Machine Learning for Search

WELCOME!

Develop
Intelligence

Axel Sirota
AI Consultant

Join Us in Making Learning Technology Easier

Our mission…

Over 16 years ago, we embarked on a journey to improve the world by making learning technology easy and accessible to everyone.

…impacts everyone daily.

And it’s working. Today, we’re known for delivering customized tech learning programs that drive innovation and transform organizations.

In fact, when you talk on the phone, watch a movie, connect with friends on social media, drive a car, fly on a plane, shop online, and order a latte with your mobile app, you are experiencing the impact of our solutions.

Over The Past Few Decades, We've Provided

Over **62,300,000** expert-led learning hours

In 2019 Alone, We Provided

Training to over **13,500** engineers

Over **120** active trainers, with an average of over two decades of experience each.

Programs in **30** countries

Upskilling and Reskilling Offerings

Develop Intelligence

Intimately customized learning experiences just for your teams.

Workshop	2-3 day upskilling experiences
Fast Track	5-day reskilling experiences
Learning Spike	1-day technology overviews
Target Topics	90-minute instructor-led micro-learnings
Hack-a-thon	Learn and build an MVP in 2-3 days

- BACK END DEVELOPMENT
- BIG DATA
- CLOUD COMPUTING
- DEVOPS
- FRONT END DEVELOPMENT
- MACHINE LEARNING
- MOBILE APP DEVELOPMENT
- SOFTWARE ENGINEERING
- SYSTEM ADMINISTRATION

AND MANY OTHER TRENDING TECHNOLOGIES

World Class Practitioners

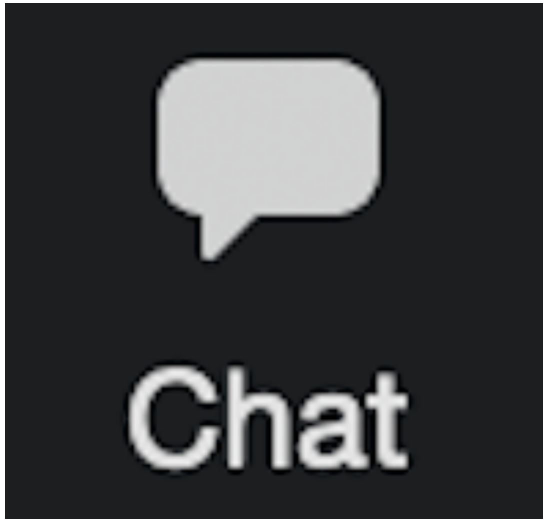
Note About Virtual Trainings

Note About Virtual Trainings

What we want

...what we've got

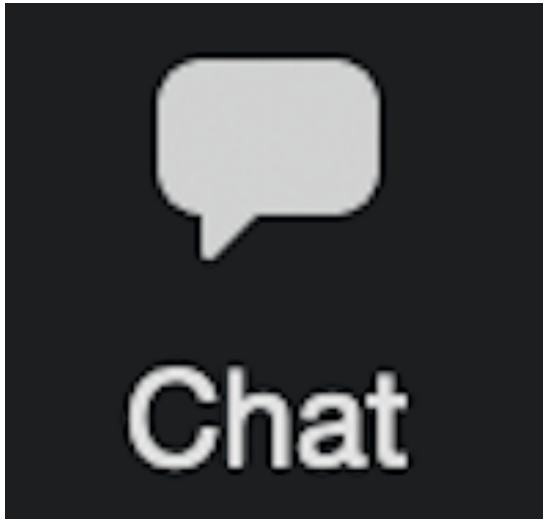
Virtual Training Expectations for You



Virtual Training Expectations for You

Arrive on time / return on time

Mute unless speaking



Use chat or ask questions verbally

Virtual Training Expectations for Me

Virtual Training Expectations for Me

I pledge to:

Virtual Training Expectations for Me

Develop Intelligence

I pledge to:

- Make this as interesting and interactive as possible

Virtual Training Expectations for Me

I pledge to:

- Make this as interesting and interactive as possible
- Ask questions in order to stimulate discussion

Virtual Training Expectations for Me

Develop Intelligence

I pledge to:

- Make this as interesting and interactive as possible
- Ask questions in order to stimulate discussion
- Use whatever resources I have at hand to explain the material

Virtual Training Expectations for Me

Develop Intelligence

I pledge to:

- Make this as interesting and interactive as possible
- Ask questions in order to stimulate discussion
- Use whatever resources I have at hand to explain the material
- Try my best to manage verbal responses so that everyone who wants to speak can do so

Virtual Training Expectations for Me

I pledge to:

- Make this as interesting and interactive as possible
- Ask questions in order to stimulate discussion
- Use whatever resources I have at hand to explain the material
- Try my best to manage verbal responses so that everyone who wants to speak can do so
- Use an on-screen timer for breaks so you know when to be back

Prerequisites

Prerequisites

- Basic Apache Solr / Search engines

Prerequisites

- Basic Apache Solr / Search engines
- Basic Python (for scripting).

Prerequisites

Develop Intelligence

- Basic Apache Solr / Search engines
- Basic Python (for scripting).
- Basic ML concepts like training, evaluating, or classification would be useful but not mandatory

What this course is about

What this course is about

This course **is not about:**

What this course is about

This course **is not about:**

- Learning Apache Solr

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

This course **is about**:

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

This course **is about**:

- Enriching search engines (Apache Solr in this case) with Deep Learning

What this course is about

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

This course **is about**:

- Enriching search engines (Apache Solr in this case) with Deep Learning
- Learning how to apply NLP techniques in search

What this course is about

Develop Intelligence

This course **is not about:**

- Learning Apache Solr
- Machine Learning Fundamentals
- Basic Tensorflow

This course **is about**:

- Enriching search engines (Apache Solr in this case) with Deep Learning
- Learning how to apply NLP techniques in search
- Creating enriches solutions in Python for retrieval systems to **simplify** the task for users

Objectives

Objectives

At the end of this course you will be able to:

Objectives

Develop Intelligence

At the end of this course you will be able to:

- Understand all the places were NLP can help in Search

Objectives

At the end of this course you will be able to:

- Understand all the places were NLP can help in Search
- Expand queries to find for synonyms of nouns

Objectives

Develop Intelligence

At the end of this course you will be able to:

- Understand all the places were NLP can help in Search
- Expand queries to find for synonyms of nouns
- Find alternative queries to execute in parallel to increase recall

Objectives

At the end of this course you will be able to:

- Understand all the places were NLP can help in Search
- Expand queries to find for synonyms of nouns
- Find alternative queries to execute in parallel to increase recall
- Learn reranking algorithms to filter the original results by similarity

Objectives

At the end of this course you will be able to:

- Understand all the places were NLP can help in Search
- Expand queries to find for synonyms of nouns
- Find alternative queries to execute in parallel to increase recall
- Learn reranking algorithms to filter the original results by similarity
- Learn distinctive tags of documents and index them for efficient use of indexes

Structure of the Course / Course Takeaways

Structure of the Course / Course Takeaways

Develop Intelligence

- We will be using a Github repository for labs throughout the training

Structure of the Course / Course Takeaways

- We will be using a Github repository for labs throughout the training
- The labs will be held in CentOS Workstations provided by you during introductions.

Structure of the Course / Course Takeaways

Develop Intelligence

- We will be using a Github repository for labs throughout the training
- The labs will be held in CentOS Workstations provided by you during introductions.
- These workstations are accessed by the browser with user and password

Structure of the Course / Course Takeaways

- We will be using a Github repository for labs throughout the training
- The labs will be held in CentOS Workstations provided by you during introductions.
- These workstations are accessed by the browser with user and password
- Download the code and instructions for labs here: https://github.com/axel-sirota/ml-solr-course

Structure of the Course / Course Takeaways

- We will be using a Github repository for labs throughout the training
- The labs will be held in CentOS Workstations provided by you during introductions.
- These workstations are accessed by the browser with user and password
- Download the code and instructions for labs here: https://github.com/axel-sirota/ml-solr-course
- The slides were sent by mail

Agenda

Develop Intelligence

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

Text Generation

- RNN
- GRU and LSTM
- Transformers

NER

- Spacy NER

Next Steps

- Closing thoughts

Agenda

Develop Intelligence

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

NER

- Spacy NER

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

Text Generation

- RNN
- GRU and LSTM
- Transformers

Next Steps

- Closing thoughts

Introduction

Who are we?

Introductions

- One by one introduce yourself:
  A. Name
  B. Role
  C. Previous experience in Machine Learning and NLP
  D. What do you want to learn from this class?
  E. Do you want to share a fun fact of yourself?

Who am I?

Who am I?

– Microsoft Certified Trainer

Who am I?

– Microsoft Certified Trainer

– Author, Instructor, and Editor at Pluralsight, Develop Intelligence, and O'Reilly Media

Who am I?

Develop Intelligence

– Microsoft Certified Trainer

– Author, Instructor, and Editor at Pluralsight, Develop Intelligence, and O’Reilly Media

– AI and Cloud Consultant

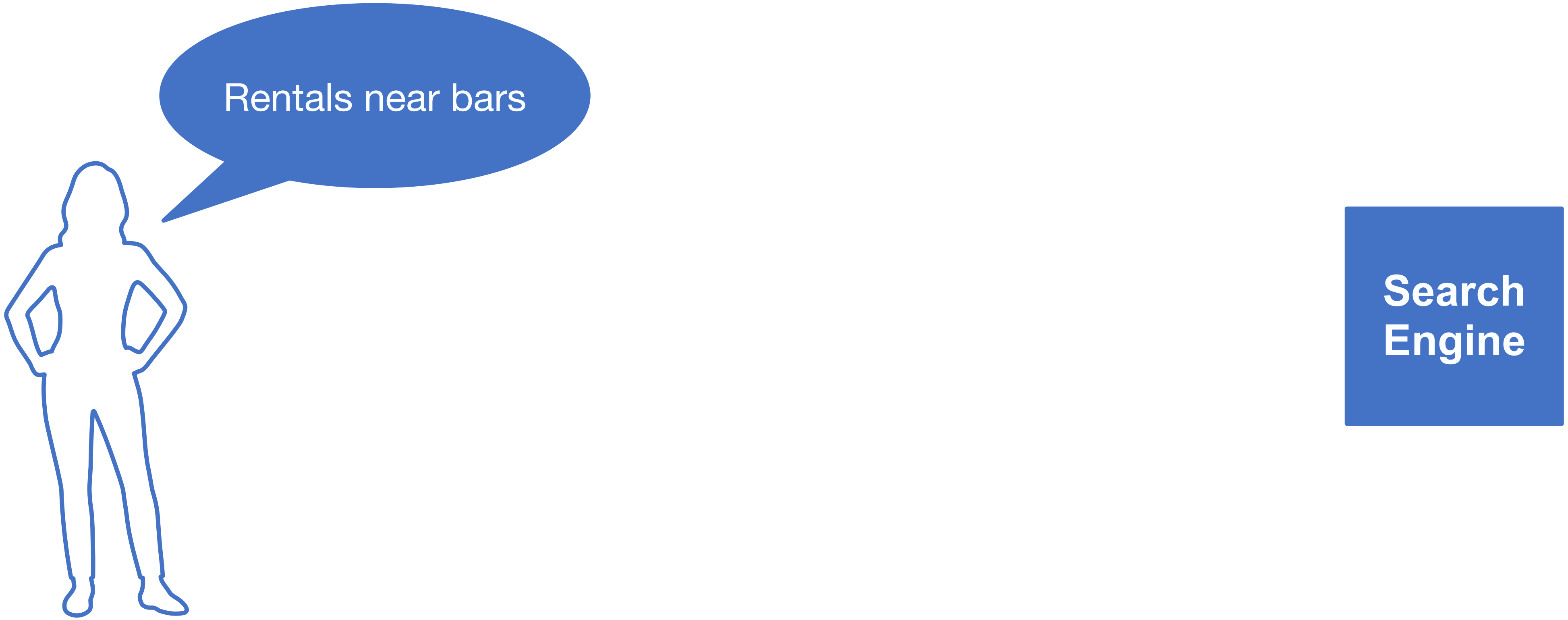
QR to my Pluralsight courses

QR to my O’Reilly trainings

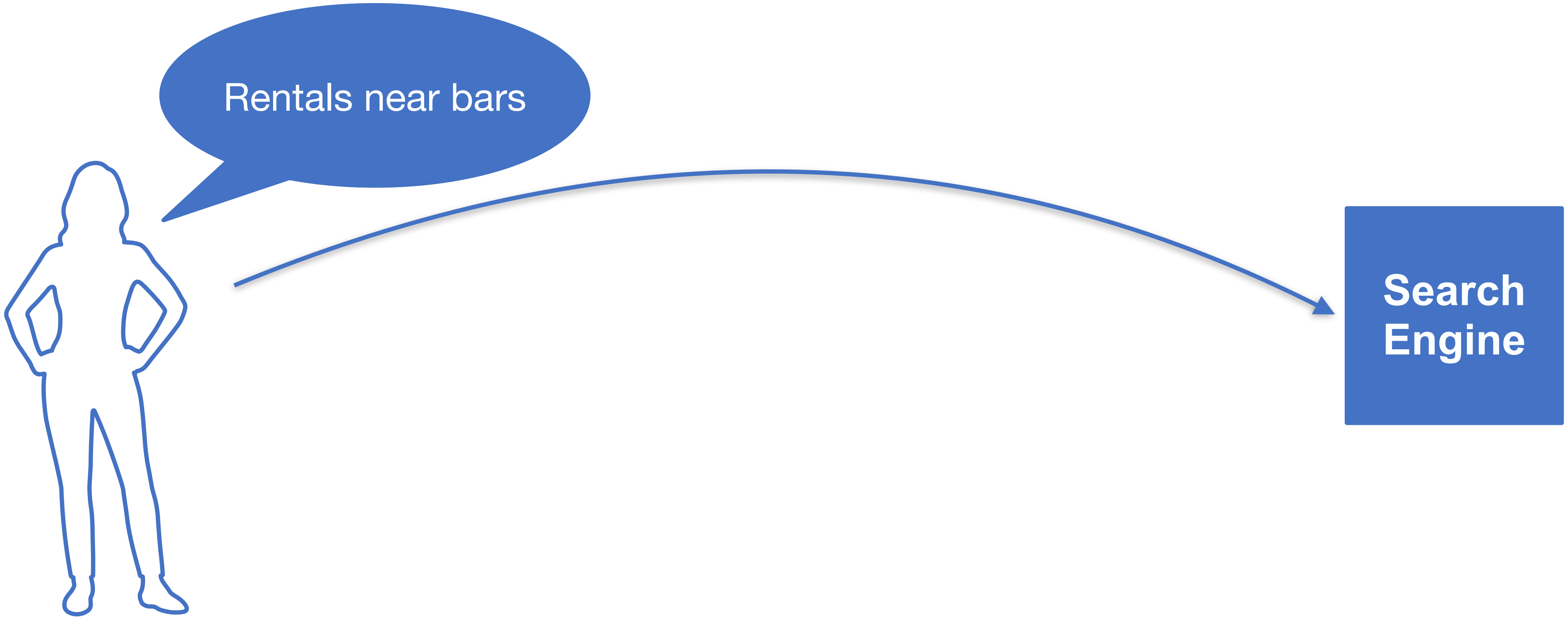
Introduction

Overview of Apache Solr and Neural Search

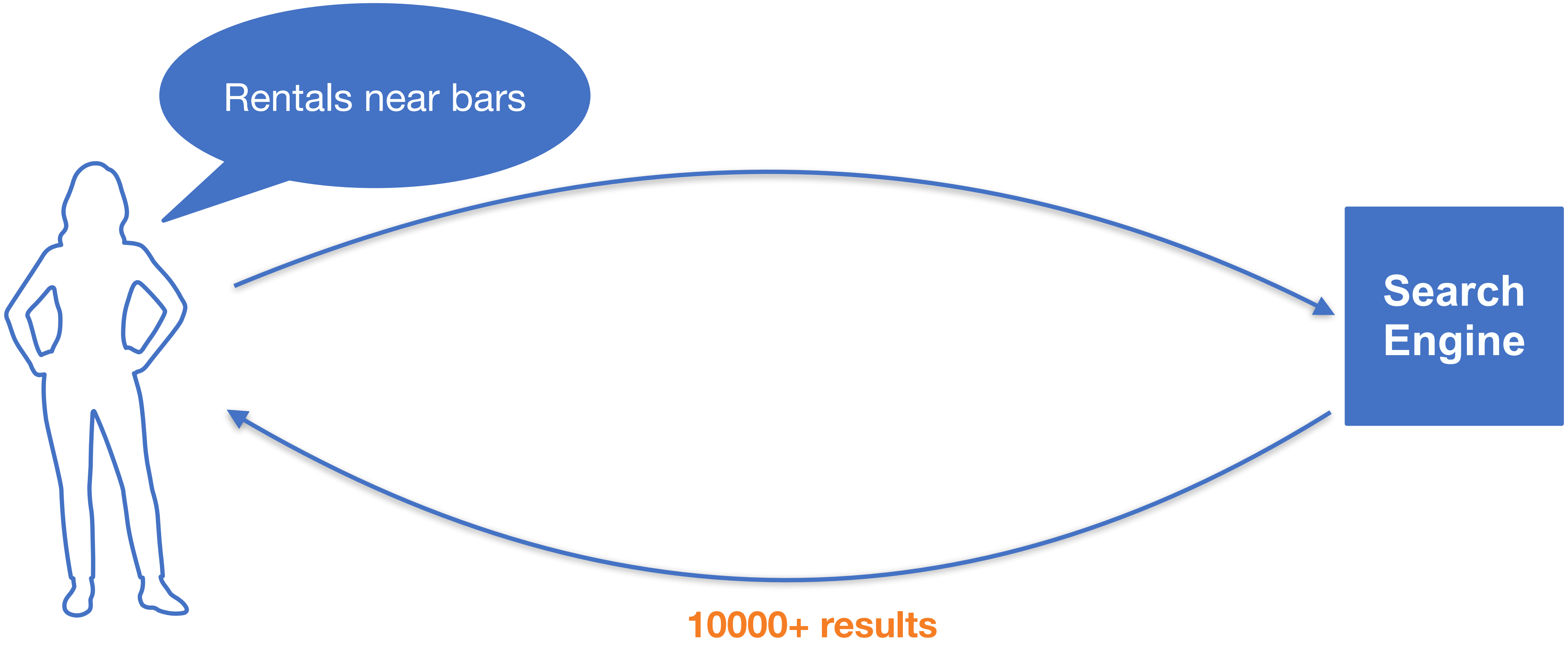
Case Study: Looking for rentals



Case Study: Looking for rentals

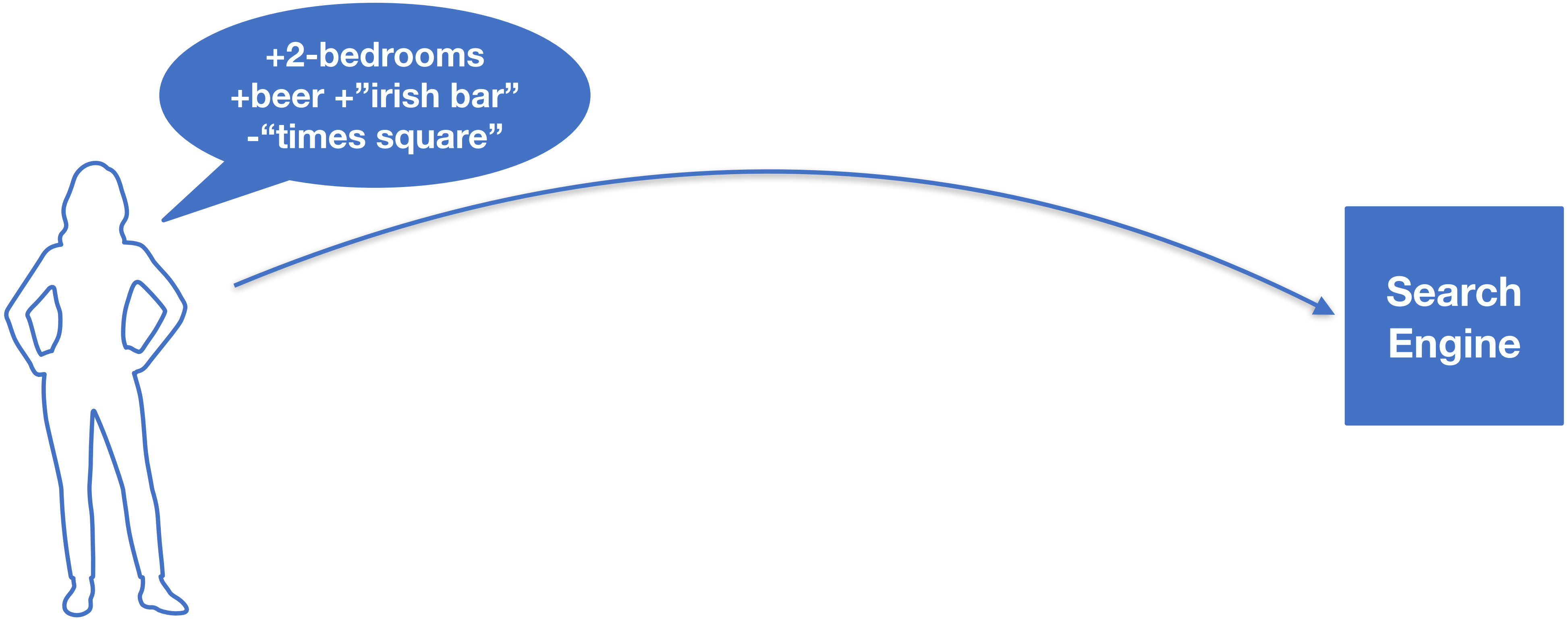


Case Study: Looking for rentals

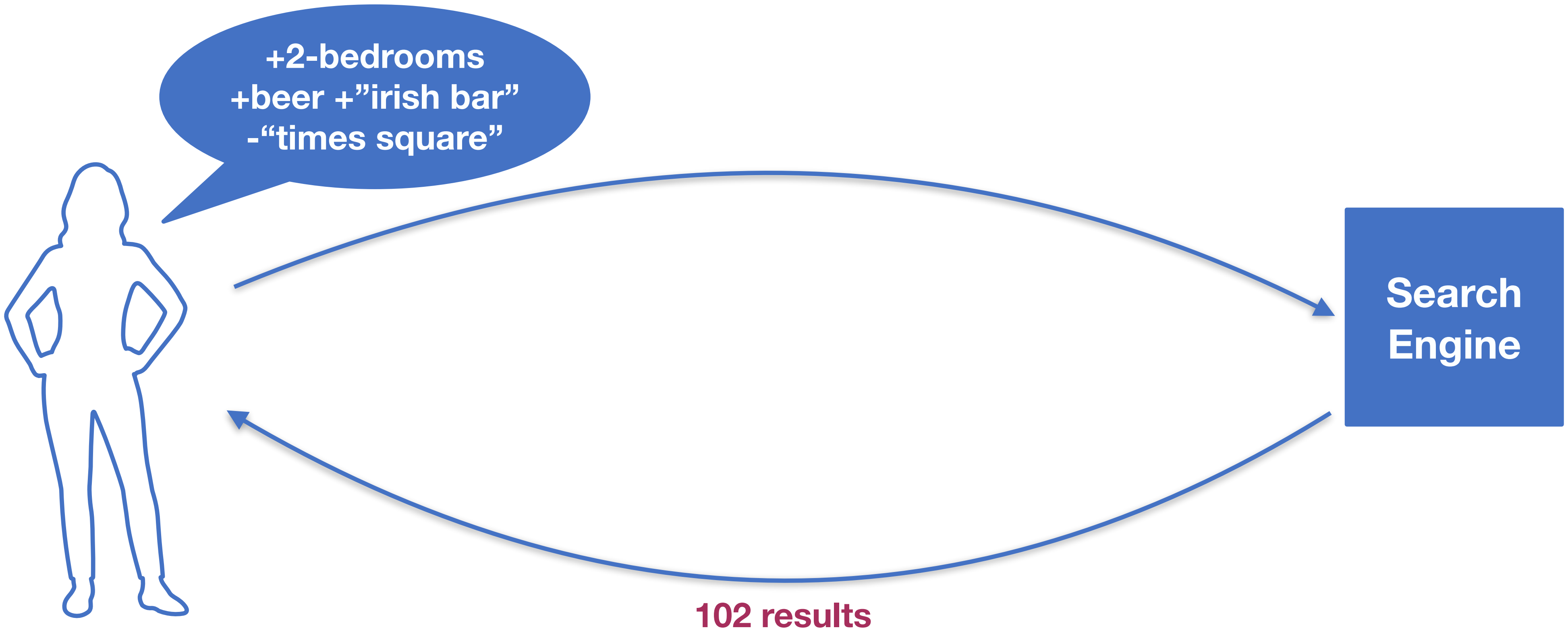


Case Study: Looking for rentals

Case Study: Looking for rentals



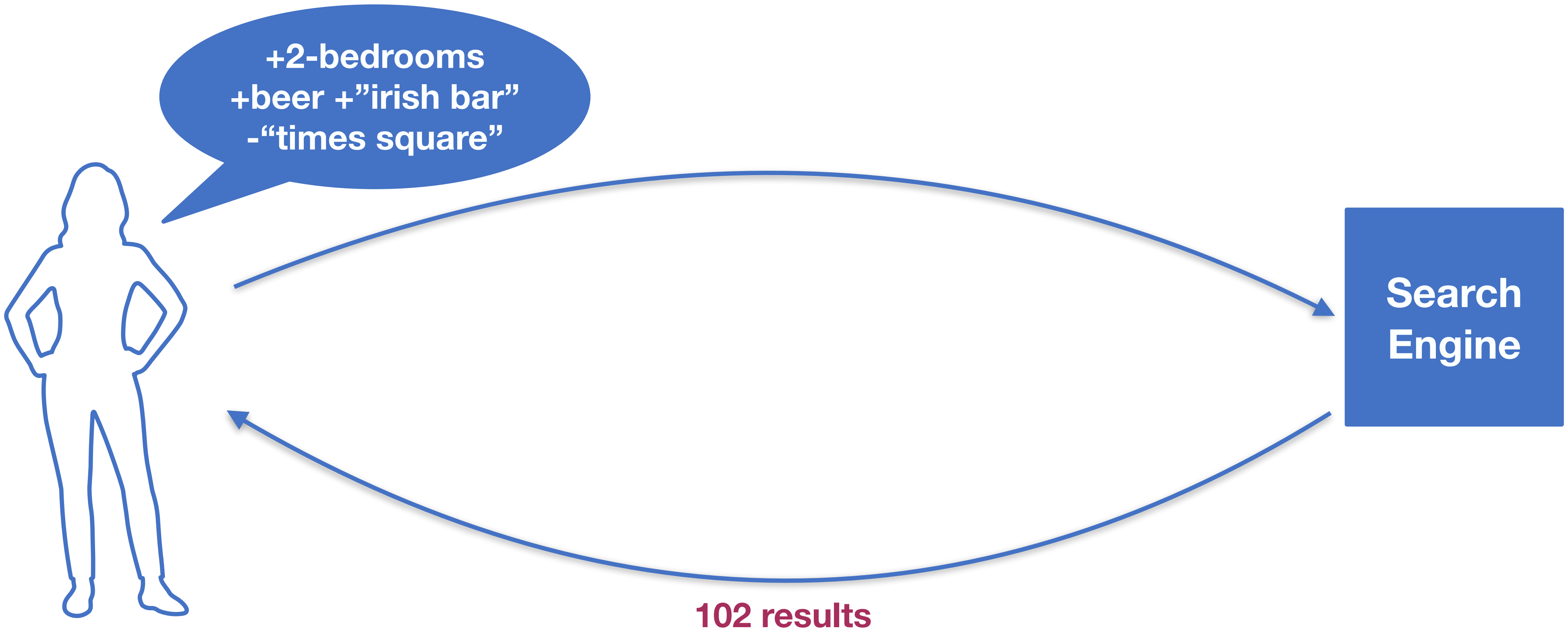
Case Study: Looking for rentals



102 results

None were exactly what she looked for!

Case Study: Looking for rentals



102 results

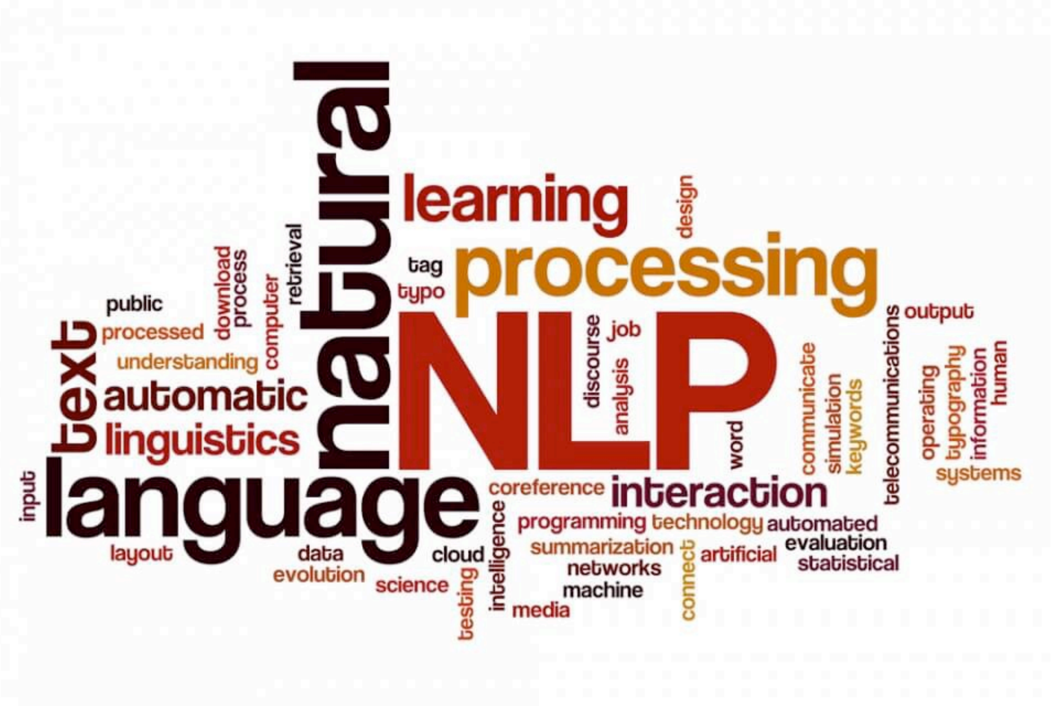
None were exactly what she looked for!

You have to search through the list of results, maybe filter with a map (if such functionality exists)

Case Study: Looking for rentals

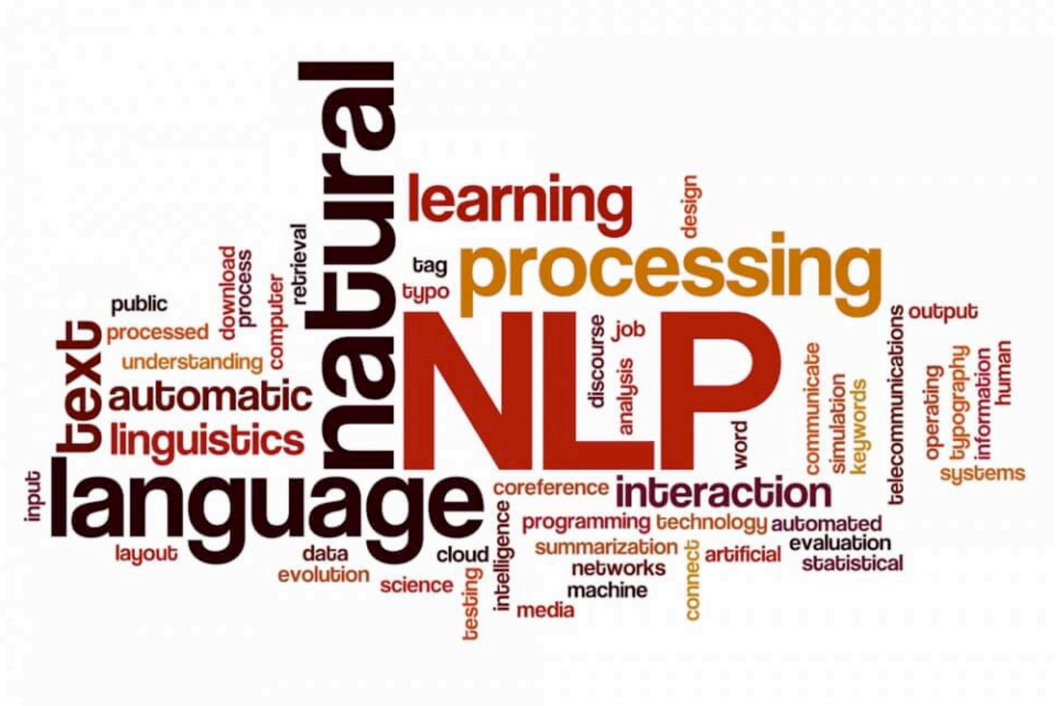
Neural Search comes to solve this!

What is Neural Search?



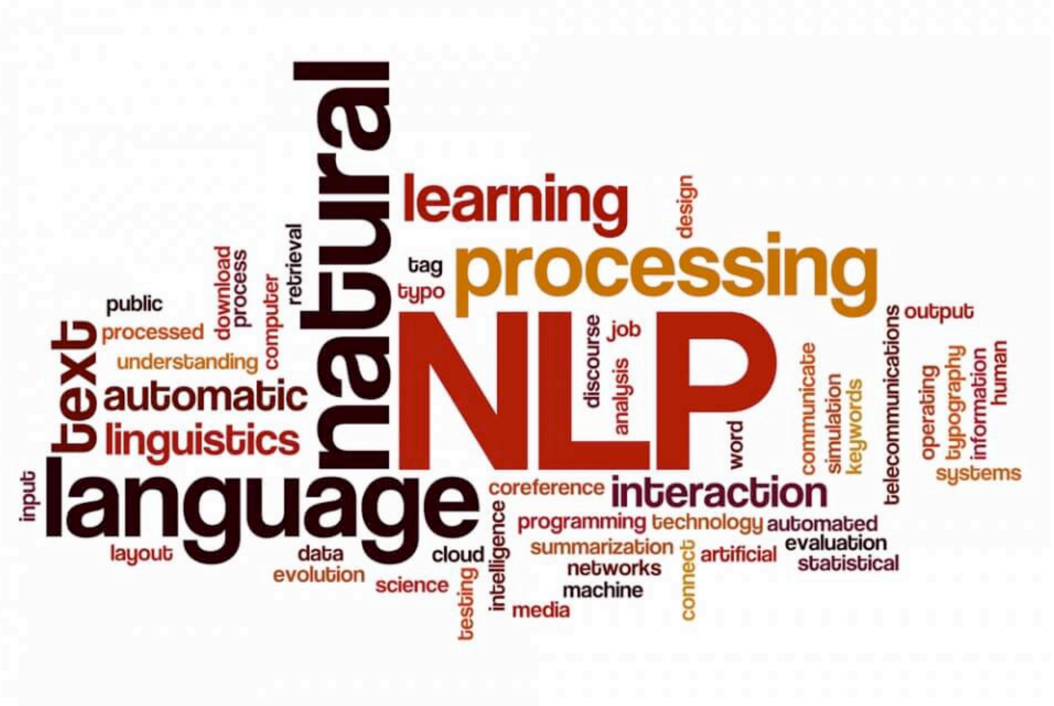
What is Neural Search?

- Applying NLP to this process



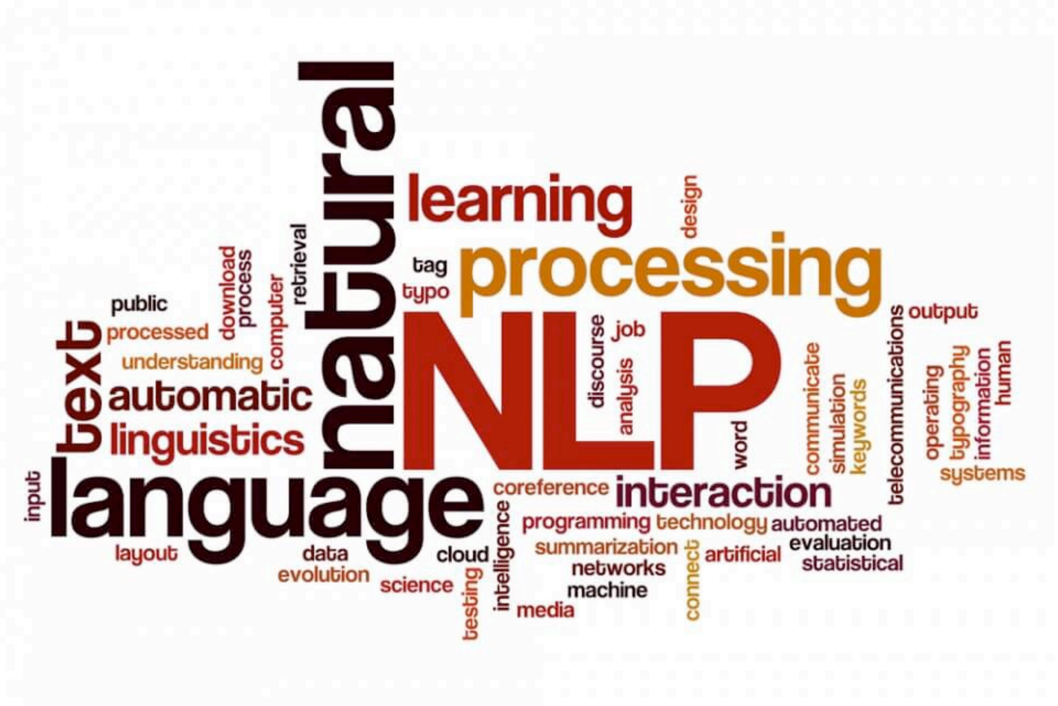
What is Neural Search?

- Applying NLP to this process
- Enrich queries and indexes

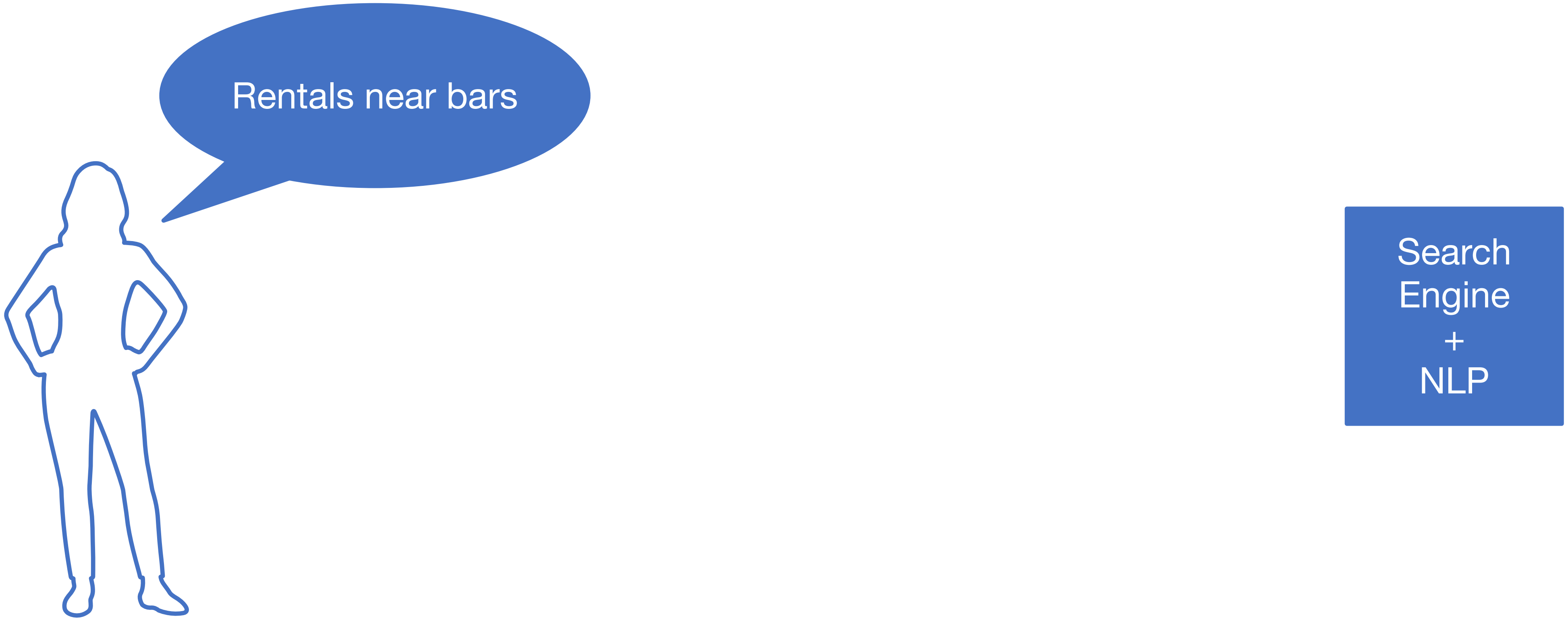


What is Neural Search?

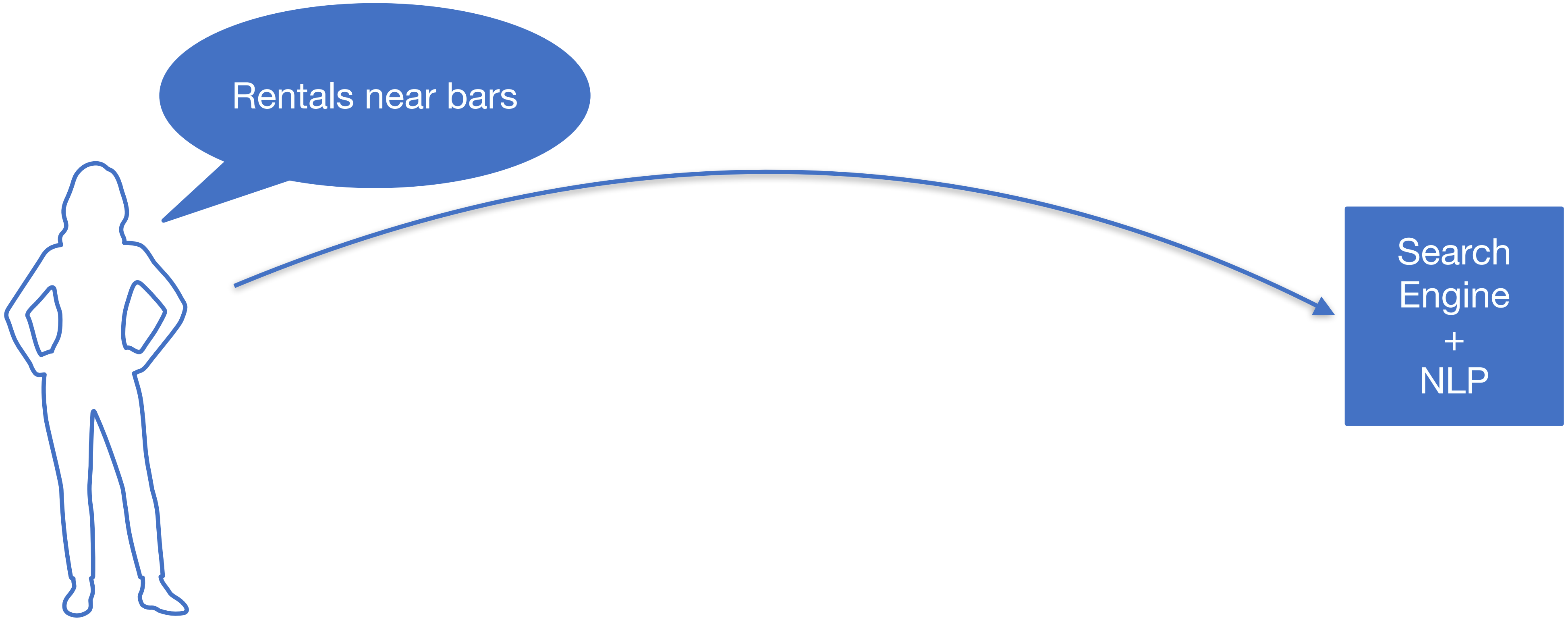
- Applying NLP to this process
- Enrich queries and indexes
- Make the first query return the results with good precision and recall



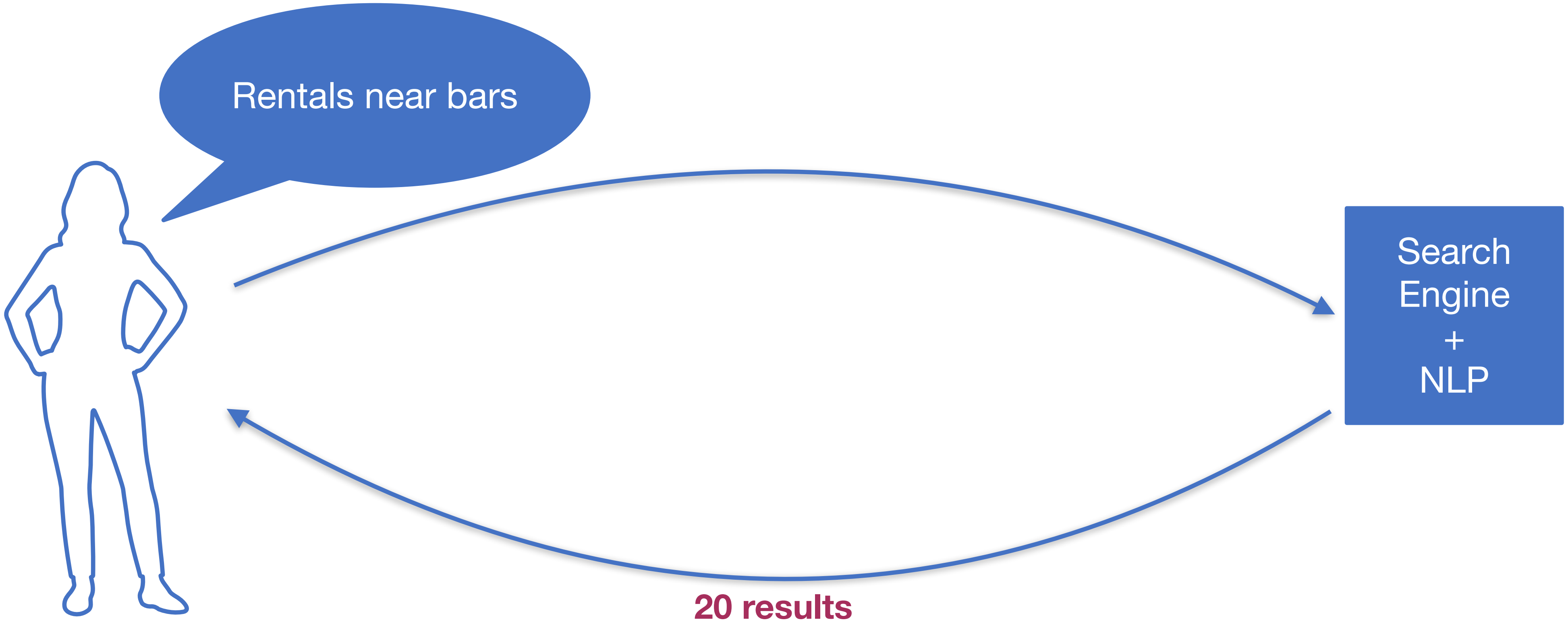
Case Study: Looking for rentals



Case Study: Looking for rentals



Case Study: Looking for rentals



NLP Areas

NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect

NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect
- Machine Translation

NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect
- Machine Translation
- Summarisation: Tag summarisation, phrase summarisation

NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect
- Machine Translation
- Summarisation: Tag summarisation, phrase summarisation
- Named Entity Recognition (NER)

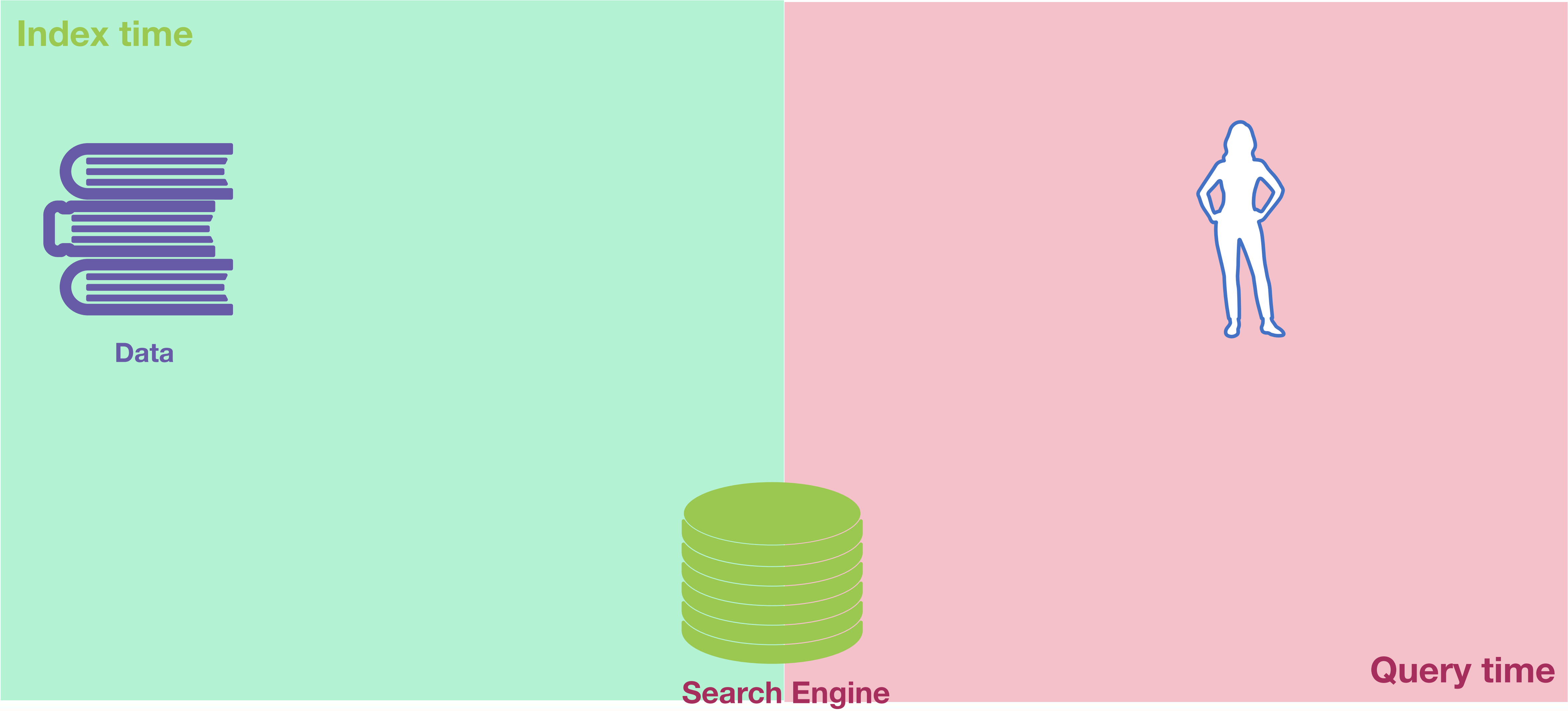
NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect
- Machine Translation
- Summarisation: Tag summarisation, phrase summarisation
- Named Entity Recognition (NER)
- Sentiment analysis

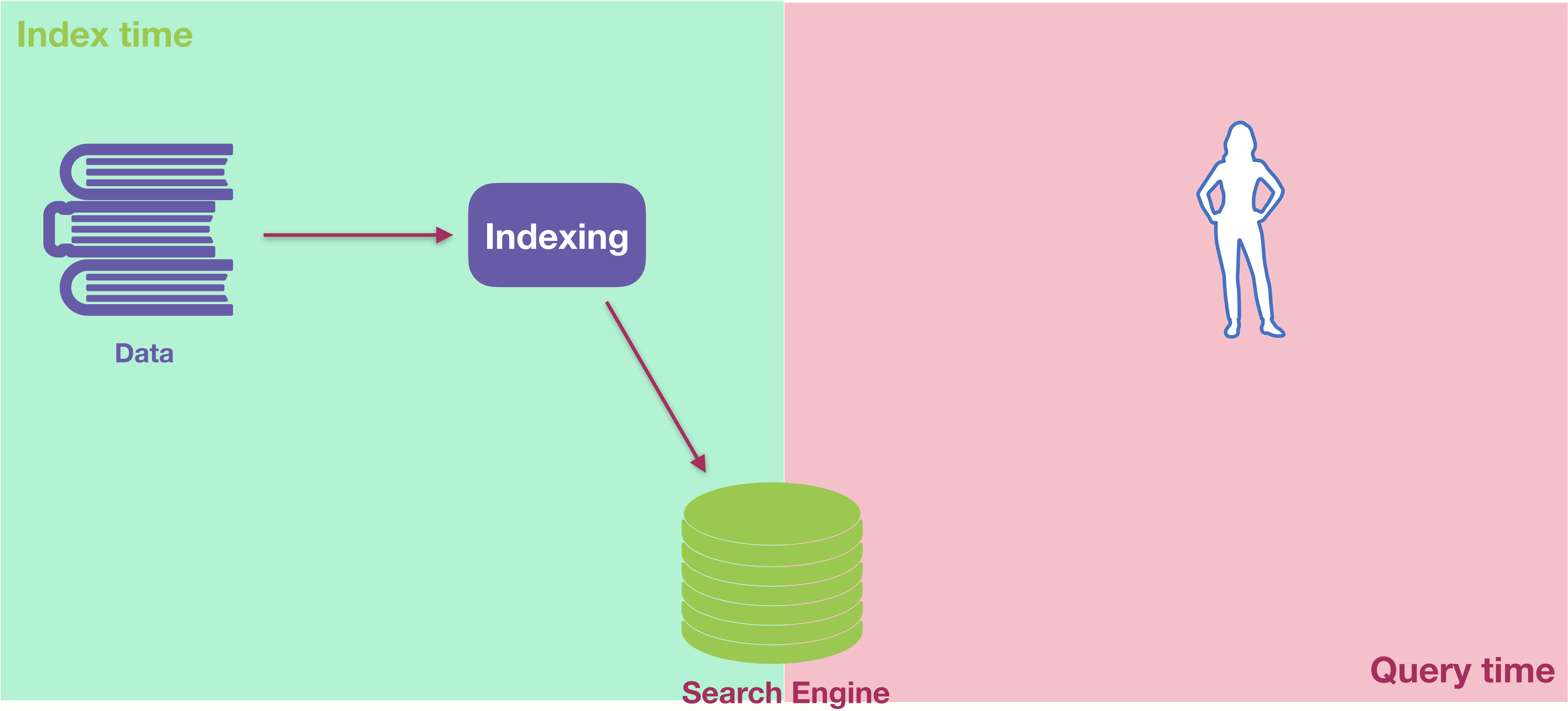
NLP Areas

- Contextual Search: Synonyms, Autocomplete, Autocorrect
- Machine Translation
- Summarisation: Tag summarisation, phrase summarisation
- Named Entity Recognition (NER)
- Sentiment analysis
- Q&A

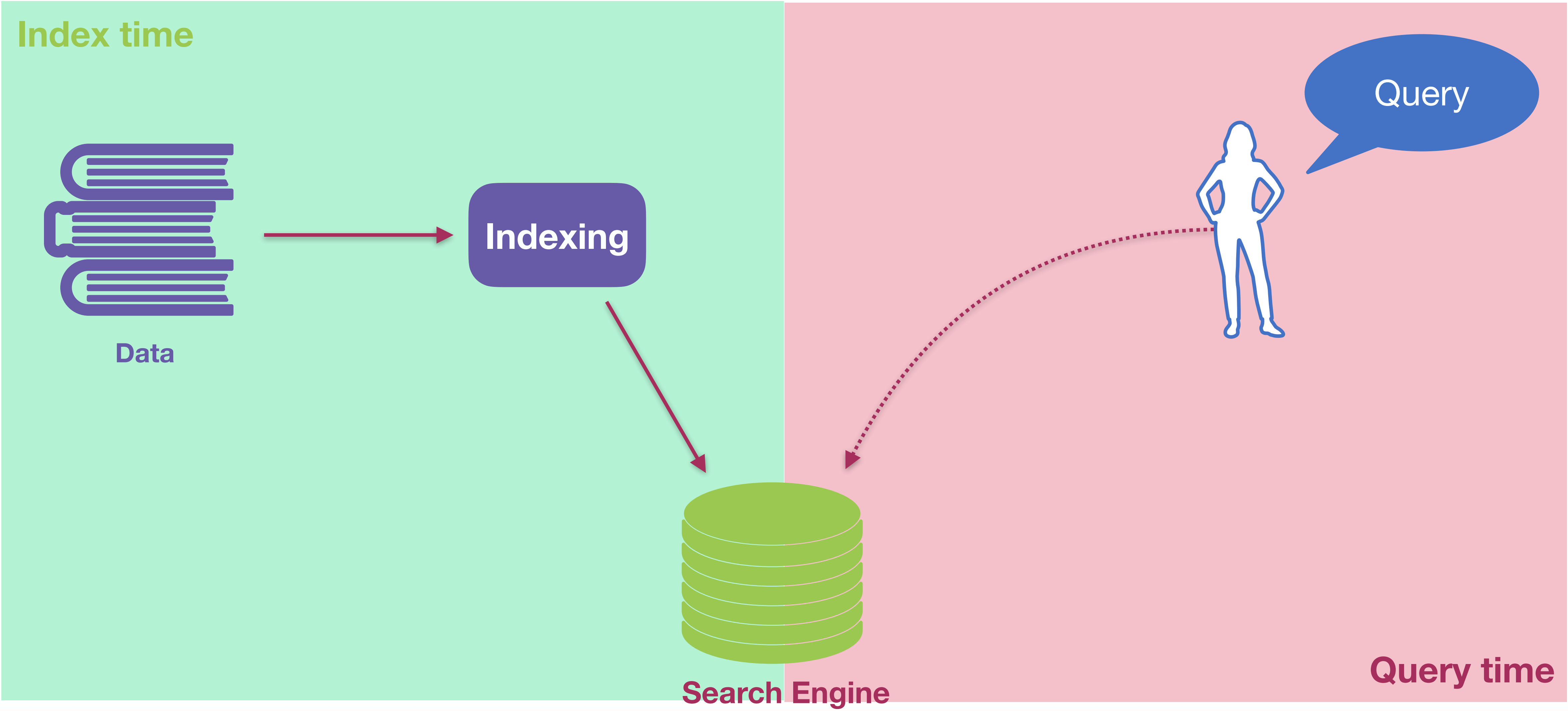
Where can NLP help Search Engines?



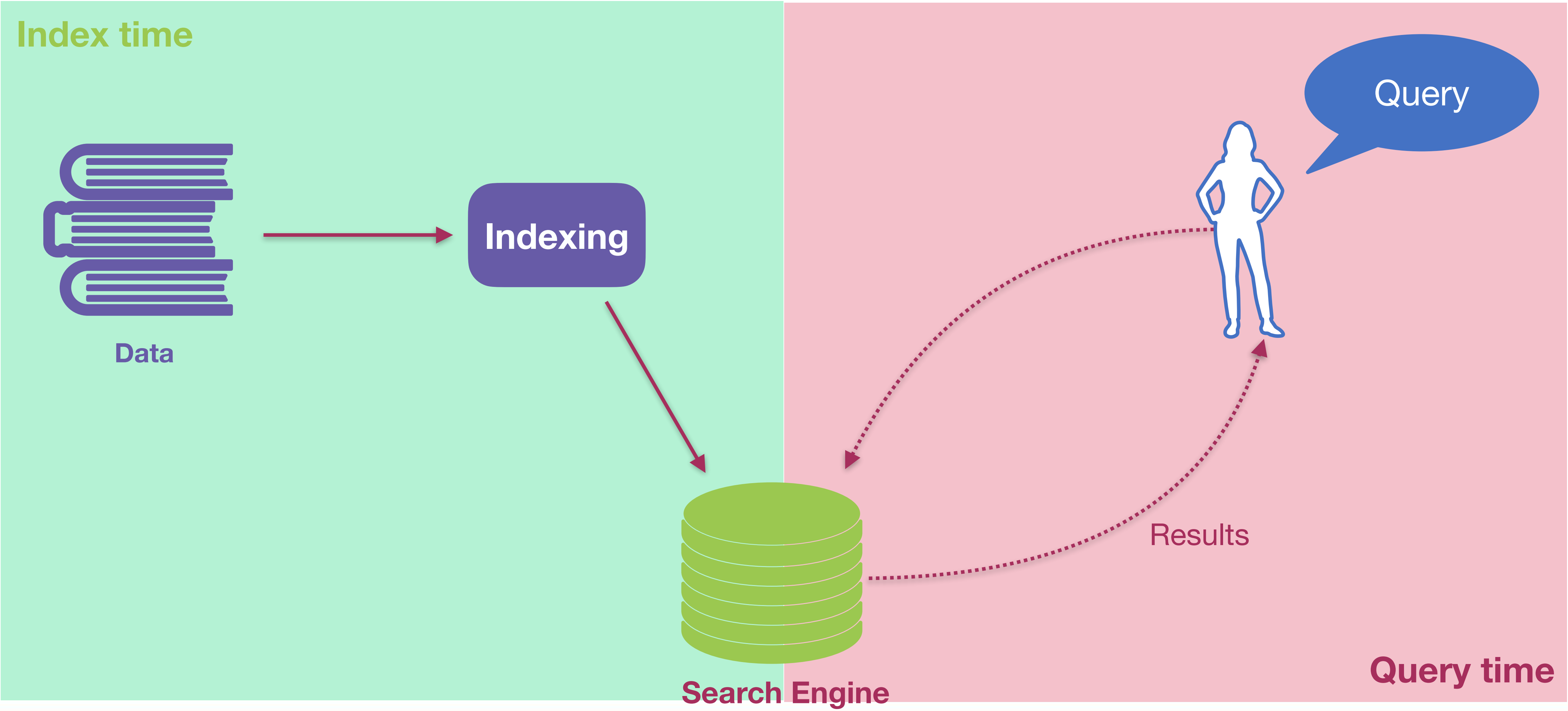
Where can NLP help Search Engines?



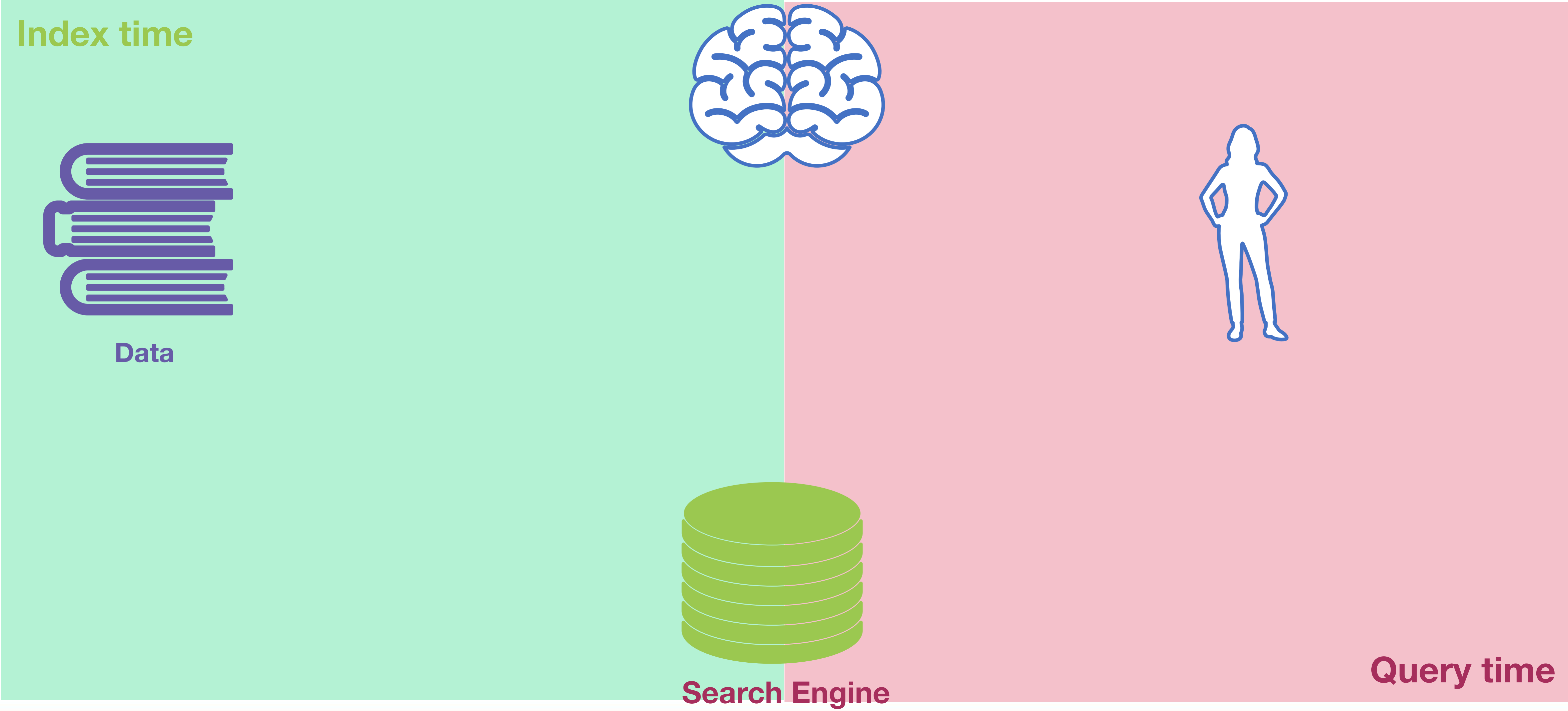
Where can NLP help Search Engines?



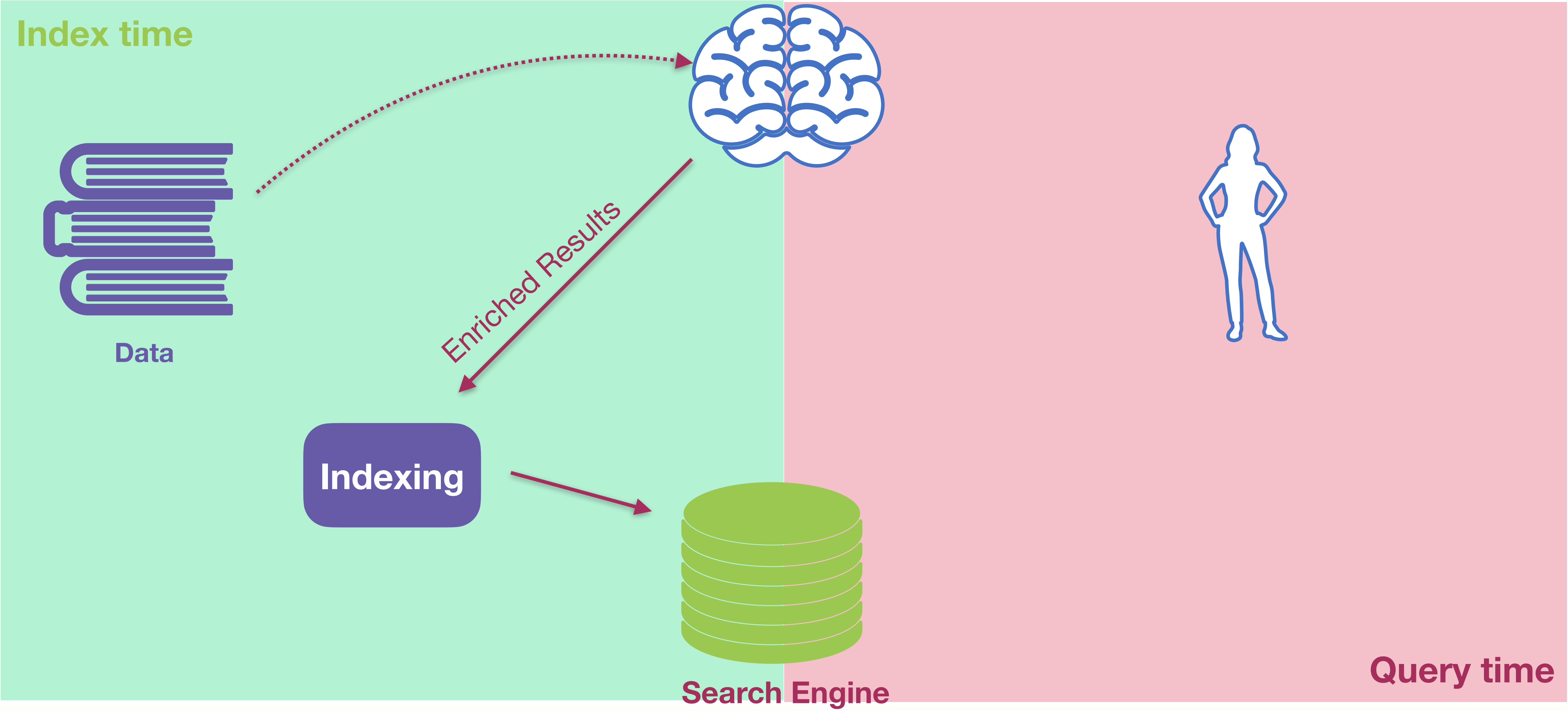
Where can NLP help Search Engines?



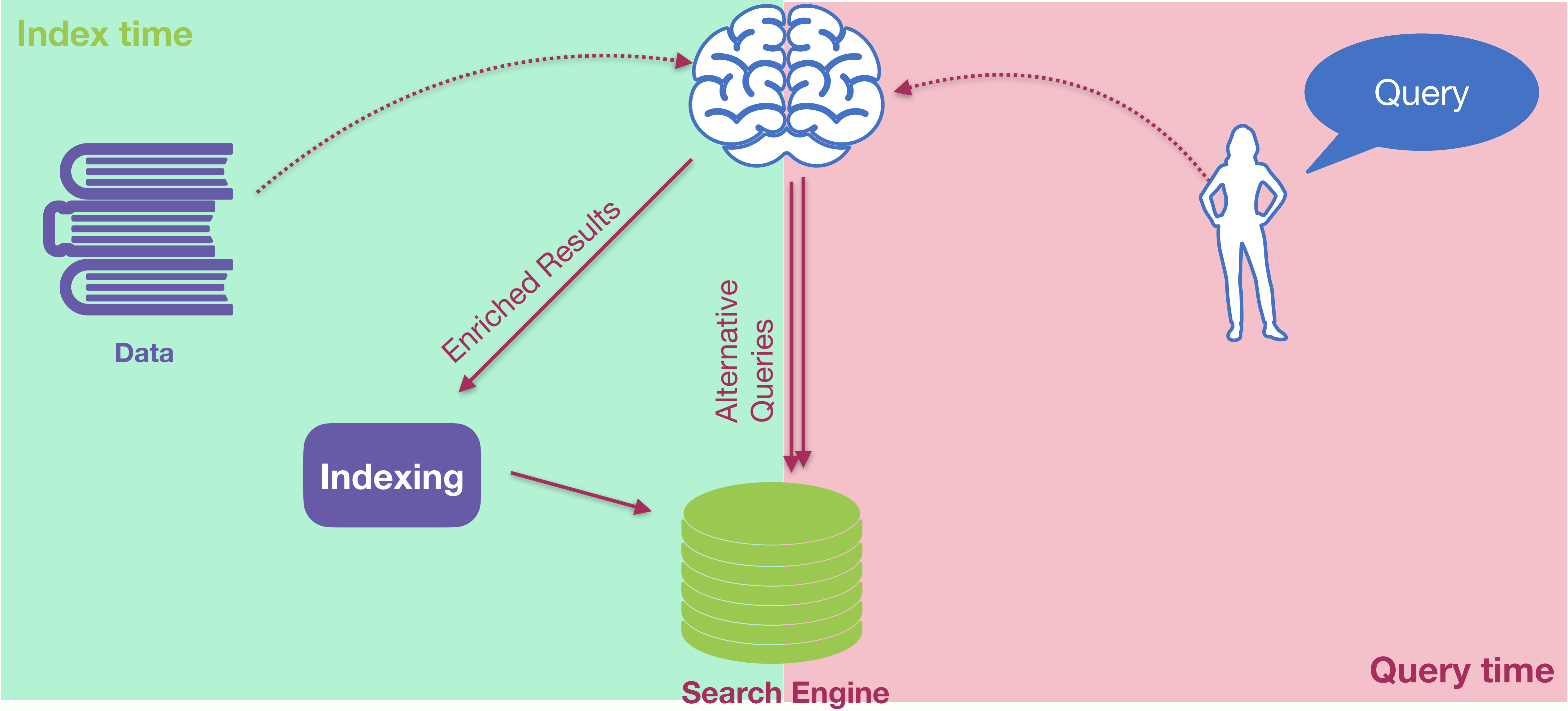
Where can NLP help Search Engines?



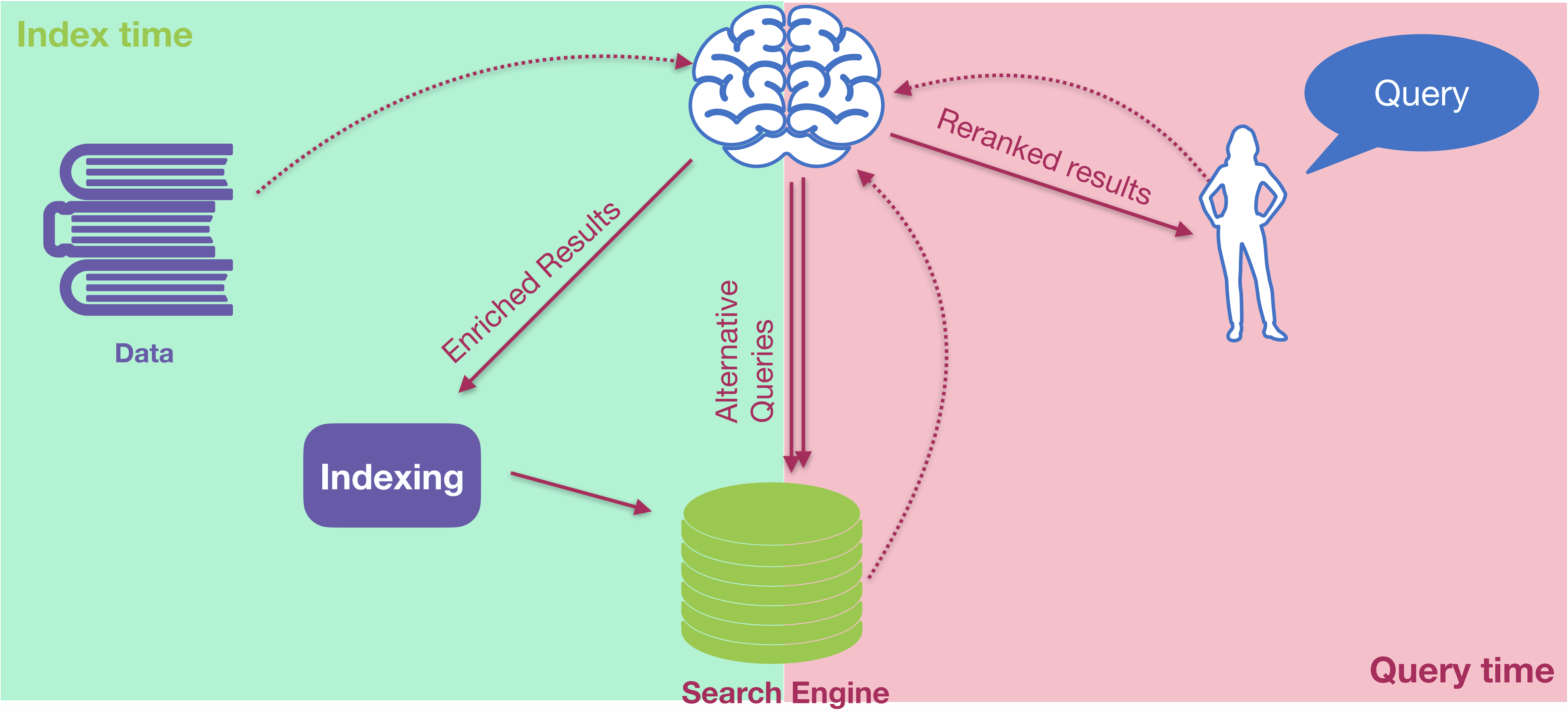
Where can NLP help Search Engines?



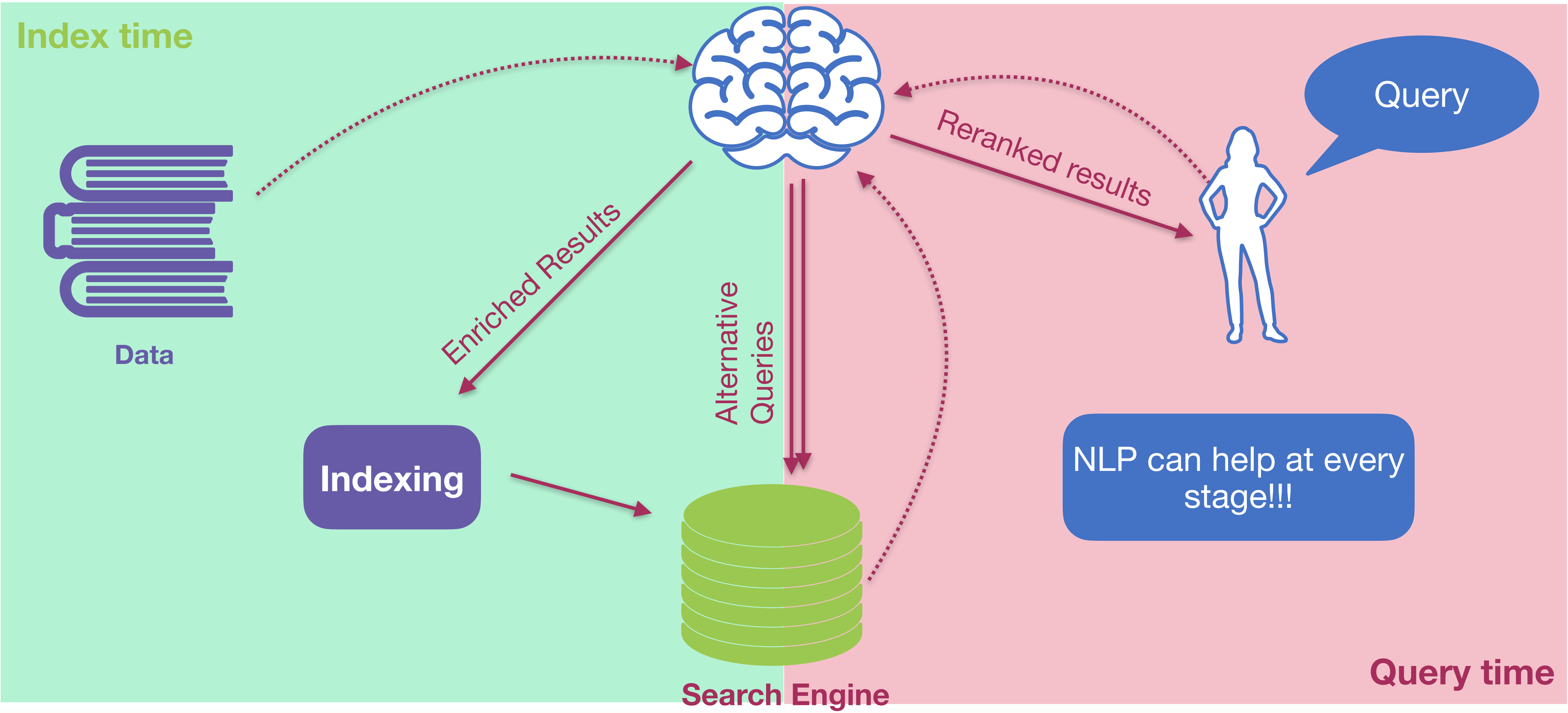
Where can NLP help Search Engines?



Where can NLP help Search Engines?

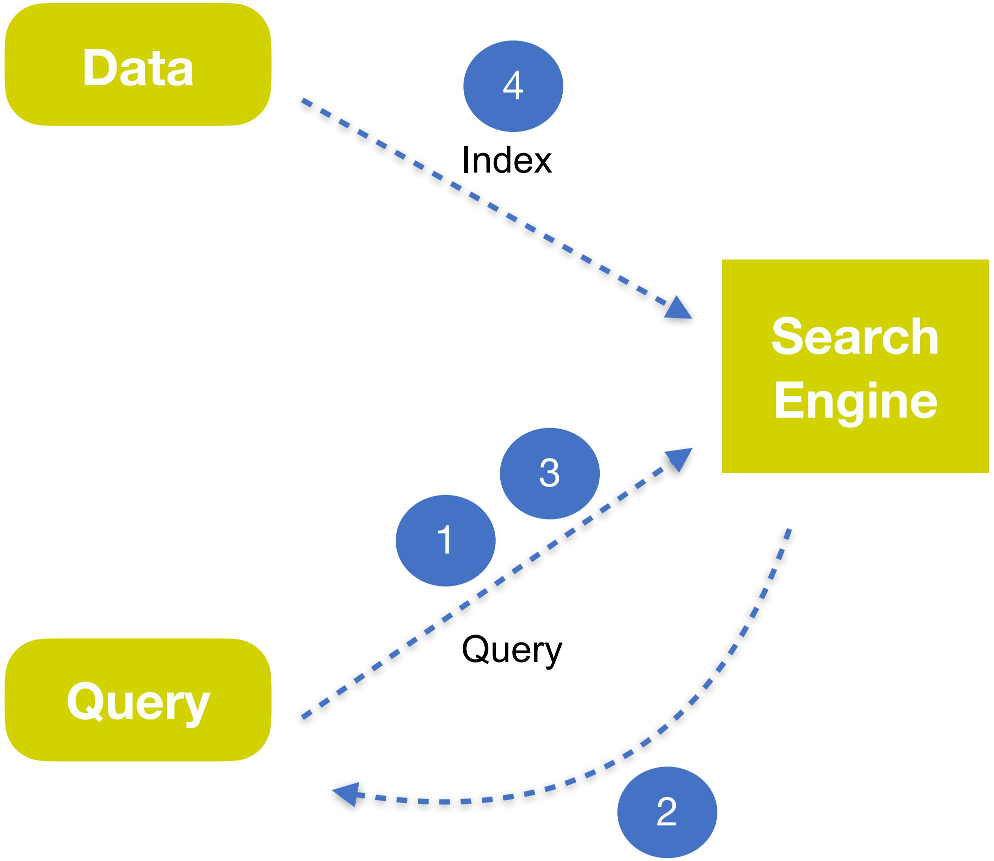


Where can NLP help Search Engines?

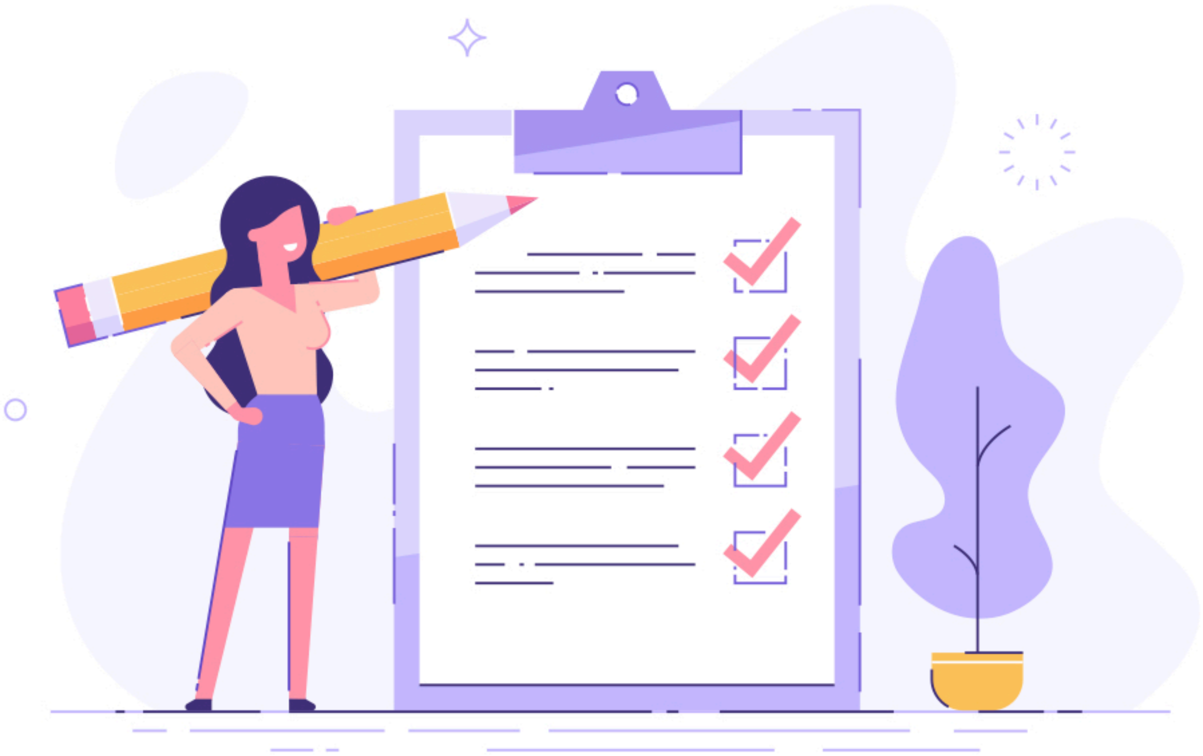


Where can NLP help Search Engines?

1. Synonym Expansion
2. Reranking results
3. Alternative Queries
4. NER

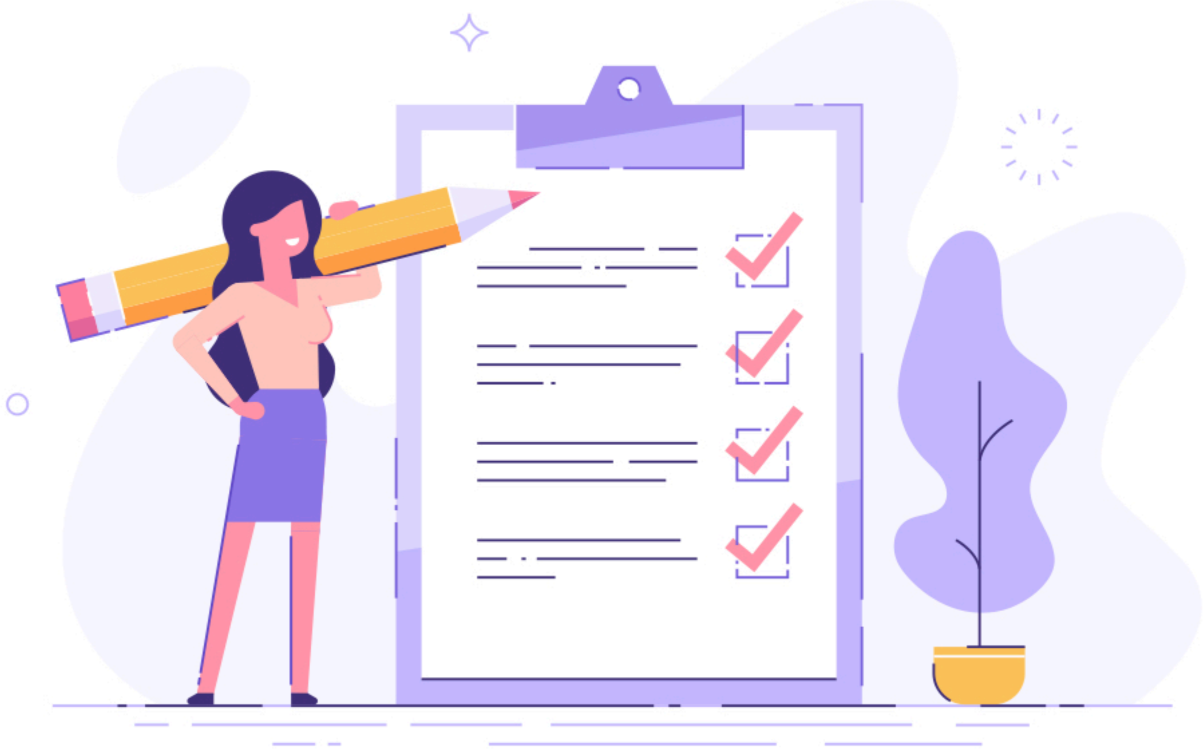


LAB: Interacting with PySolr



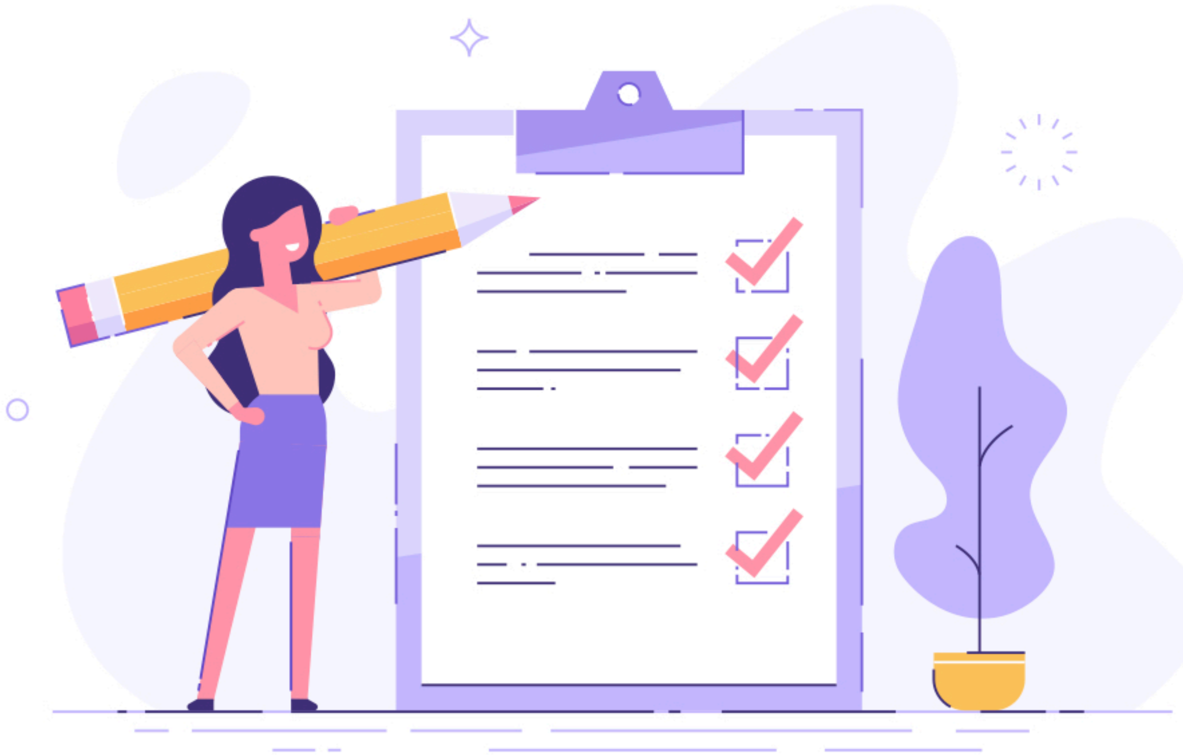
LAB: Interacting with PySolr

- First interactions with PySolr



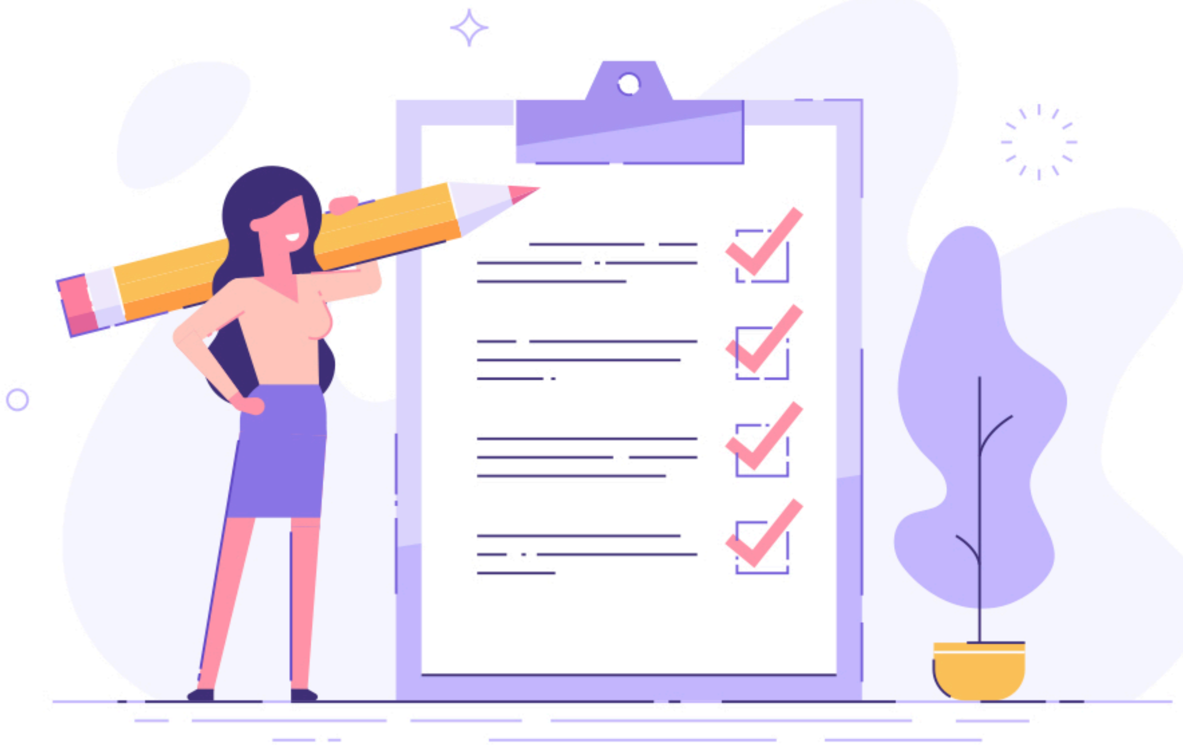
LAB: Interacting with PySolr

- First interactions with PySolr
- Environment Setup



LAB: Interacting with PySolr

- First interactions with PySolr
- Environment Setup
- Index the Airbnb dataset



Summary

Summary

Develop Intelligence

- Neural Search is the subject of using NLP techniques to search engines

Summary

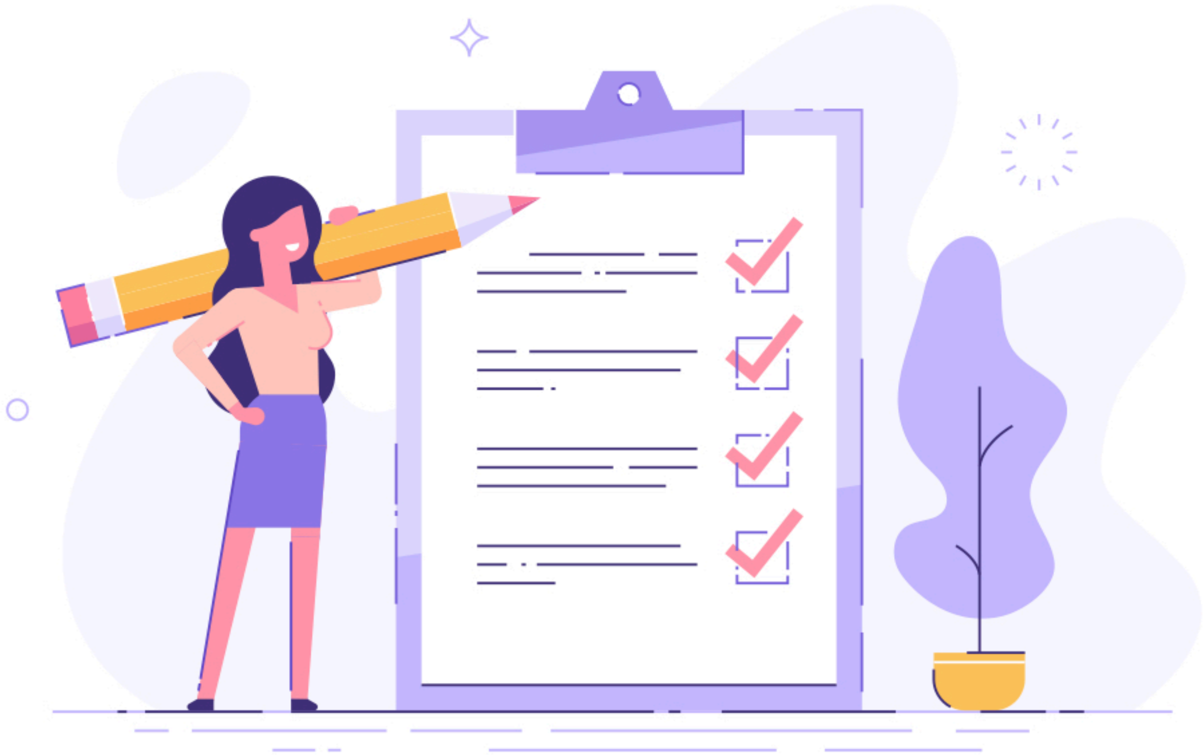
Develop Intelligence

- Neural Search is the subject of using NLP techniques to search engines
- We may use neural networks at **index time, query time, or ranking time.**

Summary

- Neural Search is the subject of using NLP techniques to search engines
- We may use neural networks at **index time, query time, or ranking time.**
- The main goal is to simplify life **for users**.

Discussion Time



Discussion Time

Can you think of another use cases for NLP within your organisation?

Discussion Time

Can you think of another use cases for NLP within your organisation?

Allocated time: 15 minutes

Break: 15 minutes

Develop
Intelligence

Word2Vec

Develop Intelligence

Synonym expansion

Agenda

Develop Intelligence

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

NER

- Scapy NER
- BERT

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

Text Generation

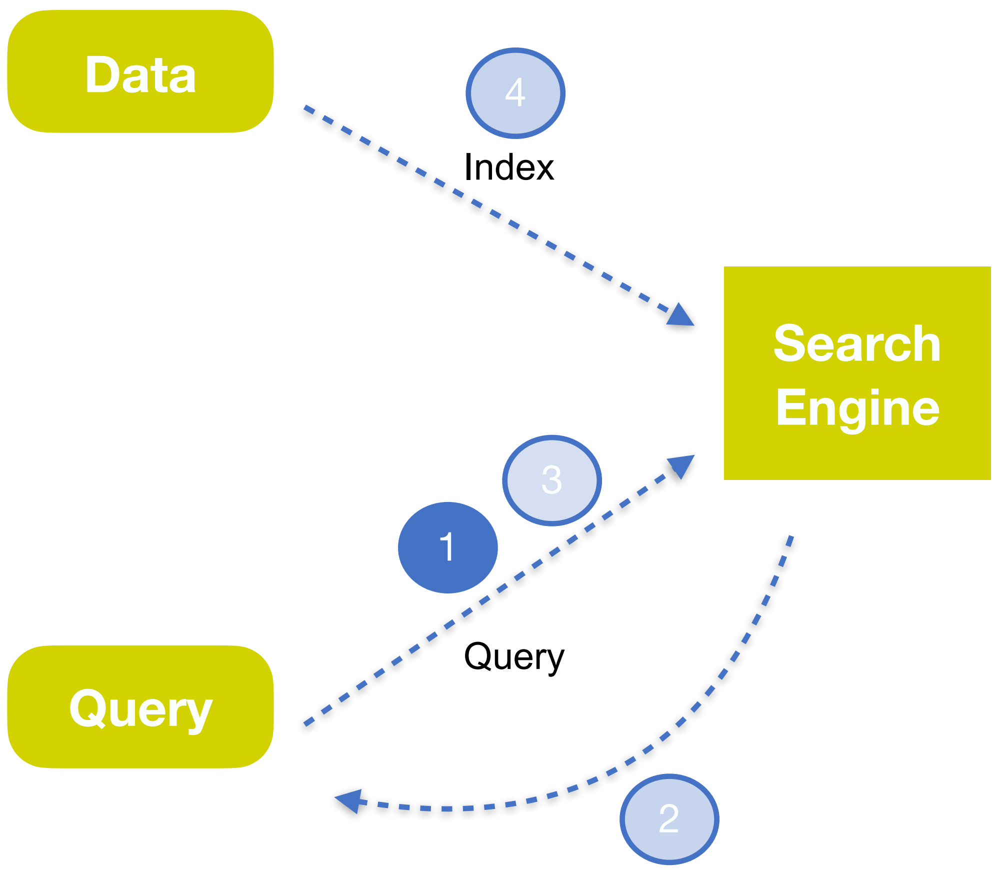
- RNN
- GRU and LSTM
- Transformers

Next Steps

- Closing thoughts

What are we going to do now?

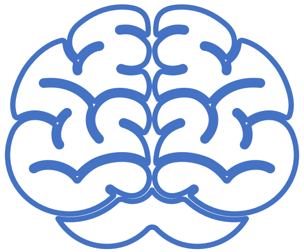
1. Synonym Expansion
2. Reranking results
3. Alternative Queries
4. NER



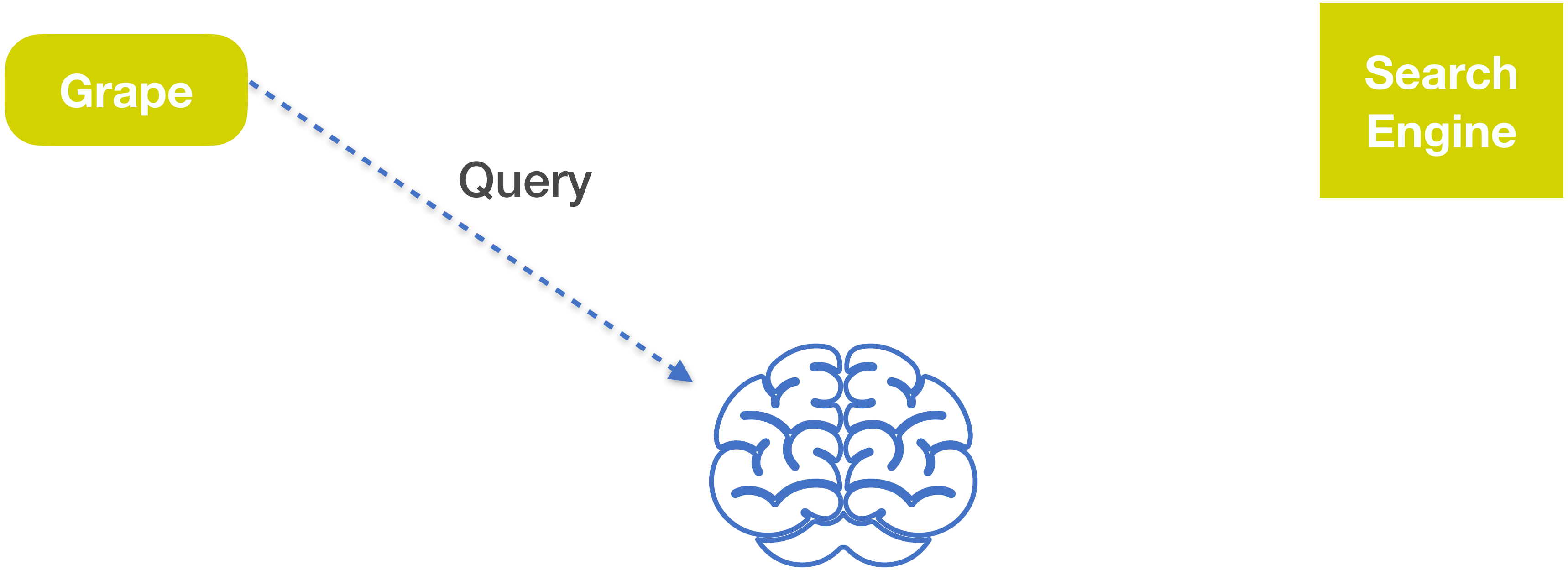
What are we going to do now?

Grape

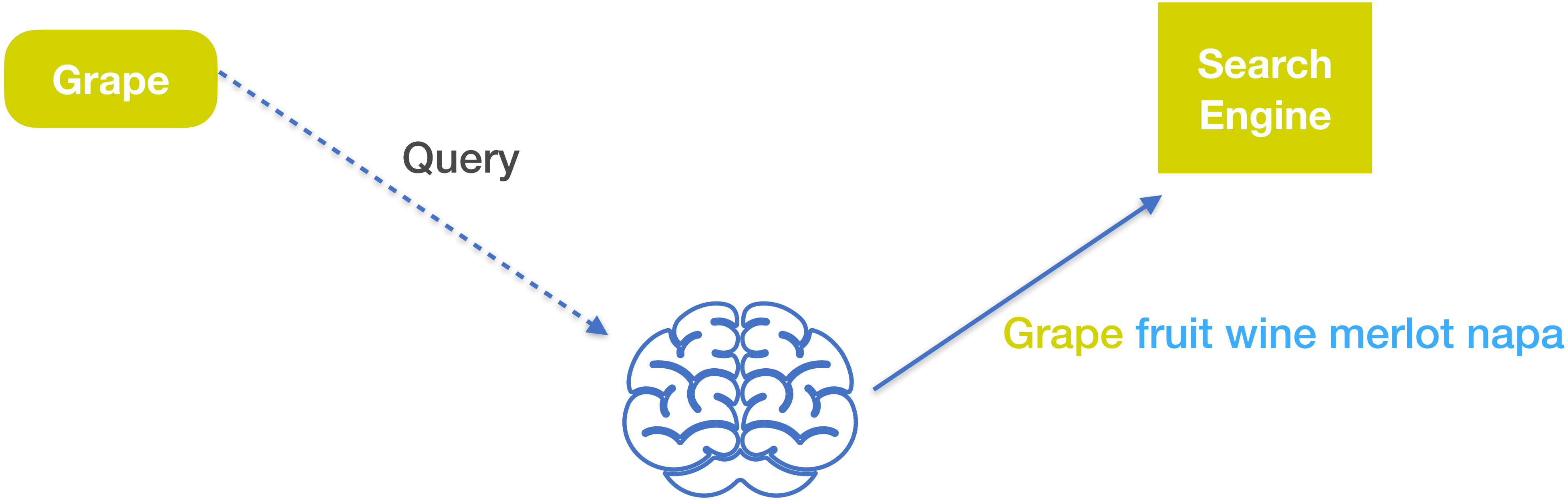
Search
Engine



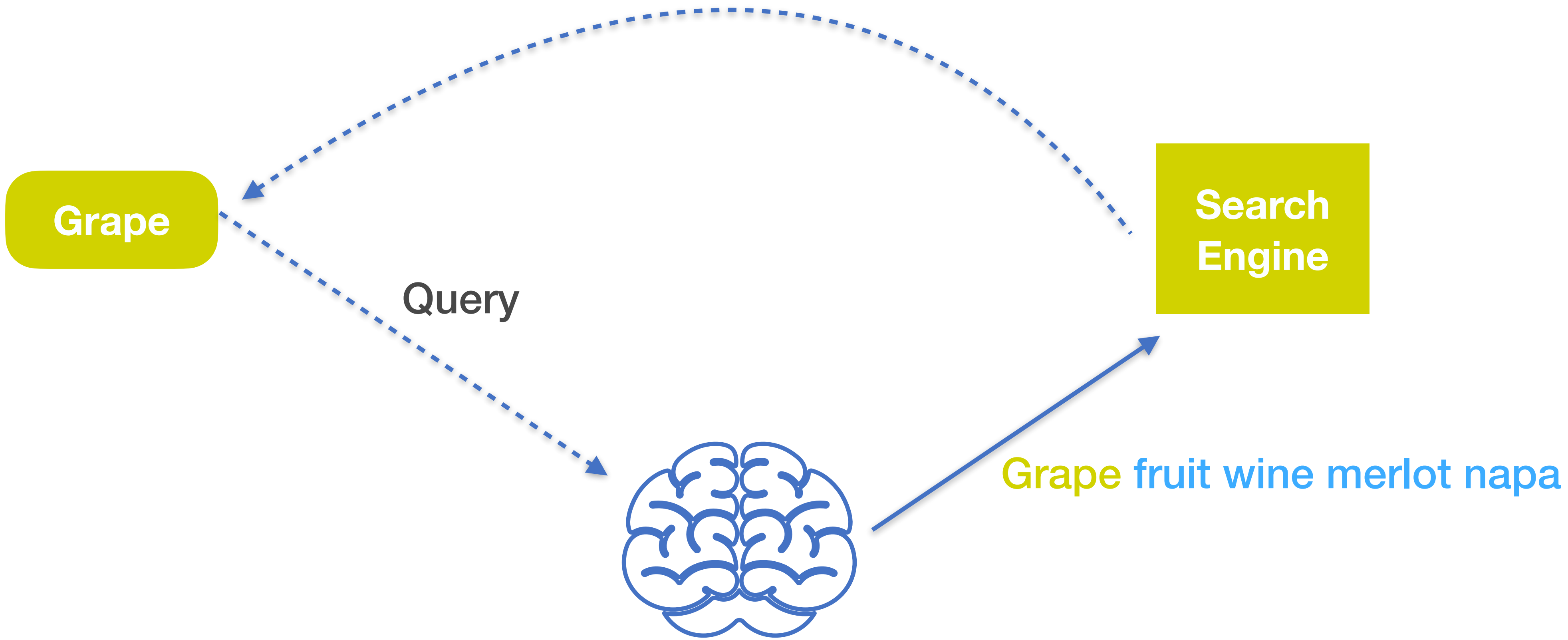
What are we going to do now?



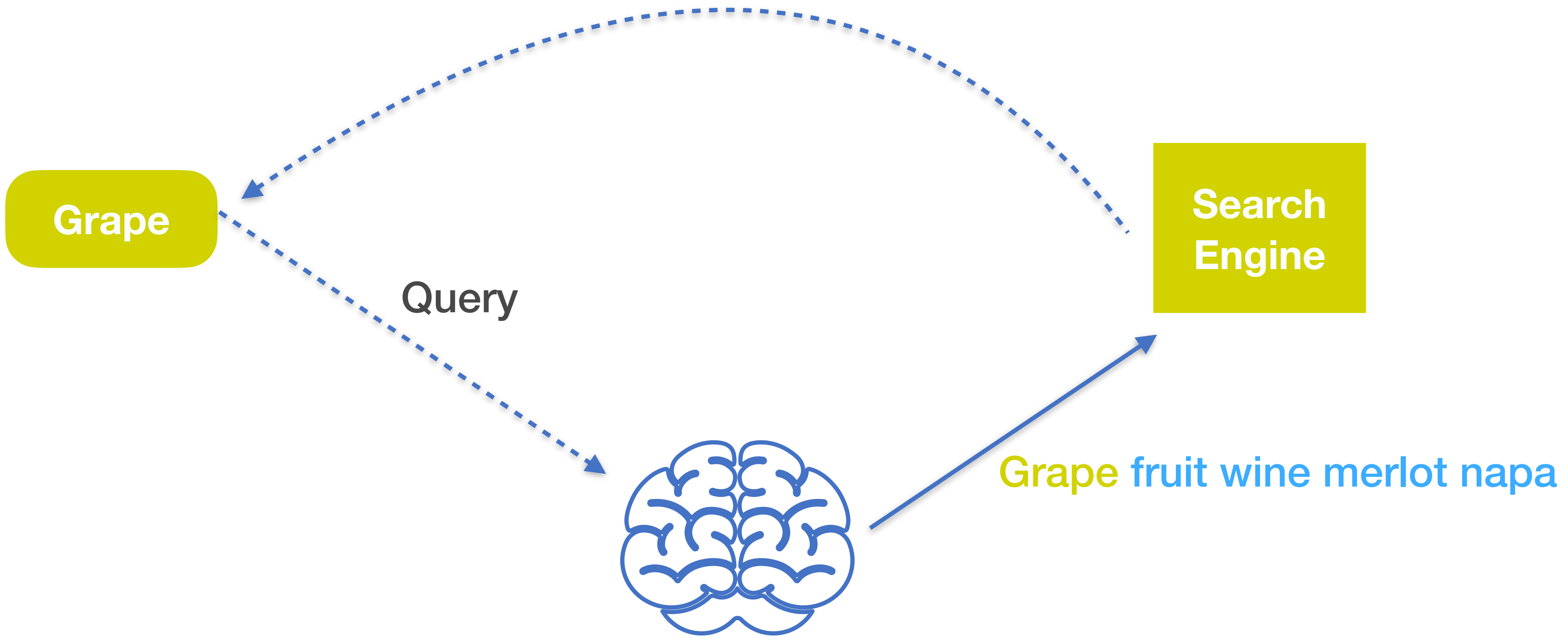
What are we going to do now?



What are we going to do now?



What are we going to do now?



Synonyms increase the chances of returning the result the user wants

How do we represent words?

Spain

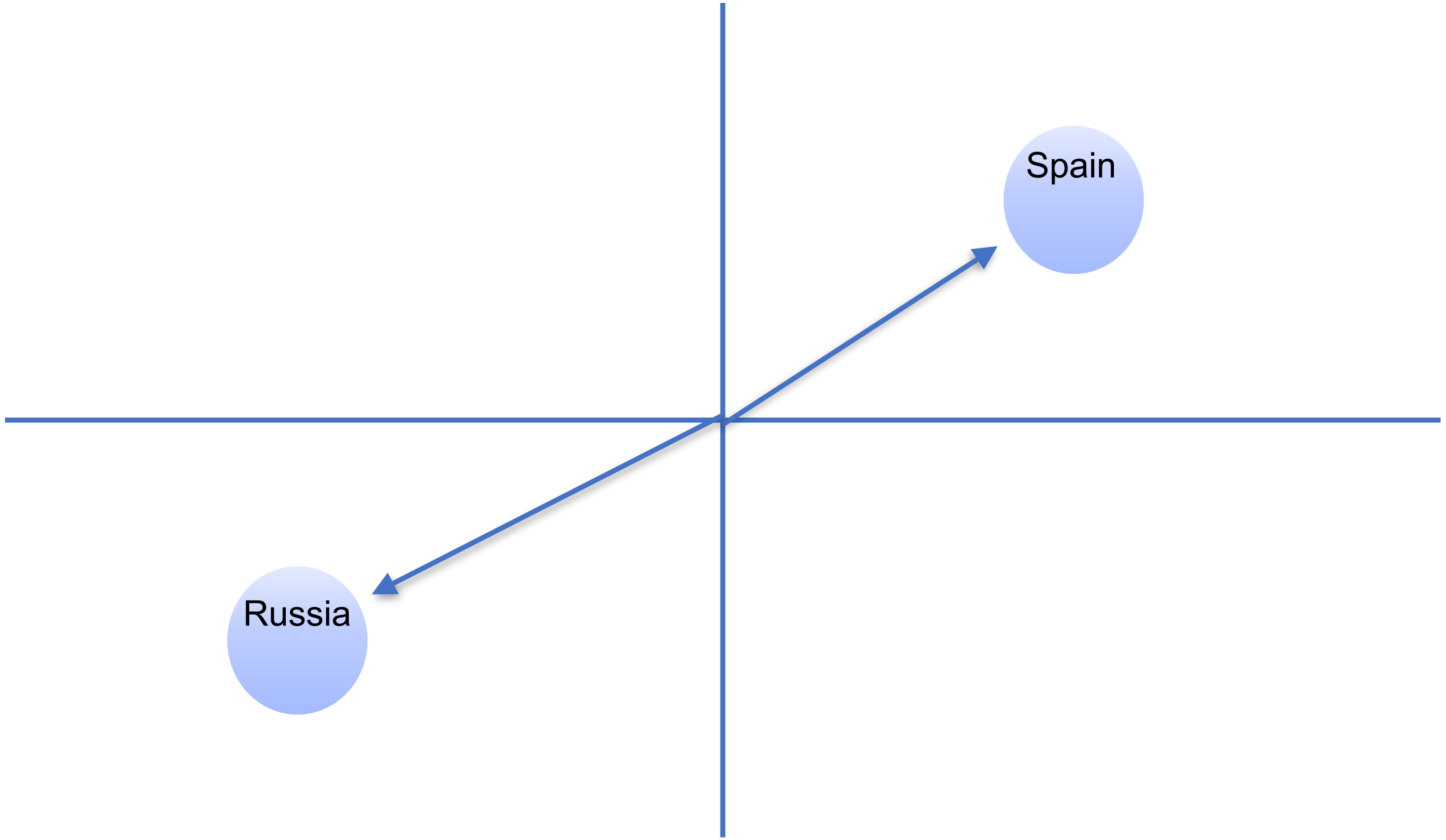
Russia

How do we represent words?

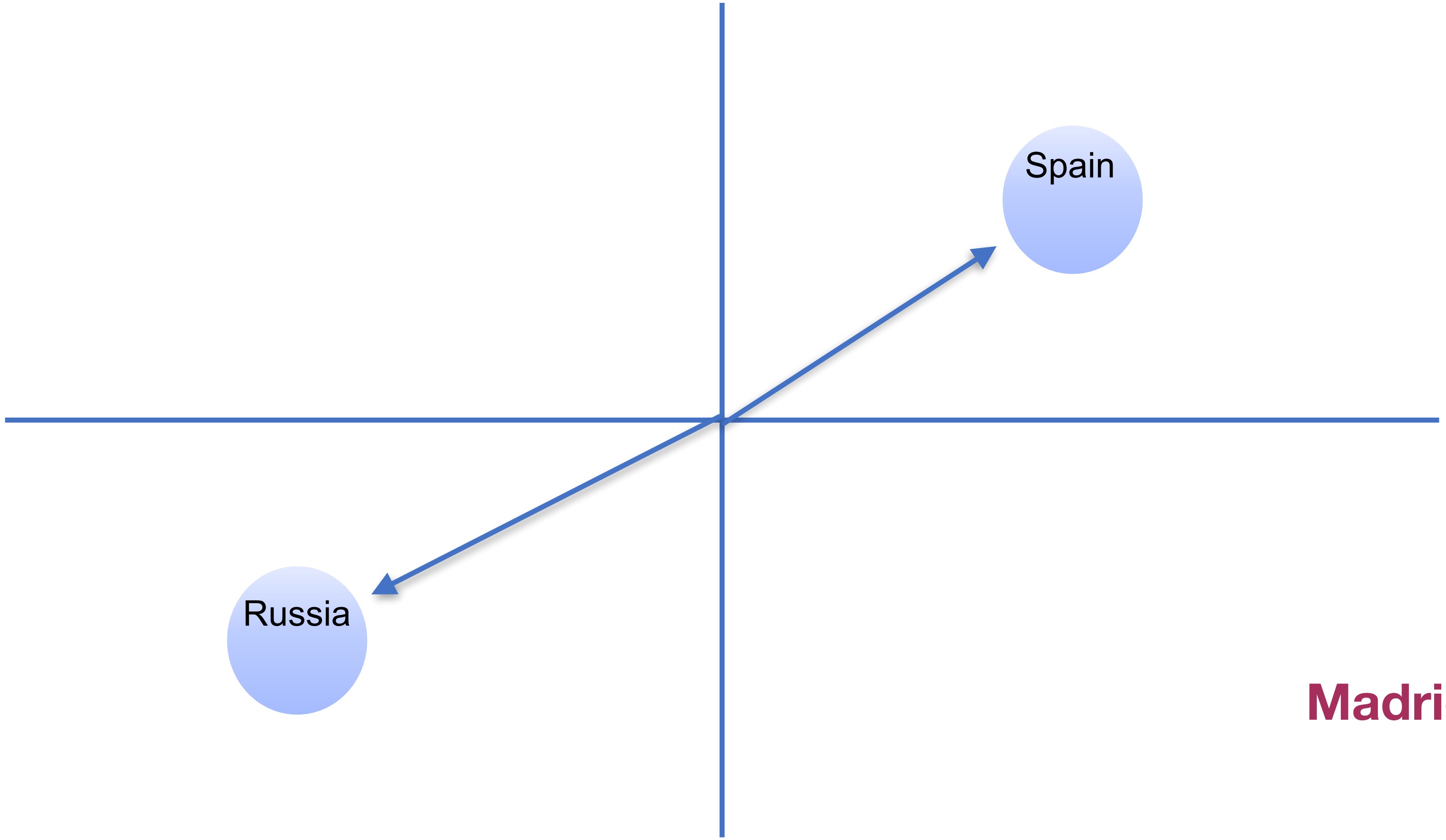
Spain → [1, 0]

Russia → [-1, 1]

How do we represent words?

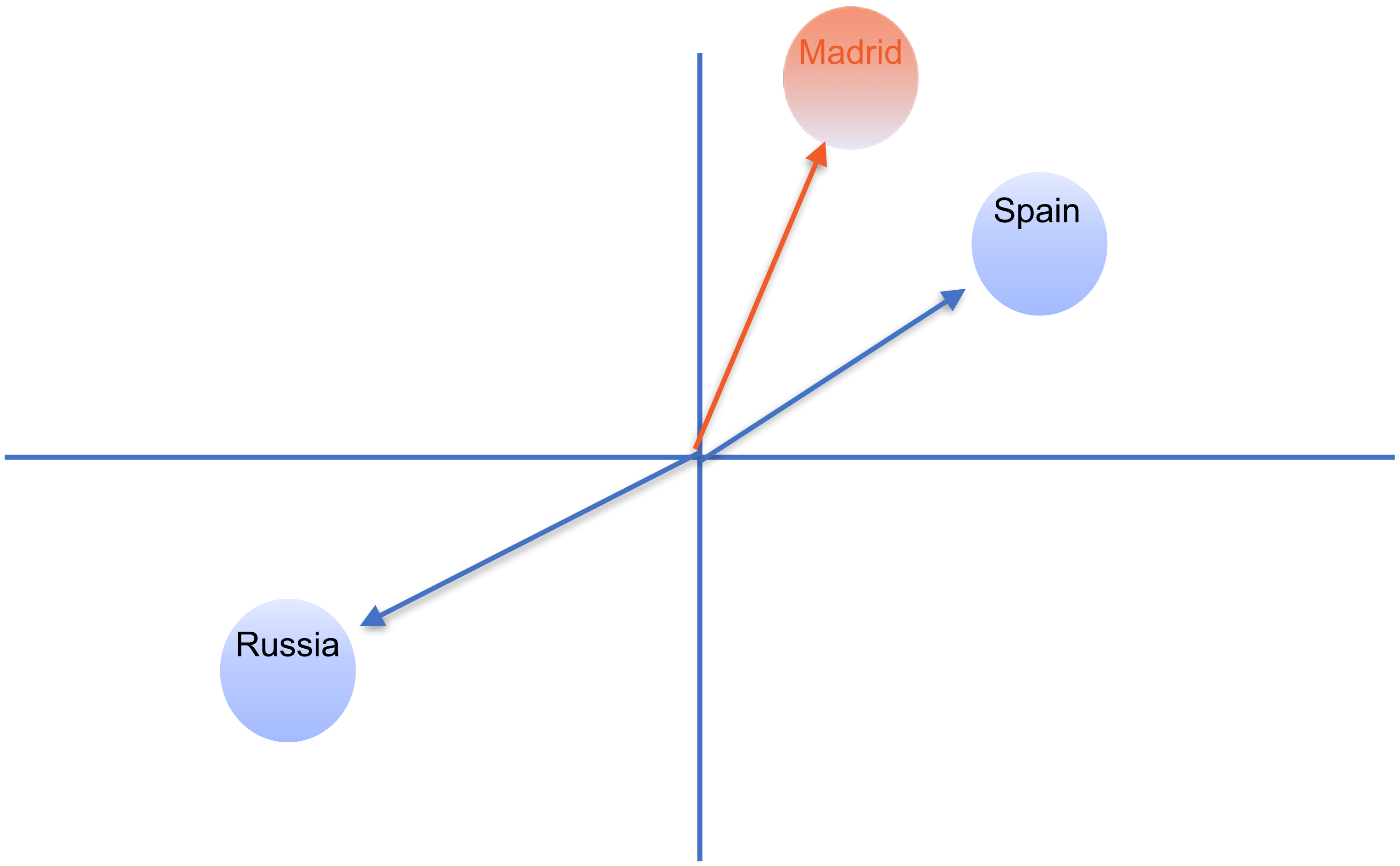


How do we represent words?

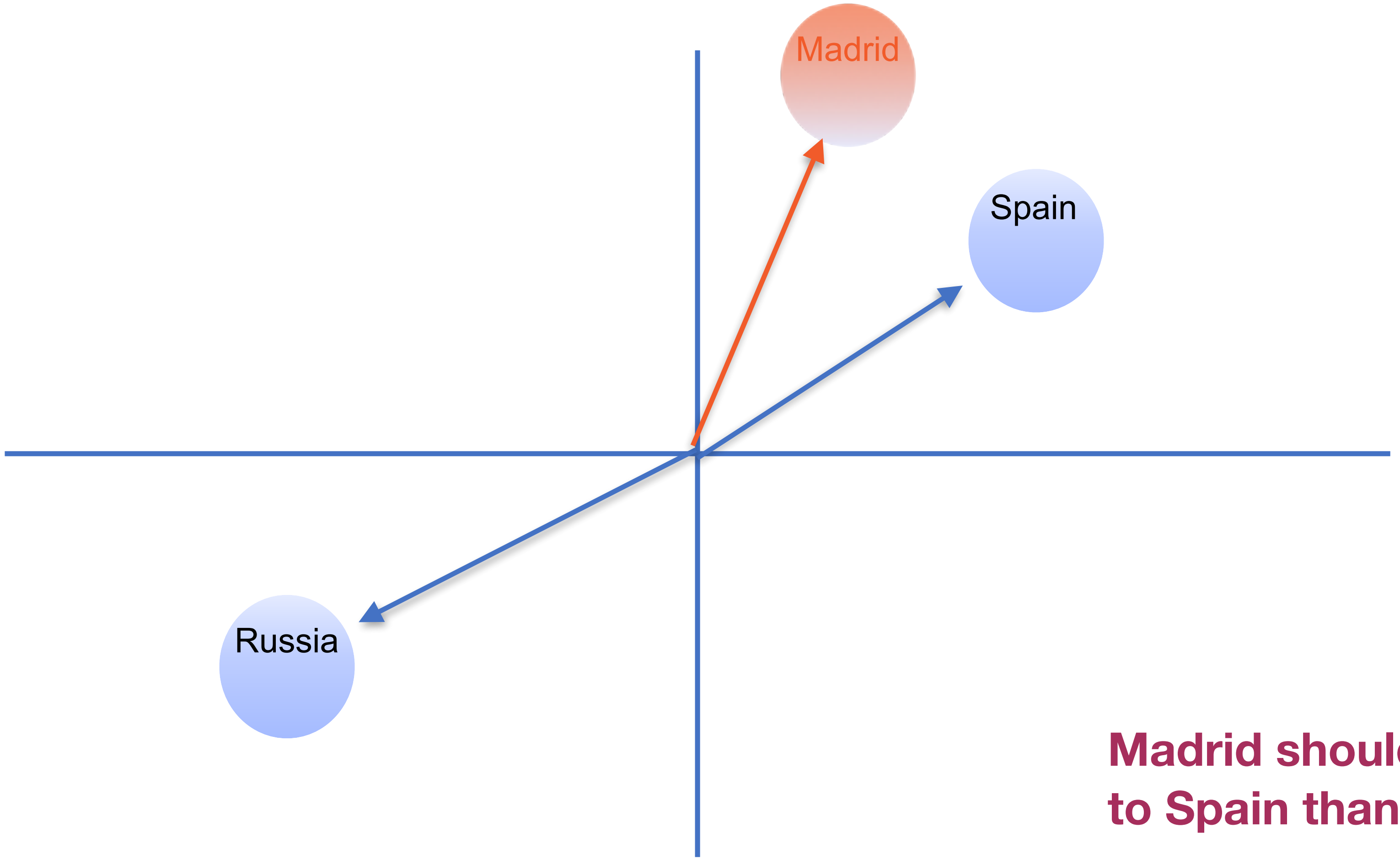


Madrid??

How do we represent words?

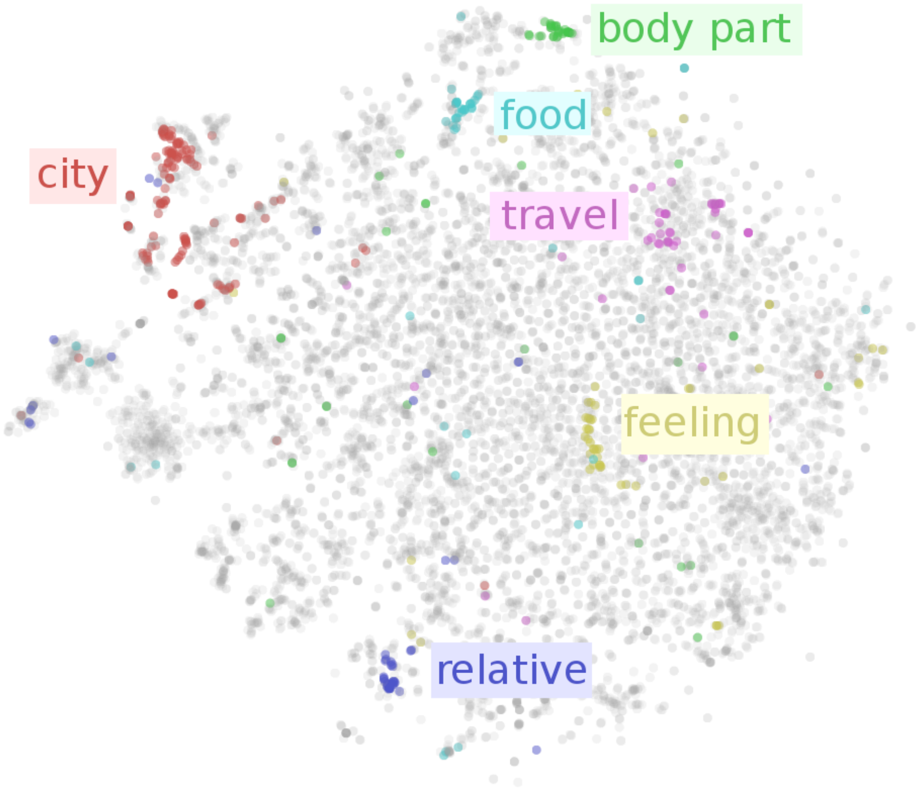


How do we represent words?

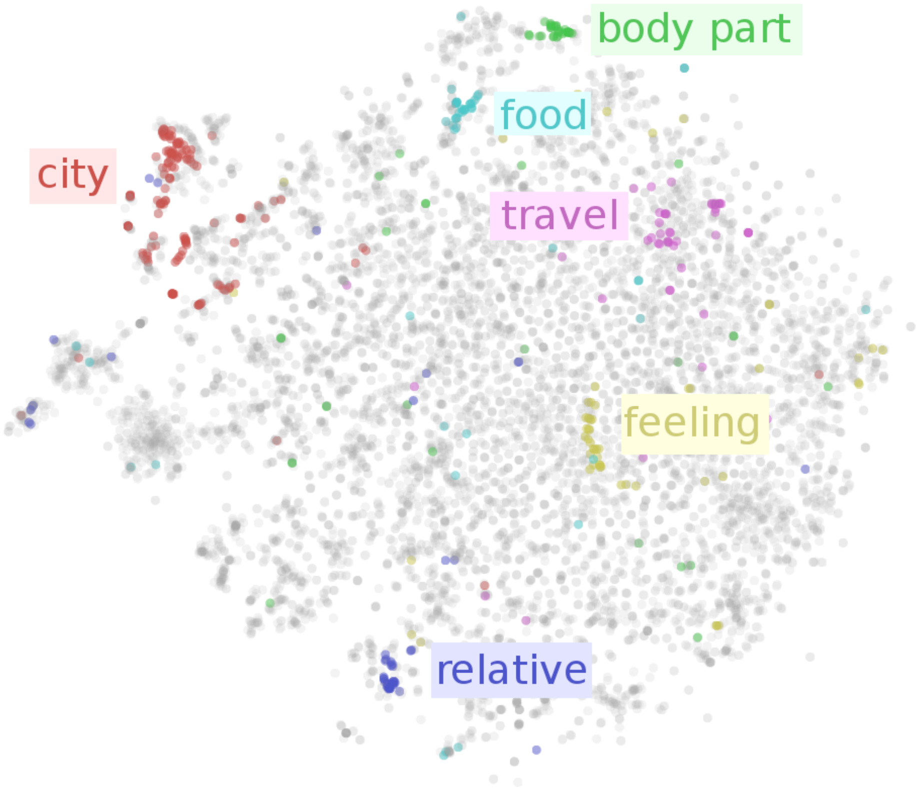


Madrid should be closer to Spain than Russia

How do we represent words?

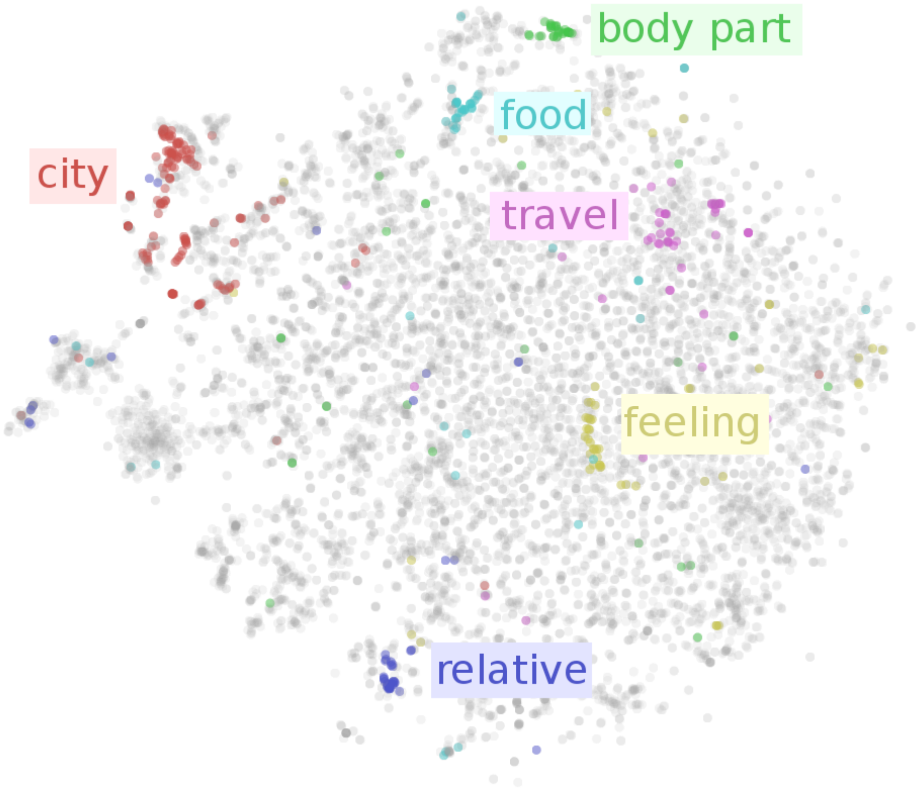


How do we represent words?



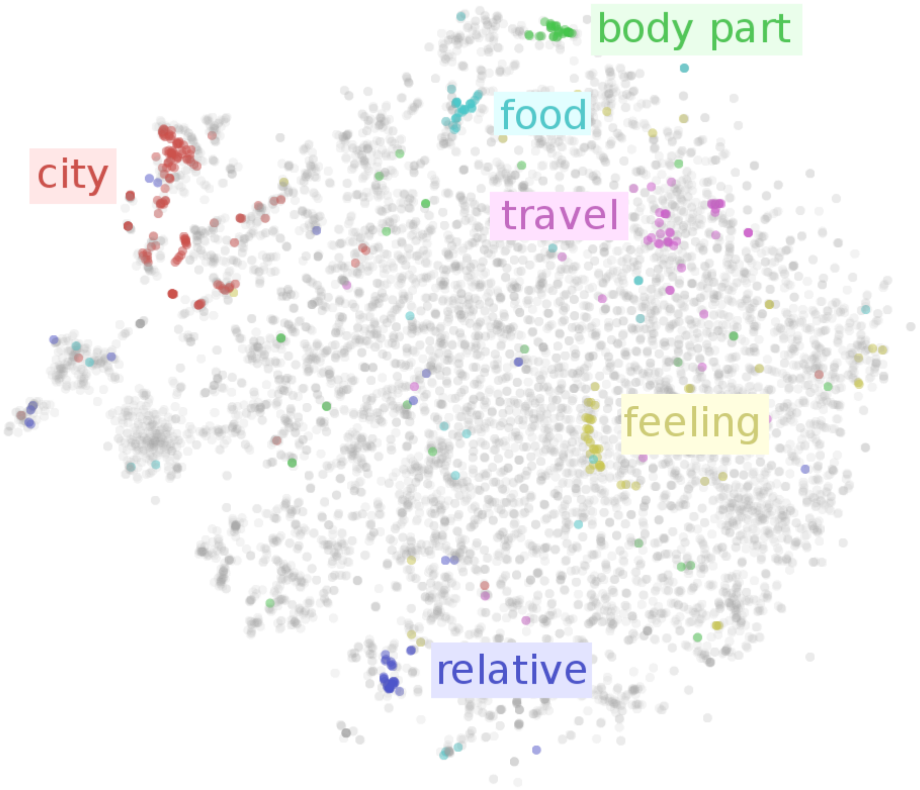
- This mapping, or function, from words to sequences of numbers is called a **word representation or embedding.**

How do we represent words?



- This mapping, or function, from words to sequences of numbers is called a **word representation or embedding.**
- If we have a good word representation, then we can find similar words! And use those synonyms to expand the query.

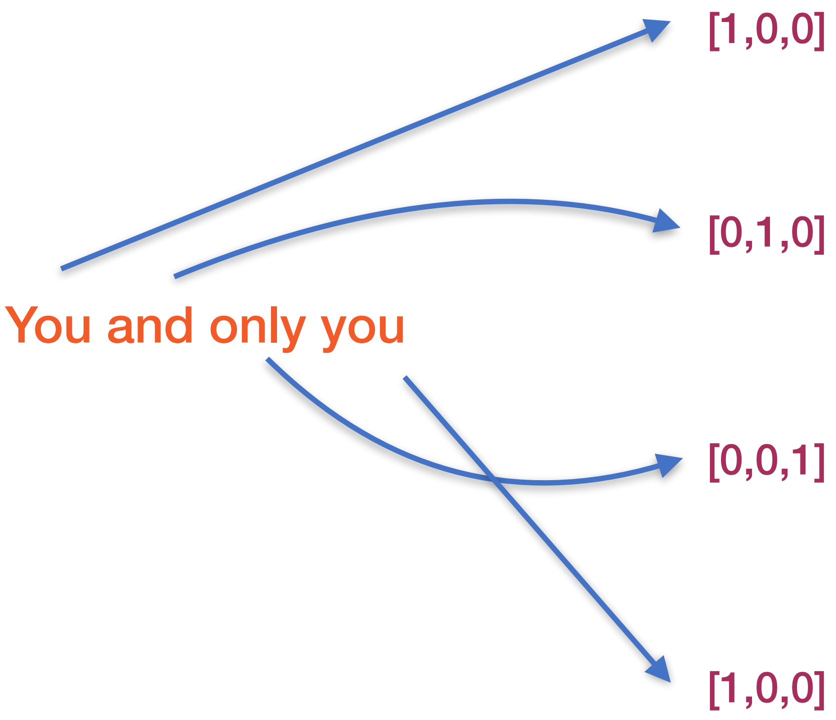
How do we represent words?



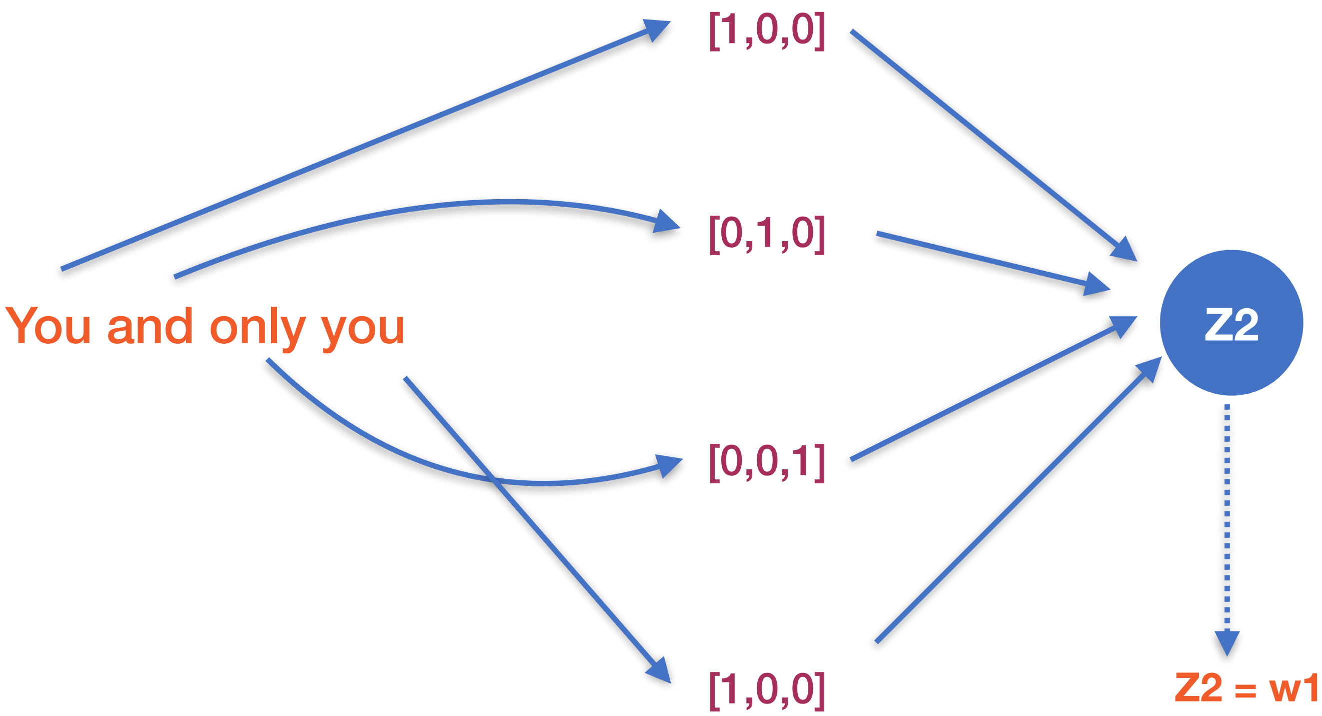
- This mapping, or function, from words to sequences of numbers is called a **word representation or embedding.**
- If we have a good word representation, then we can find similar words! And use those synonyms to expand the query.
- It has been shown that **neural networks** are great to find good embeddings

Neural Networks

You and only you

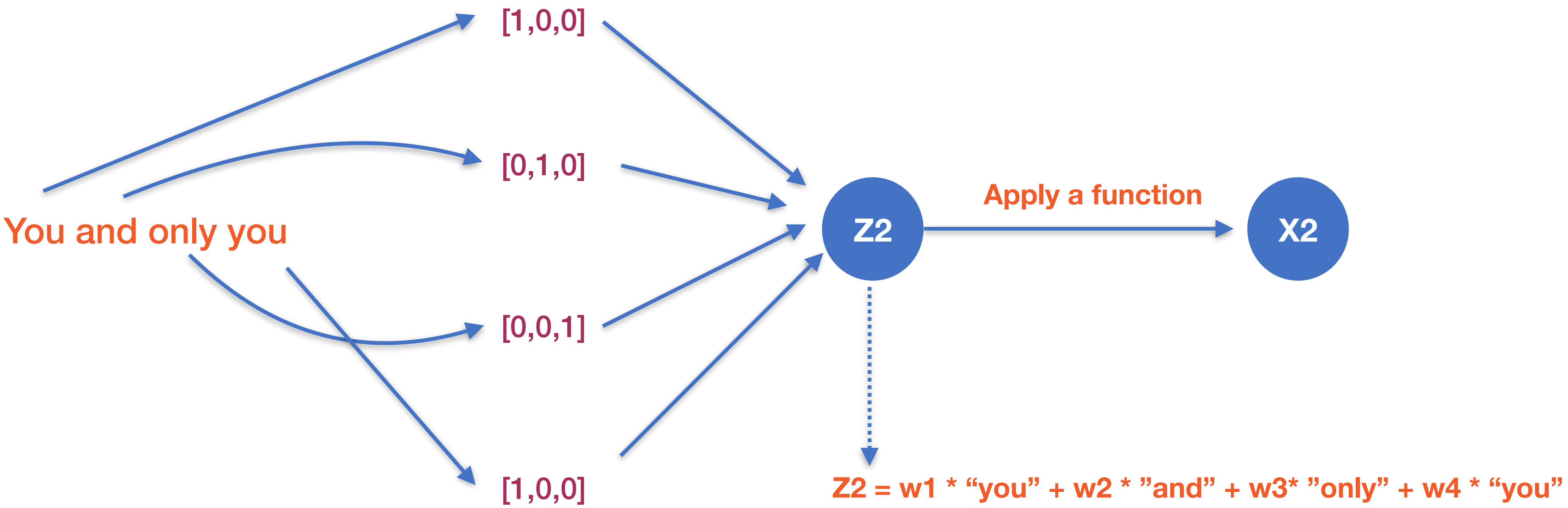

[1,0,0]
[0,1,0]
You and only you
[0,0,1]
[1,0,0]

Neural Networks

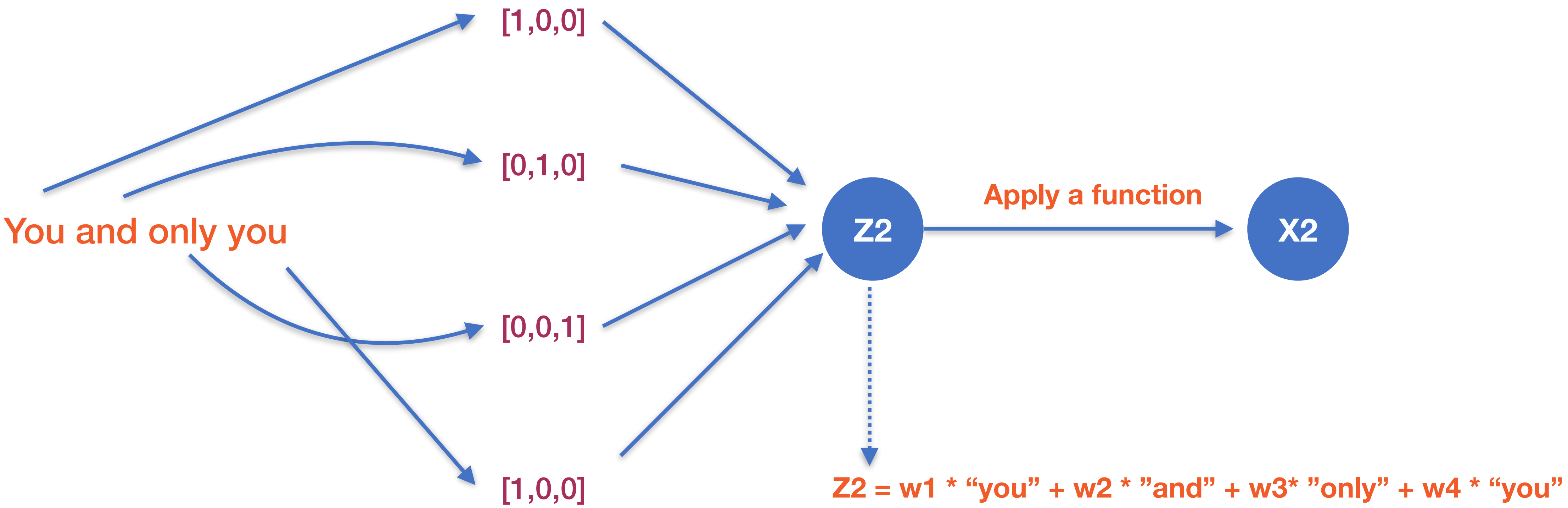


Z2 = w1 * “you” + w2 * ”and” + w3* ”only” + w4 * “you”

Neural Networks

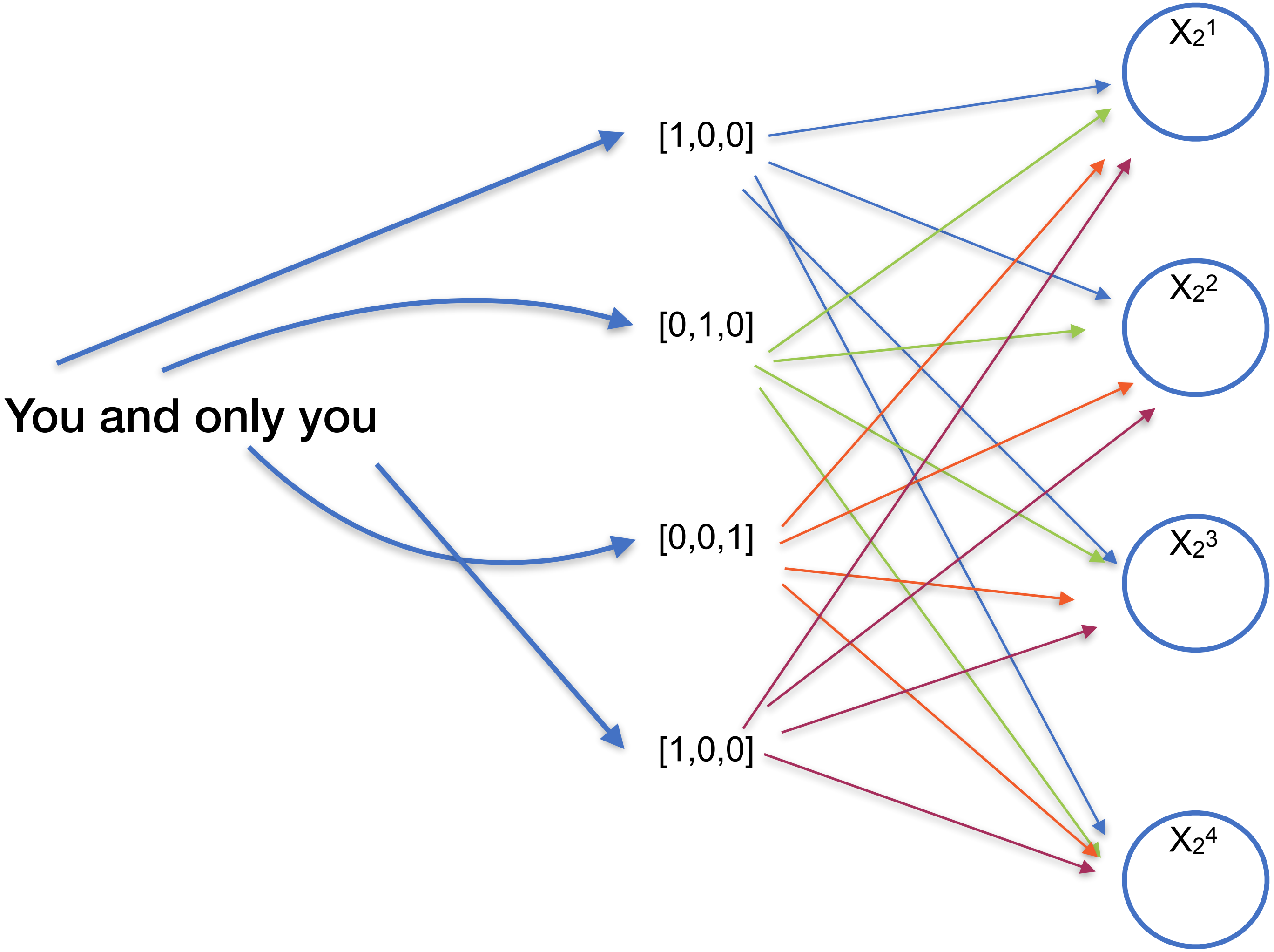


Neural Networks



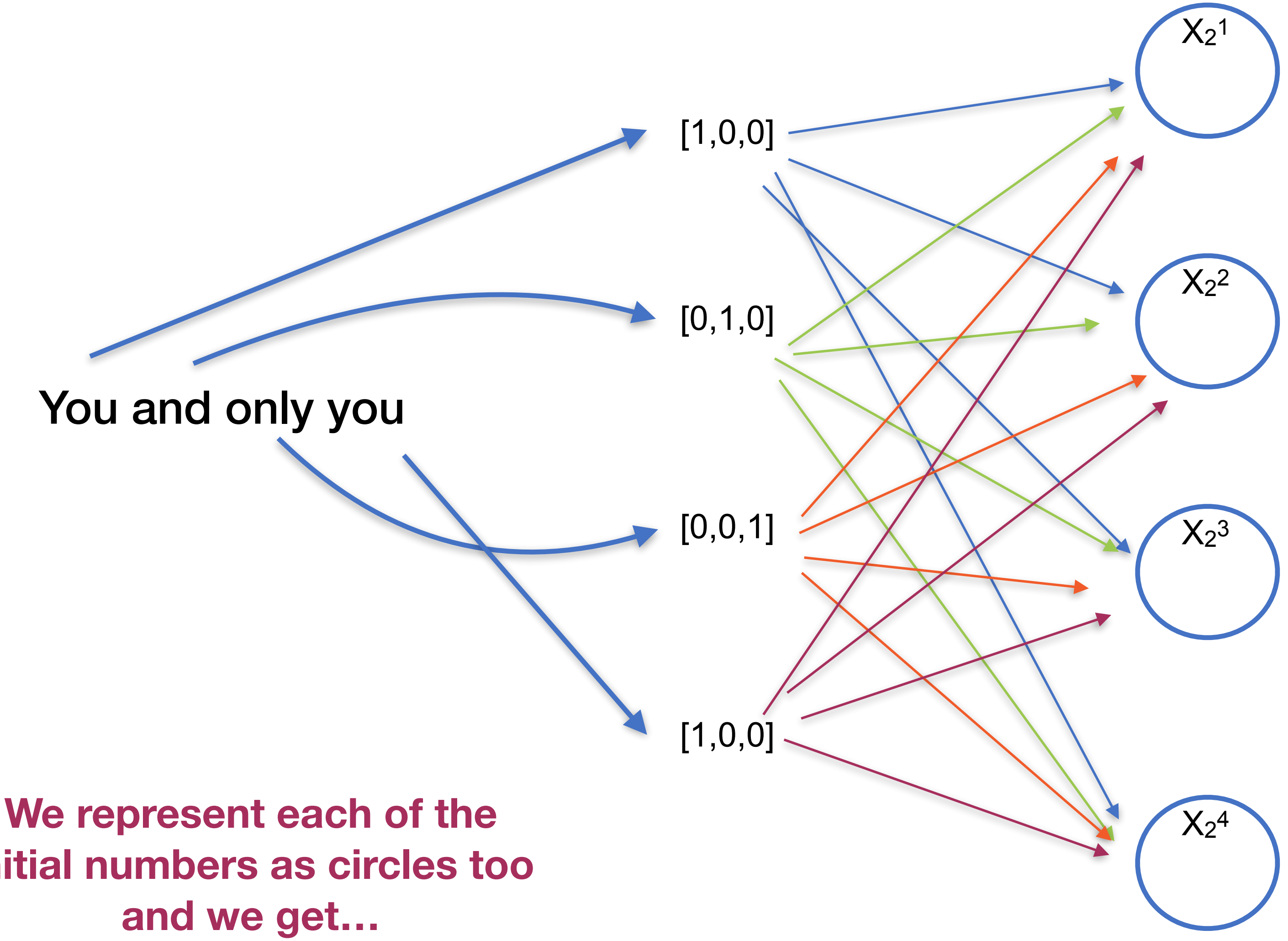
This would create a 1 dimensional representation

Neural Networks



This would create a 4 dimensional representation

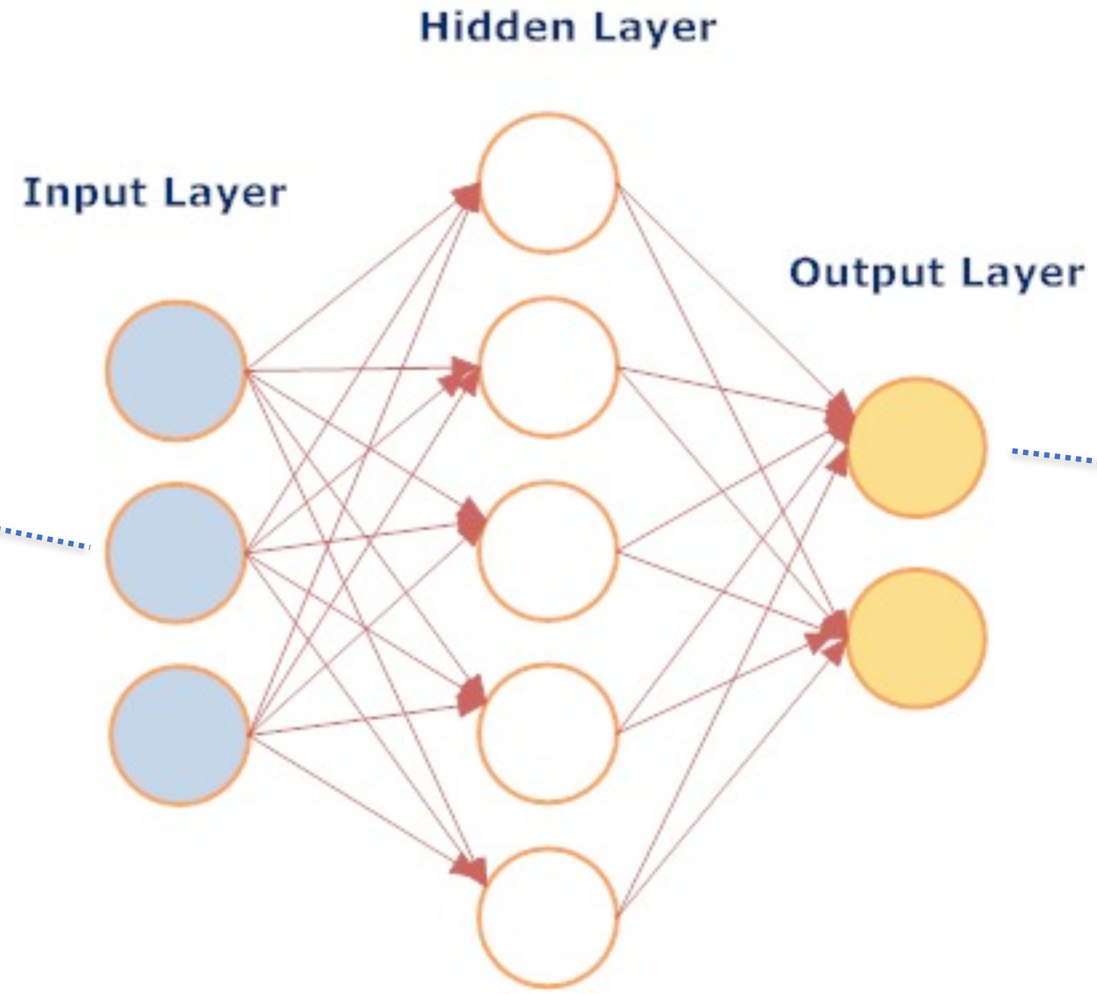
Develop Intelligence



We represent each of the initial numbers as circles too and we get…

This would create a 4 dimensional representation

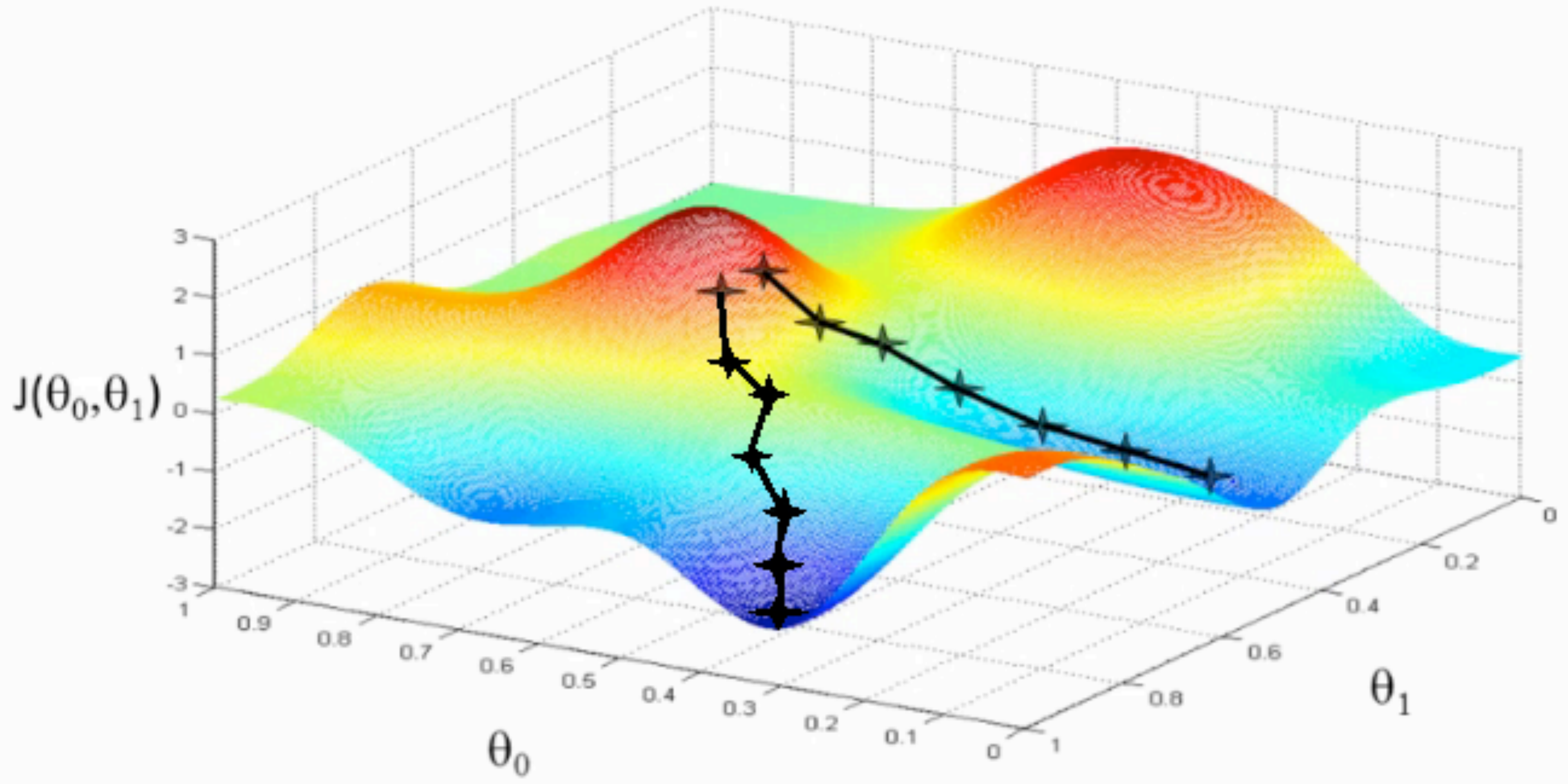
Neural Networks



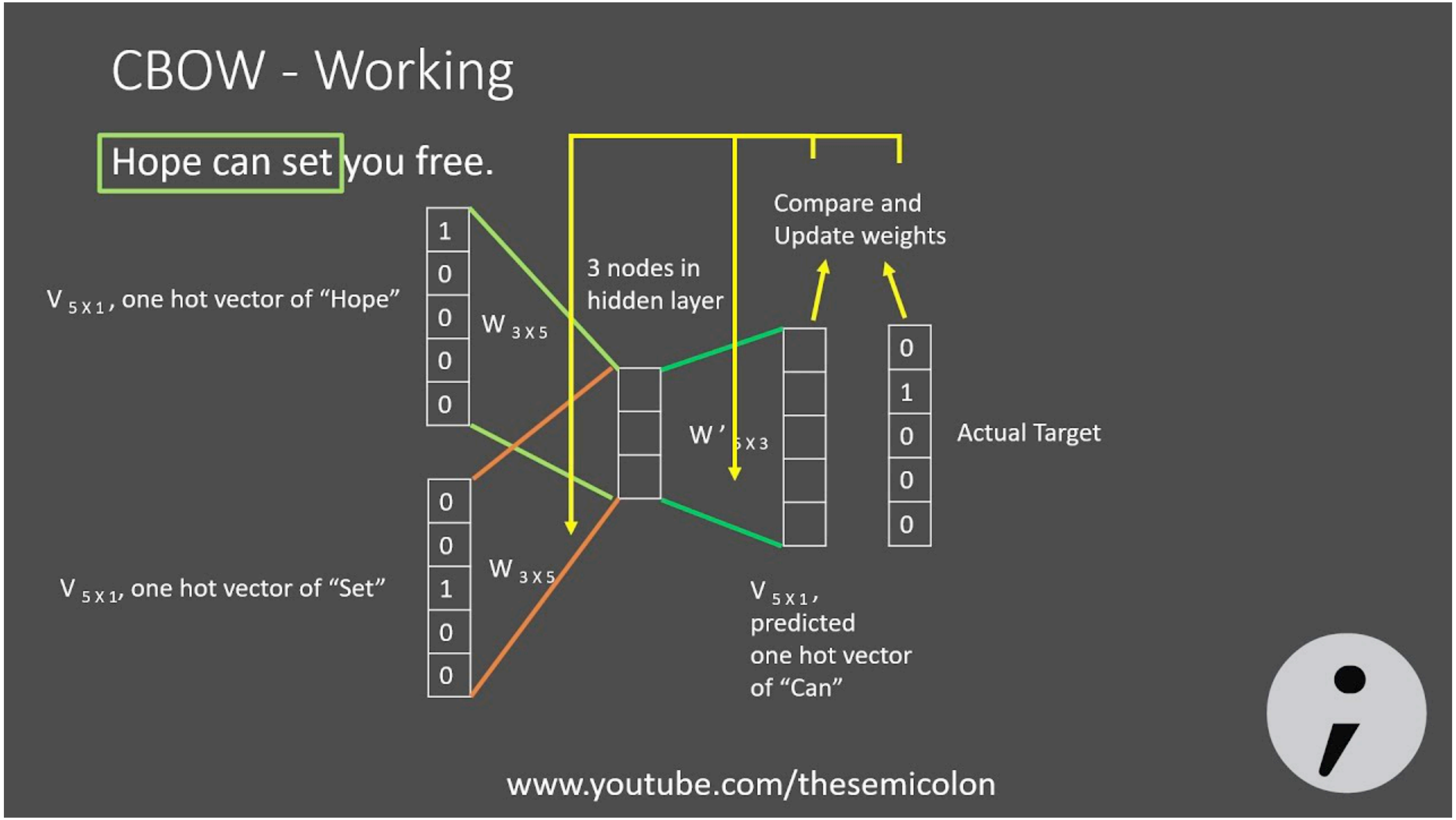
Each input word is a circle, for convenience

Here we added another “layer” of doing the same, increasing complexity and prediction power

Neural Networks: Training

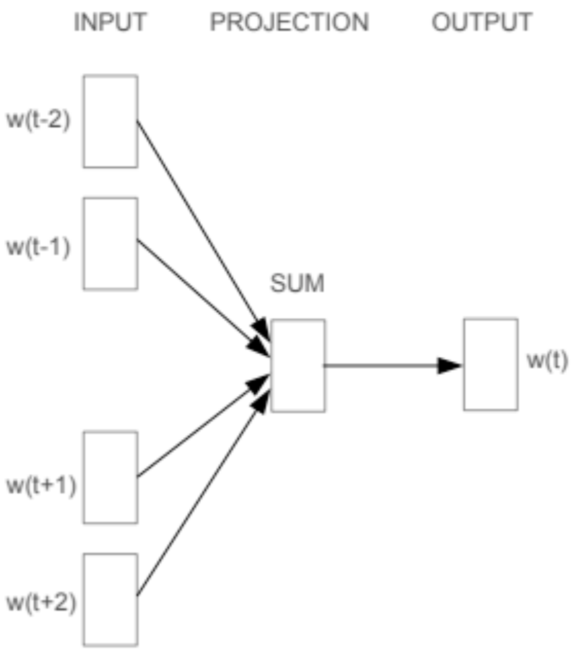


CBOW



CBOW - Working
Hope can set you free.
V 5 X 1 , one hot vector of "Hope"
1
0
0
0
0
W 3 X 5
3 nodes in
hidden layer
Compare and
Update weights
W ' 5 X 3
0
1
0
0
0
Actual Target
V 5 X 1, one hot vector of "Set"
0
0
1
0
0
W 3 X 5
V 5 X 1 ,
predicted
one hot vector
of "Can"
www.youtube.com/thesemicolon

Why we do this?

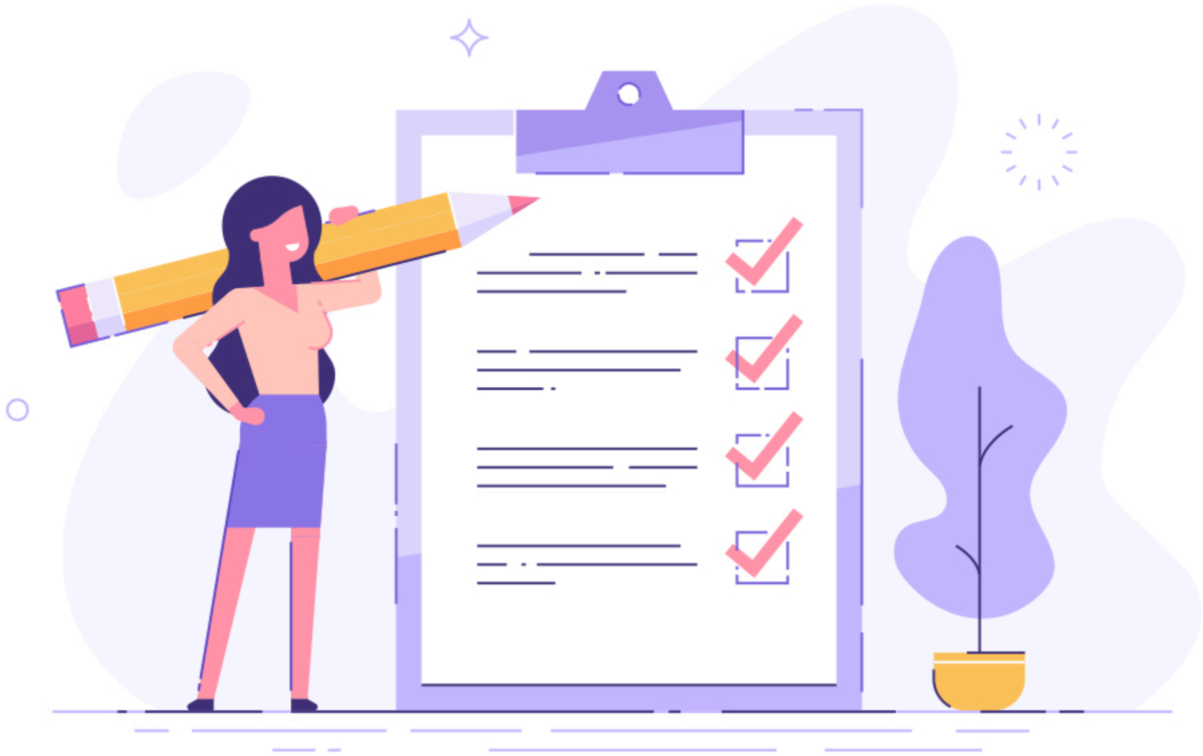


CBOW

EMBEDDING

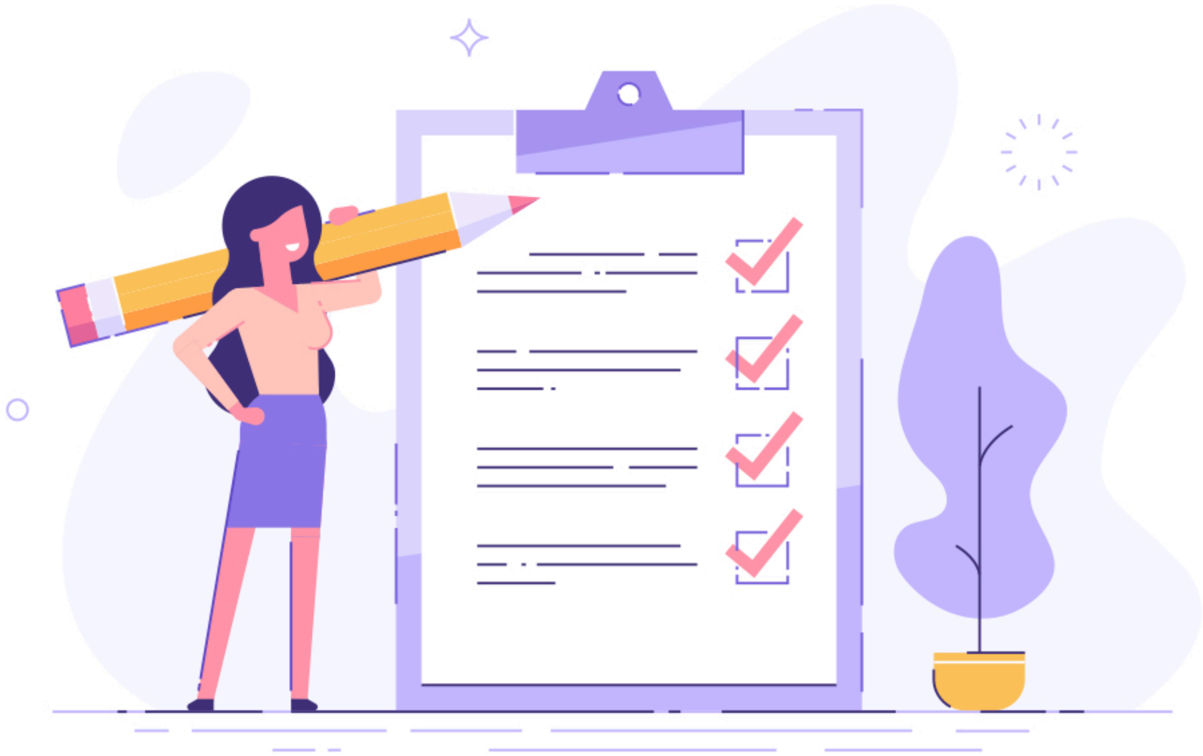
- This way, we can later get those *w* that mapped the words into this n-dimensional representation. **That is our embedding.**
- We used this trick with windows to be able to **train it** in an unlabelled fashion
- It works incredibly well

LAB: Training CBOW



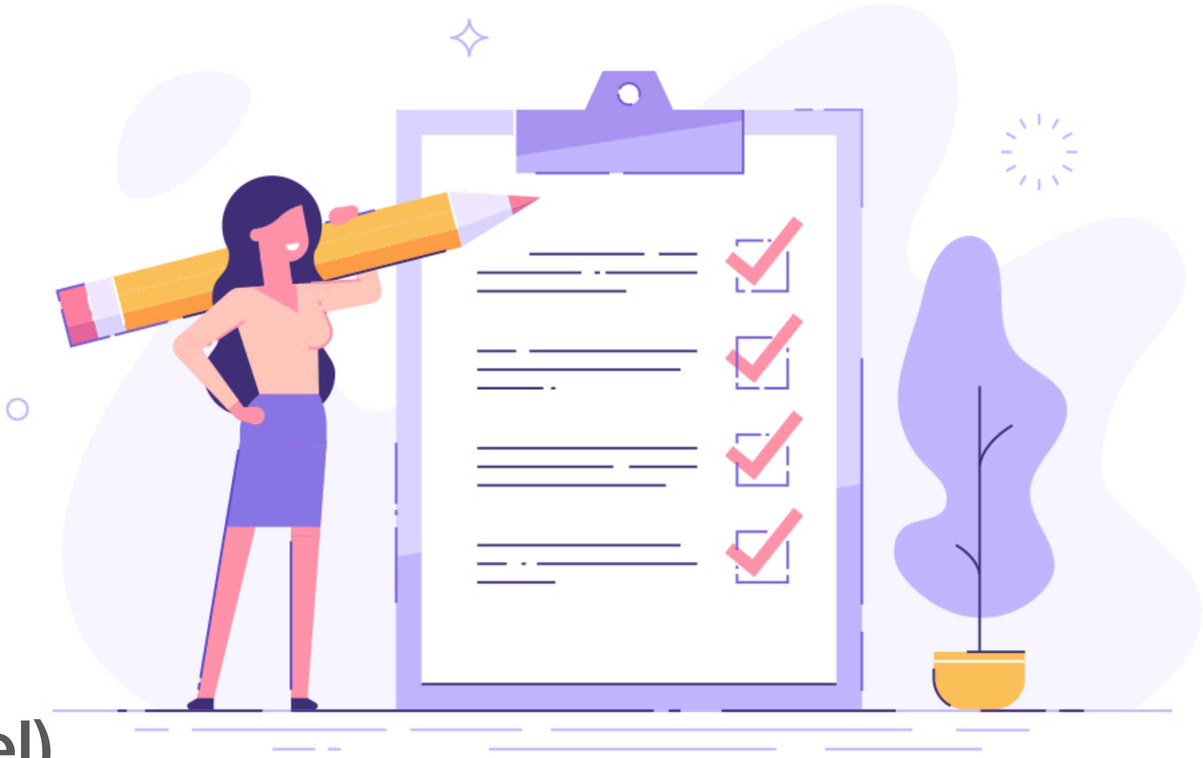
LAB: Training CBOW

- Train your first embedding using Keras



LAB: Training CBOW

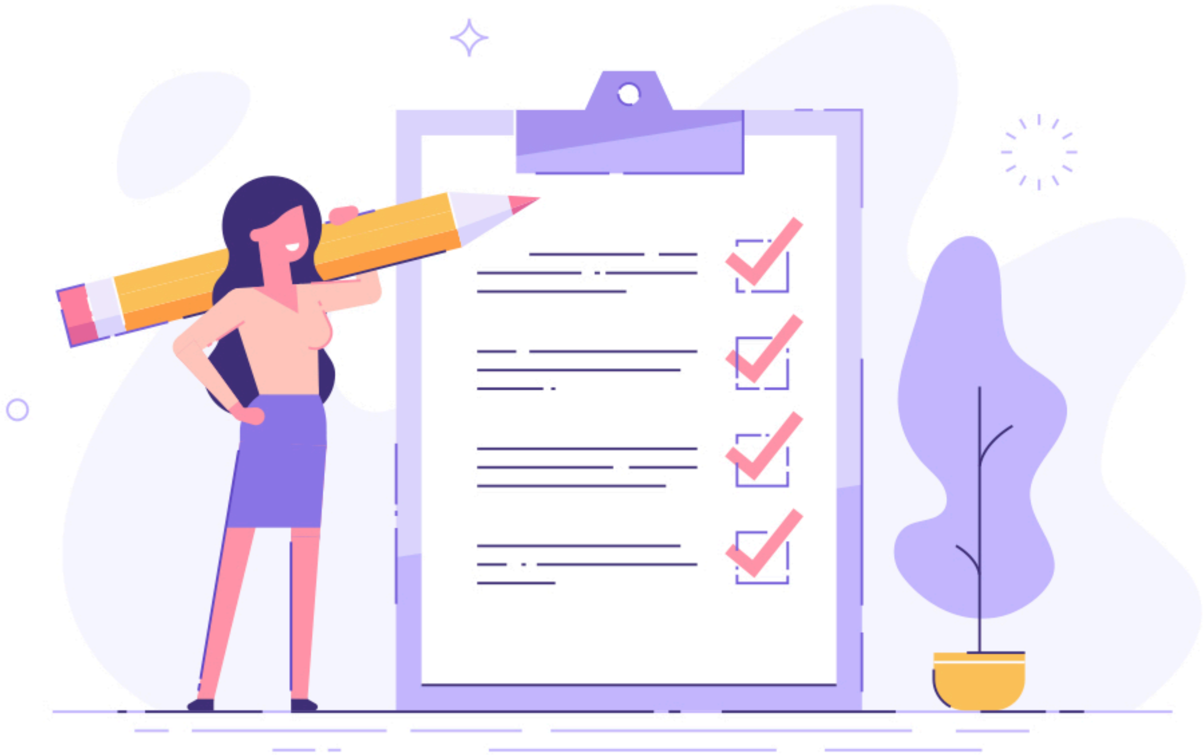
- Train your first embedding using Keras
- Transform text into a generator of training sequences



Allocated time: 30+ minutes

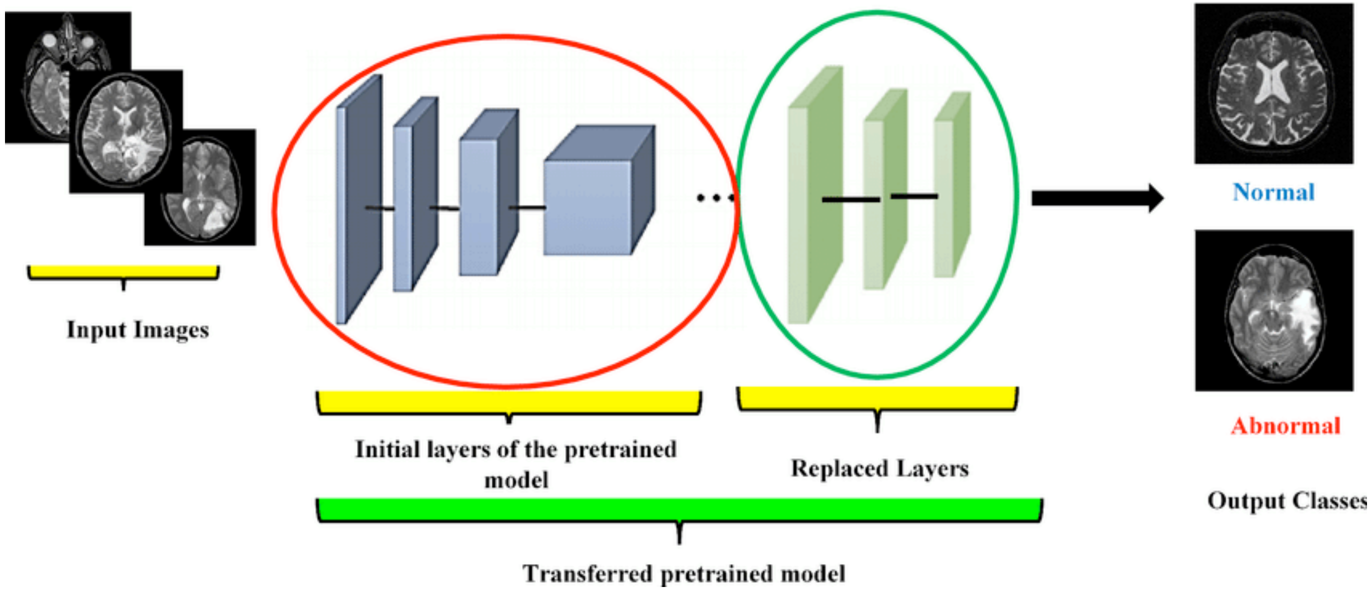
(depending on Keras comfort level)

Pulse Check



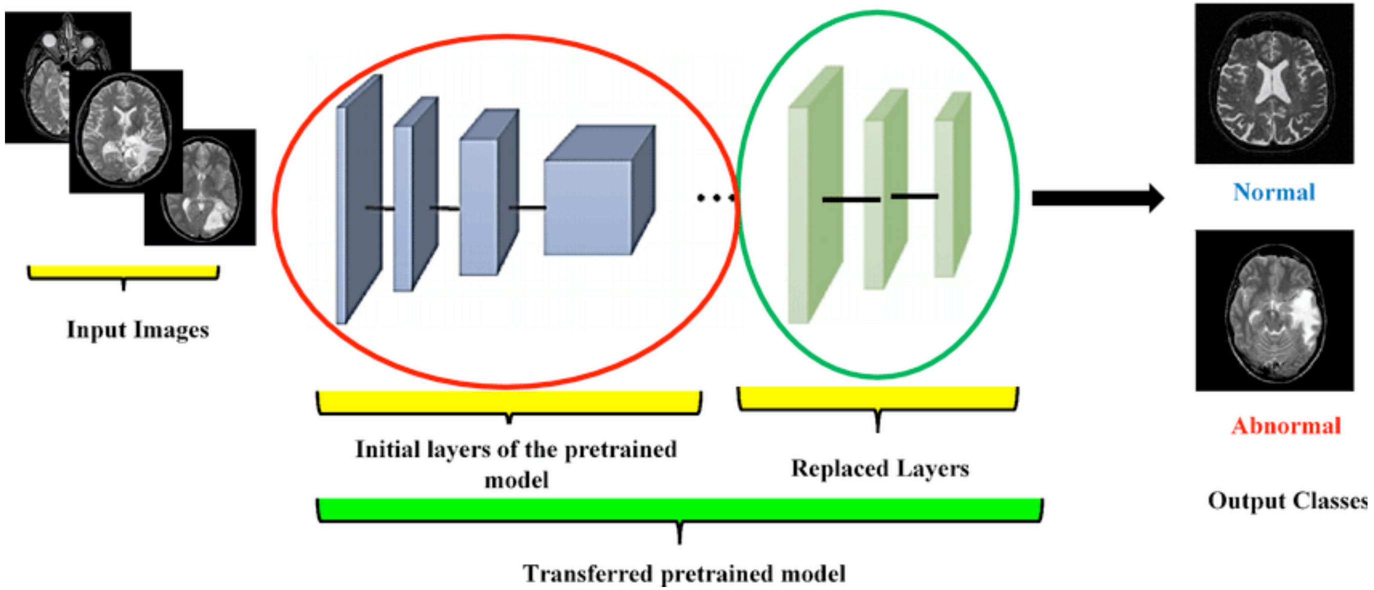
Lunch Break: 45 minutes

Can we do better?



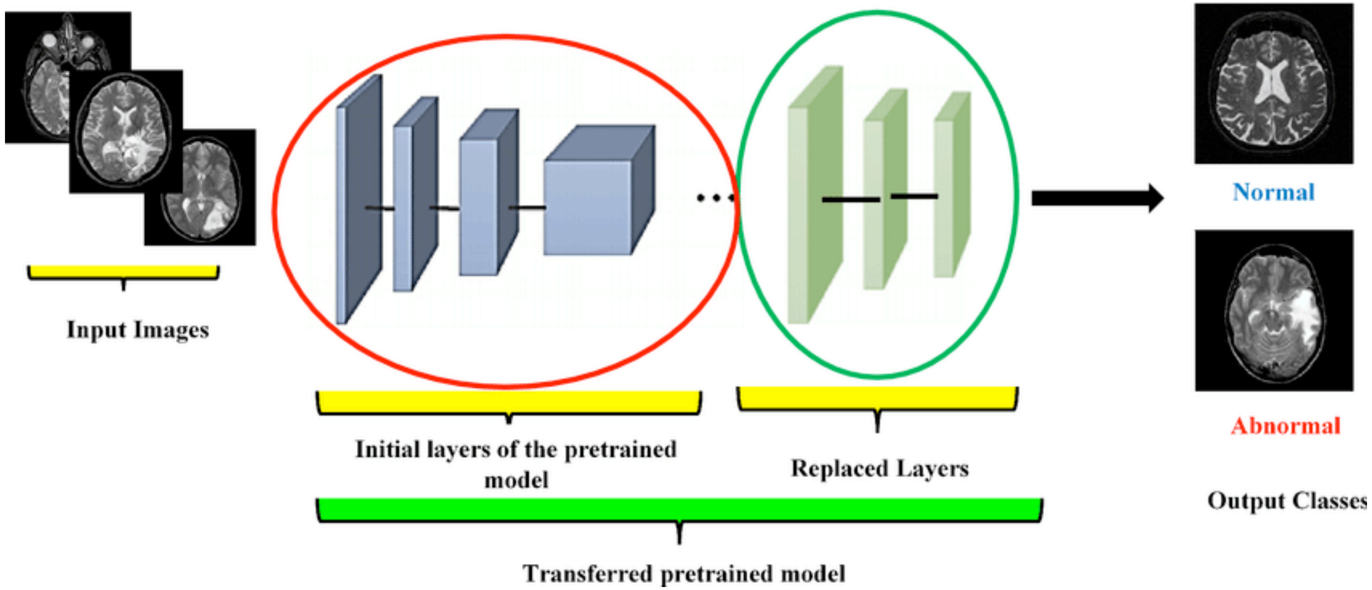
Can we do better?

- Researchers found that one can do many things to improve



Can we do better?

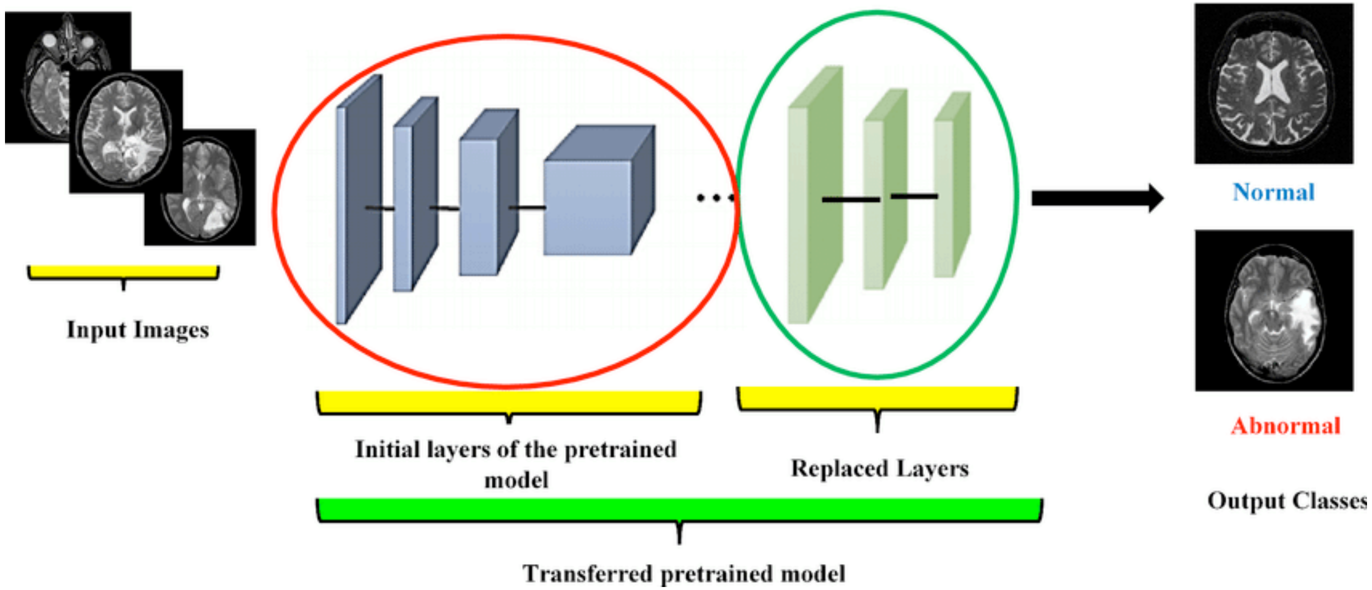
- Researchers found that one can do many things to improve



Can we do better?

- Researchers found that one can do many things to improve

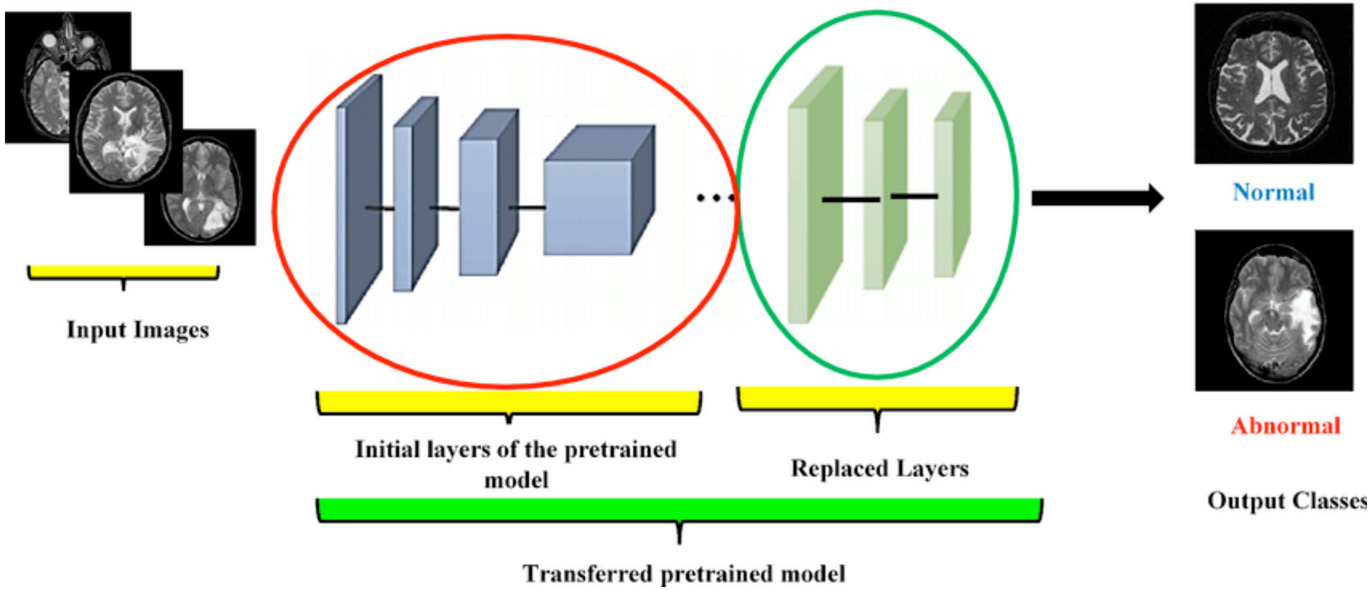
- As layers “stack up” one can reuse a pretrained model on the first layers and just train the rest.



Can we do better?

- Researchers found that one can do many things to improve

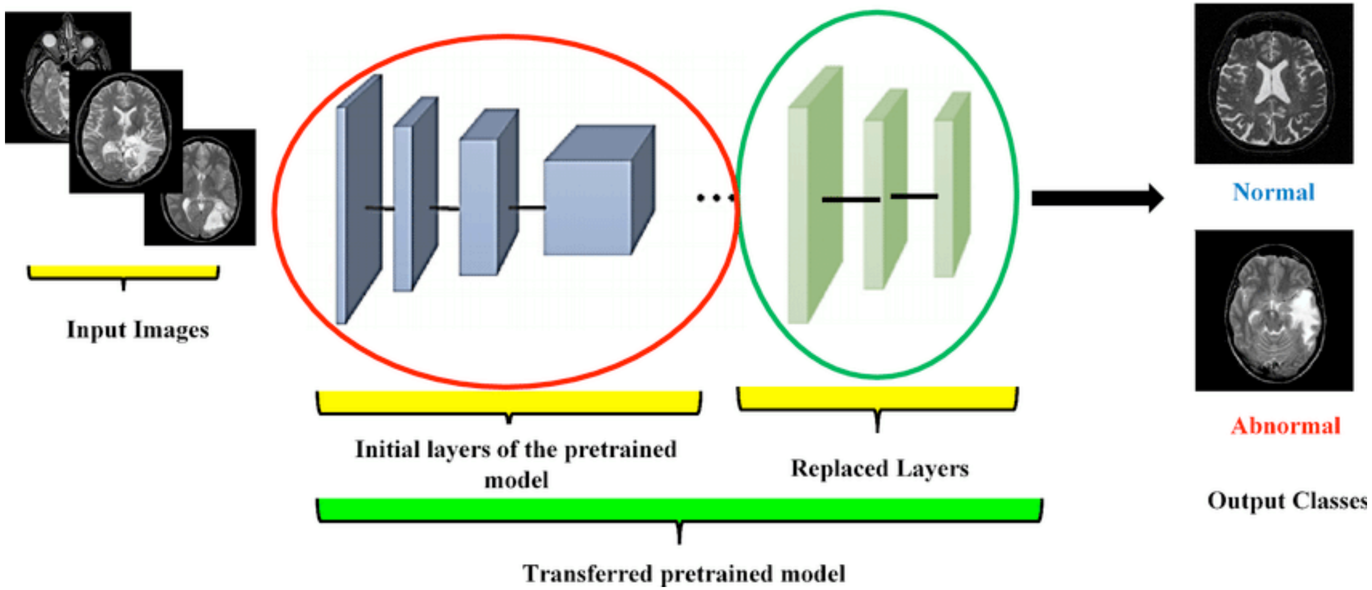
- As layers “stack up” one can reuse a pretrained model on the first layers and just train the rest.



Can we do better?

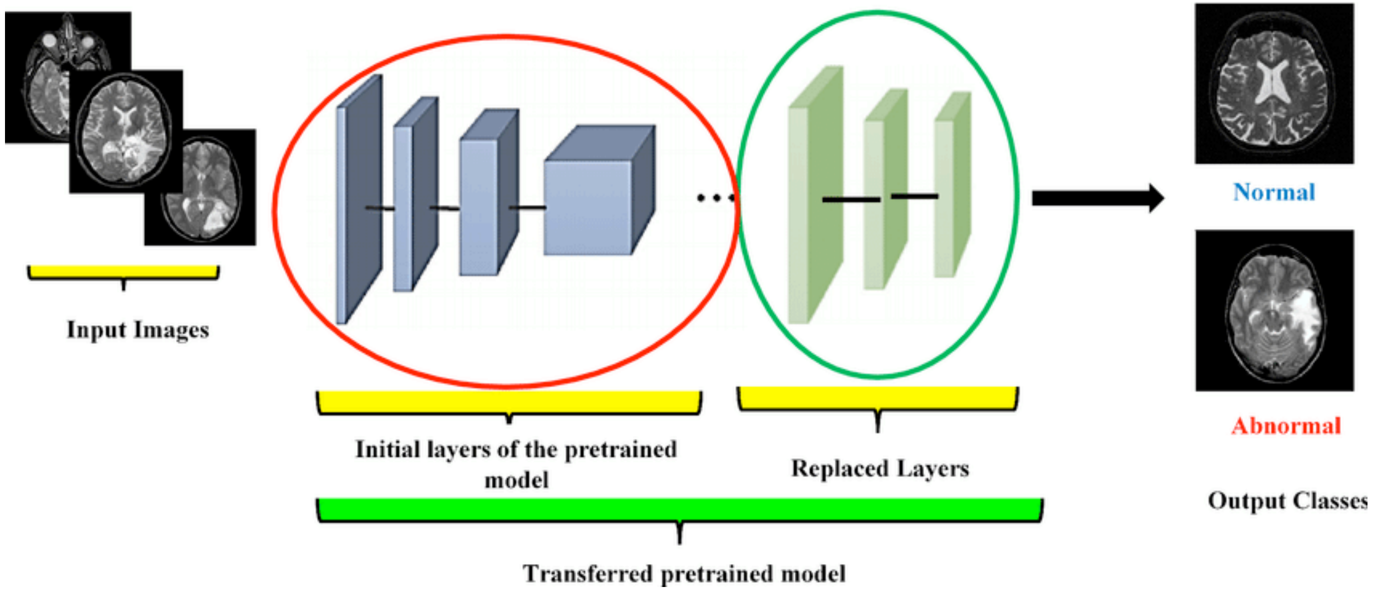
Develop Intelligence

- Researchers found that one can do many things to improve
- As layers “stack up” one can reuse a pretrained model on the first layers and just train the rest.
- Not only it speeds up, but makes it possible to do magic with little data



Can we do better?

- Researchers found that one can do many things to improve
- As layers “stack up” one can reuse a pretrained model on the first layers and just train the rest.
- Not only it speeds up, but makes it possible to do magic with little data



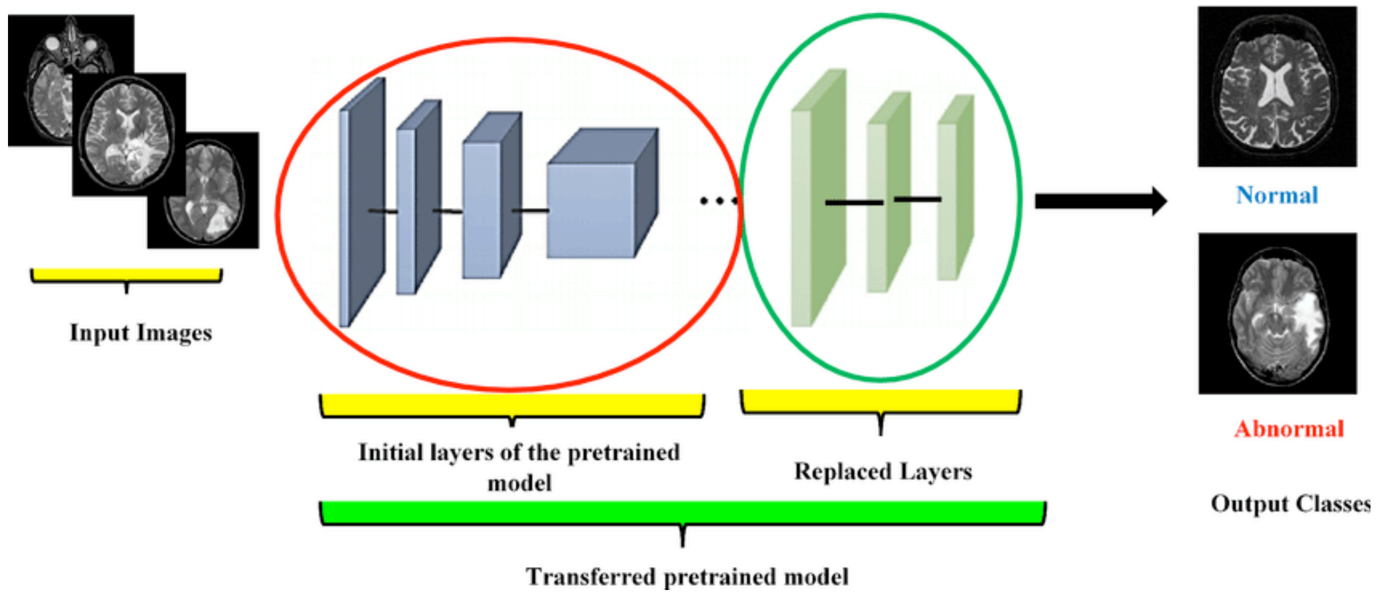
Can we do better?

- Researchers found that one can do many things to improve

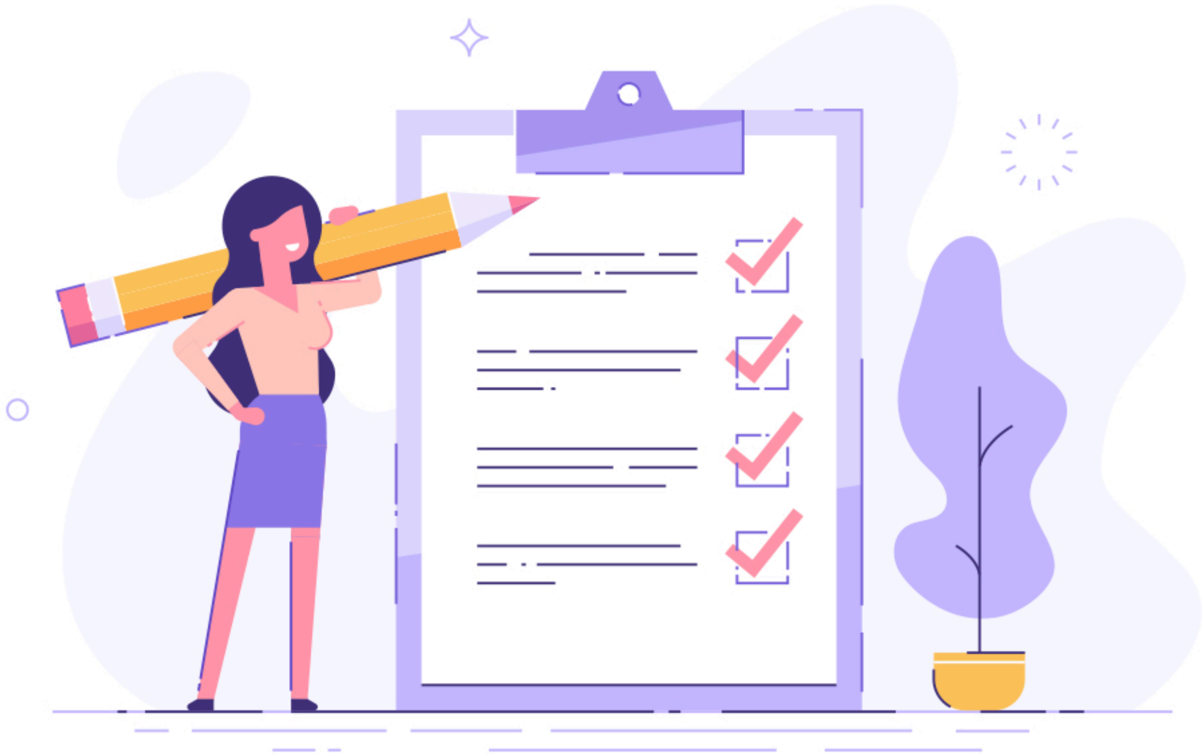
- As layers “stack up” one can reuse a pretrained model on the first layers and just train the rest.

- Not only it speeds up, but makes it possible to do magic with little data

- This is called fine-tuning

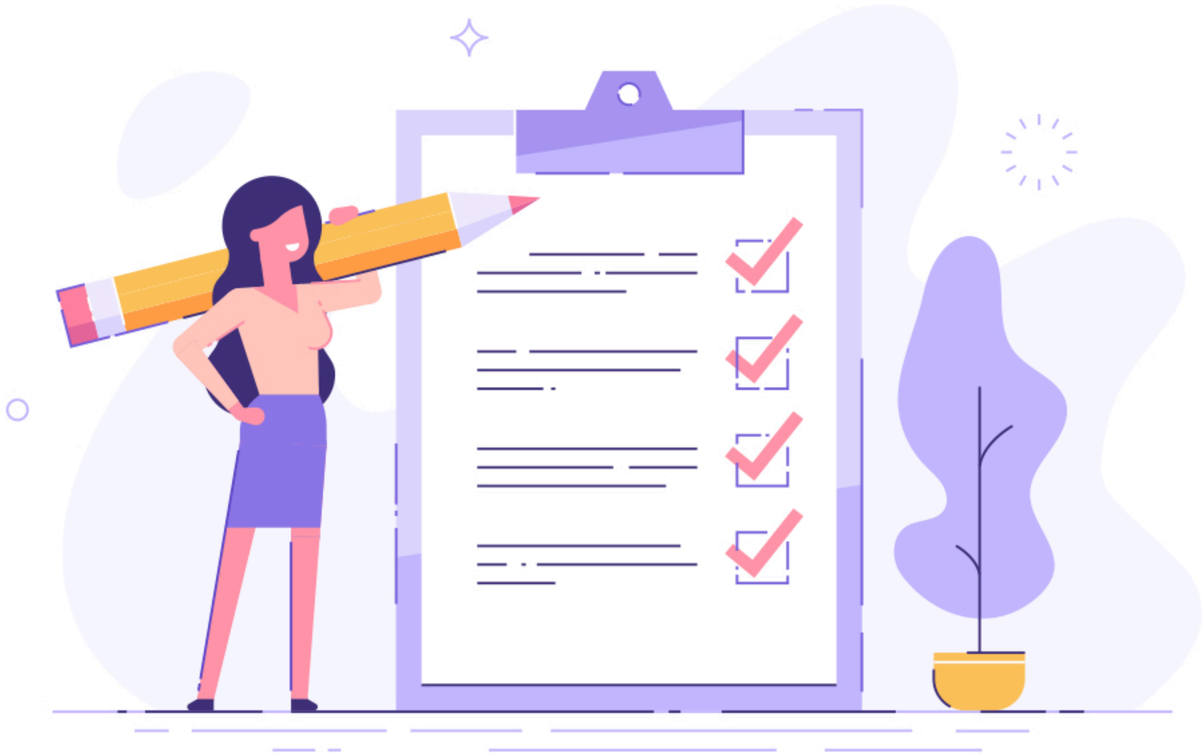


LAB: Using GloVe pre trained Embedding



LAB: Using GloVe pre trained Embedding

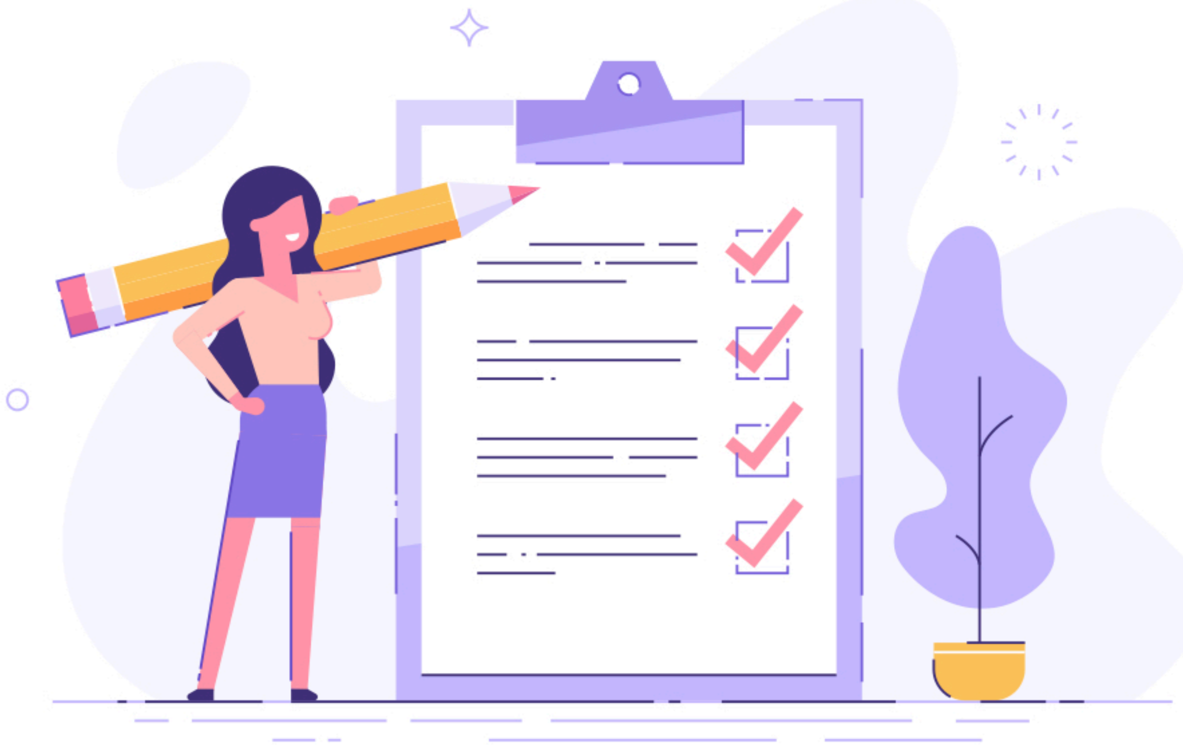
- Fine tune GloVe to our Airbnb dataset



LAB: Using GloVe pre trained Embedding

- Fine tune GloVe to our Airbnb dataset
- Verify how easy it is!

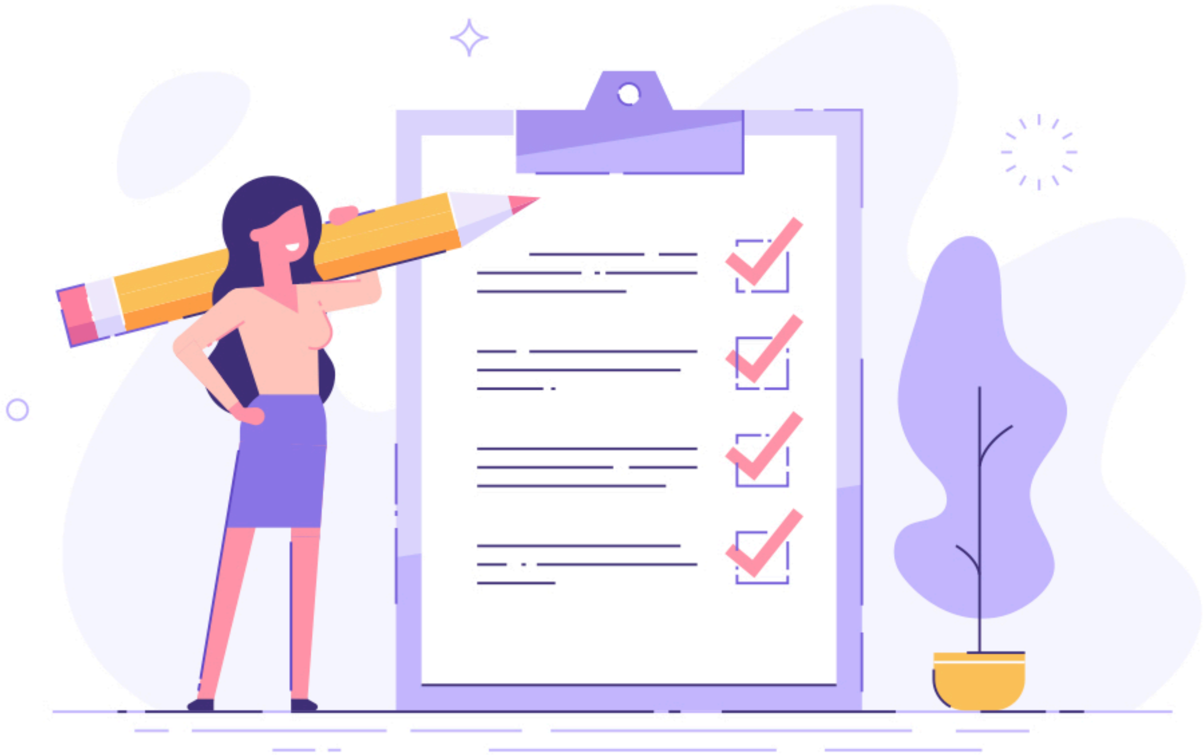
Allocated time: 20 minutes



LAB: Synonym Expansion

- Use our model to leverage synonyms in Solr

Allocated time: 15 minutes



Summary

Summary

- Synonym expansion helps recall by expanding the query with synonyms of the nouns

Summary

- Synonym expansion helps recall by expanding the query with synonyms of the nouns
- This way, even if the user doesn't input the correct term, it still gets the result

Summary

- Synonym expansion helps recall by expanding the query with synonyms of the nouns
- This way, even if the user doesn't input the correct term, it still gets the result
- Under the hood is a word embedding to find similar words on a vocabulary.

Summary

- Synonym expansion helps recall by expanding the query with synonyms of the nouns
- This way, even if the user doesn’t input the correct term, it still gets the result
- Under the hood is a word embedding to find similar words on a vocabulary.
- Known word embedding models are CBOW and Skip-Gram

Mini Break: 5 minutes

Develop
Intelligence

Ranking

Develop Intelligence

Doc2Vec

Agenda

Develop Intelligence

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

NER

- Scapy NER
- BERT

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

Text Generation

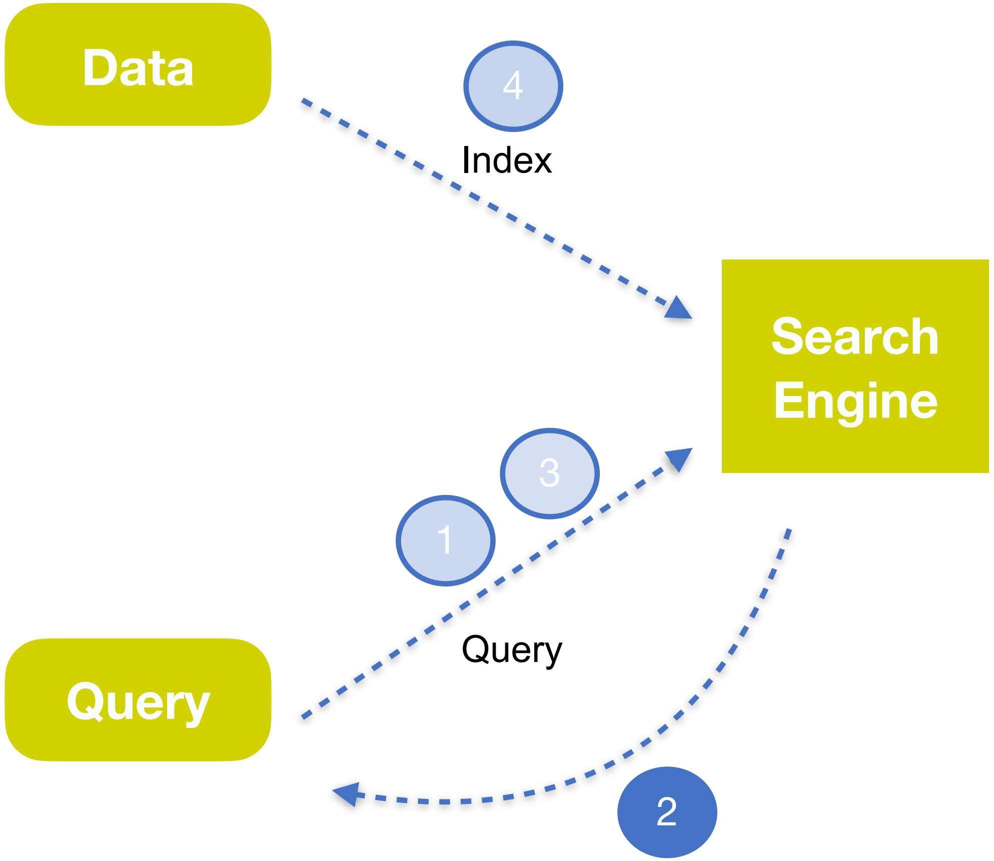
- RNN
- GRU and LSTM
- Transformers

Next Steps

- Closing thoughts

What are we going to do now?

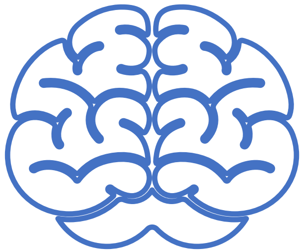
1. Synonym Expansion
2. Reranking results
3. Alternative Queries
4. NER



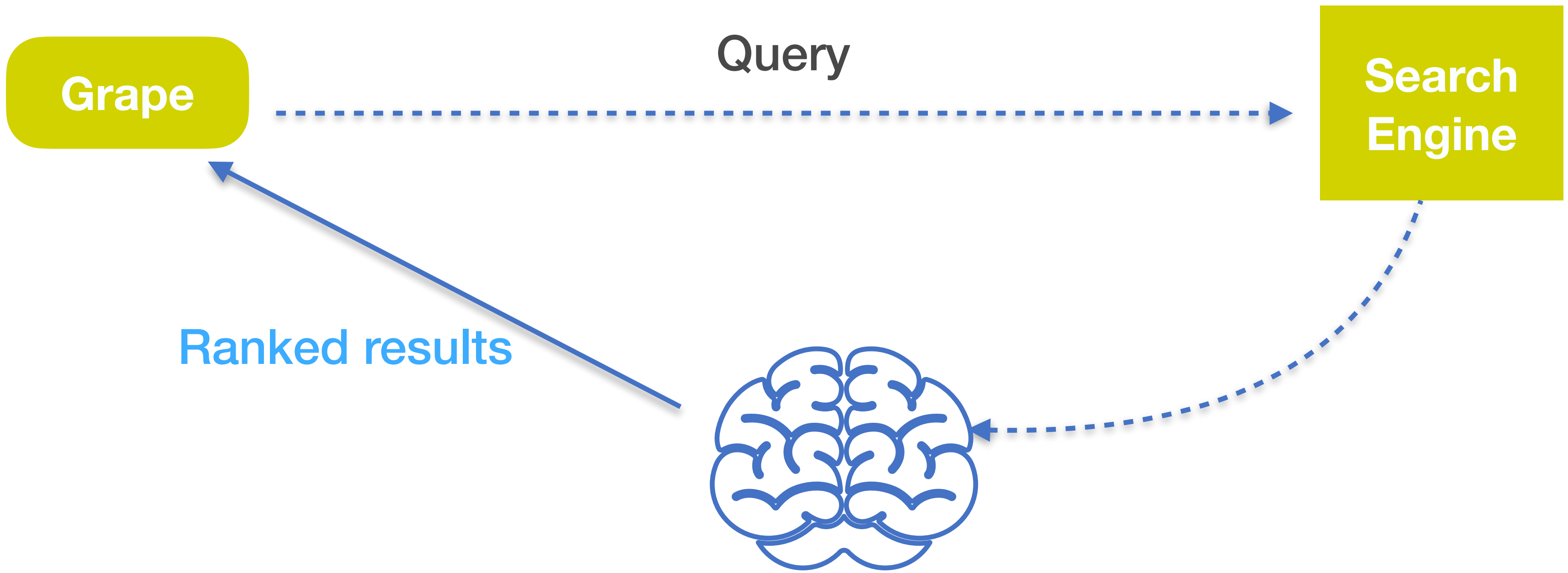
What are we going to do now?

Grape

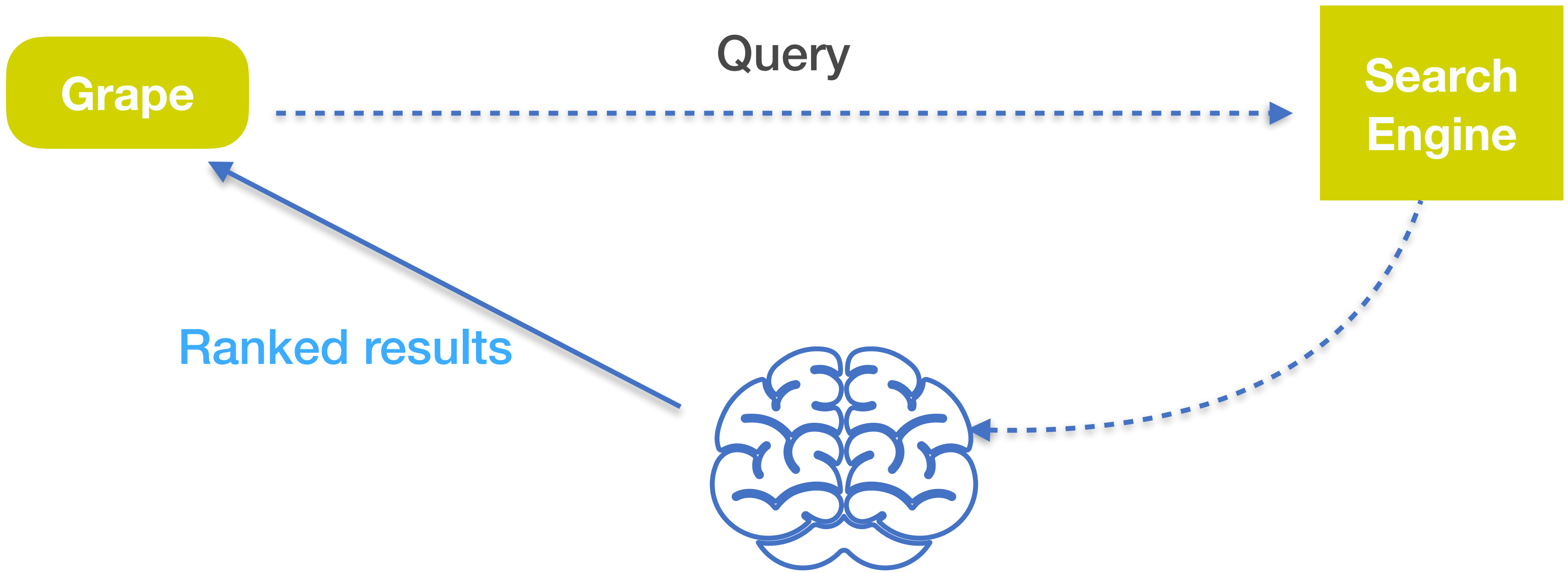
Search Engine



What are we going to do now?



What are we going to do now?



Reranking allows to get the best documents first related to the dataset and query

How can we rank using NLP?

Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

Which one do you think Lucene returns as first?

How can we rank using NLP?

Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

Which one do you think Lucene returns as first?

How can we rank using NLP?

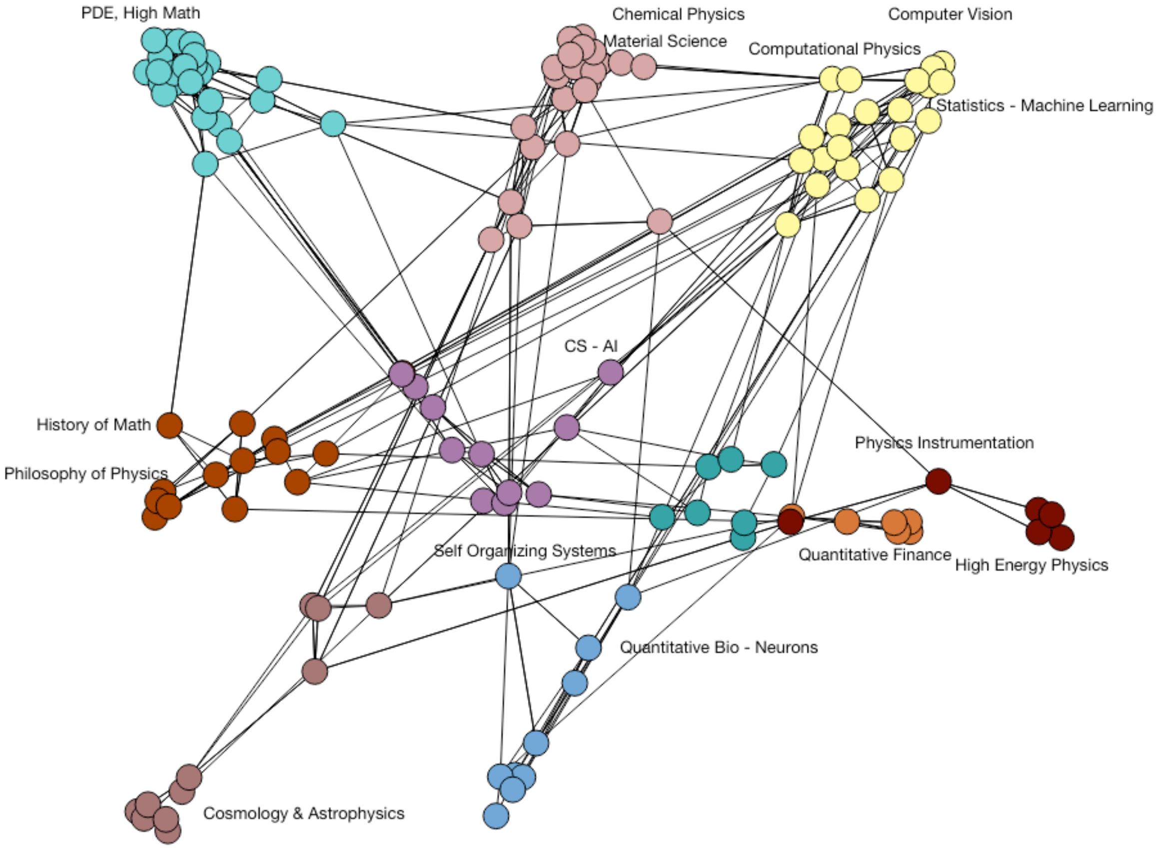
Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

Which one do you think Lucene returns as first?

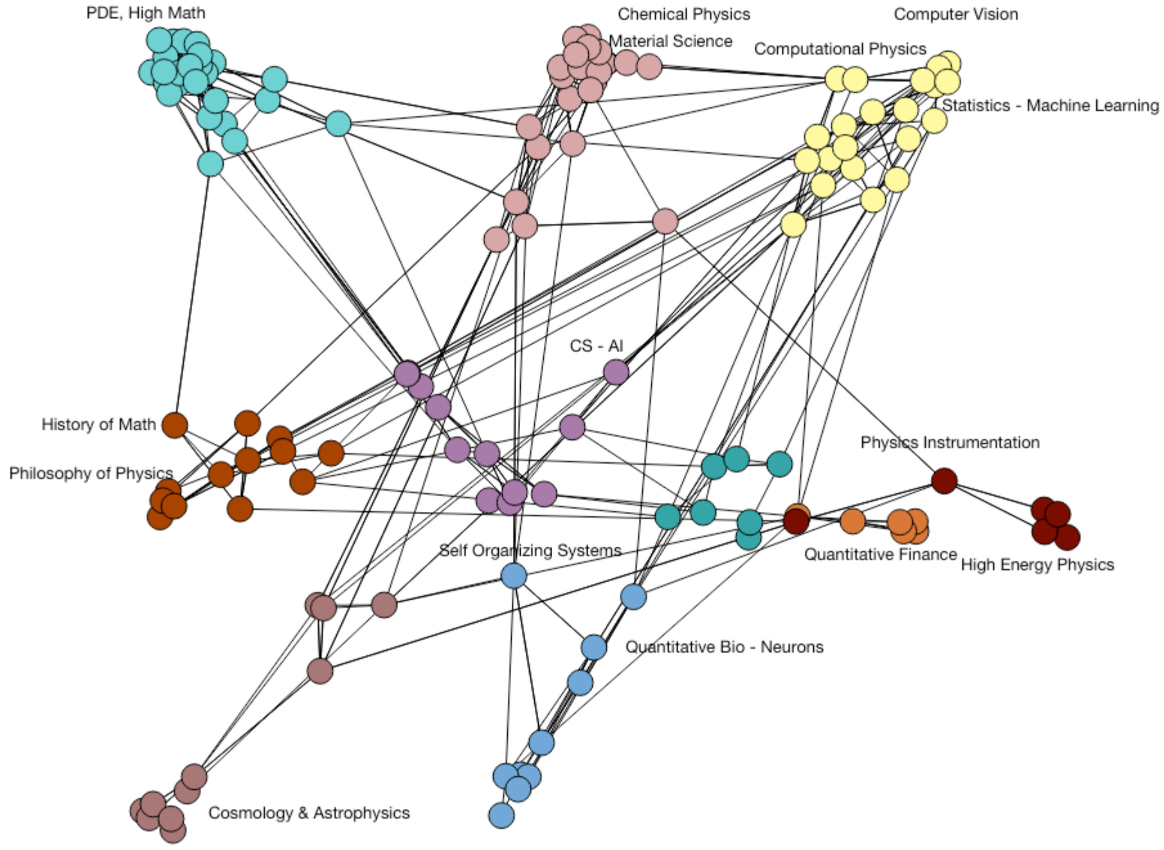
Introducing Doc2Vec



Introducing Doc2Vec

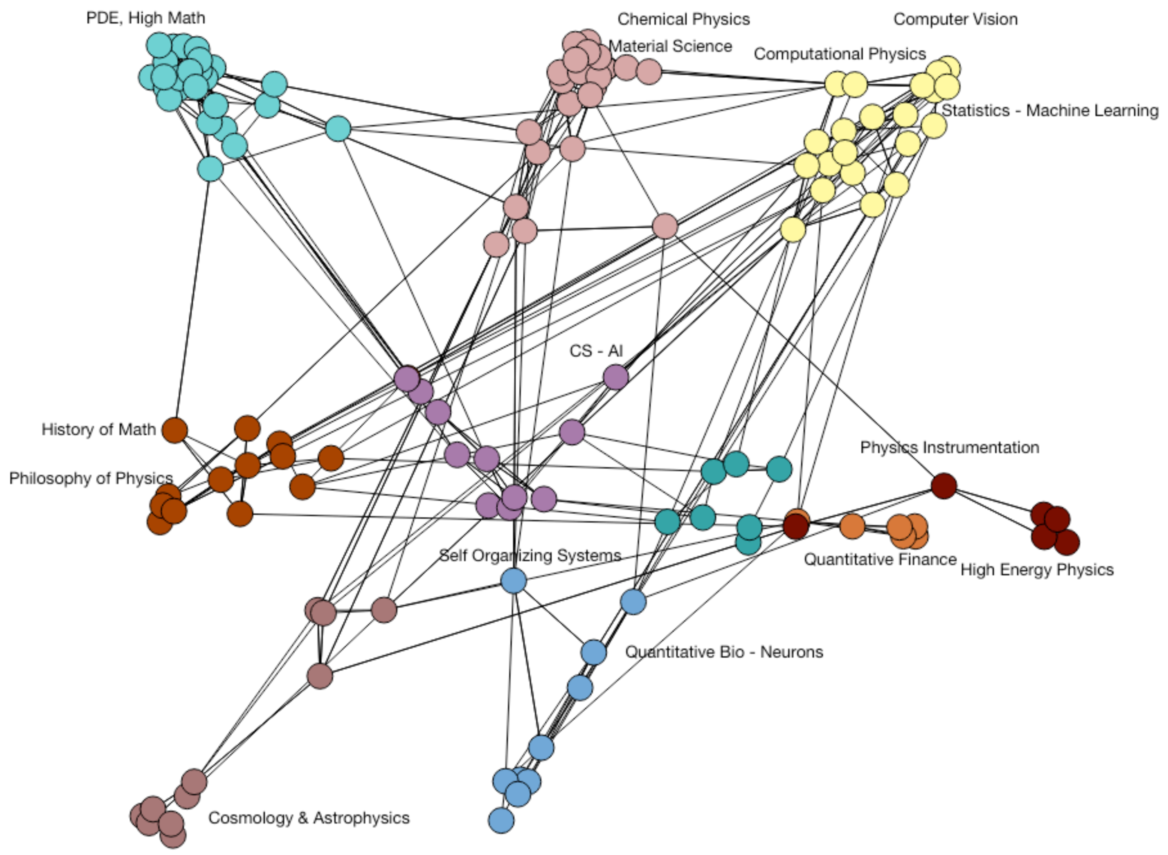
Develop Intelligence

- Similarly to before, it would be good to have a function that maps a **whole document** into numbers



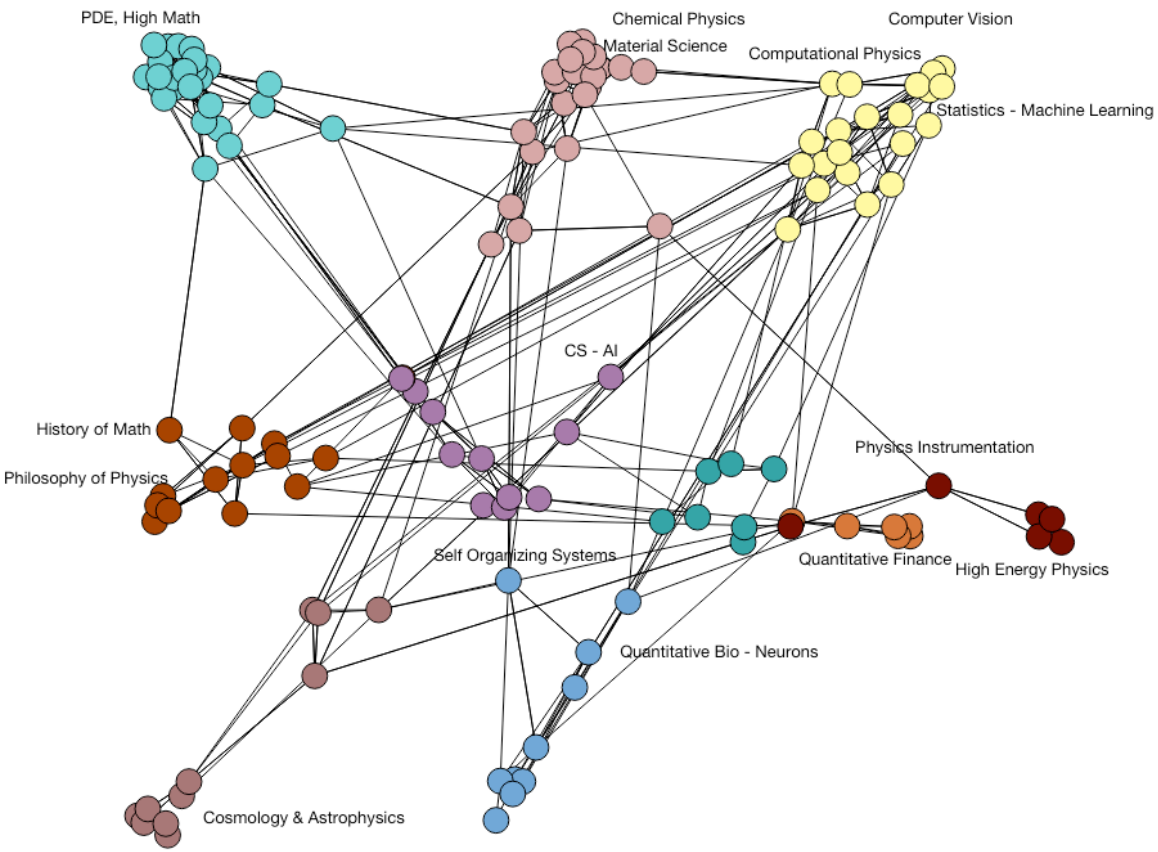
Introducing Doc2Vec

- Similarly to before, it would be good to have a function that maps a **whole document** into numbers
- This way we could find similarity between the query and each result **with respect to our specific dataset**

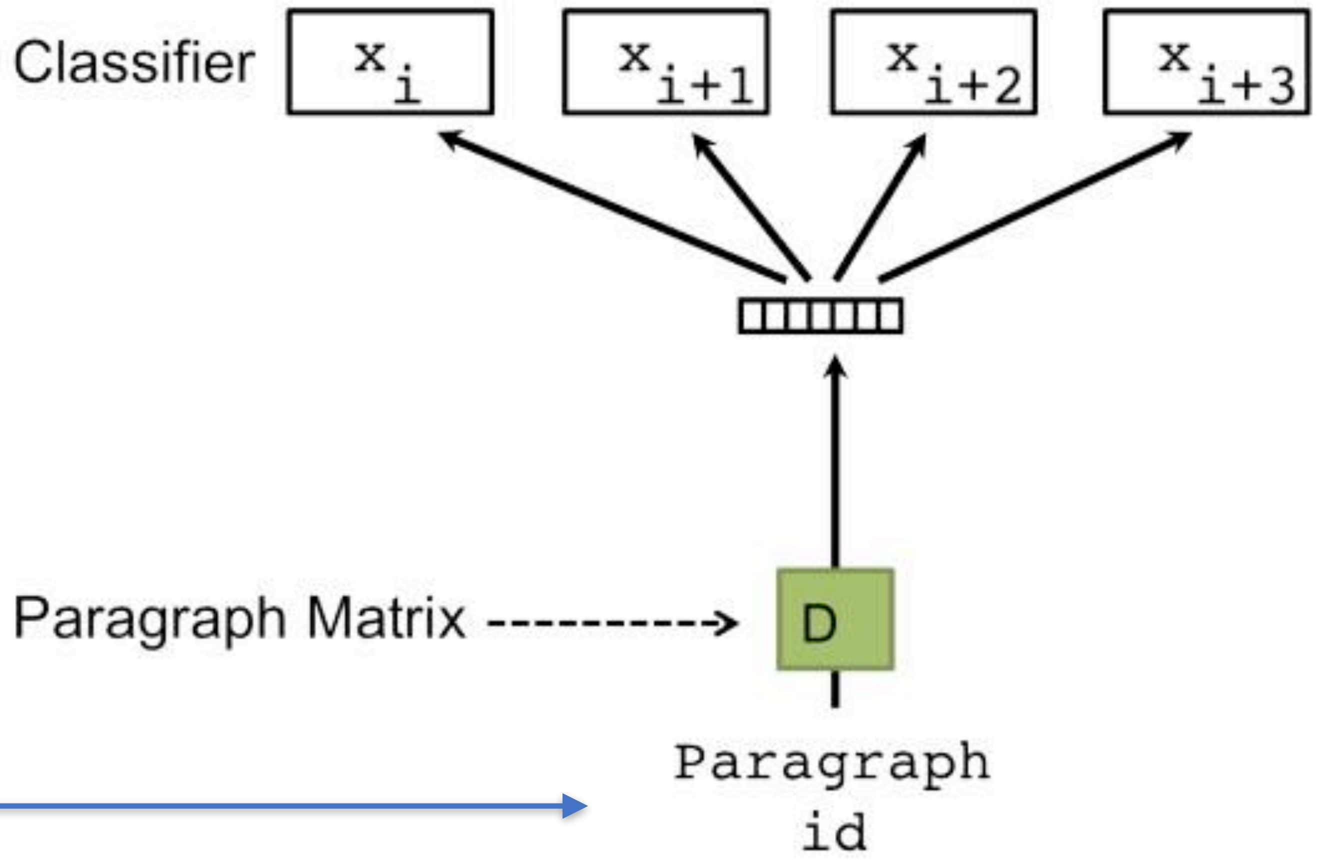


Introducing Doc2Vec

- Similarly to before, it would be good to have a function that maps a **whole document** into numbers
- This way we could find similarity between the query and each result **with respect to our specific dataset**
- This means that the same query and result would have different similarity on different datasets, which is what we want!

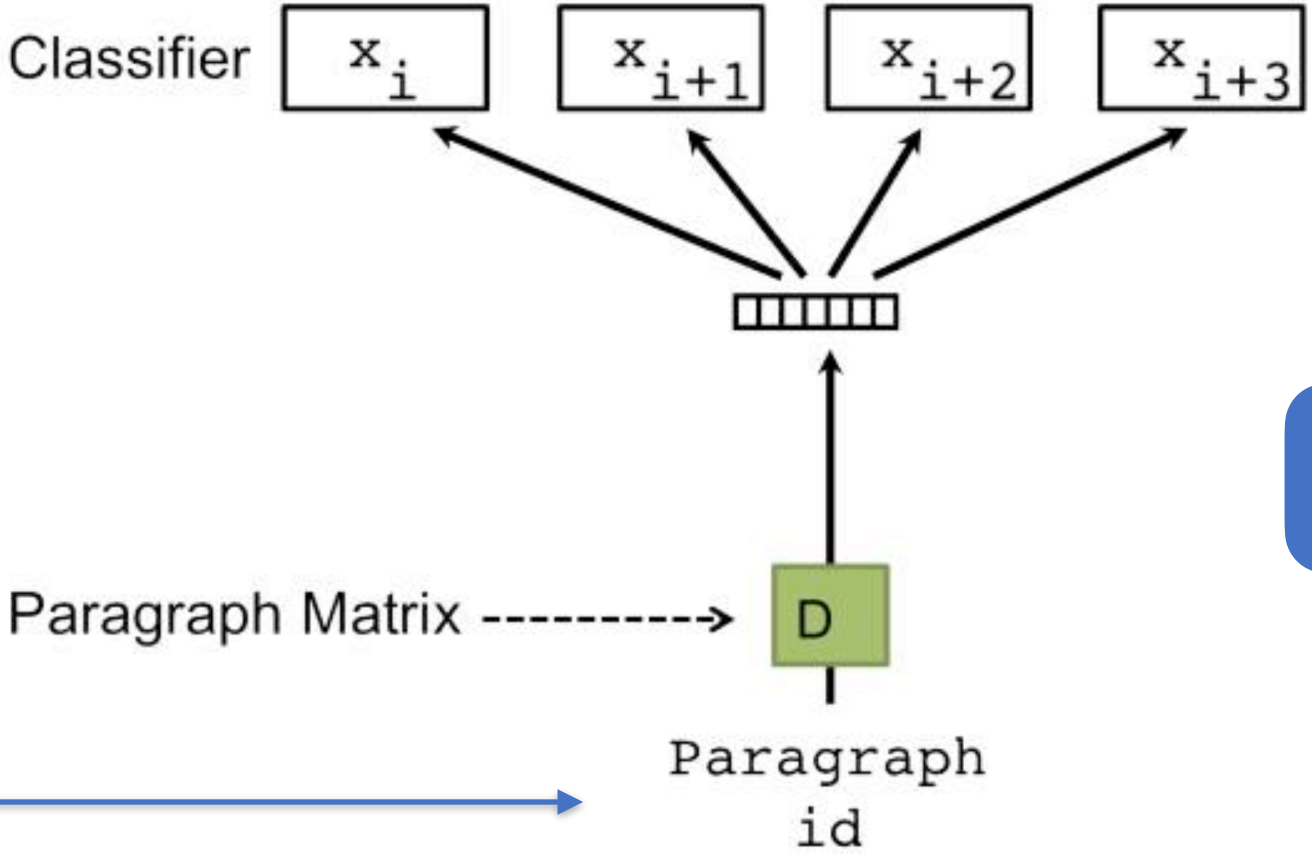


DBOW



Same as before, but assign an ID per document

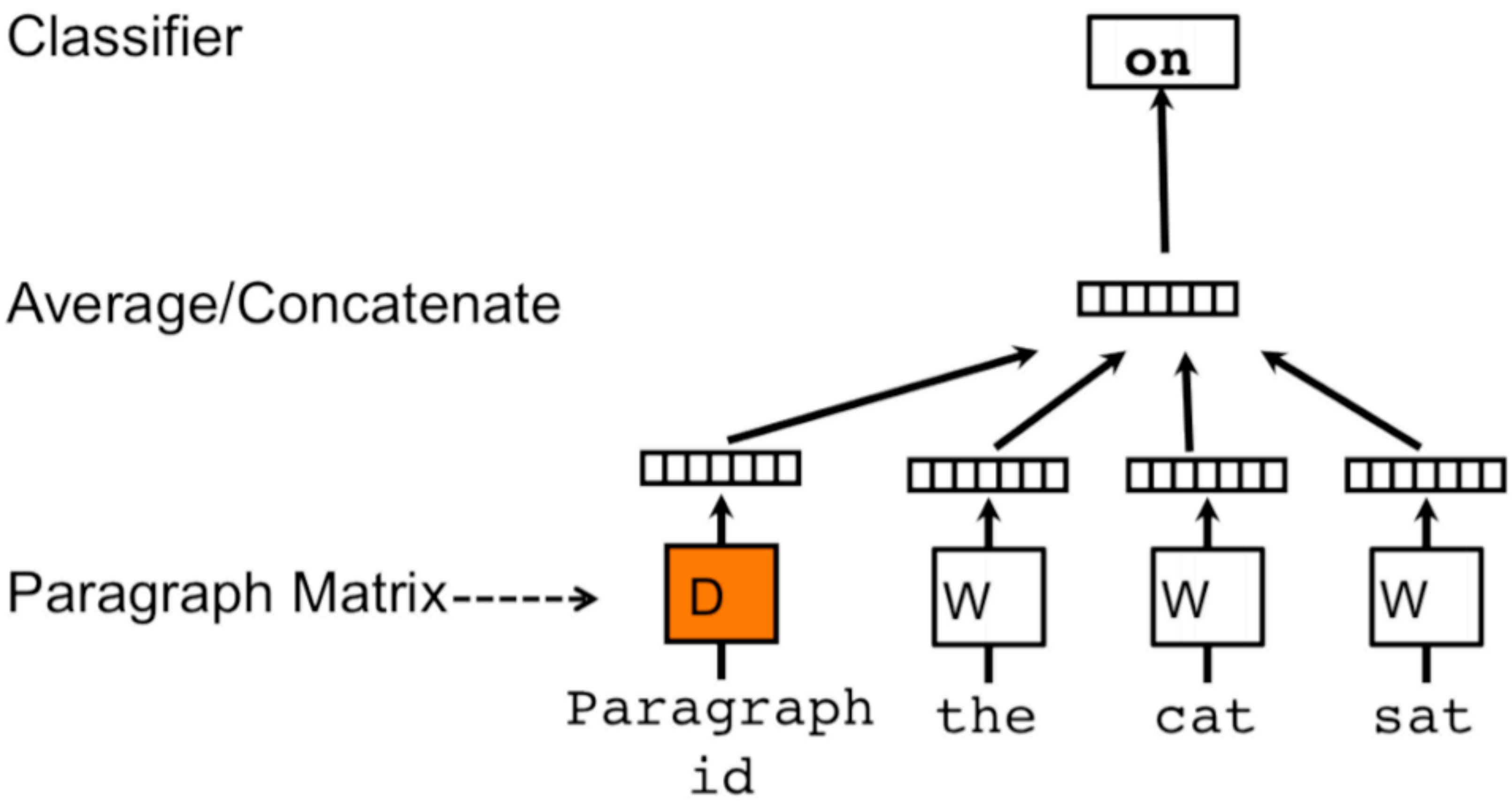
DBOW



Works surprisingly well!

Same as before, but assign an ID per document

PV-DM

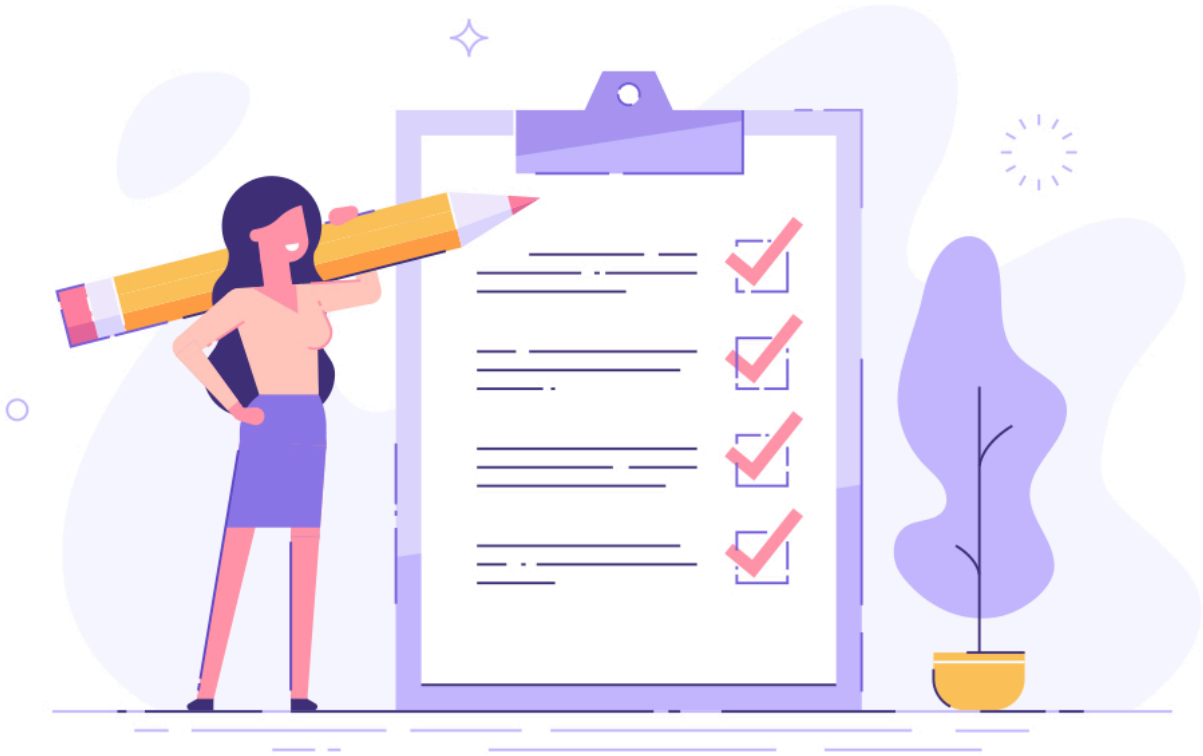


Classifier
Average/Concatenate
Paragraph Matrix----->
on
D
W
W
W
Paragraph id
the
cat
sat

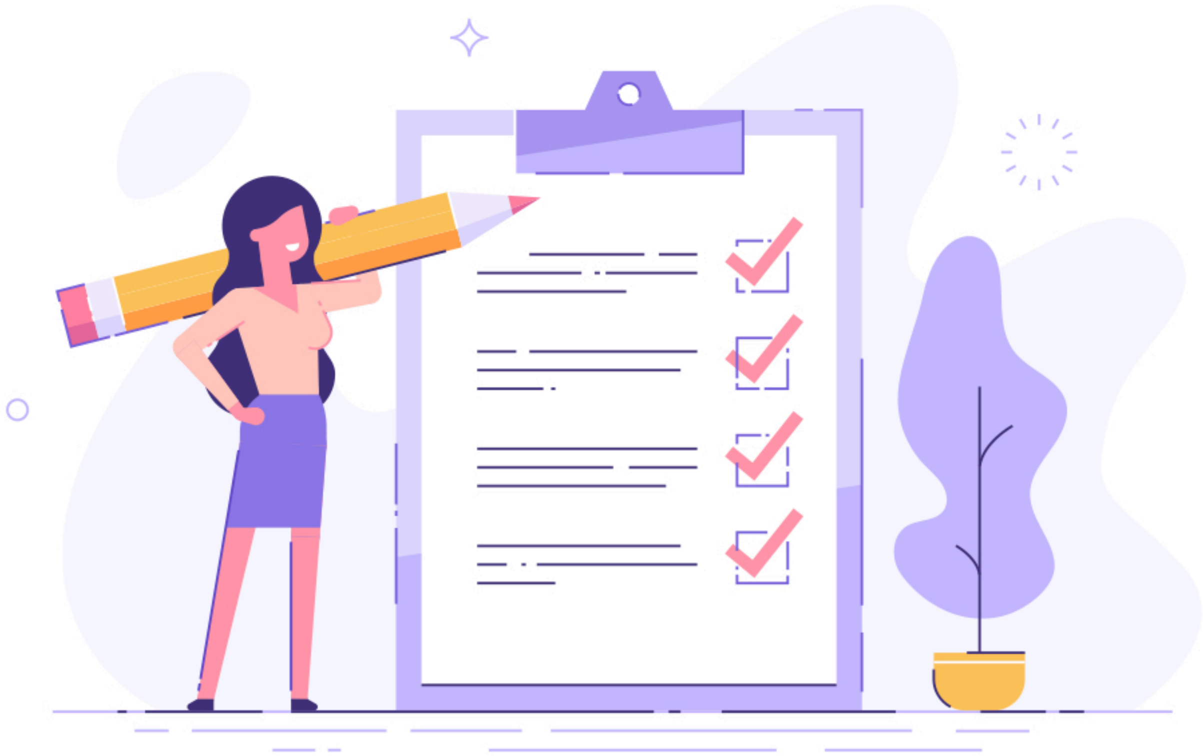
LAB: Train DBOW using GenSim

- Train a Doc2Vec model using GenSim

Allocated time: 20 minutes



Pulse Check



How can we rank using NLP?

Develop Intelligence

Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

How can we rank using NLP?

Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

A Good Doc2Vec algorithm would find the first 2 more similar to the query

How can we rank using NLP?

Develop Intelligence

Search Engine

2-bedroom apartment midtown wine

1. Beautiful condo close to wine shops in 5th avenue
2. 1 Bedroom apartment in the heart of Brooklyn close to boutique with ancient wines
3. Wine wine wine wine wine wine

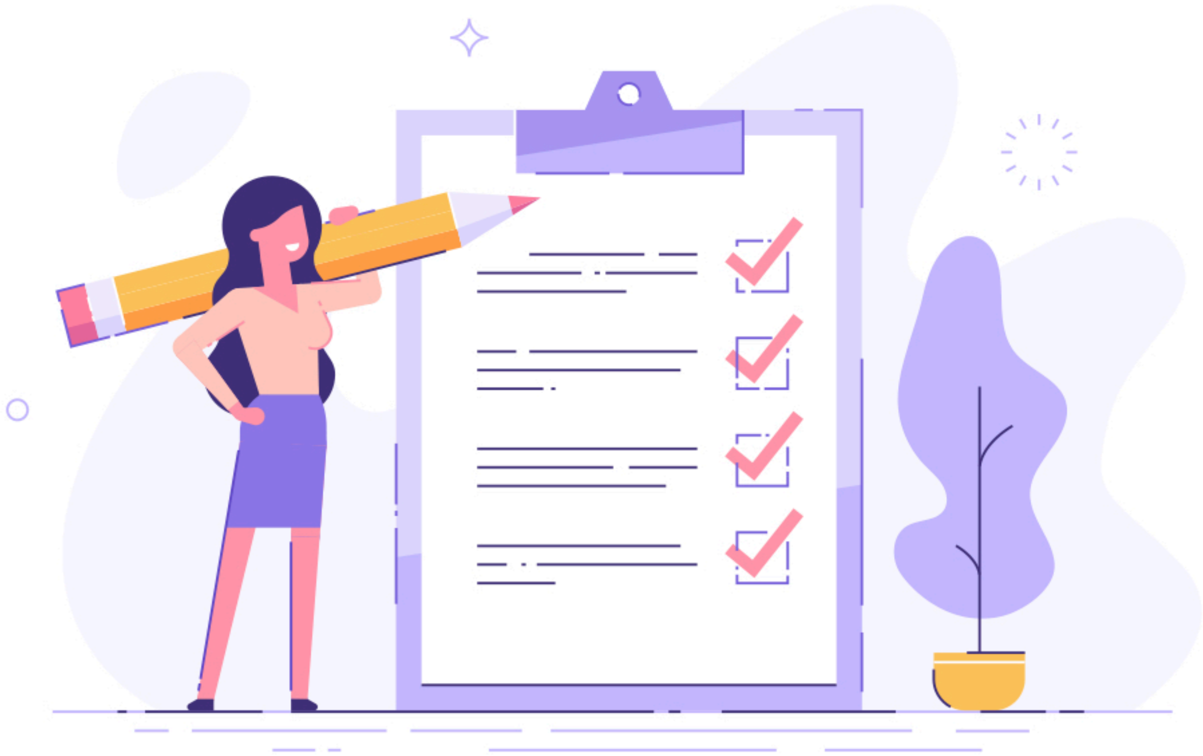
A Good Doc2Vec algorithm would find the first 2 more similar to the query

We reorder based on similarity!!

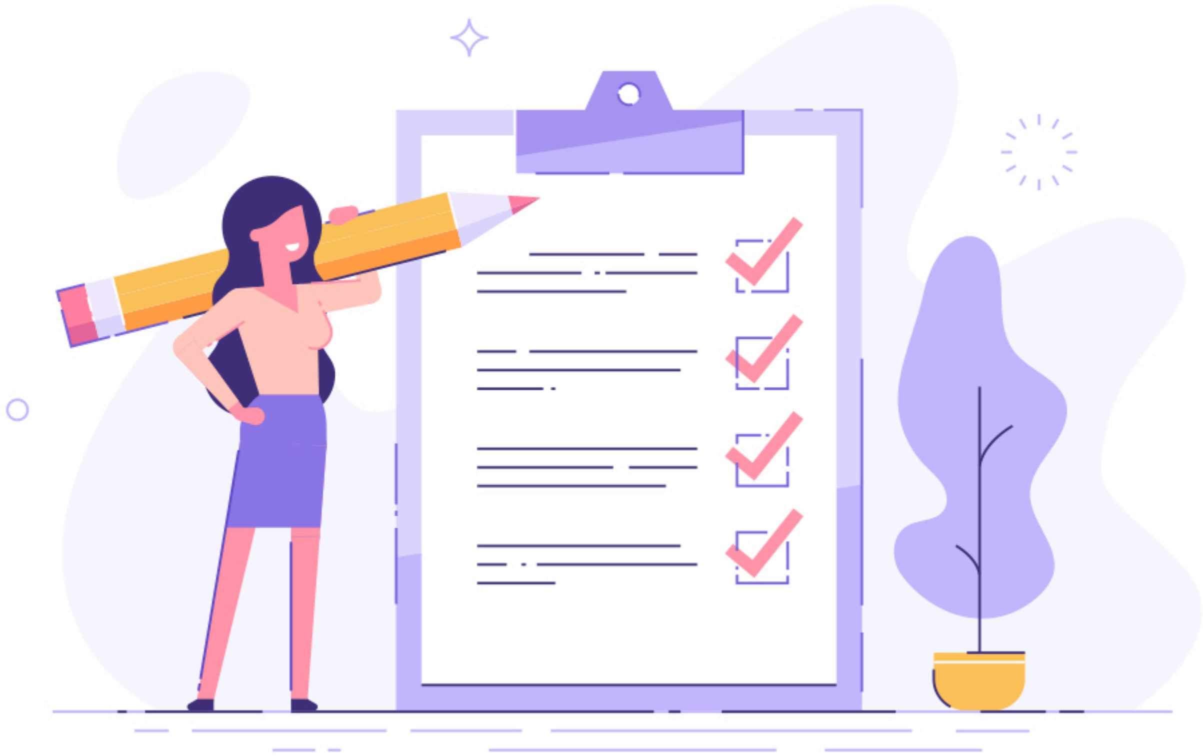
LAB: Reranking

- Create a reranking algorithm in Solr

Allocated time: 15 minutes

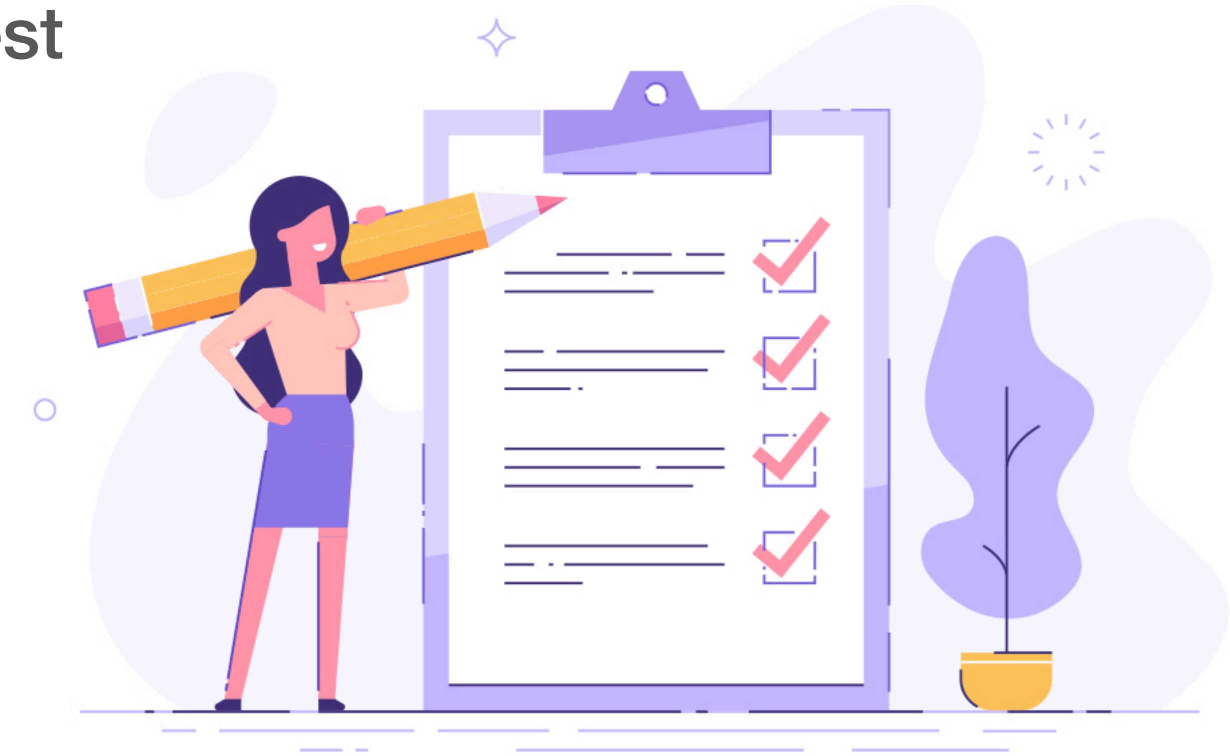


Discussion Time



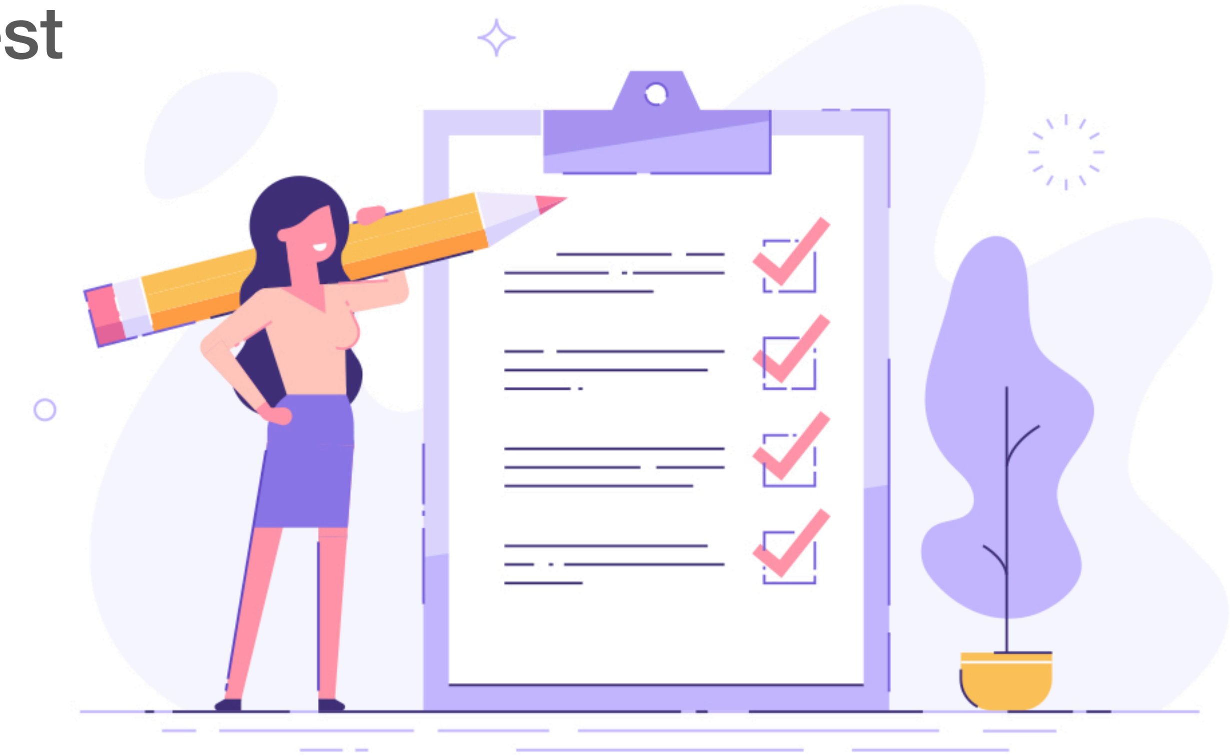
Discussion Time

Which strategy do you think is best to “rank by relevance” the documents returned by a query?



Discussion Time

Which strategy do you think is best to “rank by relevance” the documents returned by a query?



Discussion Time

Which strategy do you think is best to “rank by relevance” the documents returned by a query?

Allocated time: 15 minutes

Summary

Summary

- A way to rank documents is using doc2vec models

Summary

- A way to rank documents is using doc2vec models
- This way we get more similar results to what we wanted

Summary

- A way to rank documents is using doc2vec models
- This way we get more similar results to what we wanted
- However this is the start, after this step we could add an LTR model

Summary

- A way to rank documents is using doc2vec models
- This way we get more similar results to what we wanted
- However this is the start, after this step we could add an LTR model
- Known doc2vec models are DBOW and PV-DM

End of Day 1

Develop
Intelligence

Text Generation

RNN

Agenda

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

NER

- Spacy

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

Text Generation

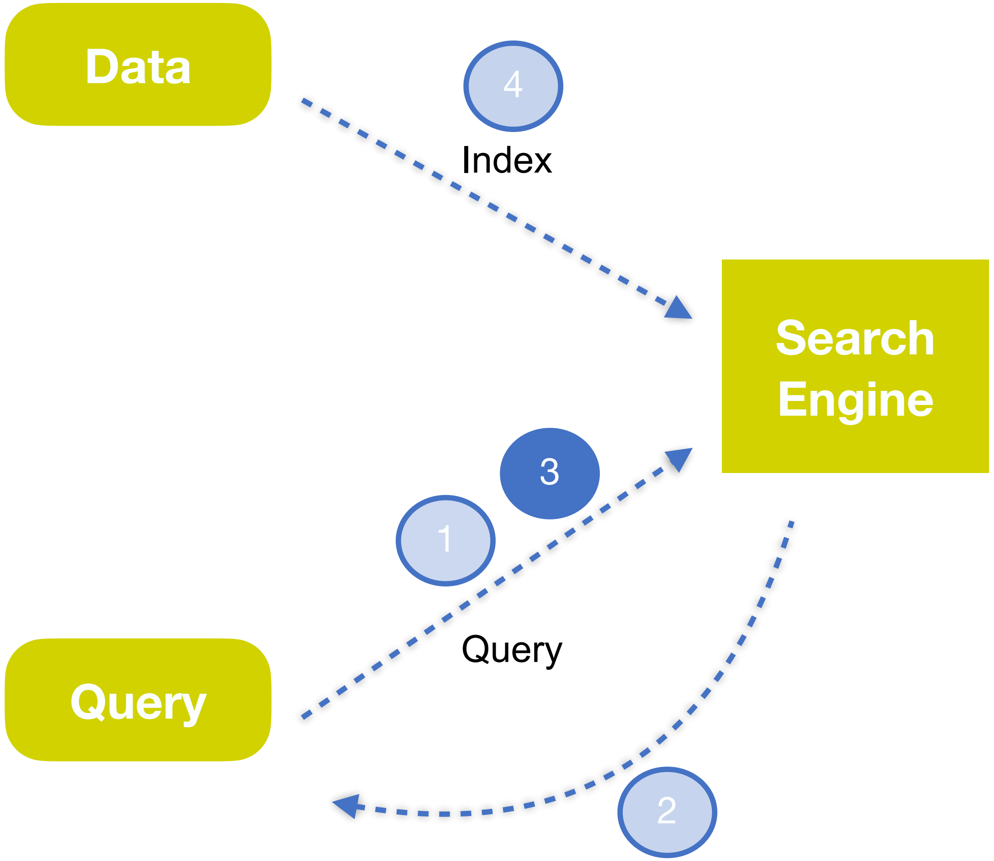
- RNN
- GRU and LSTM
- Transformers

Next Steps

- Closing thoughts

What are we going to do now?

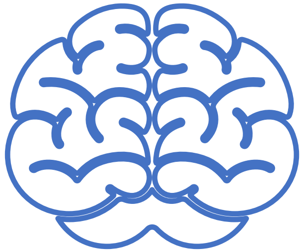
1. Synonym Expansion
2. Reranking results
3. Alternative Queries
4. NER



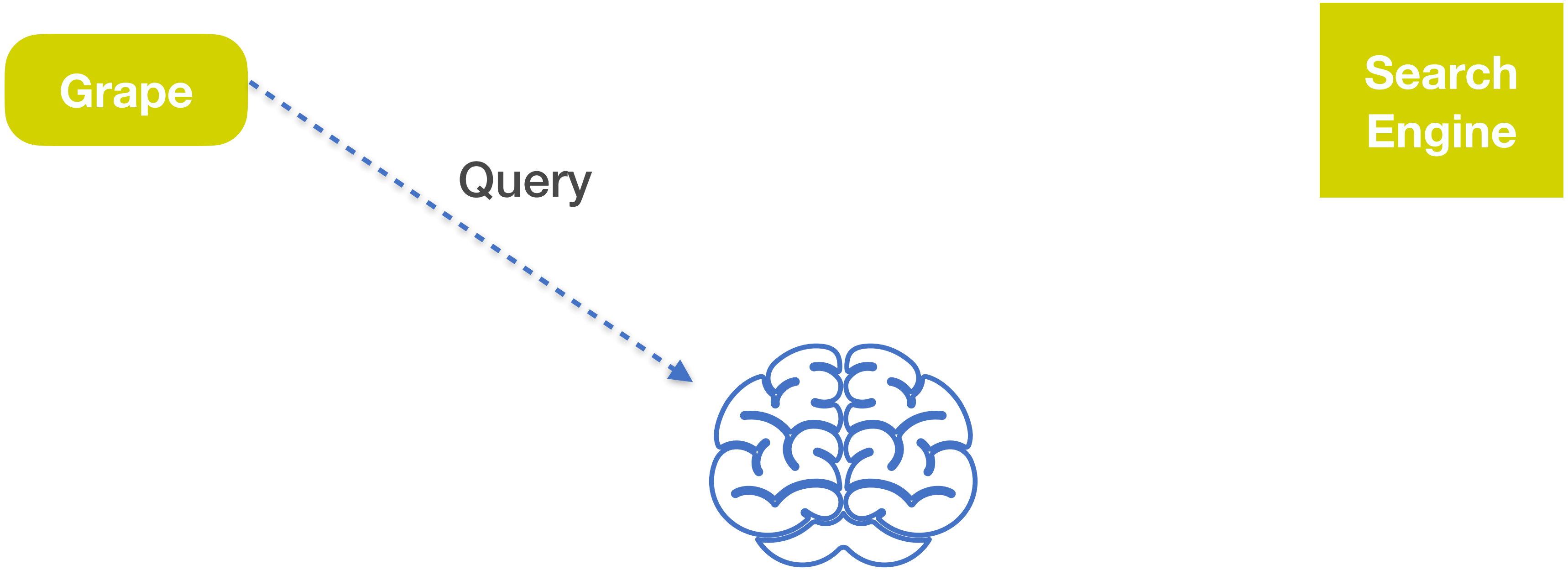
What are we going to do now?

Grape

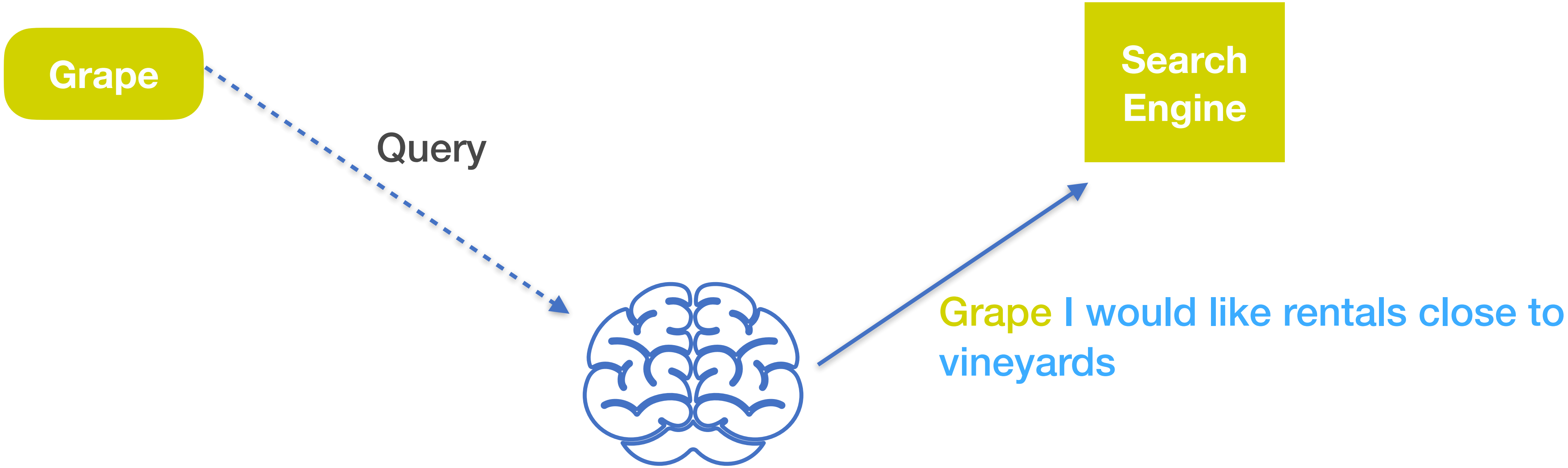
Search Engine



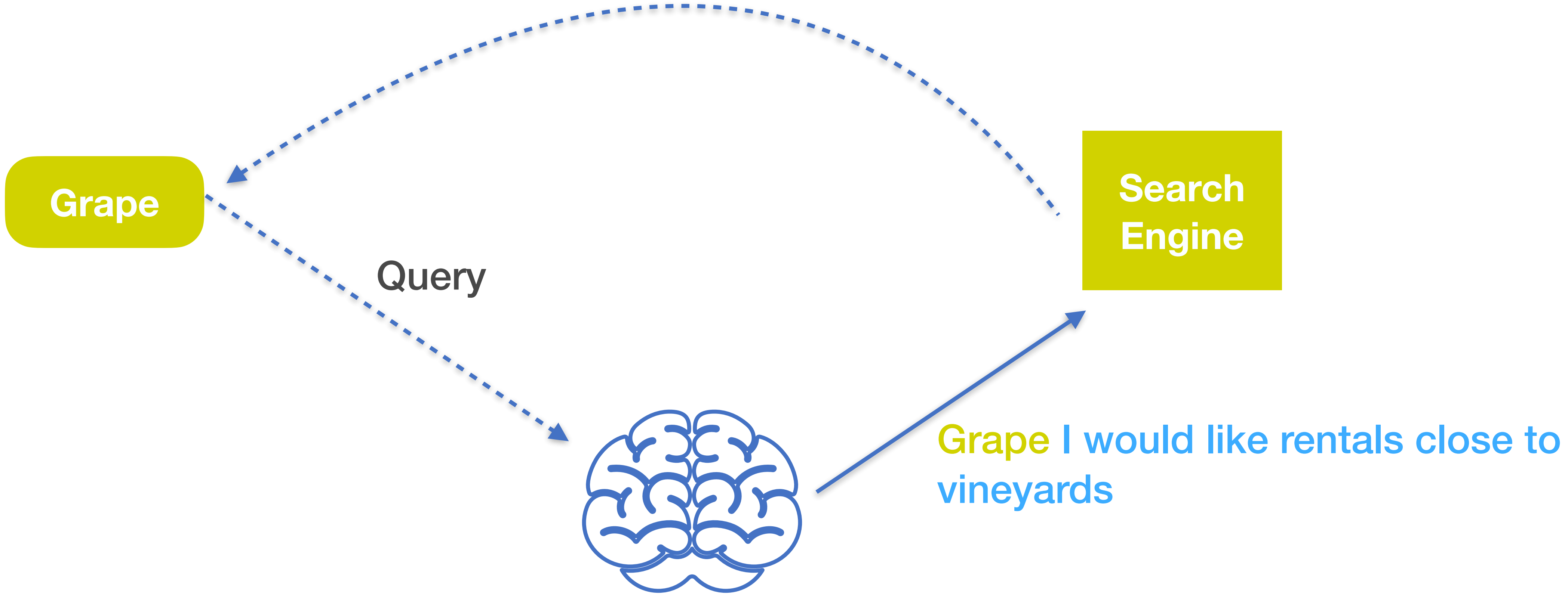
What are we going to do now?



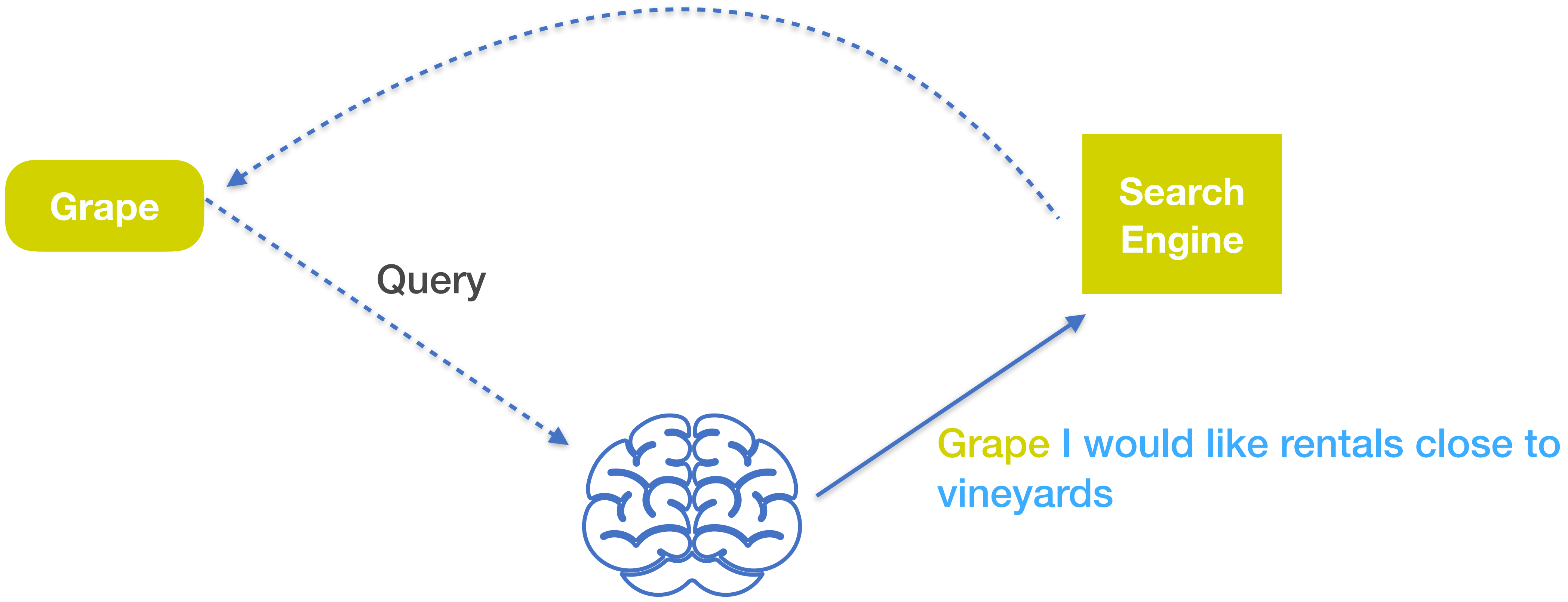
What are we going to do now?



What are we going to do now?



What are we going to do now?

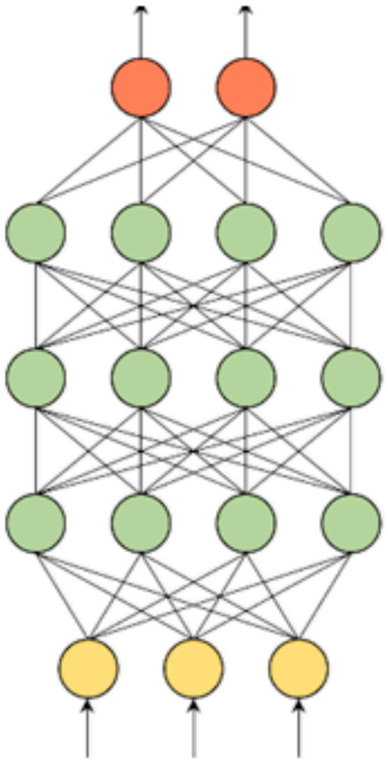


Alternative Queries simplifies the task for the user, the whole phrase is created by the NN

How do we generate text?

Probabilities over char set

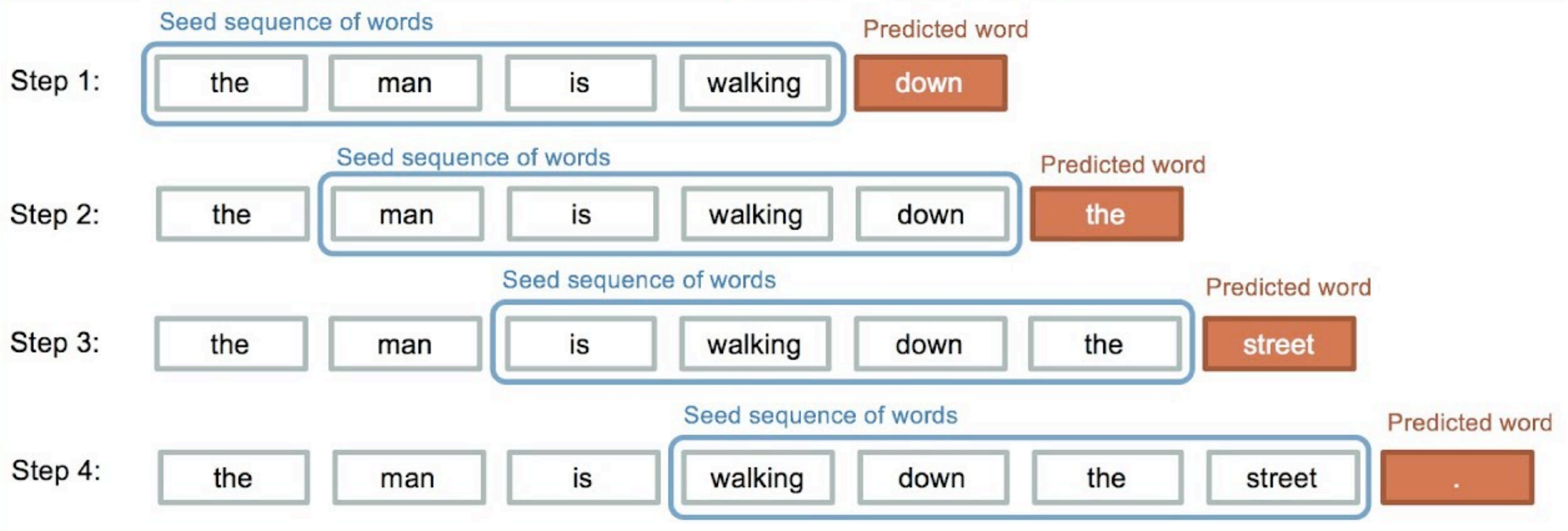
	a	b	c	d	e	f	g	...	z
0.01	0.02	0.36	0.25	0.02	0.001	0.22	0.001	...	0.06



Train Input from Corpus

I		w	a	n	t		t	o		b	a	k	e	

How do we generate text?



Problems

Problems

- If we used a normal NN, we would need a window of fixed size to predict the next character/word

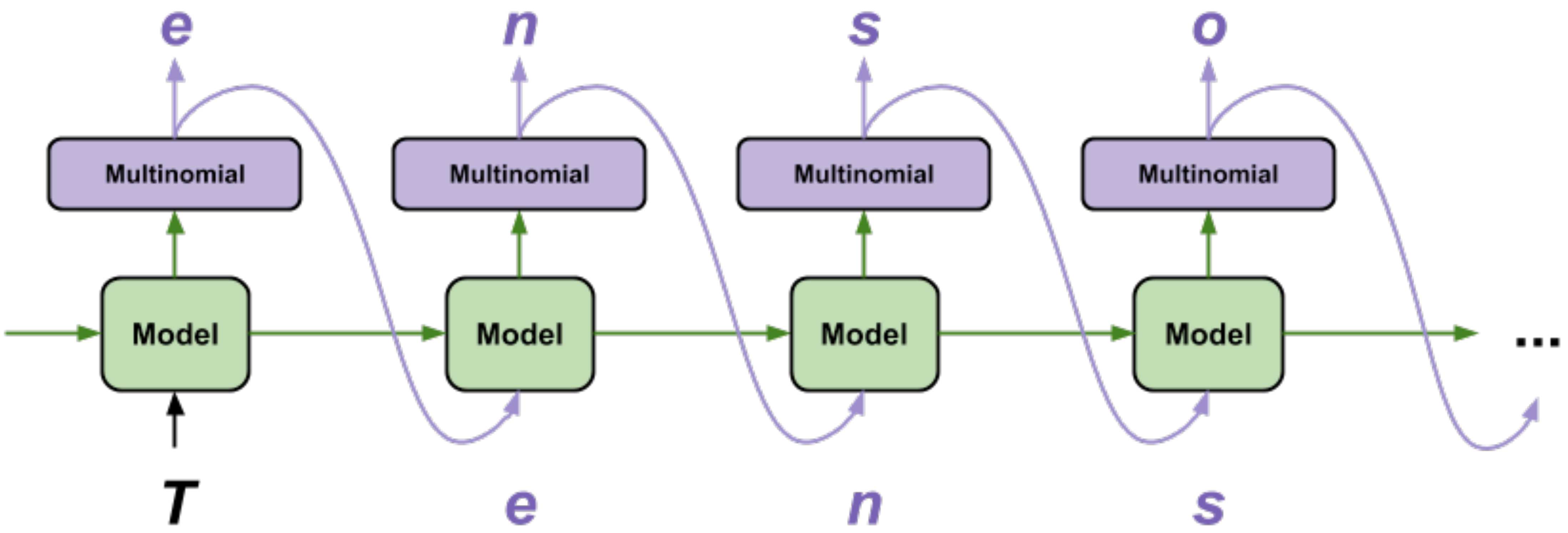
Problems

- If we used a normal NN, we would need a window of fixed size to predict the next character/word
- But in text, sometimes the core meaning comes at the end, however long it is the sequence.
  - *“Hospitals are sued by 7 foot doctors”*
  - *Local high school dropouts cut in half*

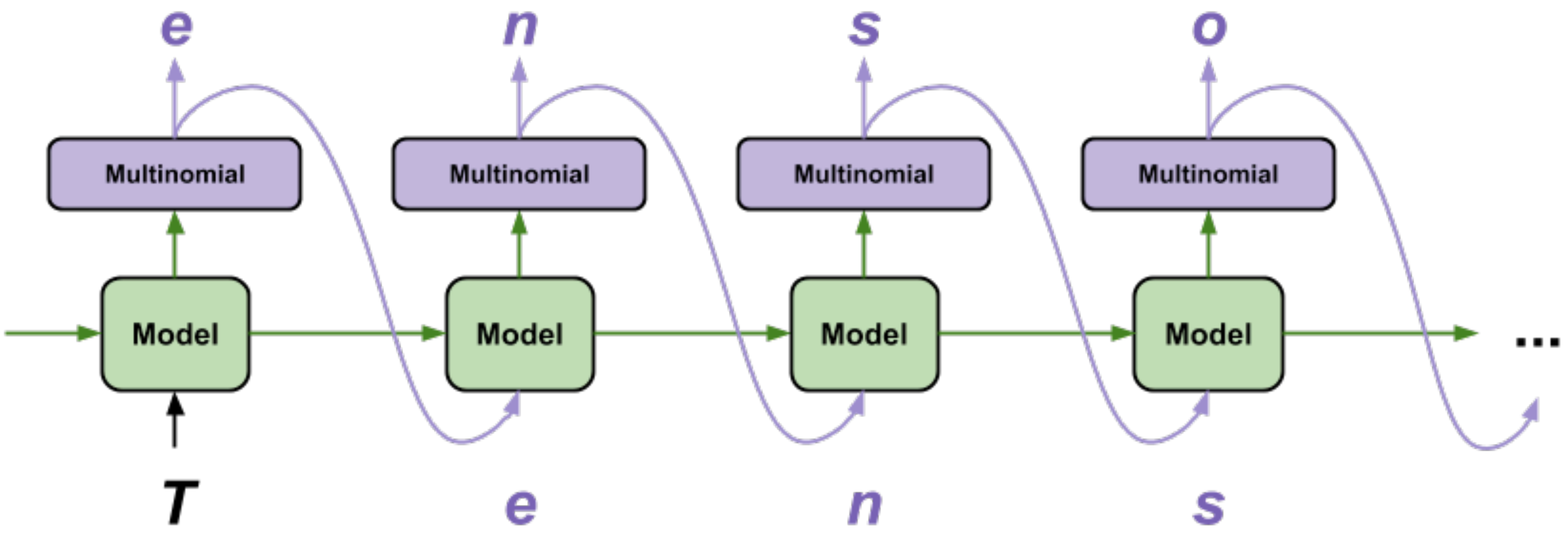
Problems

- If we used a normal NN, we would need a window of fixed size to predict the next character/word
- But in text, sometimes the core meaning comes at the end, however long it is the sequence.
  - *“Hospitals are sued by 7 foot doctors”*
  - *Local high school dropouts cut in half*
- **These problems are and for normal NN because they cannot remember!**

Character level text generation “passing memory”

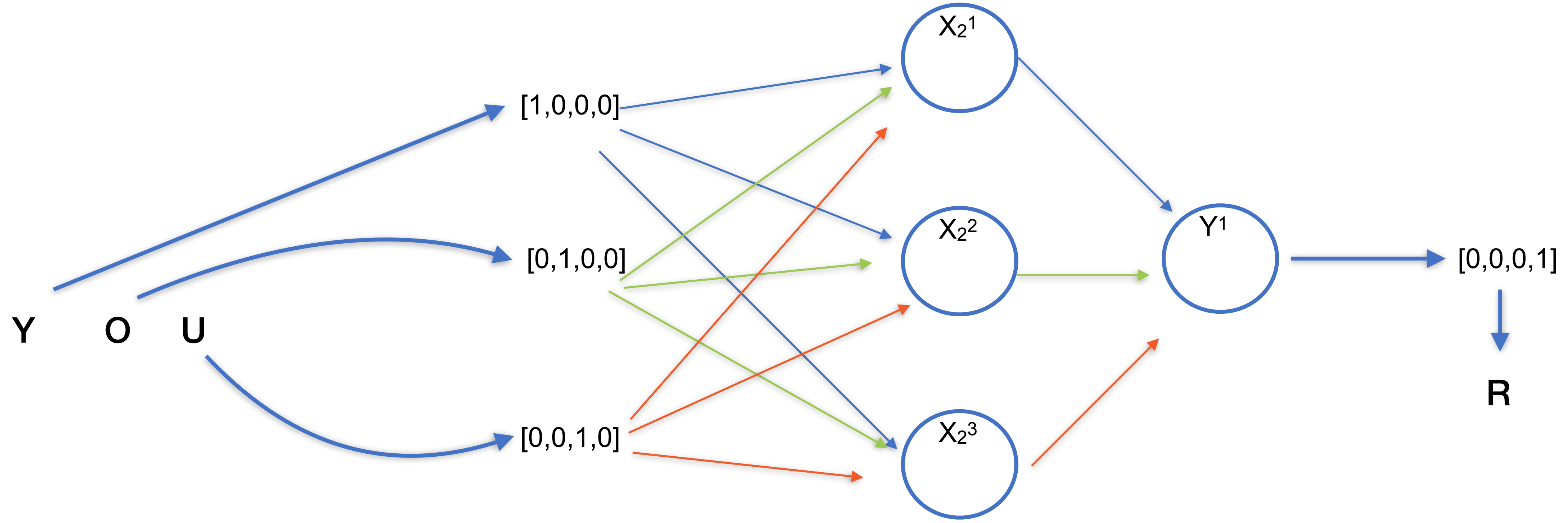


Character level text generation "passing memory"

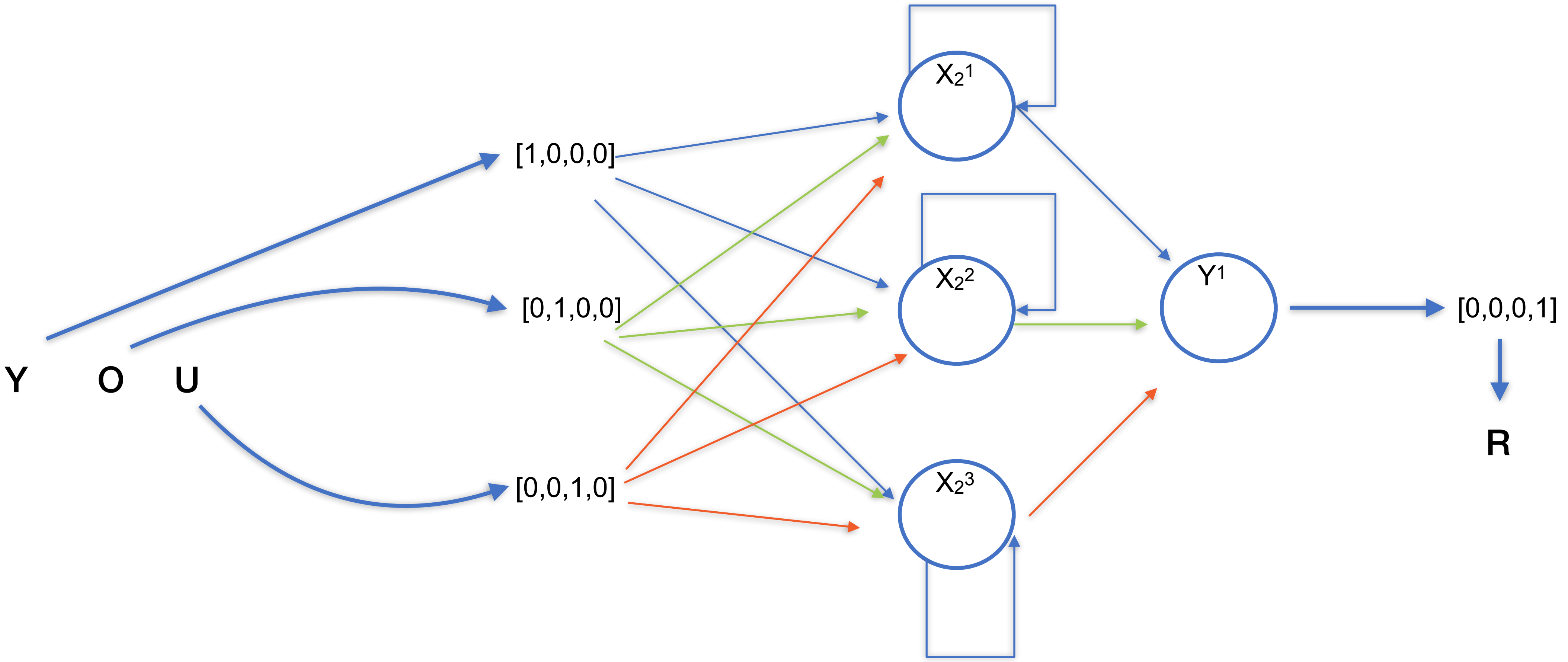


The "key" is to pass the state as input to remember

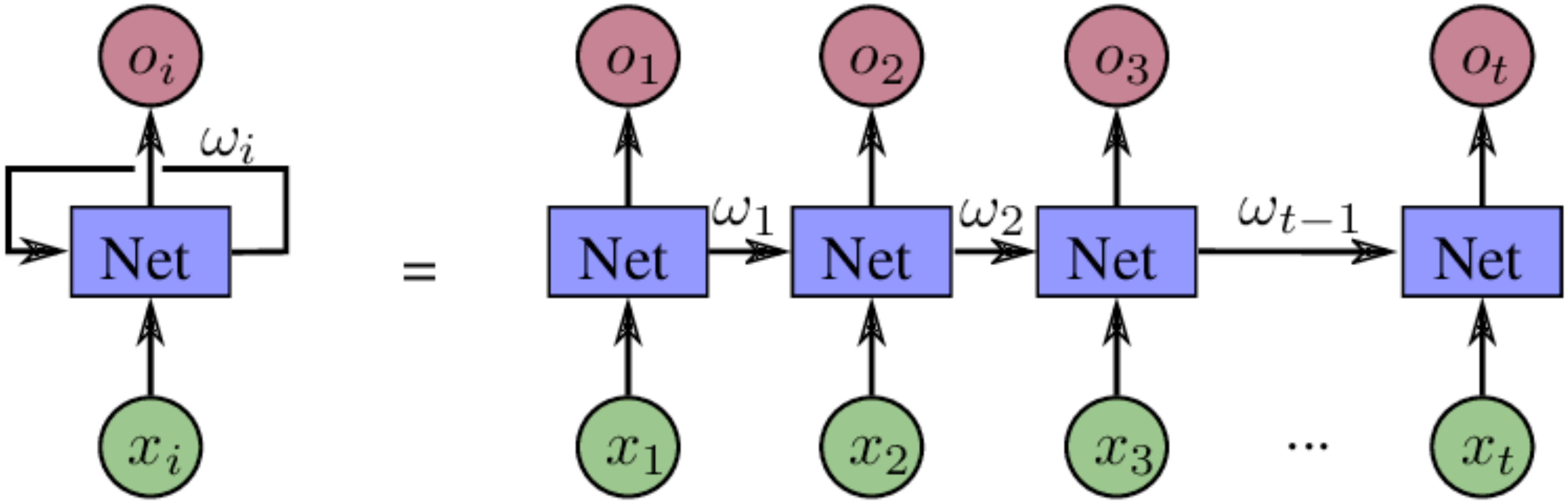
Recursive Neural Networks



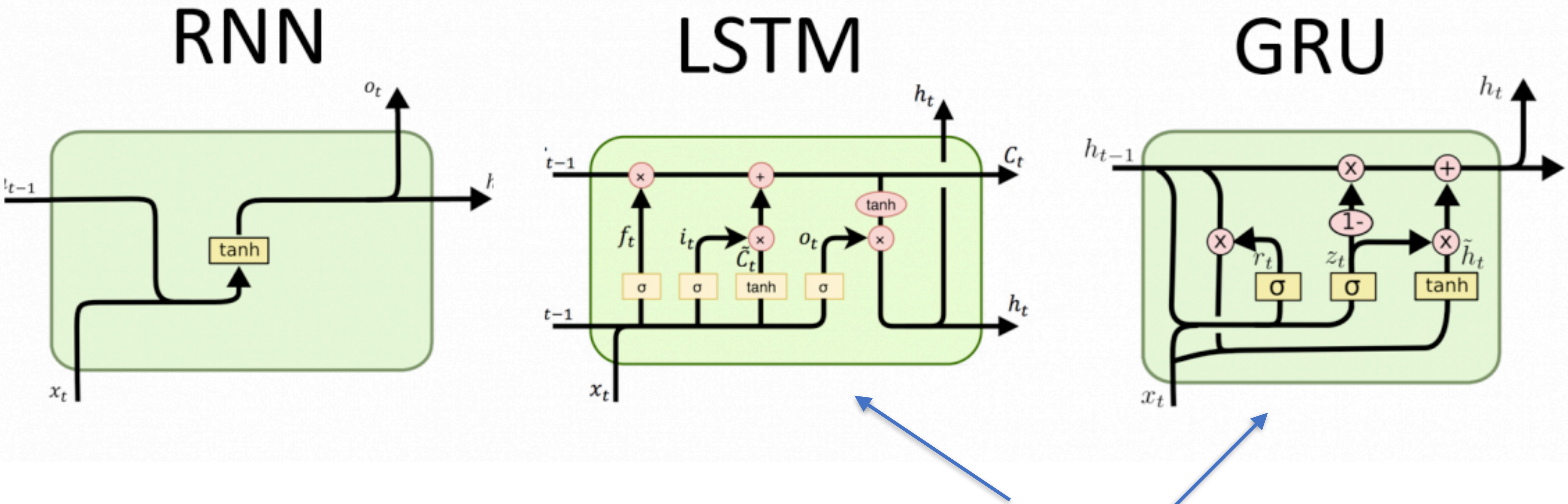
Recursive Neural Networks



Recursive Neural Networks



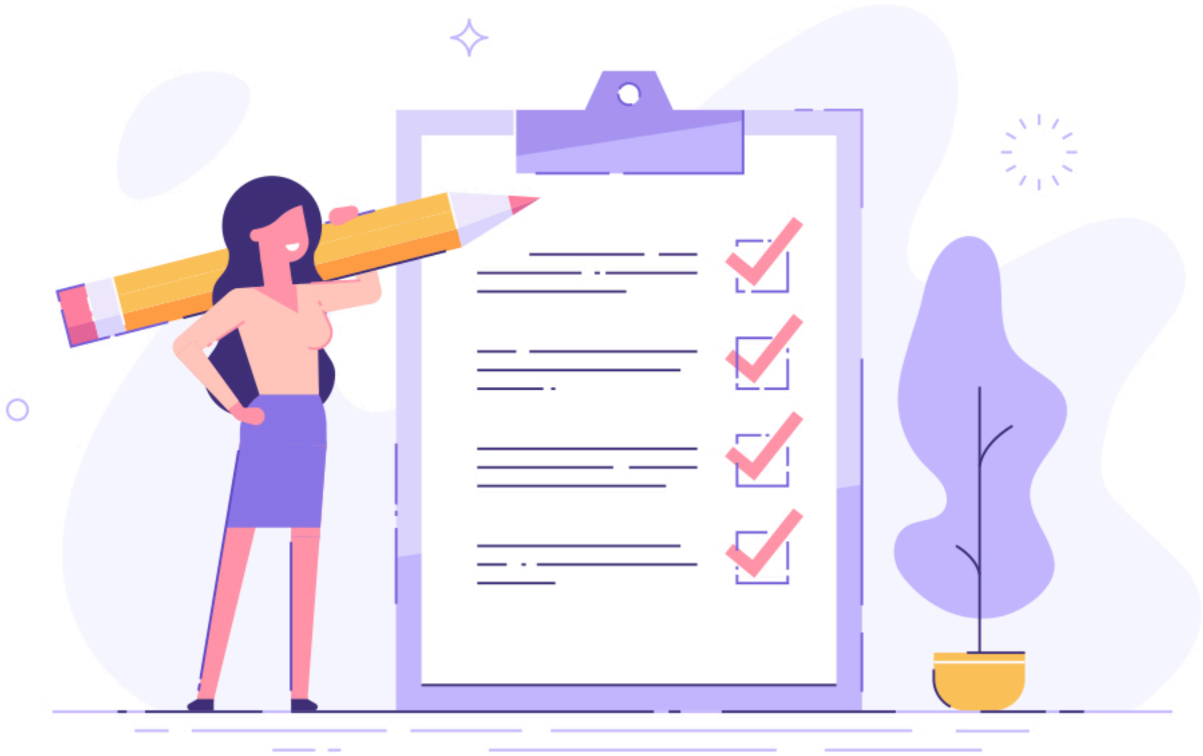
Variations of RNNs



These extra gates provide memory and forgetfulness

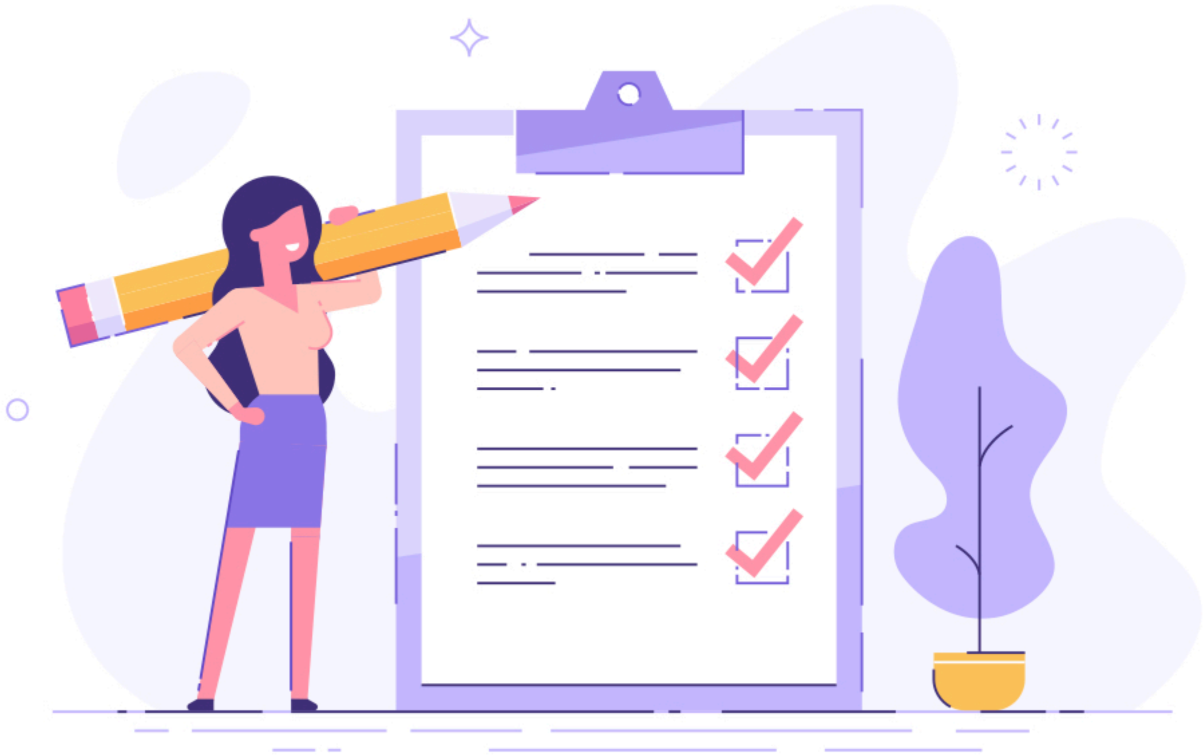
LAB: Train an RNN

- Train an RNN from scratch
- Put attention to the dataset preparation and how to predict



Allocated time: 30/45 minutes

Pulse Check

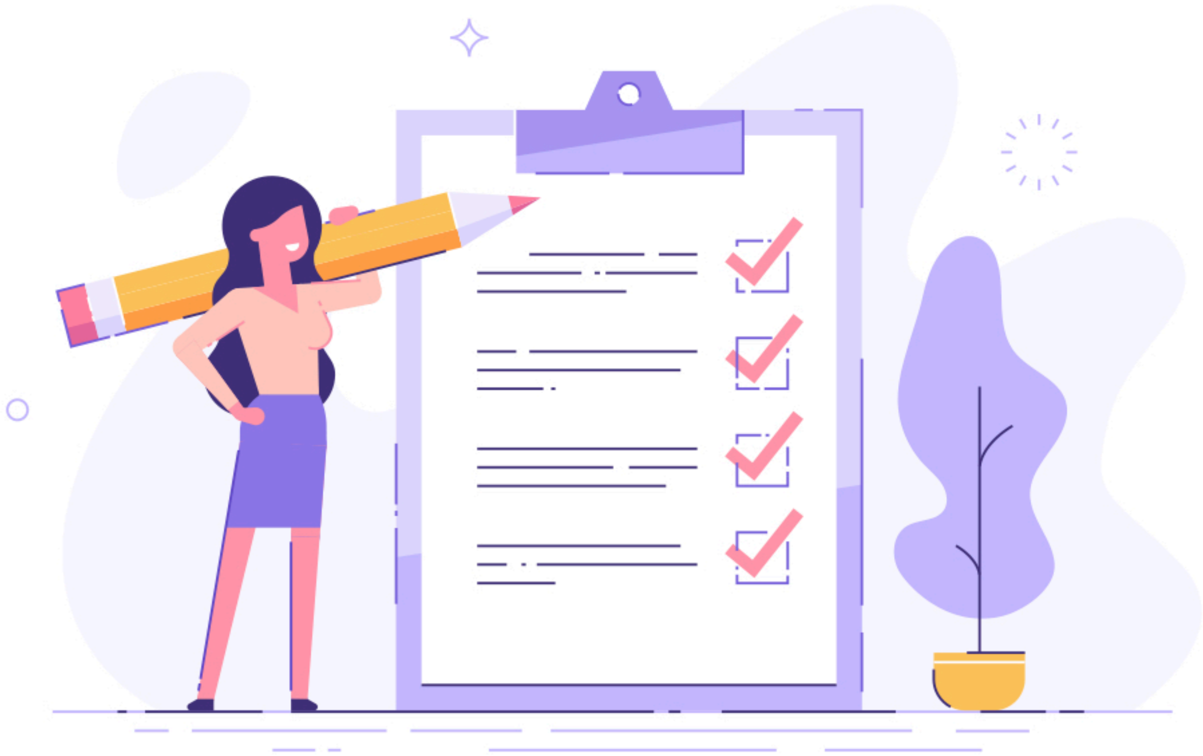


Break: 10 minutes

Develop Intelligence

LAB: Alternative Queries

- Create alternative queries in Solr with our model

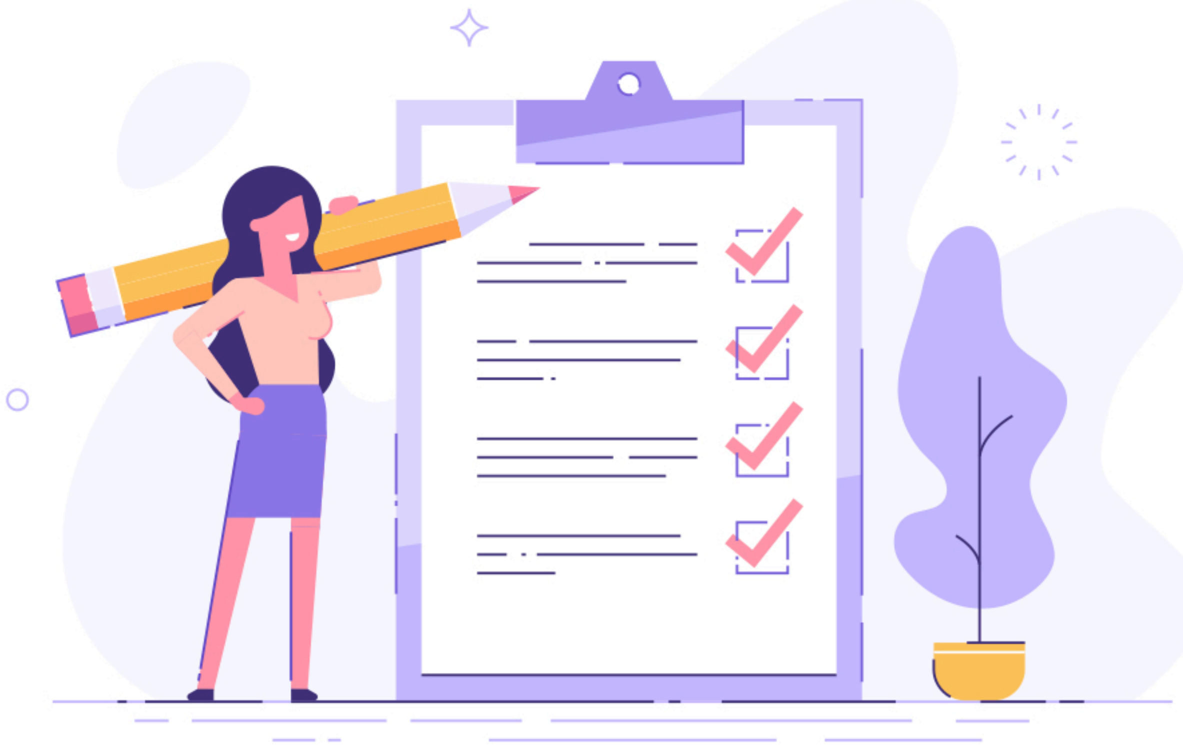


Allocated time: 15 minutes

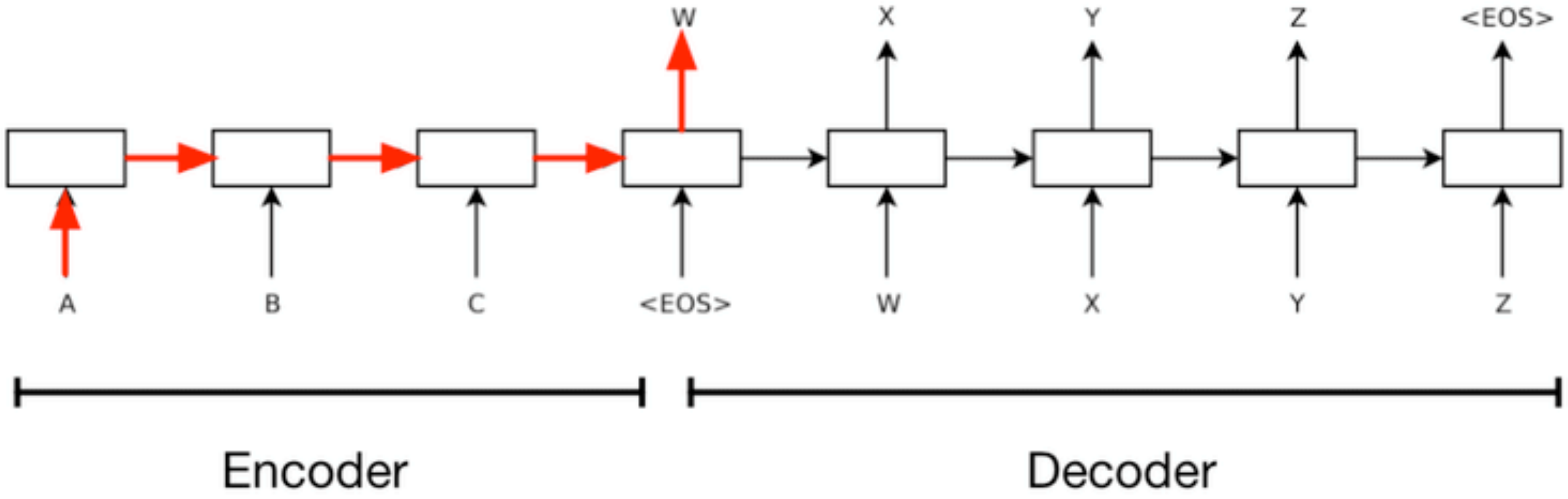
Discussion Time

Why do you think the model performed as it did? Any idea to improve it?

Allocated time: 10 minutes

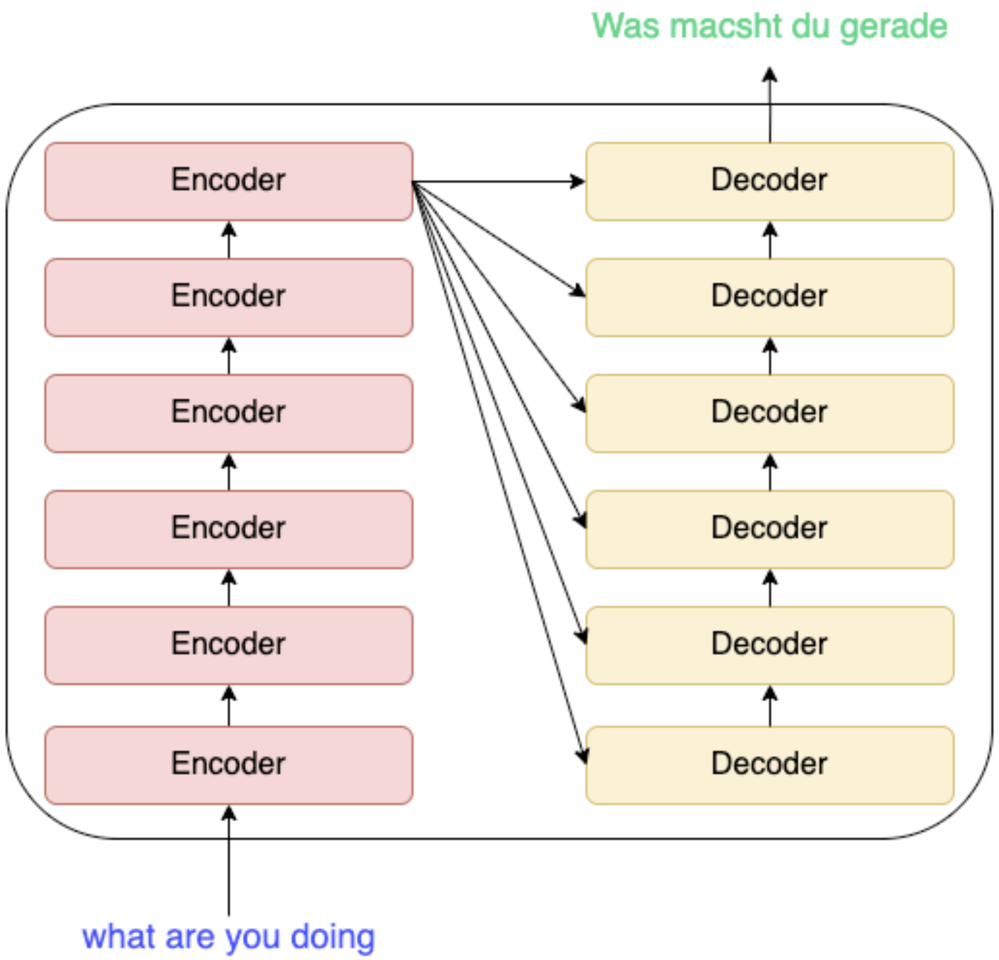


A note on Transformers



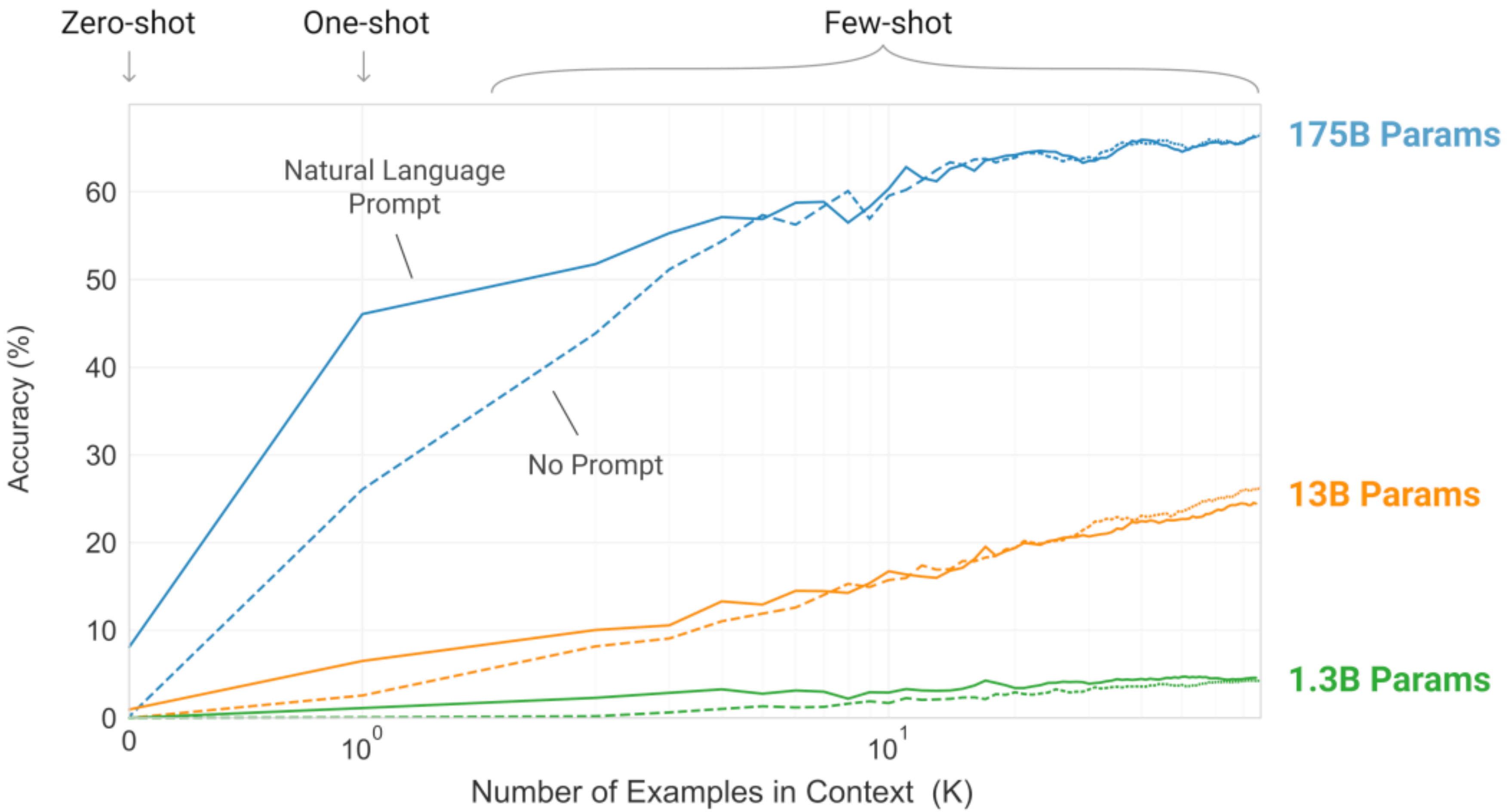
A note on Transformers

They act as black boxes that perform sequence-sequence tasks, such as text generation

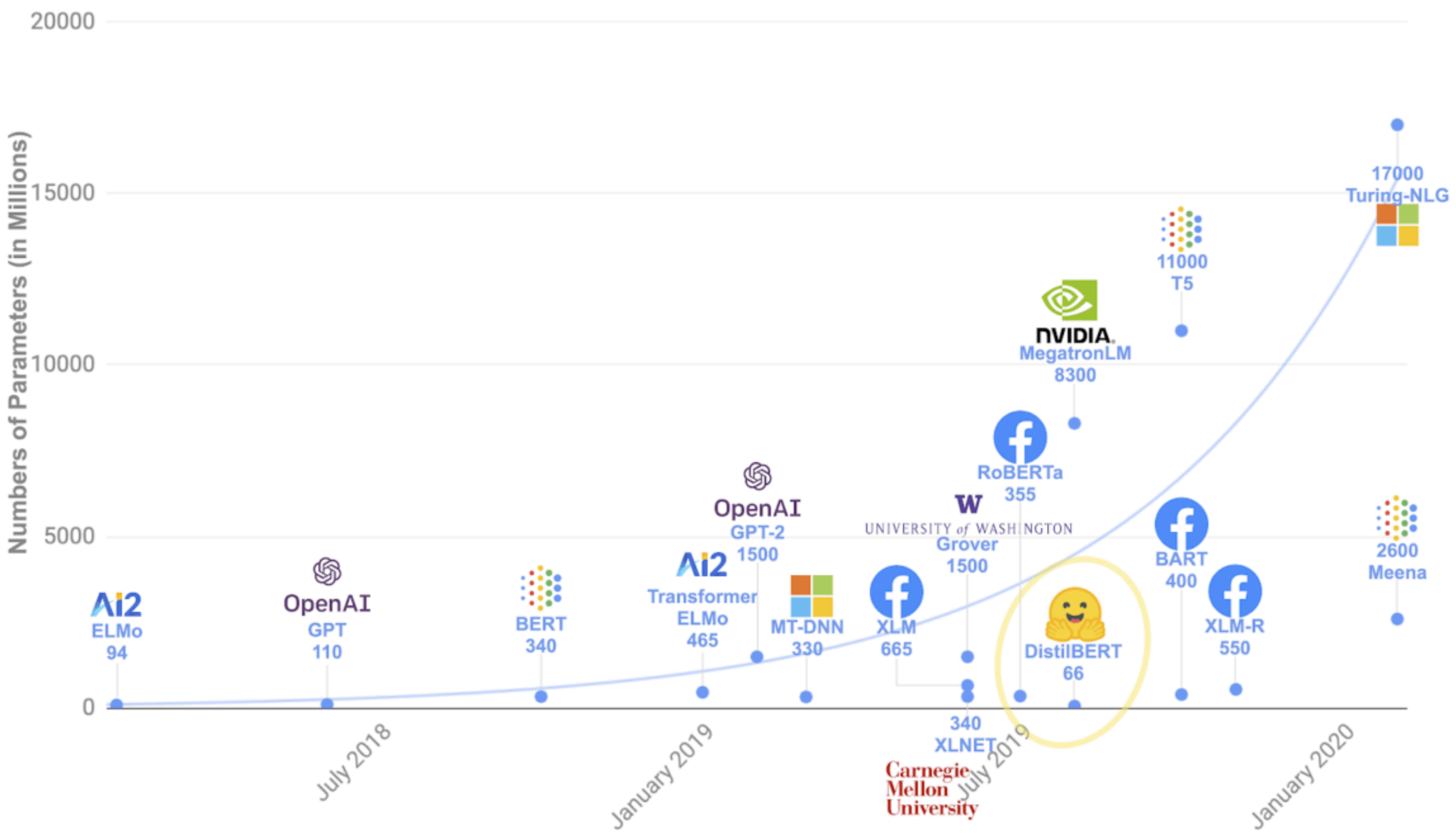


A transformer

Bigger vs Better



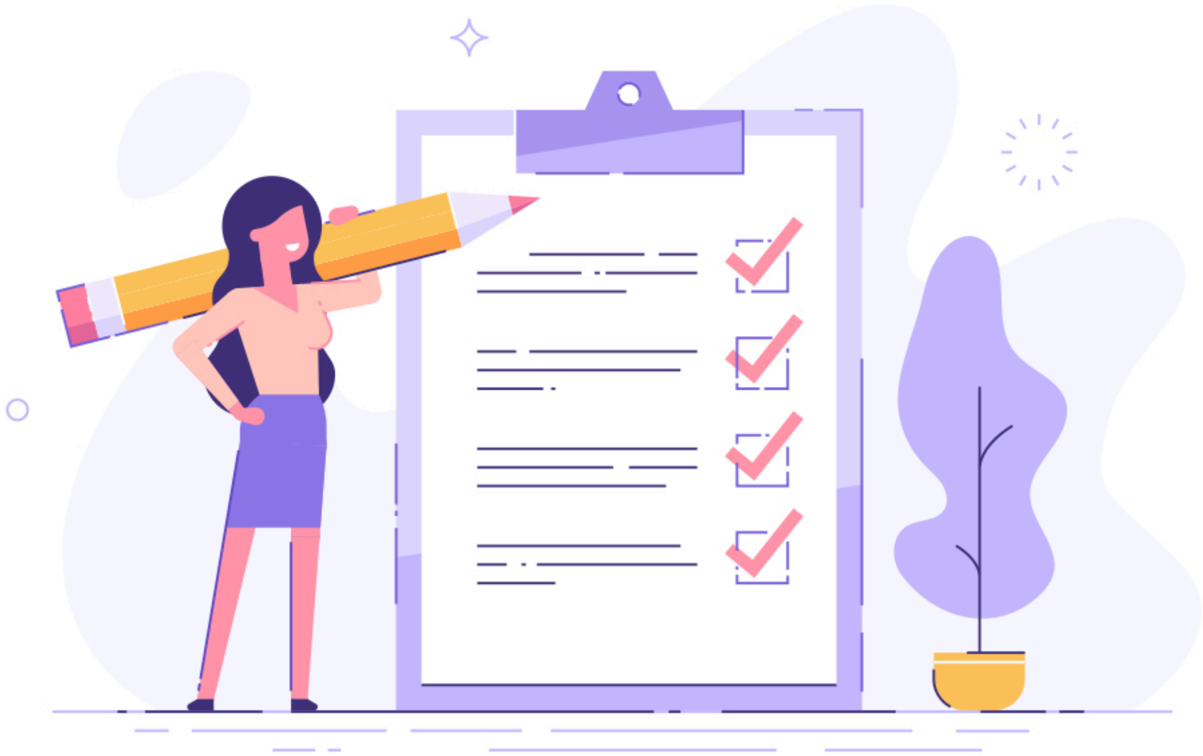
Beware of $$$



20000
15000
10000
5000
0
Numbers of Parameters (in Millions)
Ai2
ELMo
94
OpenAI
GPT
110
BERT
340
Ai2
Transformer
ELMo
465
OpenAI
GPT-2
1500
MT-DNN
330
XLM
665
UNIVERSITY of WASHINGTON
Grover
1500
340
XLNET
Carnegie
Mellon
University
RoBERTa
355
DistilBERT
66
NVIDIA
MegatronLM
8300
11000
T5
BART
400
XLM-R
550
17000
Turing-NLG
2600
Meena
July 2018
January 2019
July 2019
January 2020

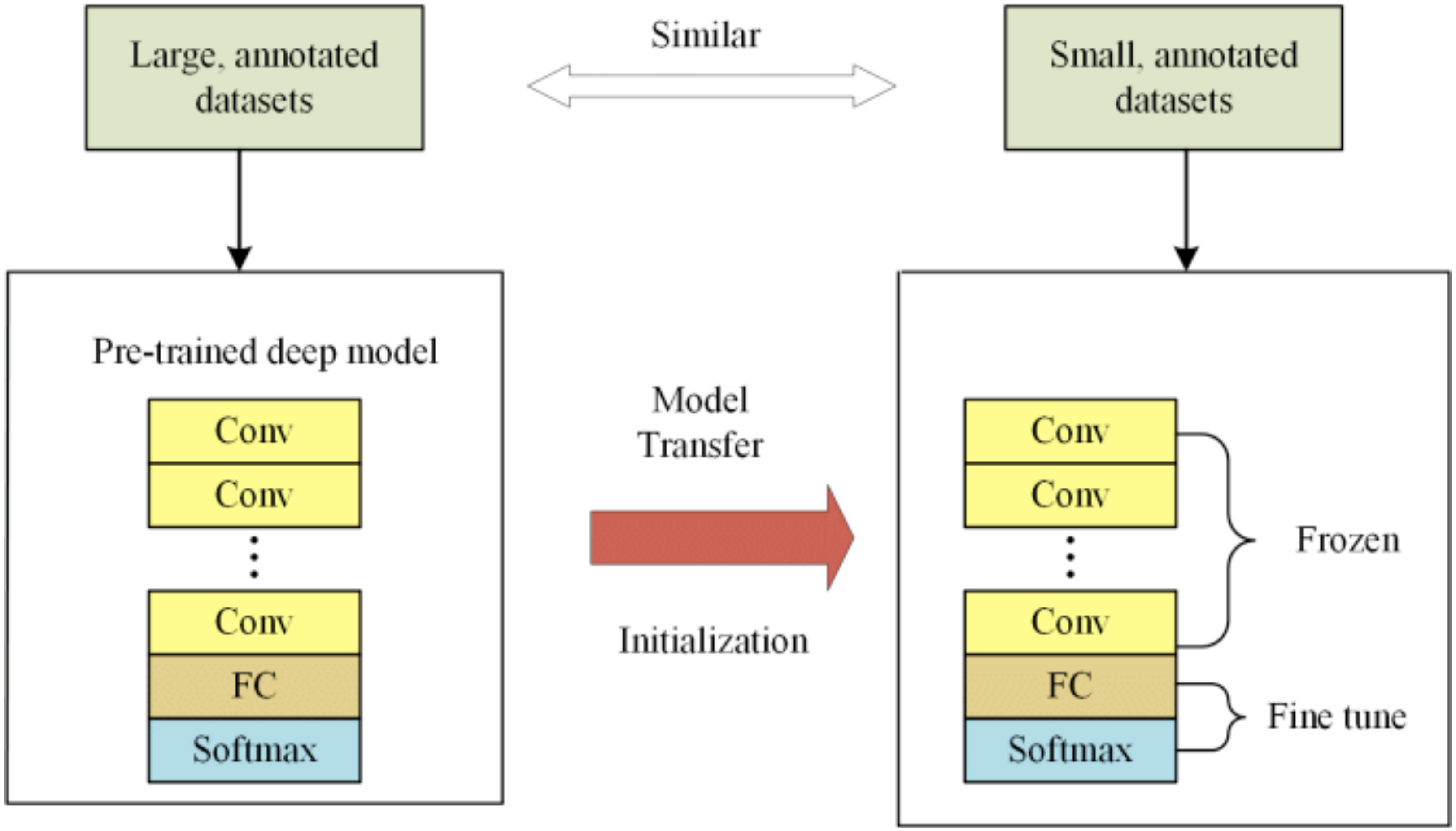
LAB: Using Transformers

- Use GPT-2 to generate the query expansion

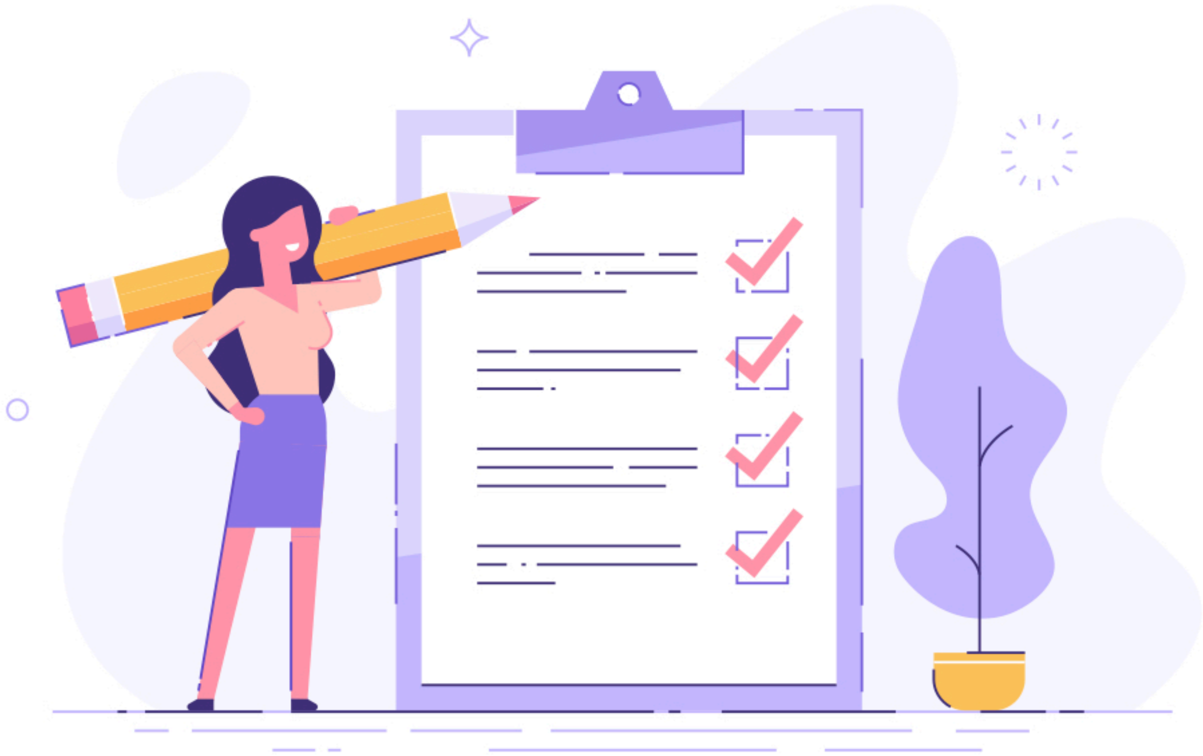


Allocated time: 20 minutes

A note on Transfer Learning



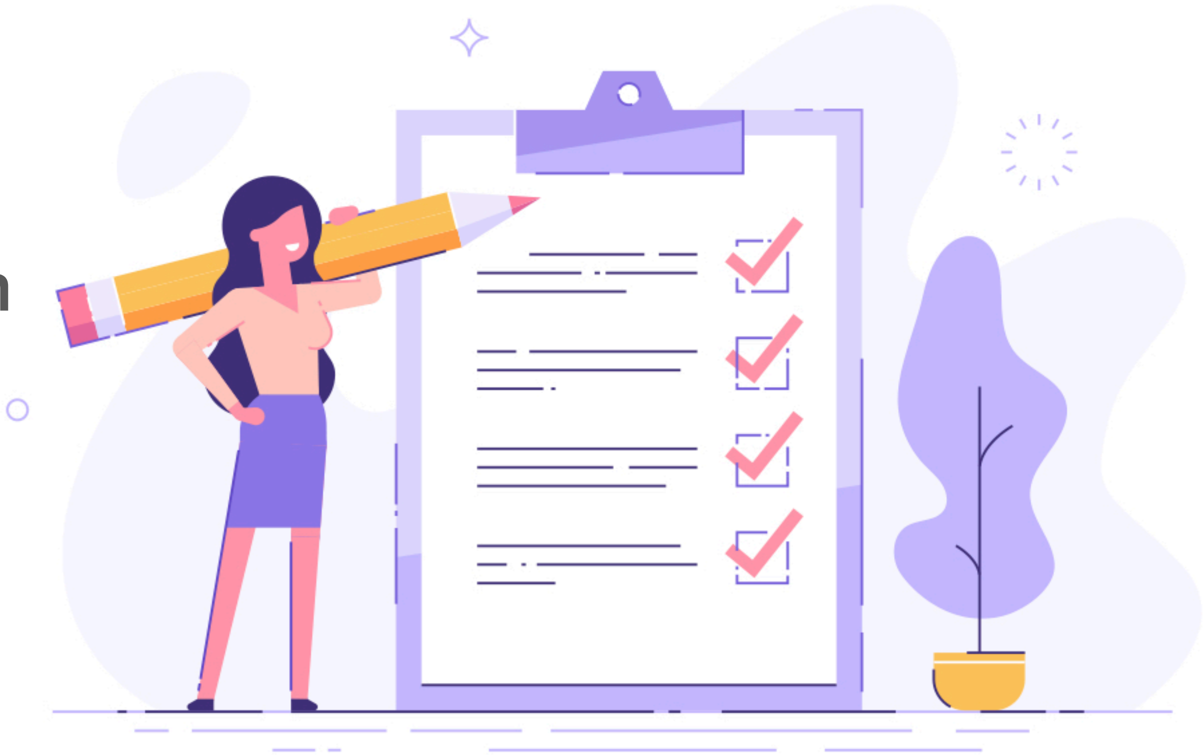
Discussion Time



Discussion Time

Develop Intelligence

I leave you as homework to use transfer learning on our GPT-2 text generation to improve it further with the Airbnb context



Summary

Summary

- Using text generation we can generate alternative queries to run

Summary

- Using text generation we can generate alternative queries to run
- This way we can make the task easier for users by allowing them to be lazy

Summary

- Using text generation we can generate alternative queries to run
- This way we can make the task easier for users by allowing them to be lazy
- Under the hoods, we need a model with recursive structure to maintain memory of the state

Summary

- Using text generation we can generate alternative queries to run
- This way we can make the task easier for users by allowing them to be lazy
- Under the hoods, we need a model with recursive structure to maintain memory of the state
- Using pre-trained transformers may help a lot on your task

Summary

Develop Intelligence

- Using text generation we can generate alternative queries to run
- This way we can make the task easier for users by allowing them to be lazy
- Under the hoods, we need a model with recursive structure to maintain memory of the state
- Using pre-trained transformers may help a lot on your task
- If we have time at the end, we can discuss about distillation! (Please remind me)

NER

NER

Agenda

Develop Intelligence

Introduction

- Introductions
- Expectations
- Overview of Apache Solr
- What is Neural Search?

Ranking

- Doc2Vec
- PD-DM and DBOW
- Learning to Rank

NER

- Spacy

Synonyms

- Neural Networks
- Word2vec
- CBOW and Skip-Gram

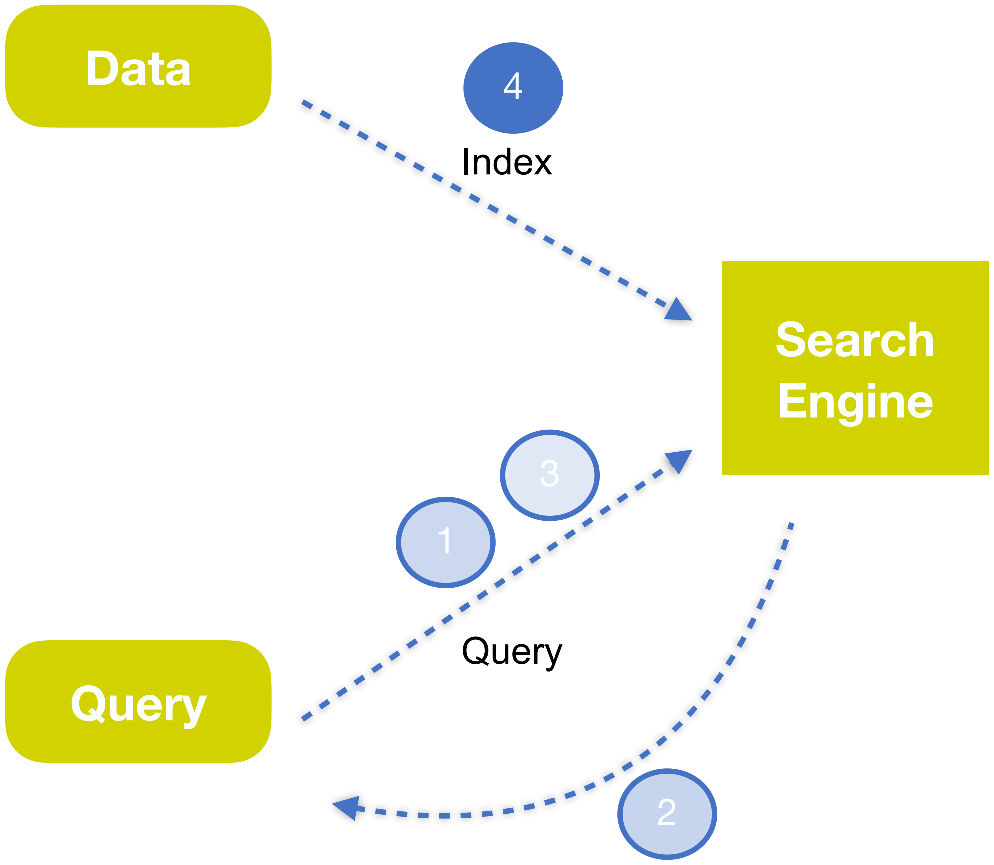
Text Generation

- RNN
- GRU and LSTM
- Transformers

Closing Thoughts

What are we going to do now?

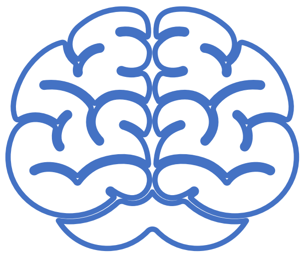
1. Synonym Expansion
2. Reranking results
3. Alternative Queries
4. NER



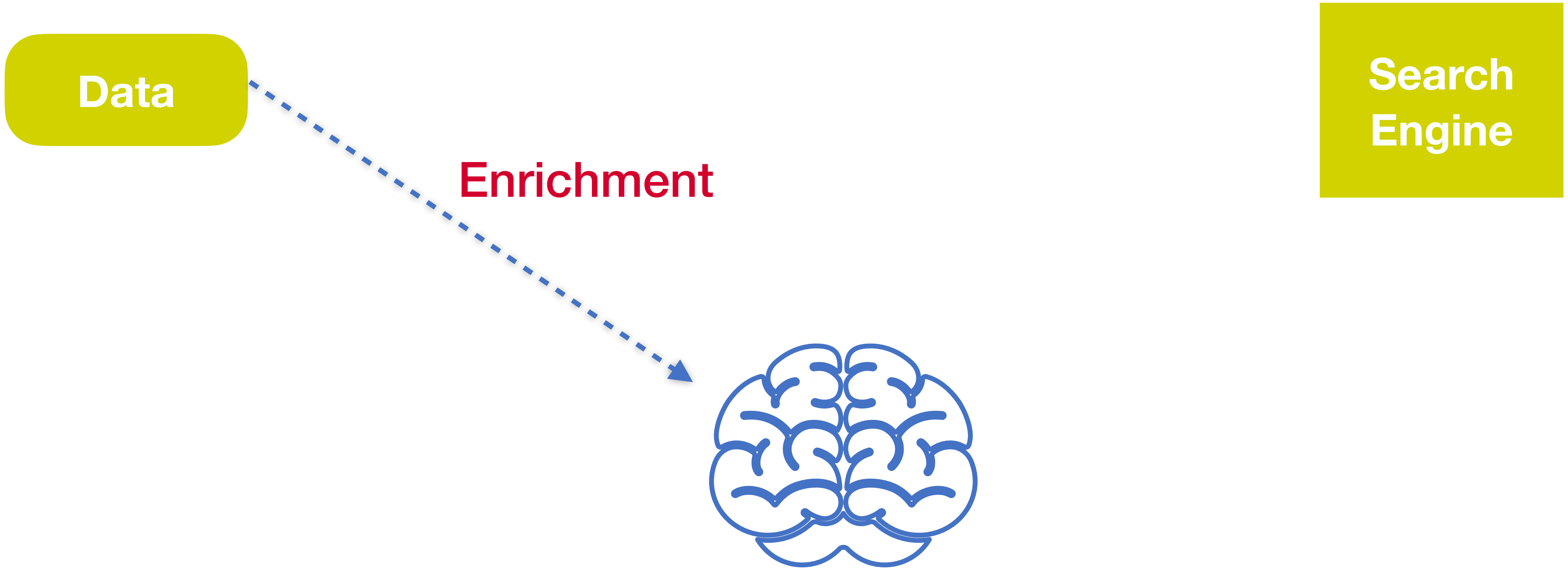
What are we going to do now?

Data

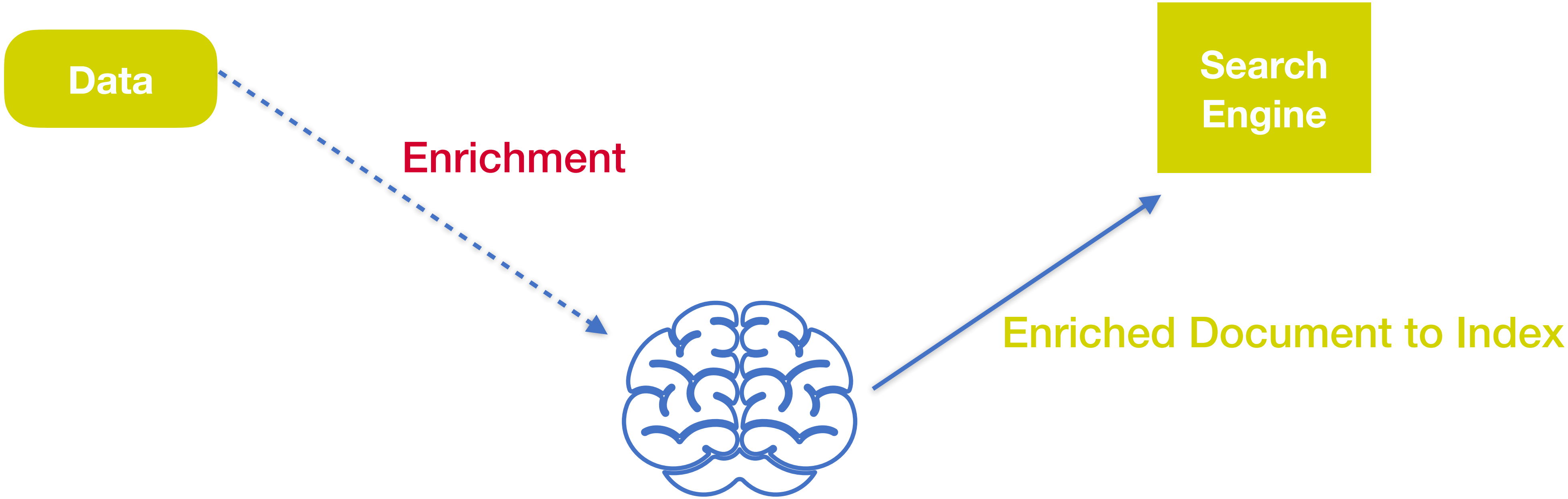
Search Engine



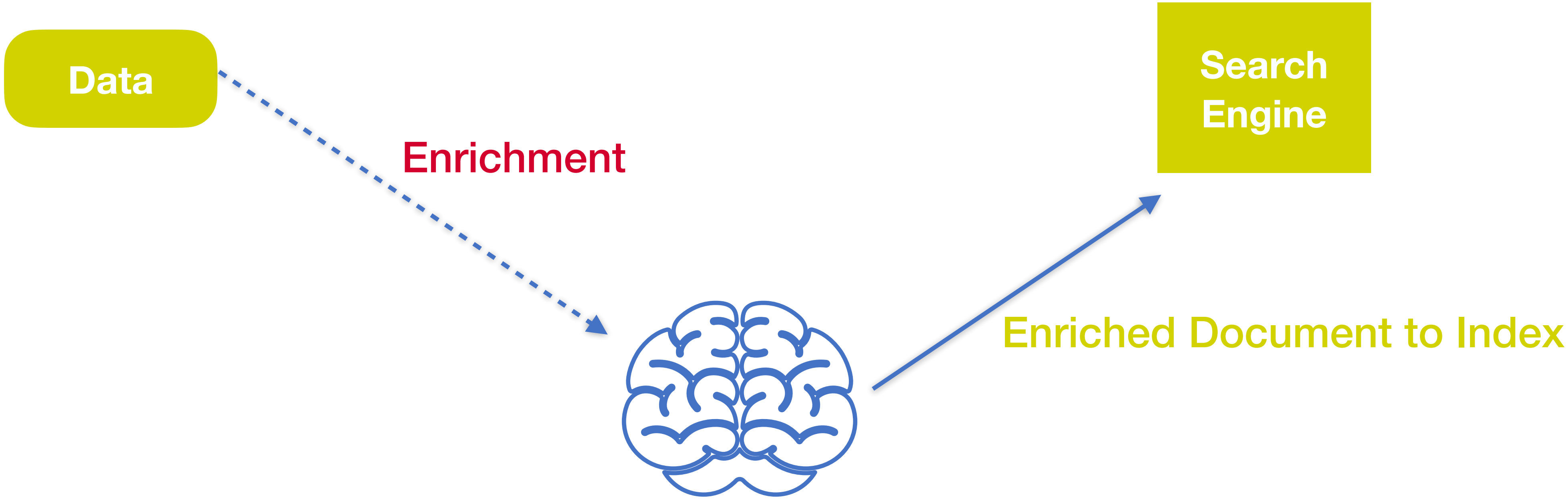
What are we going to do now?



What are we going to do now?



What are we going to do now?



NER helps extract useful tags from text and have only those enriches fields be indexable and retrievable, making the index more efficient.

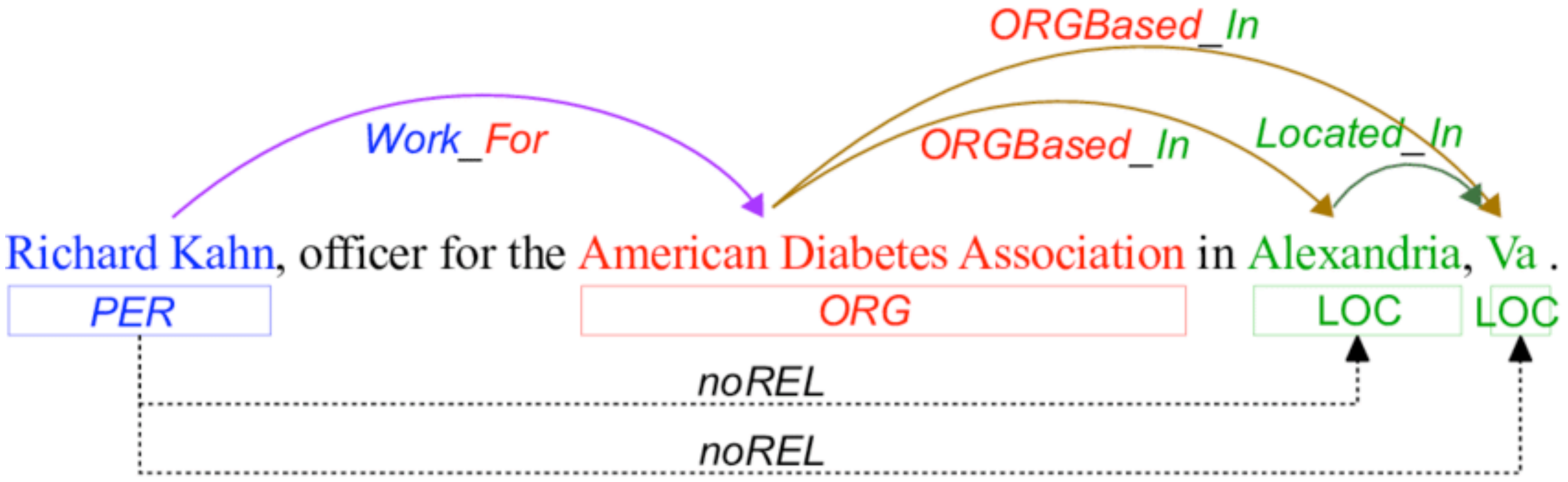
Named Entity Recognition

When Sebastian Thrun PERSON started at Google ORG in 2007 DATE , few people outside of the company took him seriously. “I can tell you very senior CEOs of major American NORP car companies would shake my hand and turn away because I wasn’t worth talking to,” said Thrun PERSON , now the co-founder and CEO of online higher education startup Udacity, in an interview with Recode ORG earlier this week DATE .

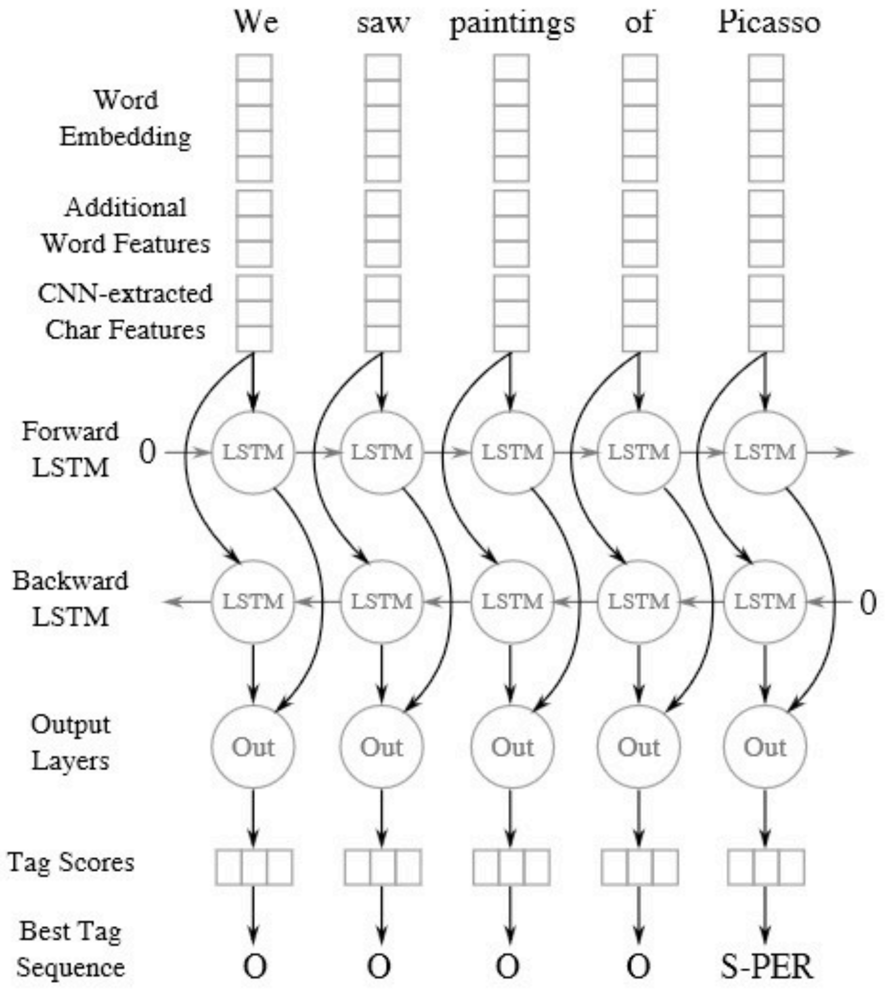
A little less than a decade later DATE , dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

All of these tags extract information from the text

Named Entity Recognition



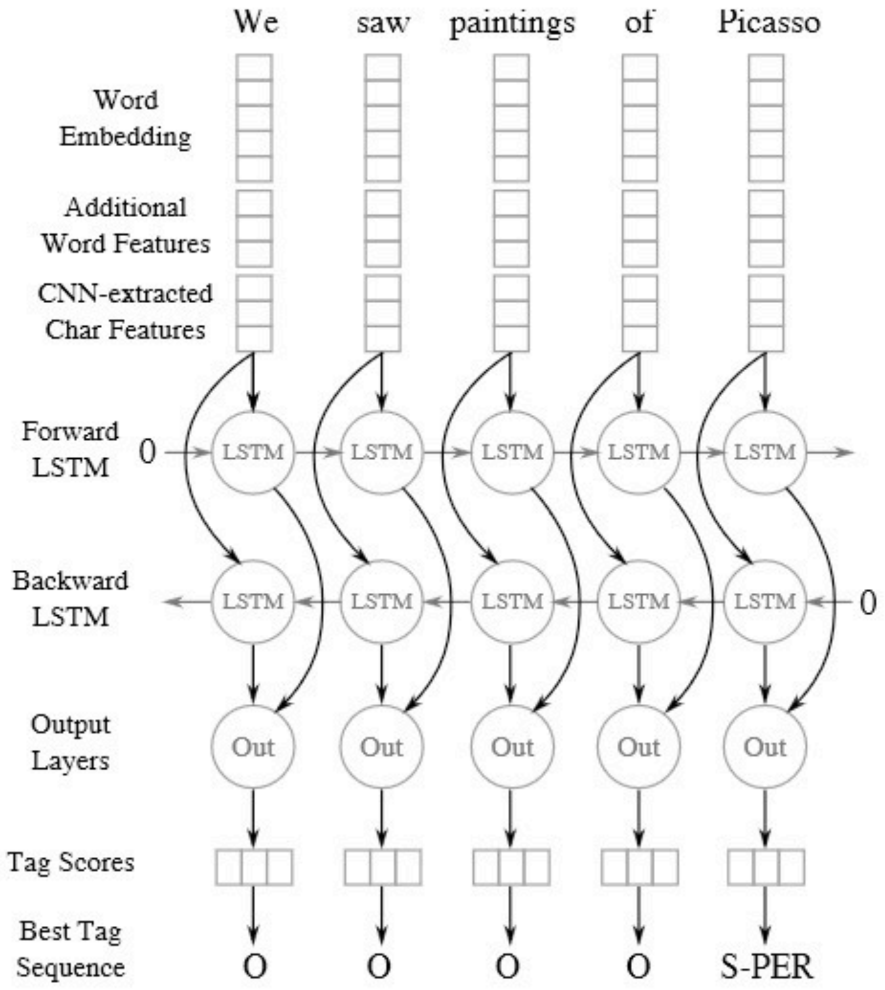
Bi-LSTM



We
saw
paintings
of
Picasso
Word Embedding
Additional Word Features
CNN-extracted Char Features
Forward LSTM
0
LSTM
Backward LSTM
Output Layers
Out
Tag Scores
Best Tag Sequence
O
O
O
O
S-PER

Bi-LSTM

We need a forward and backward pass because some tags at the beginning only make sense after reading the whole sentence



Spacy

Develop Intelligence

```
import spacy
from spacy import displacy

nlp = spacy.load("en_core_web_sm")
doc = nlp("I went to London to see Elizabeth")
html = displacy.render([doc], style="ent")
```

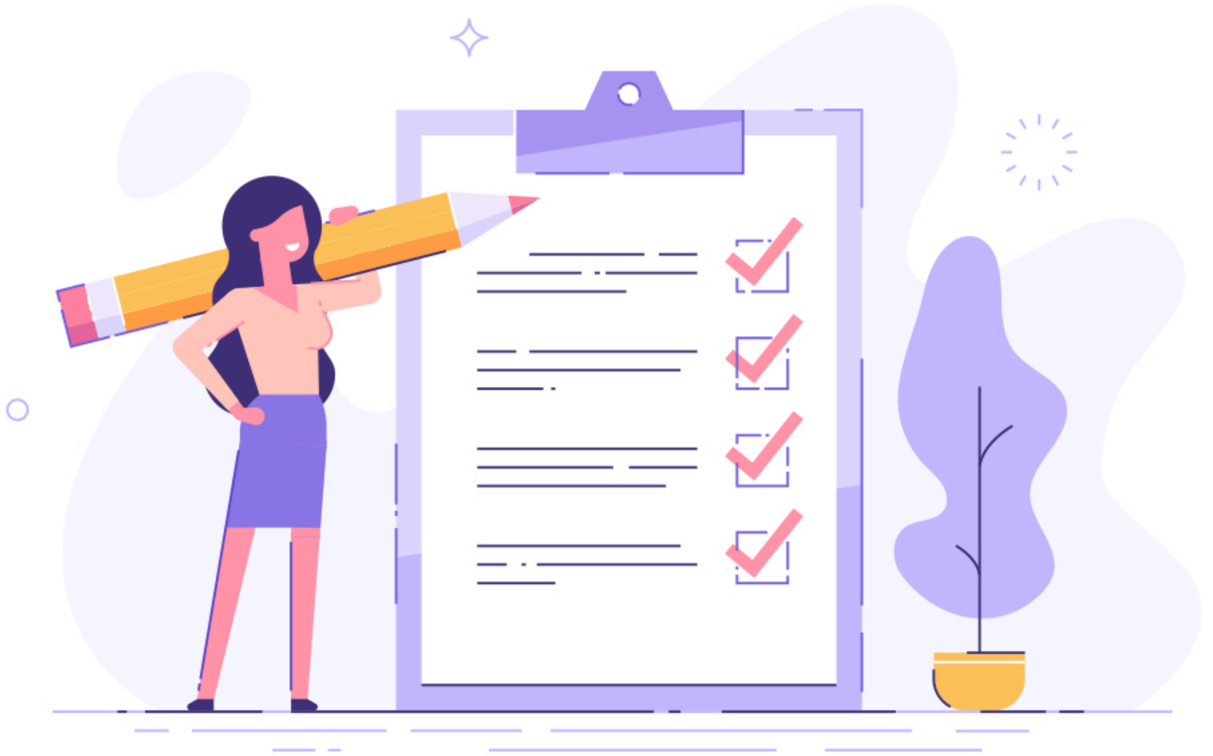
executed in 488ms, finished 14:01:44 2021-03-19

I went to London **GPE** to see Elizabeth **PERSON**

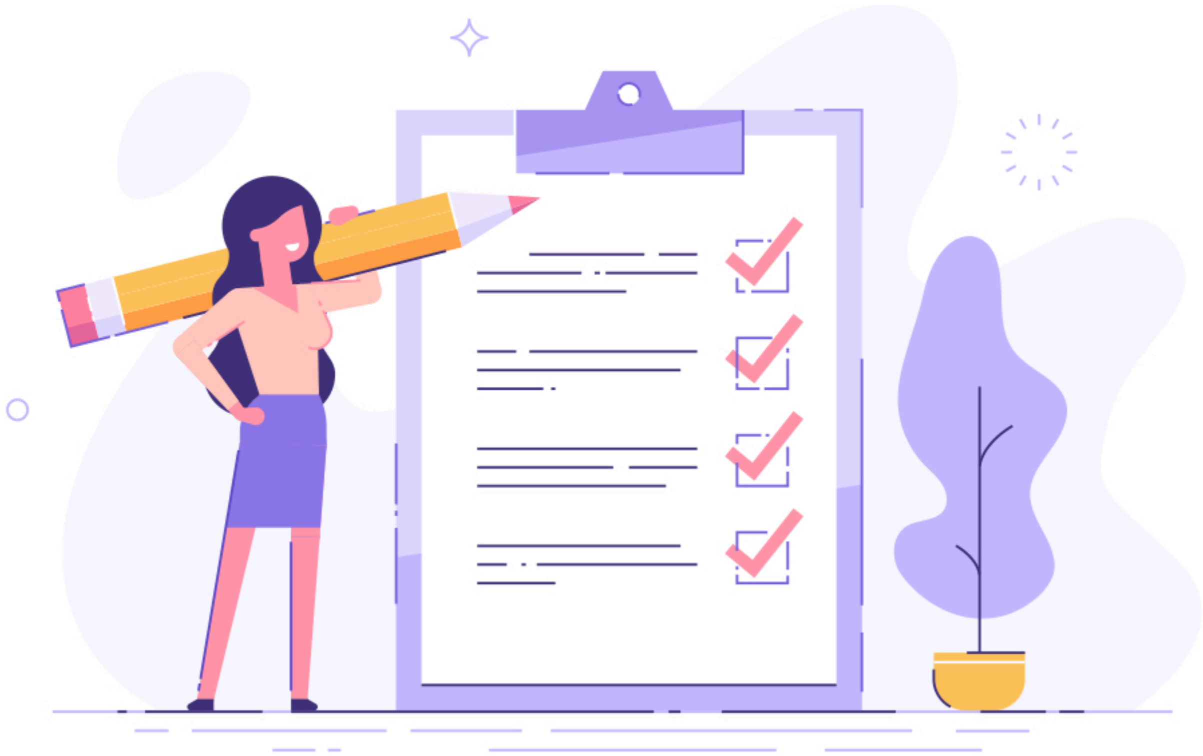
LAB: Enriching documents with NER

- Create a tags field with NER extracted tags
- Verify most useful queries still work

Allocated time: 20 minutes

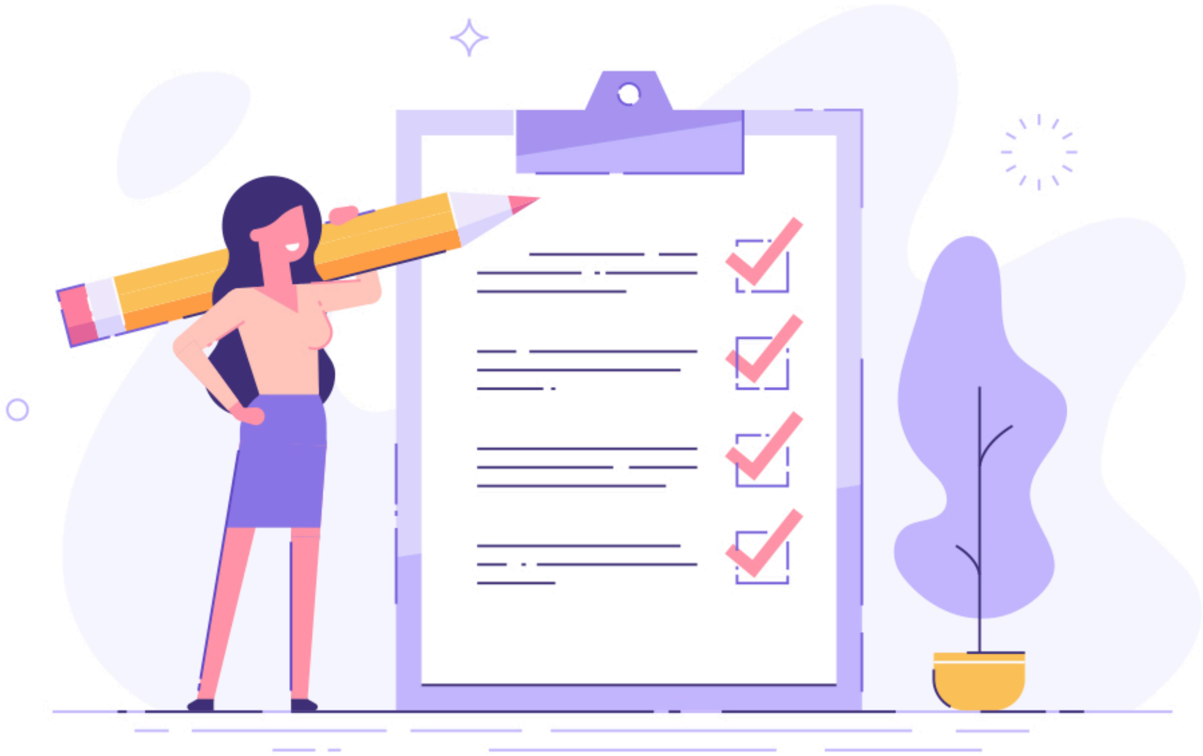


Pulse Check



Discussion Time

Do you have any use cases where NER may help you?



Break: 15 minutes

Develop
Intelligence

Bonus Labs

- Feel free to use the remaining time to revisit the old core labs
- Or enjoy the 3 extra labs covering some extra topics that relate (LTR, Bi-LSTM on Sentiment, and Bi-LSTM on NER)
- I will be here for any questions!

Please Fill in Survey

THANK YOU

Develop
Intelligence