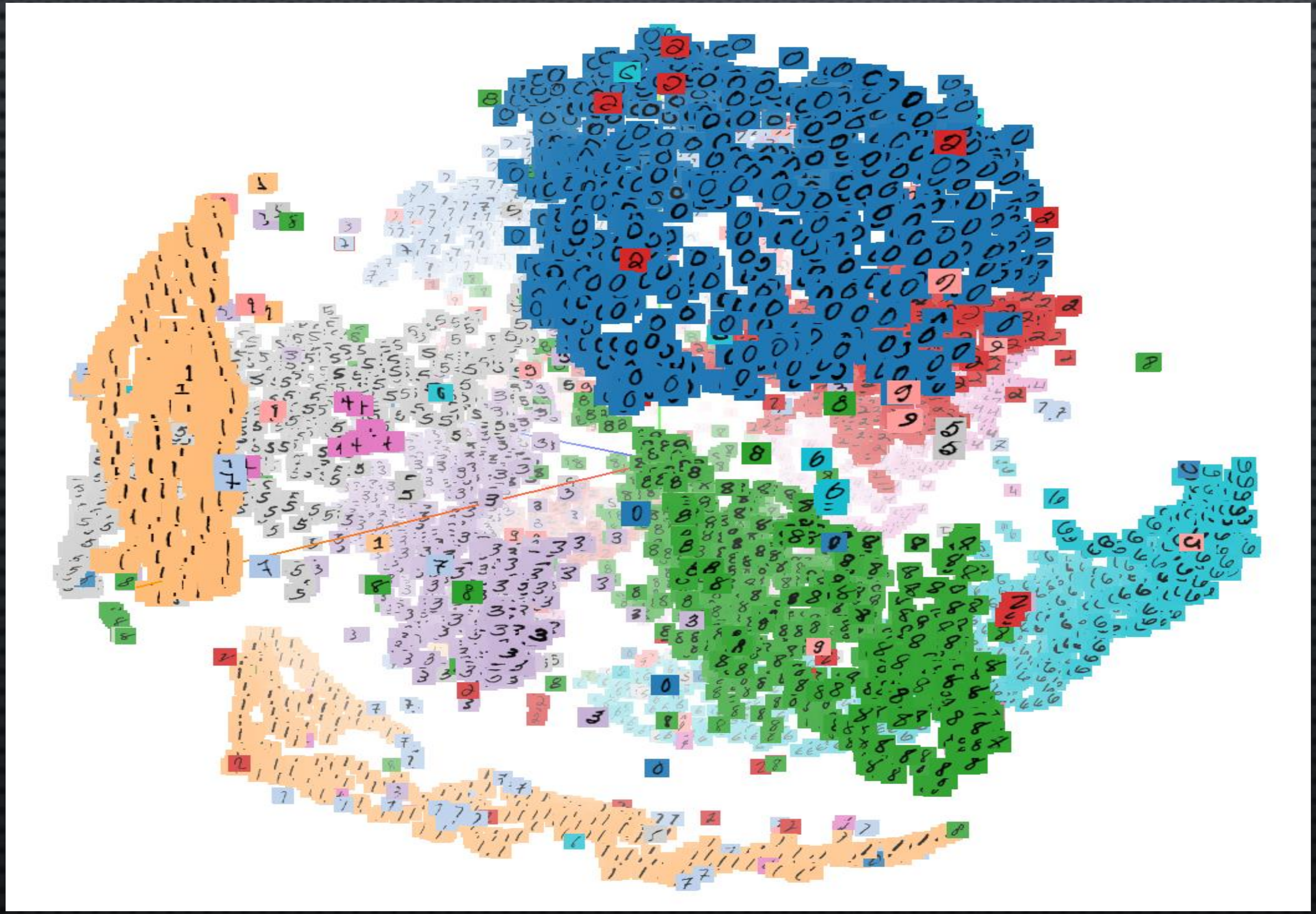
WORD EMBEDDINGS





https://jalammar.github.io/illustrated-word2vec/

OTHER EMBEDDINGS



http://projector.tensorflow.org/

OTHER EMBEDDINGS

- Images
- Text
- Entities
- Speech2vec
- **Doc2vec**
- Graph2vec
- Tweet2vec

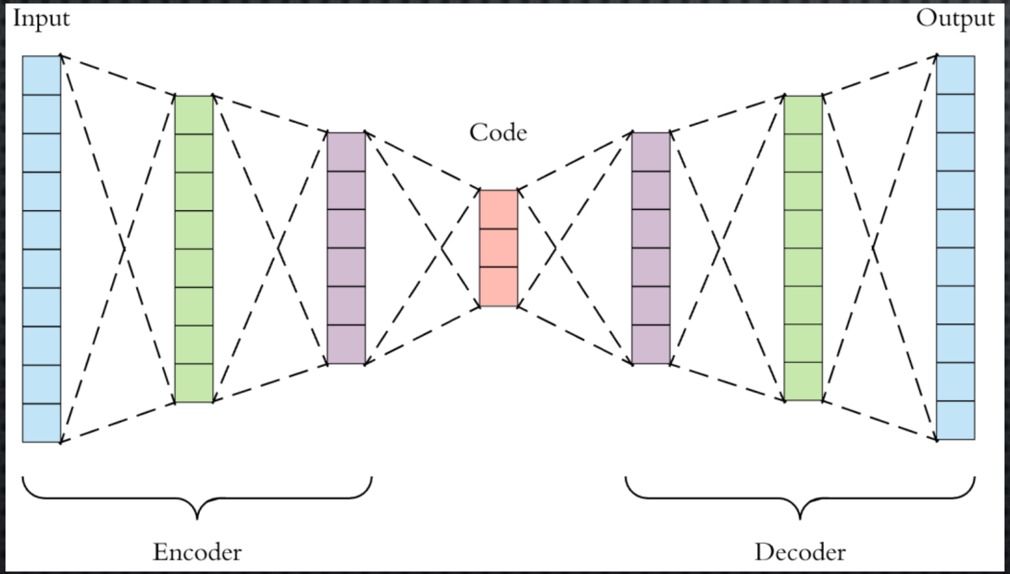
… many others

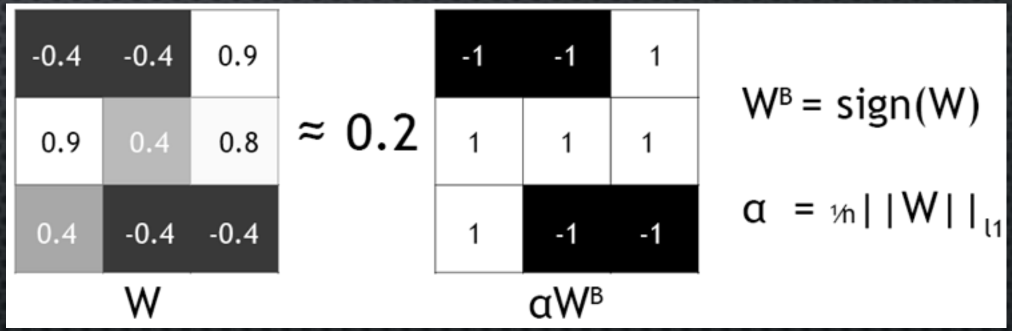
VECTOR SPACE -> FEATURE SPACE

- Embeddings are mysterious and not explainable
- SPINE: SParse Interpretable Neural Embeddings, 23 Nov 2017 (https://arxiv.org/abs/1711.08792)
- Vector space -> Sparse binary space

VECTOR SPACE -> FEATURE SPACE

SPINE: SParse Interpretable Neural Embeddings





https://software.intel.com/sites/default/files/managed/c0/e0/webops10048-fig1-binarization-procedure.png

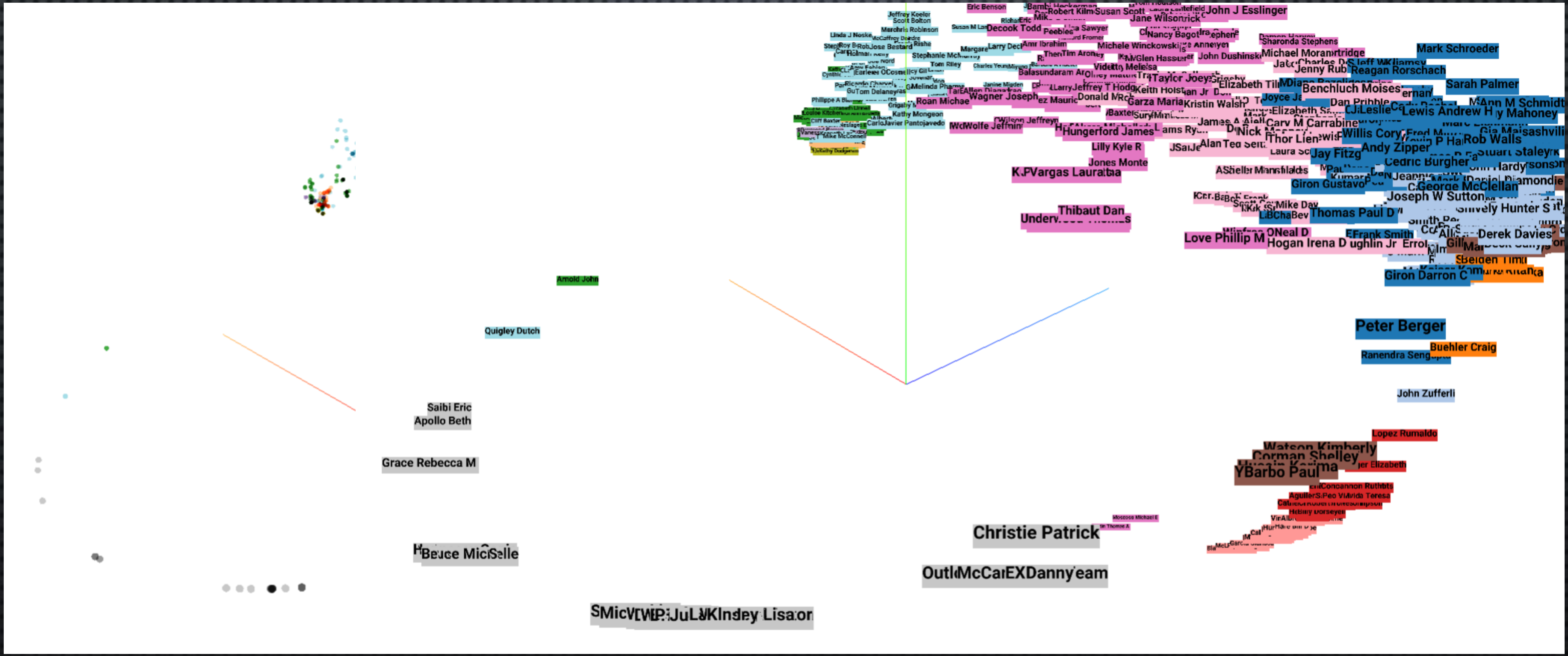
https://towardsdatascience.com/applied-deep-learning-part-3-autoencoders-1c083af4d798

Knowledge Vectorspace

- Create knowledge vector space
  - Doc2vec / custom and mixed embedding
  - Topic model for people, documents, reports, files..

- From context / Topic read point in space
  - Recommend data objects
  - Verify current choice
  - Find missing "crucial" objects

KNOWLEDGE VECTORSPACE



THANKS!

- Questions & comments?
- Muntis.Rudzitis@gmail.com