

An aerial photograph showing a vast expanse of broken ice floes floating on a deep blue sea. The ice floes are irregular in shape and size, ranging from small chunks to large, flat sheets. The water between the floes is a dark, deep blue, while the ice itself is a lighter, milky blue. The overall scene conveys a sense of a cold, desolate environment.

# Breaking the ice with NLP

Mobile Refresh | Wellington | 2018

**Ilia Kopylov**  
**ML Engineer @ Springload**

## A quick tour through...

- Background
- NLP
- Chatbots
- Dialogflow
- Wrap-up



An aerial photograph of a vast, frozen body of water, likely a sea or ocean. The surface is covered with a dense field of ice floes of various sizes and shapes. The ice has a light blue-grey hue, with some areas appearing more translucent than others. The floes are separated by thin, dark lines of open water or meltwater. The overall texture is rough and fragmented. In the bottom left corner, the word "Background" is written in a large, bold, black sans-serif font.

**Background**

- ***Feature***

An individual measurable property or characteristic of a phenomenon being observed

- ***Feature Engineering***

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG  
PILE OF LINEAR ALGEBRA, THEN COLLECT  
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL  
THEY START LOOKING RIGHT.





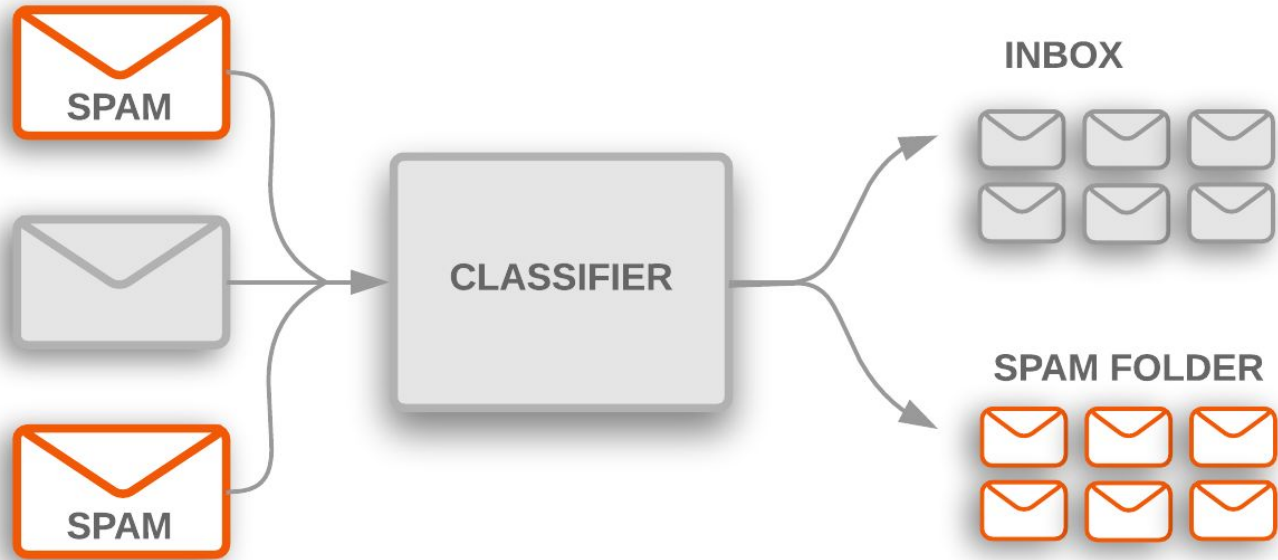


# **Natural Language Processing**

# Natural Language Processing

- Understanding
- Generation
- Speech Recognition

# Text Classification





## Bag of words

*“Is this the real life?  
Is this just fantasy?”*

[ 2, 2, 1, 1, 1, 1, 1 ]

is	2
this	2
the	1
real	1
life	1
just	1
fantasy	1

[ 2, 2, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0,  
0,  
0,  
0,  
0,  
0,  
0,  
0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ...



## Stop words

the, is, which, at, on...

## Stemming

real, reality, realistic... -> real

## Correct spelling ~~mistakes~~ mistakes

## Normalise/normalize spellings

## TF-IDF

# TF-IDF

term frequency–inverse document frequency



## TF-IDF

*Is this the real life?*

is	1
this	1
the	1
real	1
life	1
just	0
fantasy	0

*Is this just fantasy?*

is	1
this	1
the	0
real	0
life	0
just	1
fantasy	1

**Total**

is	2
this	2
the	1
real	1
life	1
just	1
fantasy	1

## TF-IDF

*Is this the real life?*

is	0.5
this	0.5
the	1
real	1
life	1
just	0
fantasy	0

*Is this just fantasy?*

is	0.5
this	0.5
the	0
real	0
life	0
just	1
fantasy	1

**Total**

is	2
this	2
the	1
real	1
life	1
just	1
fantasy	1

## TF-IDF

*Is this the real life?*

is	0.000001
this	0.000002
the	1
real	1
life	1
just	0
fantasy	0

*Is this just fantasy?*

is	0.000001
this	0.000002
the	0
real	0
life	0
just	1
fantasy	1

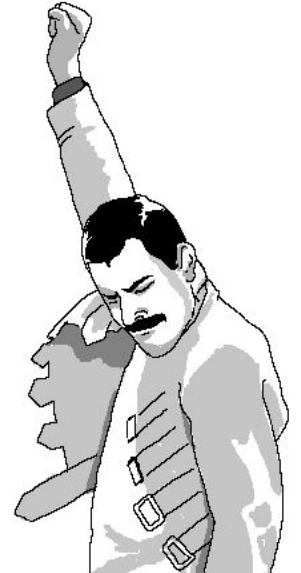

**Total**

is	10000
this	5000
the	1
real	1
life	1
just	1
fantasy	1

[illegible]



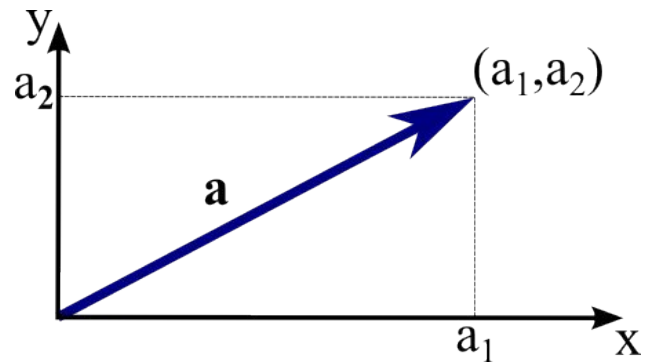
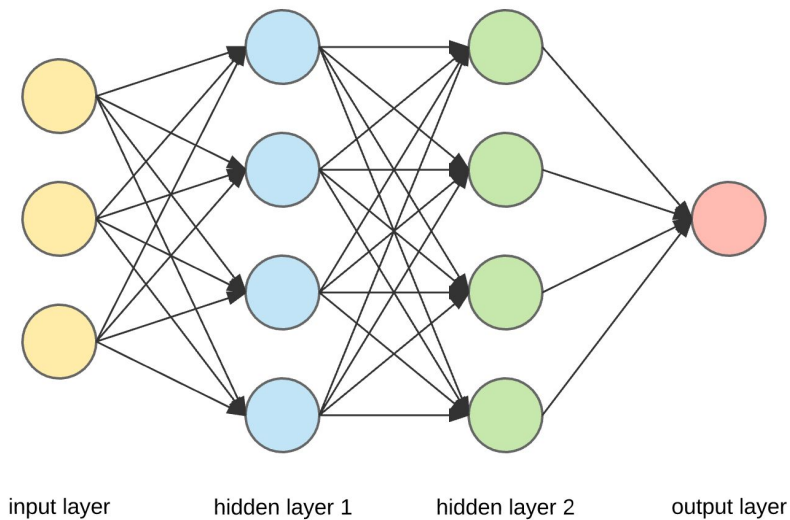
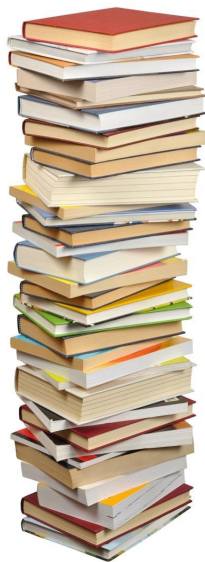
[ 0.5, 0.5, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ...



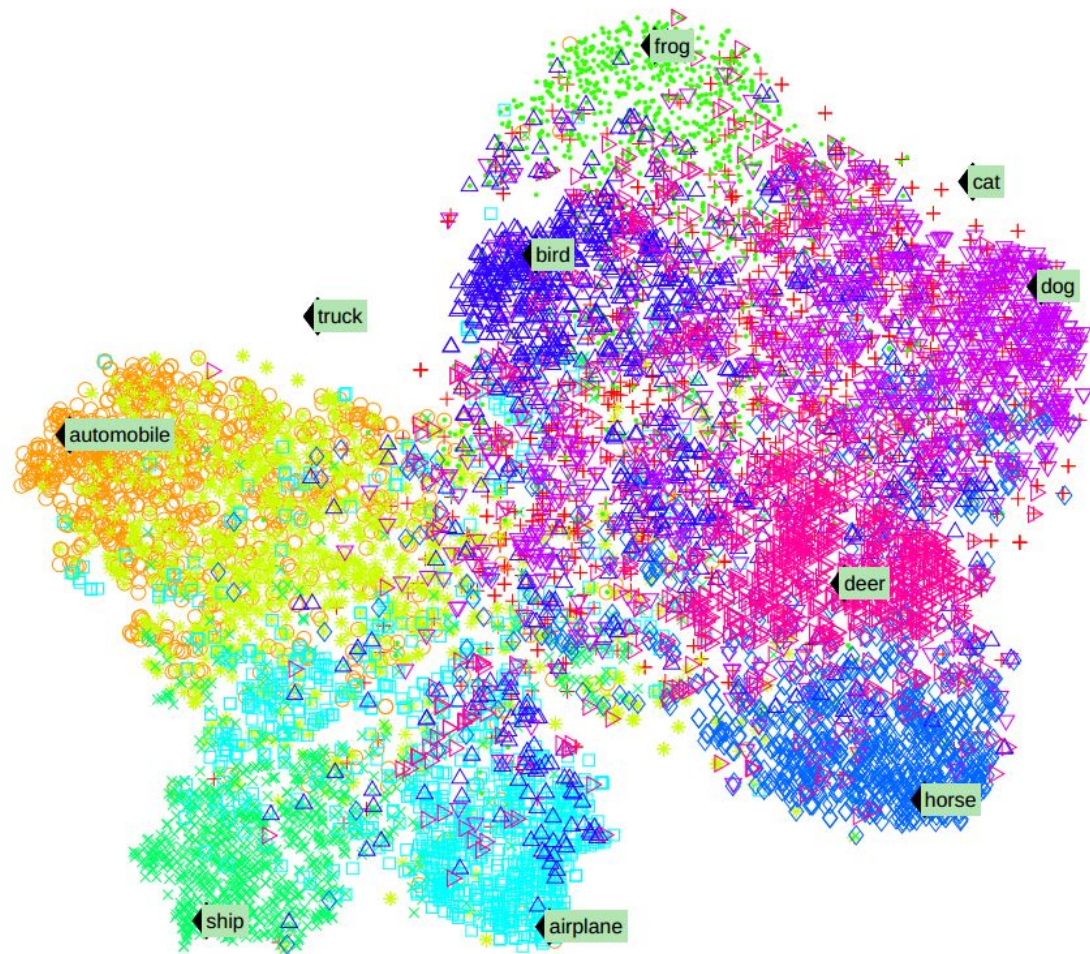
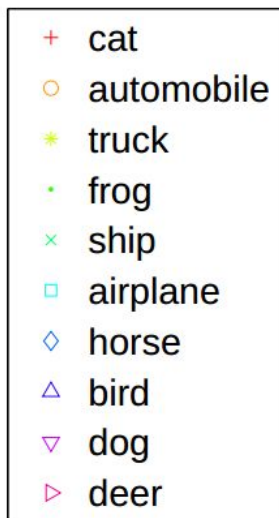
[ 0.5, 0.5, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  
0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 ...

## Sparse Vector Models

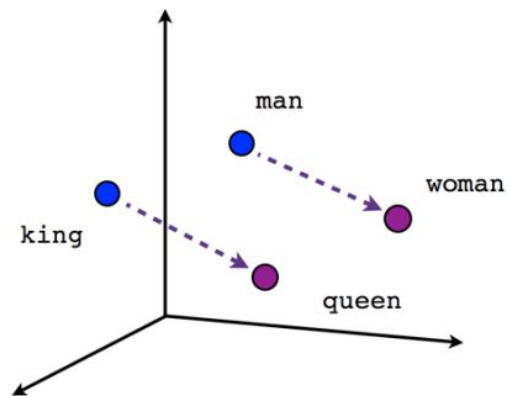
# Dense Vector Models



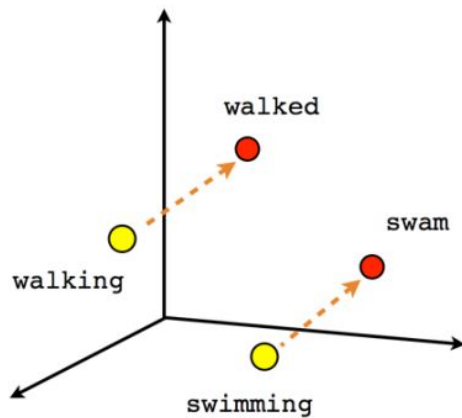
**Word vectors**  
a.k.a.  
**Word embeddings**



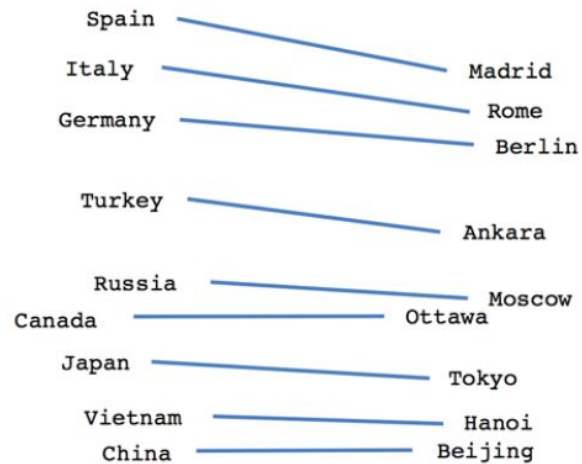




Male-Female



Verb tense



Country-Capital

An aerial photograph of a vast, frozen body of water, likely in the Arctic or Antarctic. The surface is covered with a mosaic of irregular ice floes in various shades of blue and white, separated by dark, open water. The perspective is from directly above, looking down on the ice.

# Chatbots

# Chatbots

- Purpose
- Conversation design
- Context



**Benedict Evans** ✓

@BenedictEvans

Follow



WHAT DO WE WANT?  
Natural language processing!  
WHEN DO WE WANT IT?  
Sorry, when do we want what?

1:17 PM - 22 Jan 2017

2,054 Retweets 4,073 Likes

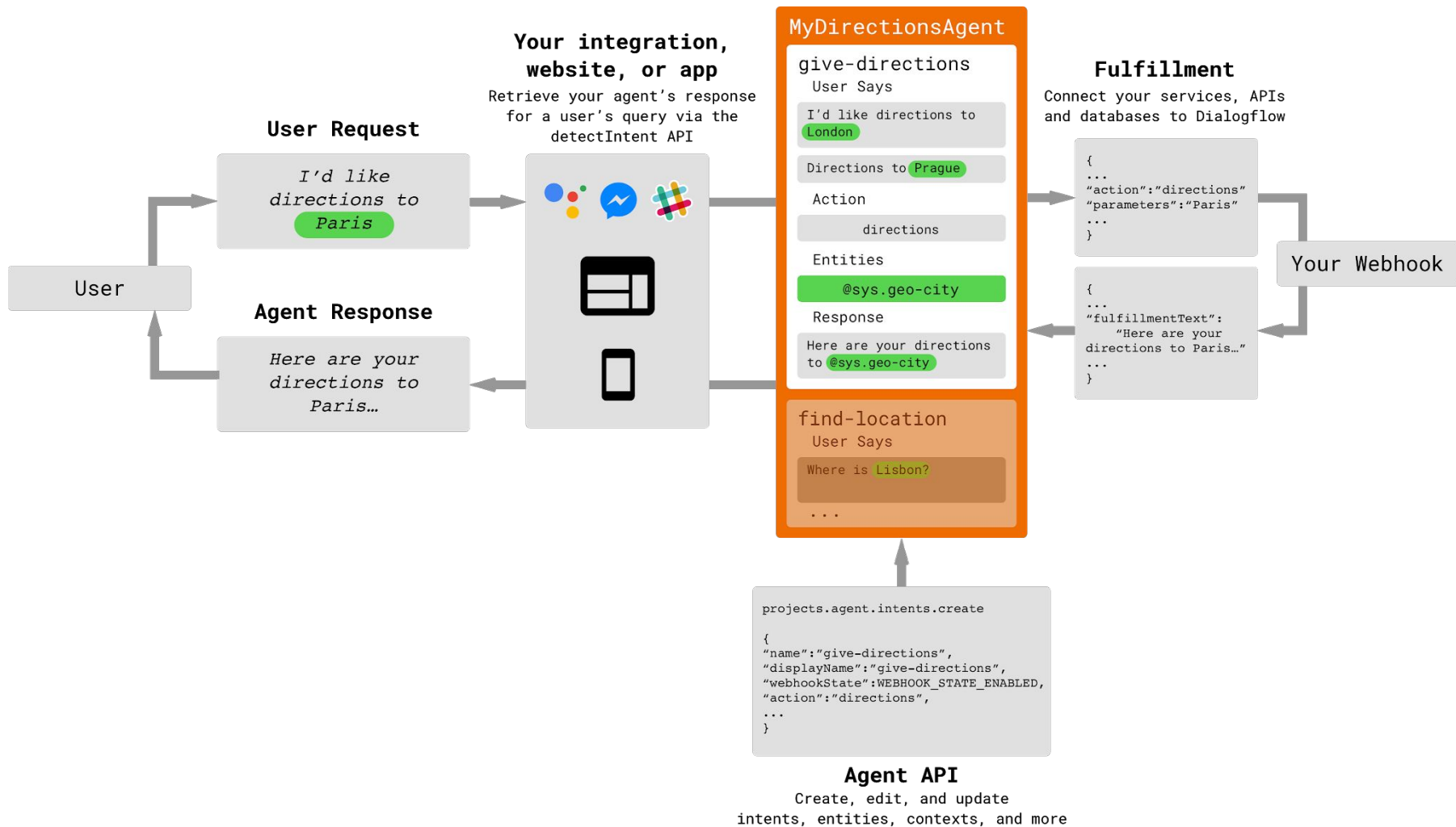




An aerial photograph of a vast, frozen body of water, likely a glacier or a large lake in a cold region. The surface is covered with numerous irregular, broken ice floes of varying sizes, creating a complex, mosaic-like pattern. The ice has a light blue-grey hue, and the gaps between the floes reveal a darker, possibly liquid water surface. The overall scene conveys a sense of cold, isolation, and natural power.

# **Dialogflow**

**formerly api.ai**





An aerial photograph of a vast, frozen body of water, likely in the Arctic or Antarctic. The surface is covered with a mosaic of ice floes of various sizes and shapes, separated by dark, open water. The ice floes have a textured, slightly uneven appearance, with some showing signs of melting or cracking. The overall color palette is a mix of light blues, whites, and dark blues, creating a cold and desolate atmosphere.

**Demo time**

**goo.gl/sHbfGQ**

iliakplv