Text-to-SPARQL

By: Shahriar Yazdipour

Dr. Nadine Steinmetz

November 2021

The SPIRIT of science
TECHNISCHE UNIVERSITÄT ILMENAU

Agenda

- Intro
- RDF and SPARQL
- Transformers
- T5
- Our Model
- Samples
- Q&A

Intro

Our Goal

- In this study we present a model, which generates SPARQL queries based on the input question given to the model.
- The Text-to-SPARQL model is a transformer encoder-decoder based architecture, pre-trained for translation task on the Google Brain T5 model.
- We applied two models' sizes based on the Seq2Seq Hugging Face T5 Transformer framework by utilizing NVIDIA GPUs.

Intro

RDF (Resource Description Framework)

- RDF is a general method of describing data by defining relationships between data objects.

- Built around the existing Web standards: XML and URL (URI).

- RDF enables effective data integration from multiple sources, detaching data from its schema.
- This allows multiple schemas to be applied, interlinked, queried as one and modified without changing the data instances.

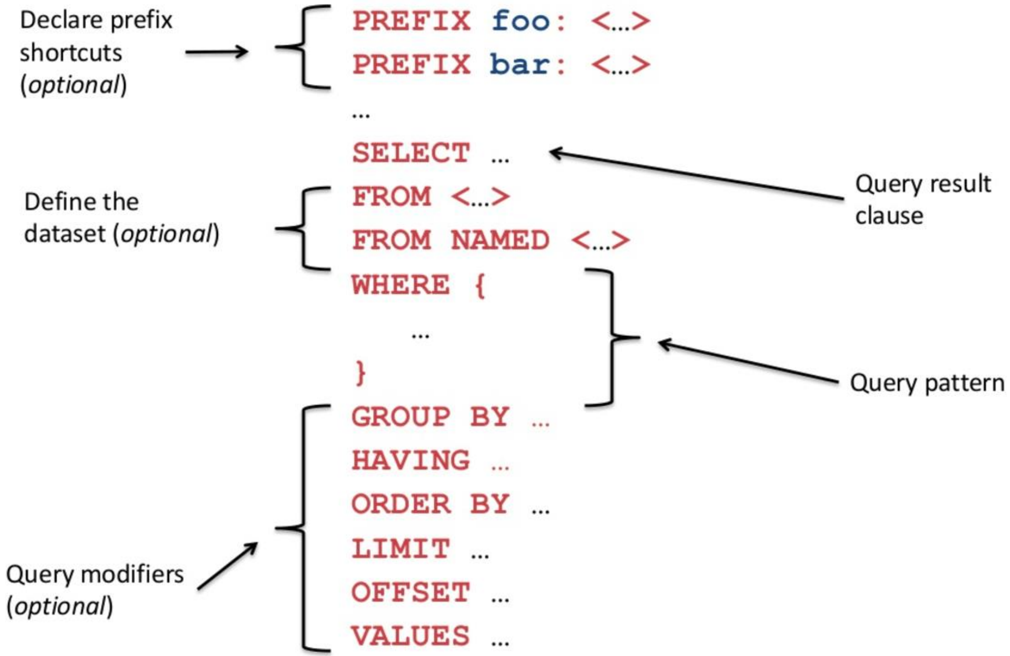
Intro

SPARQL

- SPARQL is the standard query language.
- Used for manipulating and retrieving information contained in RDF graphs
- Using it needs
    - an understanding of the entities in the field to be queried
    - knowledge about the syntax and semantics of the language
- That is why its application is commonly narrowed to a group of Semantic Web experts skilled in the query language.
- To make it more accessible, researchers are working on ways to translate questions from natural language to SPARQL over the last years.

Intro

SPARQL



Intro

SPARQL

Returns names and emails of every person in the dataset:

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
       ?email
WHERE
  {
    ?person  a          foaf:Person .
    ?person  foaf:name  ?name .
    ?person  foaf:mbox  ?email .
  }
```

Intro

Wikidata SPARQL

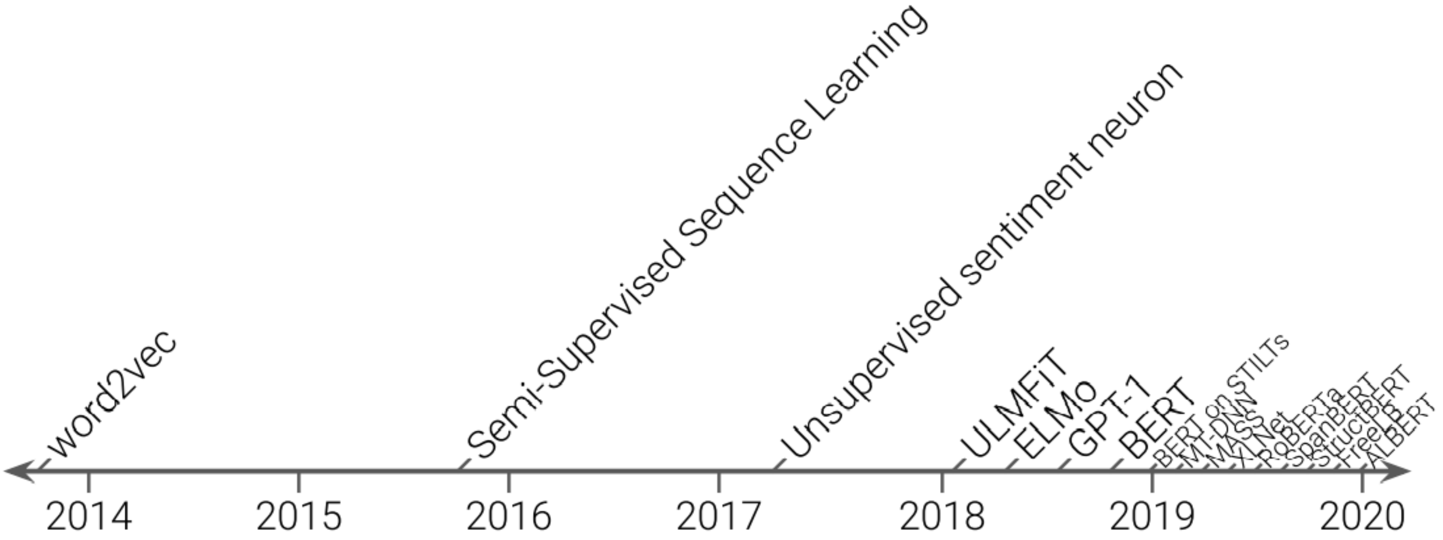
Transformers

- Many essential and excellent works start around 2015 about Word2Vec to map words and encode some meaning about them in continuous space.
- An important paper in 2017 by OpenAI (Learning to generate reviews and discovering sentiment), showed that by just training a language model on a purely unsupervised objective, the model could learn concepts that were potentially useful for downstream tasks.
- In 2018, ULMFiT, which took the recipe from semi-supervised sequence learning, added some tweaks, figured out how to get it working better, and got some noble results with a similar pipeline, pre-training a language model, fine-tuning on a downstream task.

Transformers

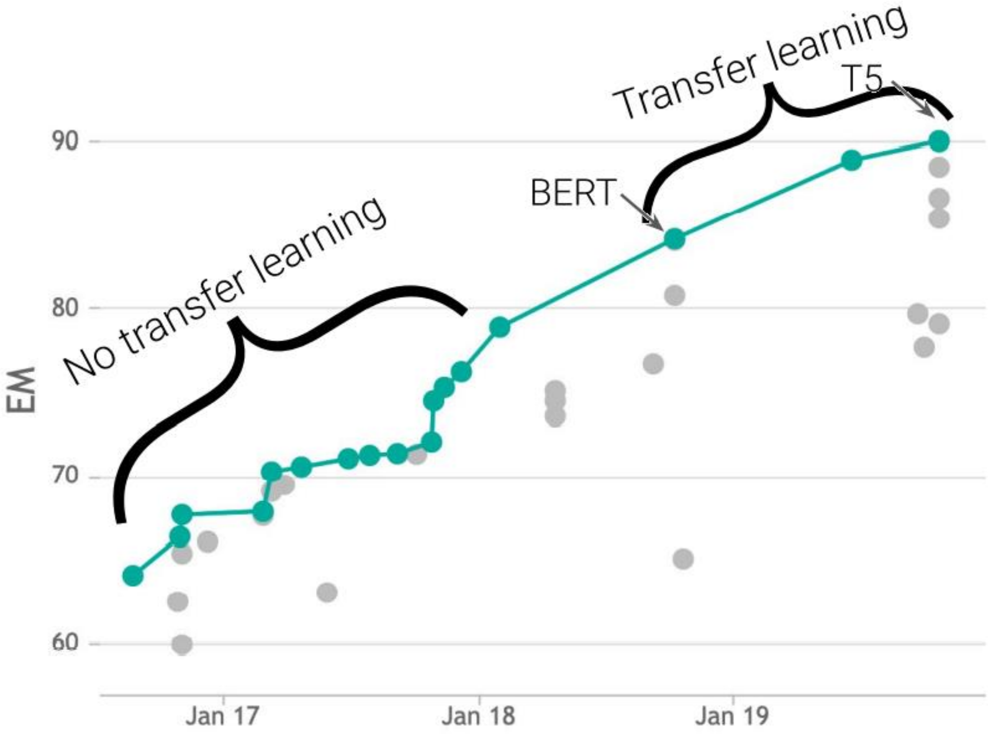
- ELMo showed that we could get significantly better performance by using a bi-directional language model.
- GPT1 showed we can get good performance by using a transformer with a language model.
- Finally, in 2018, BERT convinced many researchers that a bi-directional transformer could get outstanding performance.

Transformers



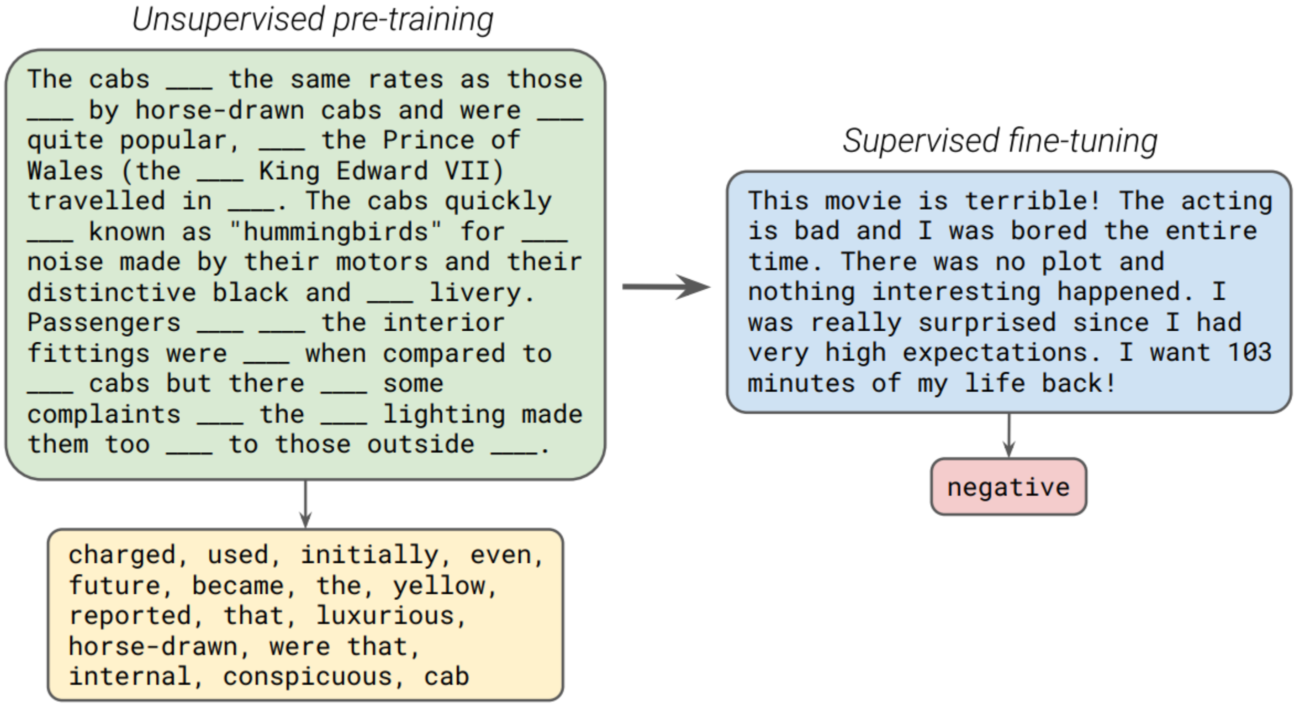
Source: Transfer Learning for NLP: T5 and Beyond at Montreal Institute for Learning Algorithms Tea Talk & Spotify Research Seminar, 2020.

SQuAD Exact Match score



Source: https://paperswithcode.com/sota/question-answering-on-squad11-dev

Transformers



Source: Transfer Learning for NLP: T5 and Beyond at Montreal Institute for Learning Algorithms Tea Talk & Spotify Research Seminar, 2020.

T5

Exploring Transfer Learning with T5: the Text-To-Text Transfer Transformer from Google Brain

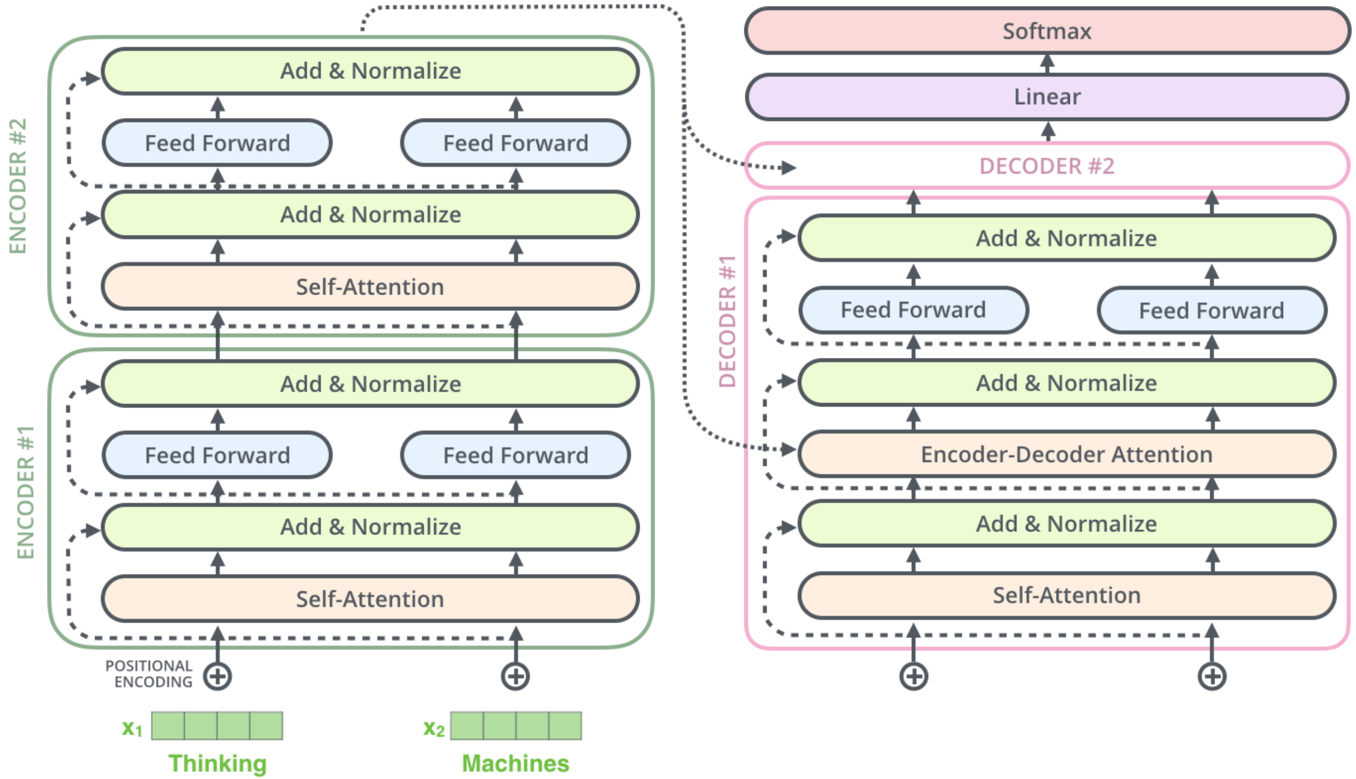
Source: https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html (February, 2020)

T5

- T5 Model implemented by Raffel et al. (2020) at Google Brain
- Sequence-to-Sequence framework
- Uses the BERT encoder-decoder architecture
- Using C4 Dataset for pretraining

- Originally T5 was introduced with five pre-trained models
    - Small (60 million parameters)
    - Base(220 million parameters)
    - Large(770 million parameters)
    - 3B(3 billion parameters)
    - 11B(11 billion parameters)

Model - BERT



Source: The Illustrated Transformer – Jay Alammar – Visualizing machine learning one concept at a time. (jalammar.github.io)

Common Crawl Web Extracted Text

```
Menu

Lemon

Introduction

The lemon, Citrus Limon (l.) Osbeck, is a
species of small evergreen tree in the
flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for
culinary and non-culinary purposes
throughout the world, primarily for its juice,
which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6%
citric acid, with a ph of around 2.2, giving it
a sour taste.

Article

The origin of the lemon is unknown, though
lemons are thought to have first grown in
Assam (a region in northeast India),
northern Burma or China.
A genomic study of the lemon indicated it
was a hybrid between bitter orange (sour
orange) and citron.
```

```
Please enable JavaScript to use our site.

Home
Products
Shipping
Contact
FAQ

Dried Lemons, $3.59/pound

Organic dried lemons from our farm in
California.
Lemons are harvested and sun-dried for
maximum flavor.
Good in soups and on popcorn.

The lemon, Citrus Limon (l.) Osbeck, is a
species of small evergreen tree in the
flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for
culinary and non-culinary purposes
throughout the world, primarily for its juice,
which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6%
citric acid, with a ph of around 2.2, giving it
a sour taste.
```

```
Lorem ipsum dolor sit amet, consectetur
adipiscing elit.
Curabitur in tempus quam. In mollis et ante
at consectetur.
Aliquam erat volutpat.
Donec at lacinia est.
Duis semper, magna tempor interdum
suscipit, ante elit molestie urna, eget
efficitur risus nunc ac elit.
Fusce quis blandit lectus.
Mauris at mauris a turpis tristique lacinia at
nec ante.
Aenean in scelerisque tellus, a efficitur
ipsum.
Integer justo enim, ornare vitae sem non,
mollis fermentum lectus.
Mauris ultrices nisl at libero porta sodales in
ac orci.

function Ball(r) {
  this.radius = r;
  this.area = pi * r ** 2;
  this.show = function(){
    drawCircle(r);
  }
}
```

Colin Raffel, Answering Questions by Querying the Implicit Knowledge Base Inside T5 at AKBC 2020 Unstructured and Structured Knowledge Bases Workshop, 2020.

Colossal Clean Crawled Corpus (C4)

Menu

Lemon

Introduction

The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Article

The origin of the lemon is unknown, though lemons are thought to have first grown in Assam (a region in northeast India), northern Burma or China.
A genomic study of the lemon indicated it was a hybrid between bitter orange (sour orange) and citron.

Please enable JavaScript to use our site.

Home
Products
Shipping
Contact
FAQ

Dried Lemons, $3.59/pound

Organic dried lemons from our farm in California.
Lemons are harvested and sun-dried for maximum flavor.
Good in soups and on popcorn.

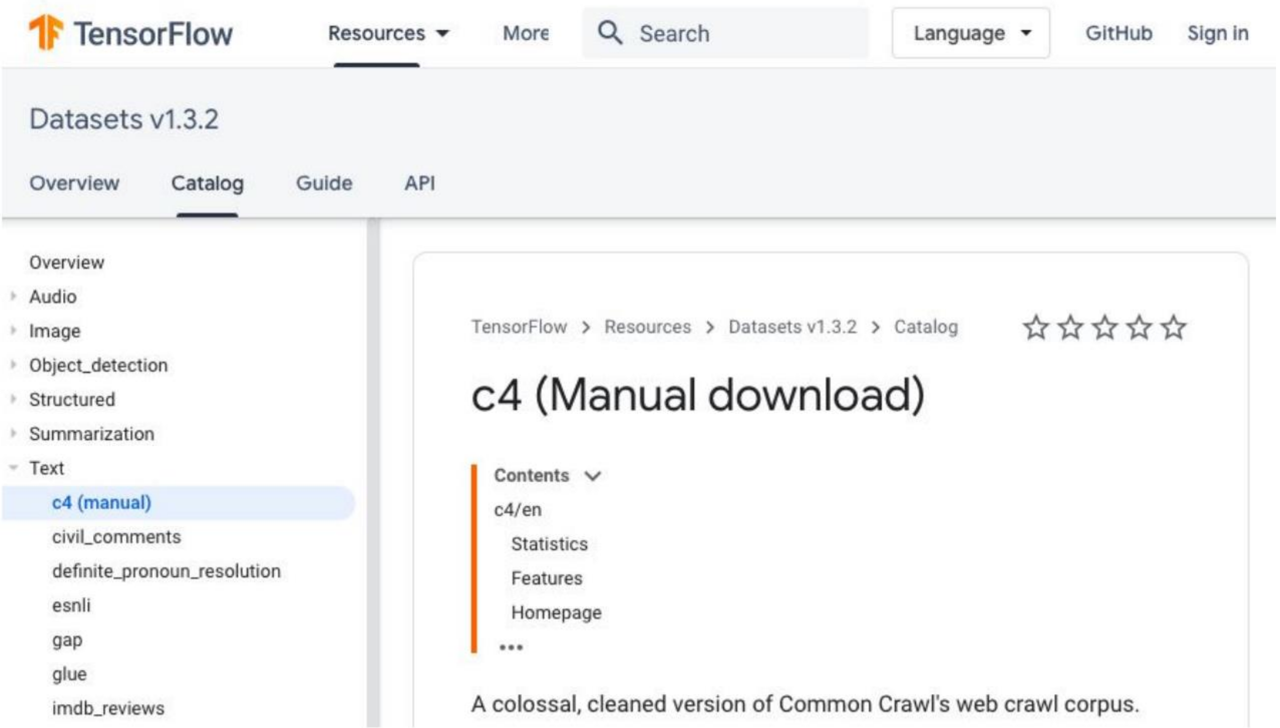
The lemon, Citrus Limon (l.) Osbeck, is a species of small evergreen tree in the flowering plant family rutaceae.
The tree's ellipsoidal yellow fruit is used for culinary and non-culinary purposes throughout the world, primarily for its juice, which has both culinary and cleaning uses.
The juice of the lemon is about 5% to 6% citric acid, with a ph of around 2.2, giving it a sour taste.

Lorem ipsum dolor sit amet, consectetur adipiscing elit.
Curabitur in tempus quam. In mollis et ante at consectetur.
Aliquam erat volutpat.
Donec at lacinia est.
Duis semper, magna tempor interdum suscipit, ante elit molestie urna, eget efficitur risus nunc ac elit.
Fusce quis blandit lectus.
Mauris at mauris a turpis tristique lacinia at nec ante.
Aenean in scelerisque tellus, a efficitur ipsum.
Integer justo enim, ornare vitae sem non, mollis fermentum lectus.
Mauris ultrices nisl at libero porta sodales in ac orci.

```
function Ball(r) {
  this.radius = r;
  this.area = pi * r ** 2;
  this.show = function(){
    drawCircle(r);
  }
}
```

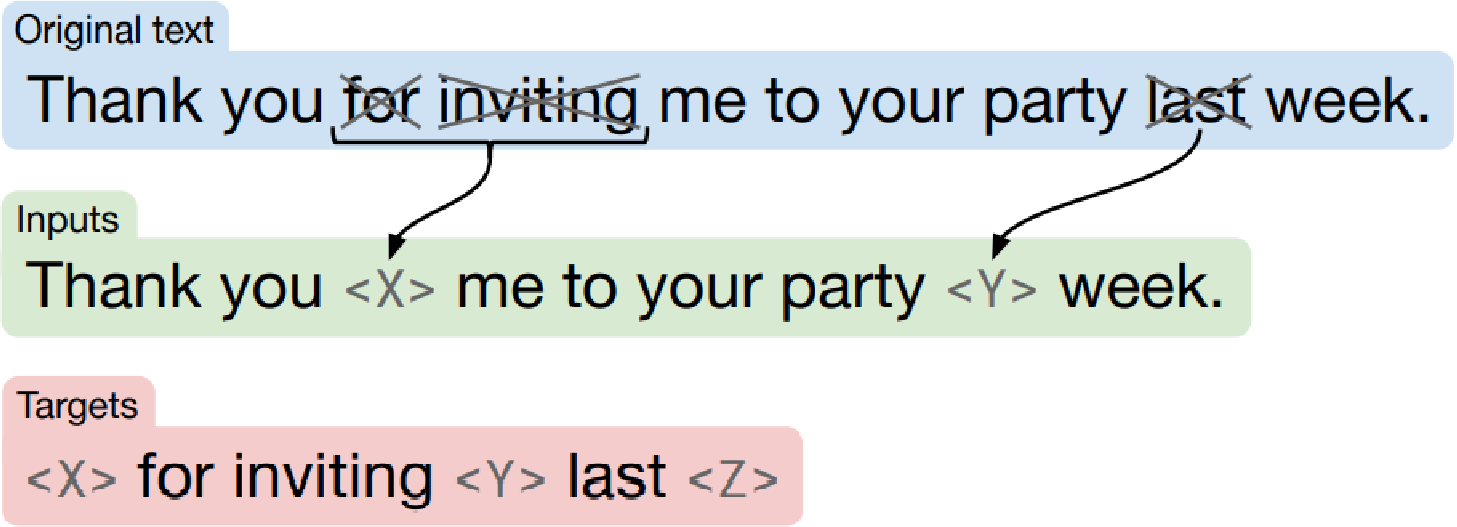
Colin Raffel, Answering Questions by Querying the Implicit Knowledge Base Inside T5 at AKBC 2020 Unstructured and Structured Knowledge Bases Workshop, 2020.

Colossal Clean Crawled Corpus (C4)



Colin Raffel, Answering Questions by Querying the Implicit Knowledge Base Inside T5 at AKBC 2020 Unstructured and Structured Knowledge Bases Workshop, 2020.

Pre-Training



Different Approaches

Encoder-decoder architecture

Architecture	Params	Cost	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
★ Encoder-decoder	$2P$	M	**83.28**	**19.24**	**80.88**	**71.36**	**26.98**	**39.82**	**27.65**
Enc-dec, shared	P	M	82.81	18.78	**80.63**	**70.73**	26.72	39.03	**27.46**
Enc-dec, 6 layers	P	$M/2$	80.88	18.97	77.59	68.42	26.38	38.40	26.95
Language model	P	M	74.70	17.93	61.14	55.02	25.09	35.28	25.86
Prefix LM	P	M	81.82	18.61	78.94	68.11	26.43	37.98	27.39

Span prediction objective

Span length	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
★ Baseline (i.i.d.)	**83.28**	19.24	80.88	71.36	**26.98**	**39.82**	**27.65**
2	**83.54**	19.39	**82.09**	**72.20**	**26.76**	**39.99**	**27.63**
3	**83.49**	**19.62**	**81.84**	**72.53**	**26.86**	39.65	**27.62**
5	**83.40**	19.24	**82.05**	**72.23**	**26.88**	39.40	**27.53**
10	82.85	19.33	**81.84**	70.44	**26.79**	39.49	**27.69**

C4 dataset

Dataset	Size	GLUE	CNNDM	SQuAD	SGLUE	EnDe	EnFr	EnRo
★ C4	745GB	83.28	**19.24**	80.88	71.36	**26.98**	**39.82**	**27.65**
C4, unfiltered	6.1TB	81.46	19.14	78.78	68.04	26.55	39.34	27.21
RealNews-like	35GB	**83.83**	**19.23**	80.39	72.38	**26.75**	**39.90**	**27.48**
WebText-like	17GB	**84.03**	**19.31**	**81.42**	71.40	**26.80**	**39.74**	**27.59**
Wikipedia	16GB	81.85	**19.31**	81.29	68.01	**26.94**	39.69	**27.67**
Wikipedia + TBC	20GB	83.65	**19.28**	**82.08**	**73.24**	**26.77**	39.63	**27.57**

Source: Exploring Transfer Learning with T5: the Text-To-Text Transfer Transformer

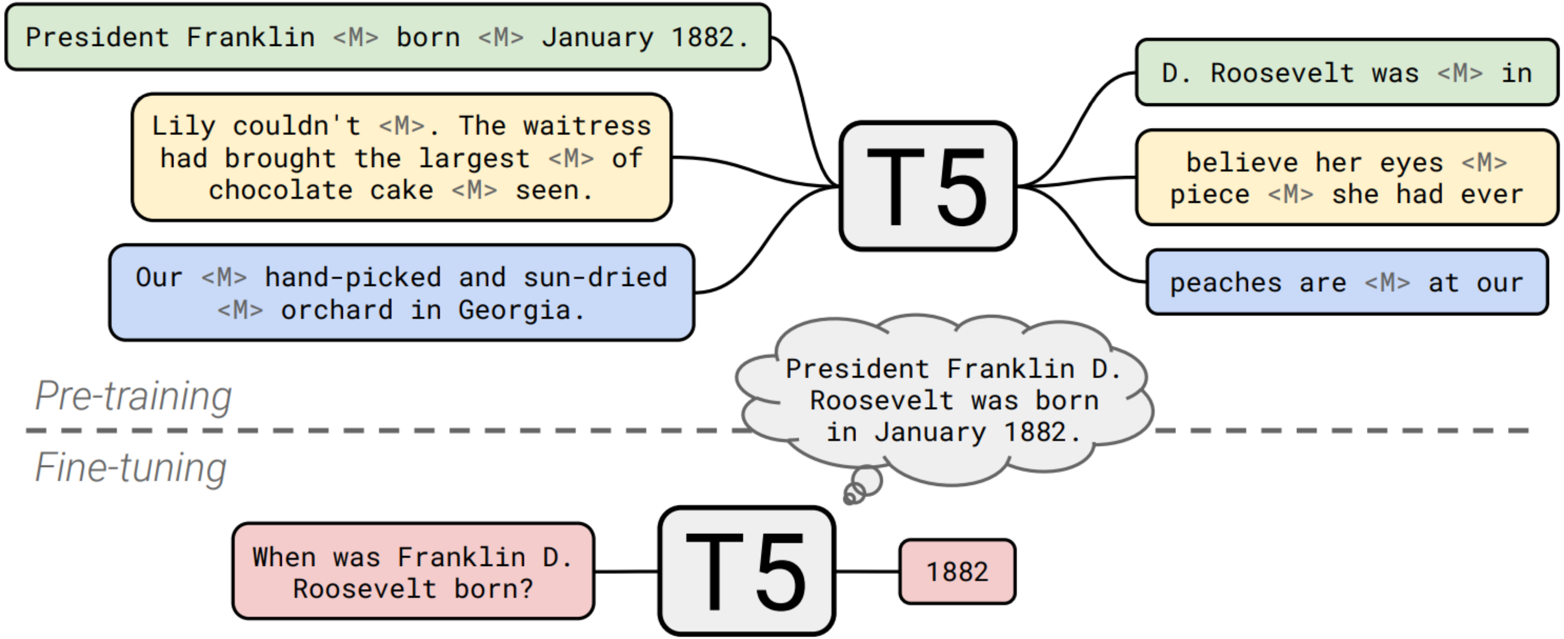
T5 Models

Model	GLUE Average	CNN/DM ROUGE-2-F	SQuAD EM	SuperGLUE Average
Previous best	89.4	20.30	90.1	84.6
T5-Small	77.4	19.56	87.24	63.3
T5-Base	82.7	20.34	92.08	76.2
T5-Large	86.4	20.68	93.79	82.3
T5-3B	88.5	21.02	94.95	86.4
T5-11B	**90.3**	**21.55**	**91.26**	**89.3**

Human score = 89.8

Source: Exploring Transfer Learning with T5: the Text-To-Text Transfer Transformer

Training



Colin Raffel, Answering Questions by Querying the Implicit Knowledge Base Inside T5 at AKBC 2020 Unstructured and Structured Knowledge Bases Workshop, 2020.

Our Model

Fine-Tuning T5 for SPAQL

- LC-QuAD v2 Dataset
- T5Tokenizer - Sentence Piece Tokenizer
- *"translate English to Sparql: " prefix*
- AdamW Optimizer
- After pre-training
    - 250MB model based on T5-Small
    - 850MB model based on T5-Base

Tools

Huggingface is all you need

Dataset

LC-QuAD v2.0

- English Question Text and its equivalent Wikidata SPARQL query
- Include Paraphrased question
- Contains "template" and other properties that we did not use
- Replace Brace with Brackets
- Convert WikiData P-IDs and Q-IDs to english vocabularies
- 48148 records after preprocessing (24074 records original)
  - 38586 training data
  - 9562 test data

Wikidata - LC-QuAD

Data Sample

	question	sparql_wikidata	paraphrased_question	template
0	What periodical literature does Delta Air Lines use as a moutpiece?	select distinct ?obj where { wd:Q188920 wdt:P2813 ?obj . ?obj wdt:P31 wd:Q1002697 }	What is Delta Air Line's periodical literature mouthpiece?	<S P ?O ; ?O instanceOf Type>
1	Who is the child of Ranavalona I's husband?	SELECT ?answer WHERE { wd:Q169794 wdt:P26 ?X . ?X wdt:P22 ?answer}	What is the name of Ranavalona I's husband's child?	C RCD xD . xD RDE ?E
2	Is it true Jeff_Bridges occupation Lane Chandler and photographer ?	ASK WHERE { wd:Q174843 wdt:P106 wd:Q1804811 . wd:Q174843 wdt:P106 wd:Q33231 }	Are Jeff Bridges and Lane Chandler both photographers?	Ask (ent-pred-obj1` . ent-pred-obj2)
3	What is the pre-requisite of phase matter of Galinstan?	SELECT ?answer WHERE { wd:Q675176 wdt:P515 ?X . ?X wdt:P156 ?answer}	What range are the papers at the Monique Genonceaux about?	E REF xF . xF RFG ?G
4	which cola starts with the letter p	SELECT DISTINCT ?sbj ?sbj_label WHERE { ?sbj wdt:P31 wd:Q134041 . ?sbj rdfs:label ?sbj_label . FILTER(STRSTARTS(lcase(?sbj_label), 'p')) . FILTER (lang(?sbj_label) = 'en') } LIMIT 25	which cola begins with the letter p	<?S P O ; ?S instanceOf Type ; starts wit character >
5	Is the right ascension of malin 1 less than 15.1398?	ASK WHERE { wd:Q4180017 wdt:P6257 ?obj filter(?obj < 15.1398) }	Does malin 1 have a right ascension lower than 15.1398?	ASK ?sbj ?pred ?obj filter ?obj = num
6	What is the complete list of records released by Jerry Lee Lewis?	select distinct ?obj where { wd:Q202729 wdt:P358 ?obj . ?obj wdt:P31 wd:Q273057 }	What is the total list of records discharged by Jerry Lee Lewis?	<S P ?O ; ?O instanceOf Type>
7	Who won the prize at the sequel of the 1885 Wimbledon Championships- Gentlemen's Singles?	SELECT ?answer WHERE { wd:Q1356316 wdt:P156 ?X . ?X wdt:P1346 ?answer}	Who won the prize at the spin-off of the 1885 Wimbledon Championships- Gentlemen's Singles?	C RCD xD . xD RDE ?E
8	Is it true that the carbon footprint of the iPhone X Max is 106?	ASK WHERE { wd:Q56599233 wdt:P5991 ?obj filter(?obj = 106) }	Does the iPhone X Max have a carbon footprint of 106?	ASK ?sbj ?pred ?obj filter ?obj = num

Dataset

KDWD

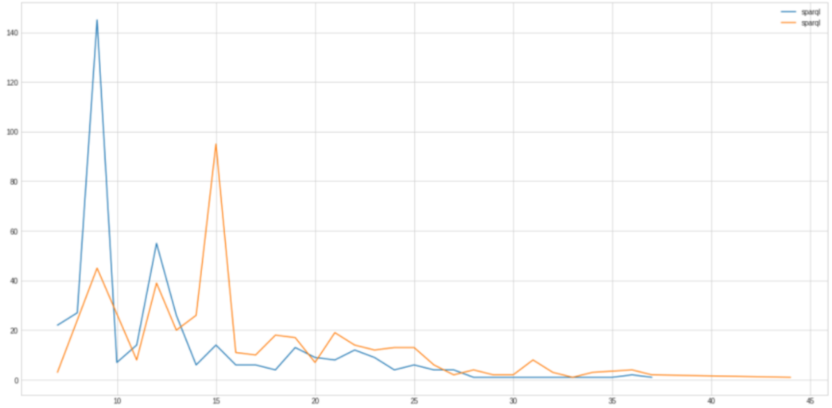
- Dataset from Kensho R&D
- 4 Gigabytes
- Contains WikiData properties and items labels and their Q-ID and P-IDs
- https://www.kaggle.com/kenshoresearch/kdwd-wikidata-small-ontology

	property_id	en_label	en_description	count
2451	2936	language used	language widely used (spoken or written) in th...	8427
333	530	diplomatic relation	diplomatic relations of the country	6057
275	463	member of	organization or club to which the subject belo...	4186
81	150	contains administrative territorial entity	(list of) direct subdivisions of an administra...	3909
15	31	instance of	that class of which this subject is a particul...	1315
23	47	shares border with	countries or administrative subdivisions, of e...	871

Dataset

QALD9

- DBPedia only dataset
- (+) Supports multiple languages
- (+) Properties are not encoded by P-ID/Q-IDs (queries more readable for humans also for LM)
- (-) Quantity of the data, less than a thousand records (not suitable for a deep learning study)
- Essential preprocessing step to reduce queries length
  - Created a small dictionary of all common referenced URIs and used them as prefixes
  - kurtosis: 11.8 to 18.0 skew: 3.1 to 4.0
  - Max char length: 461 to 388 std: 61.7 to 49.9



DBPedia – QALD

Prefix

rdf:	http://www.w3.org/1999/02/22-rdf-syntax-ns
onto:	http://dbpedia.org/ontology/
dbp:	http://dbpedia.org/property/
res:	http://dbpedia.org/resource/
skos:	http://www.w3.org/2004/02/skos/core#
dct:	http://purl.org/dc/terms/
foaf:	http://xmlns.com/foaf/0.1/
dbc:	http://dbpedia.org/resource/Category:
rdfs:	http://www.w3.org/2000/01/rdf-schema#
yago:	http://dbpedia.org/class/yago/
prop:	http://dbpedia.org/property/
owl:	http://www.w3.org/2002/07/owl#
dbpedia2:	http://dbpedia.org/property/
xsd:	http://www.w3.org/2001/XMLSchema#
db:	http://dbpedia.org/

DBPedia – QALD

Data Sample


```
{
  "dataset" : {
    "id" : "qald-9-test-multilingual"
  },
  "questions" : [ {
    "id" : "99",
    "answertype" : "resource",
    "aggregation" : false,
    "onlydbo" : true,
    "hybrid" : false,
    "question" : [ {
      "language" : "de",
      "string" : "In welcher Zeitzone liegt Salt Lake City?",
      "keywords" : "Salt Lake City,  Zeitzone"
    }, {
      "language" : "en",
      "string" : "What is the time zone of Salt Lake City?",
      "keywords" : "Salt Lake City,  time zone"
    }, {
      "language" : "nl",
      "string" : "In welke tijdzone ligt Salt Lake City?",
      "keywords" : "Salt Lake City,  tijdzone"
    } ],
    "query" : {
      "sparql" : "PREFIX res: <http://dbpedia.org/resource/> PREFIX dbp: <http://dbpedia.org/property/>
SELECT DISTINCT ?uri WHERE { res:Salt_Lake_City <http://dbpedia.org/ontology/timeZone> ?uri }"
    },
```

Hardware

- **Tesla K80 GPU**
- **11441MiB Memory**
- **FP16 mixed-precision = Off**
- **Train T5-small in more than 2 hours**
- **Train T5-base model in about 6 hours**
- **FP16 mixed-precision = On**
- **Train T5-small in 50 minutes**
- **Train T5-base model in 2 hours**

Results

In 4807 training steps and 8 batch size

- 31% Training Loss
- 23% Validation Loss
- no over-fitting
- Our aim was to make the validation loss as low as possible

Results

- Using BERTScore, we leveraged the pre-trained contextual embeddings from BERT and match words in candidate and reference sentences by cosine similarity.
- BERTScore measures precision, recall, and F1 measure, which can help evaluate different tasks.
- This approach provides better evaluation for sentences that use synonyms or similar words.

Results

Wikidata - LC-QuAD

System	Precision	Recall	F1 score
Our Model Small	0.56	0.21	0.31
Our Model Base	0.58	0.22	0.32

Results

Wikidata - LC-QuAD

System	Precision	Recall	F1 score
Our Model Small	0.56	0.21	0.31
Our Model Base	0.58	0.22	0.32

DBPedia

System	Precision	Recall	F1 score
QAmp	0.25	0.50	0.33
WQAqua	0.22	0.38	0.28
Our Model Small	0.39	0.15	0.22
Our Model Base	0.36	0.13	0.20

Results

QAmp

- 2020
- LC-QuAD DBPedia
  QALD9 Only for Benchmarking
- Unsupervised message-passing algorithm

WQAqua

- 2021
- 3 Datasets
- Monument dataset + LC-QuAD + DBNQA
- 894,499 records
- 2-layer LSTMs, ConvS2S 15-layer, FairSEQ model
- 3 GPUs

Samples

Wikidata - LC-QuAD

```
+-----------+-------------------------------------------------------------------------------------------------+
| QUESTION  |                          Who is the country for head of state of Mahmoud Abbas?                  |
+-----------+-------------------------------------------------------------------------------------------------+
|  Target   |          select distinct ?sbj where { ?sbj wdt:P35 wd:Q127998 . ?sbj wdt:P31 wd:Q6256 }          |
+-----------+-------------------------------------------------------------------------------------------------+
| RawRESULT | select distinct ?obj where [ wd:mahmoud_abbas wdt:head_of_state ?obj . ?obj wdt:instance_of wd:country ] |
+-----------+-------------------------------------------------------------------------------------------------+
|  RESULT   |          select distinct ?obj where { wd:Q127998 wdt:P35 ?obj . ?obj wdt:P31 wd:Q6256 }          |
+-----------+-------------------------------------------------------------------------------------------------+
```

Samples

Wikidata - LC-QuAD

QUESTION	Which female actress on South Park is the voice over and is used as a singer?
Target	SELECT ?answer WHERE { wd:Q16538 wdt:P725 ?answer . ?answer wdt:P106 wd:Q177220}
RawRESULT	select ?answer where [wd:south_park wdt:voice_over ?answer . ?answer wdt:influenced_by wd:singer]
RESULT	select ?answer where { wd:Q16538 wdt:Q12280274 ?answer . ?answer wdt:P737 wd:Q177220}

Samples

Wikidata - LC-QuAD

QUESTION	What was the population of Somalia in 2009-0-0?
Target	SELECT ?obj WHERE { wd:Q1045 p:P1082 ?s . ?s ps:P1082 ?obj . ?s pq:P585 ?x filter(contains(YEAR(?x),'2009')) }
RawRESULT	select ?obj where [wd:samala p:population ?s . ?s ps:population ?obj . ?s pq:point_in_time ?x filter(contains(YEAR(?x),'2009'))]
RESULT	select ?obj where { wd:Q2216714 p:P1082 ?s . ?s ps:P1082 ?obj . ?s pq:P585 ?x filter(contains(YEAR(?x),'2009')) }

Samples

Wikidata - LC-QuAD

QUESTION	Are Taiko some kind of Japanese musical instrument?
Target	ASK WHERE { <http://dbpedia.org/resource/Taiko> a <http://dbpedia.org/class/yago/WikicatJapaneseMusicalInstruments> }
RawRESULT	ask where [wd:taiko wdt:instrument wd:japanese]
RESULT	ask where { wd:Q221769 wdt:P1303 wd:Q5287 }

Samples

Wikidata - LC-QuAD

QUESTION	What is the highest mountain in Germany?
Target	PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX onto: <http://dbpedia.org/ontology/> SELECT ?uri WHERE { ?uri rdf:type onto:Mountain ; onto:elevation ?elevation ; onto:locatedInArea <http://dbpedia.org/resource/Germany> } ORDER BY DESC(?elevation) LIMIT 1
RawRESULT	select ?ent where [?ent wdt:instance_of wd:mountain . ?ent wdt:highest_mountain ?obj . ?ent wdt:country wd:germany] order BY DESC(?obj)limit 5
RESULT	select ?ent where { ?ent wdt:P31 wd:Q8502 . ?ent wdt:Q35691748 ?obj . ?ent wdt:P17 wd:Q183} order BY DESC(?obj)limit 5

Samples

DBPedia – QALD

QUESTION	Who is the mayor of Berlin?
Target	SELECT DISTINCT ?uri WHERE { res:Berlin dbp:leader ?uri }
RESULT	select distinct ?uri where [?uri dbr:Berlin dbo:head_of_government ?uri]

Samples

DBPedia – QALD

```
+-----------+-------------------------------------------------------------------+
| QUESTION  |                 What are the nicknames of San Francisco?          |
+-----------+-------------------------------------------------------------------+
|  Target   | SELECT DISTINCT ?string WHERE { res:San_Francisco foaf:nick ?string } |
+-----------+-------------------------------------------------------------------+
|  RESULT   |  select distinct ?uri where [ ?uri dbr:San_Francio dbo:name ?uri ]  |
+-----------+-------------------------------------------------------------------+
```

Samples

DBPedia – QALD

QUESTION	How many seats does the home stadium of FC Porto have?
Target	PREFIX dbo: <http://dbpedia.org/ontology/> PREFIX dbp: <http://dbpedia.org/property/> PREFIX dbr: <http://dbpedia.org/resource/> PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#> PREFIX db: <http://dbpedia.org/> SELECT ?capacity WHERE { { dbr:FC_Porto dbo:ground ?ground . ?ground dbo:capacity ?capacity } UNION { dbr:FC_Porto dbo:ground ?ground . ?ground dbp:capacity ?capacity } }
RESULT	select (COUNT(?obj) AS ?value) [dbr:Fc_Porto dbo:home_stadium ?obj]

Access

Our fine-tuned models and customized datasets are publicly available at this huggingface page

- https://huggingface.co/yazdipour/text-to-sparql-t5-small
- https://huggingface.co/yazdipour/text-to-sparql-t5-small-qald9

The codes needed to replicate all fine-tuning processes and use of the model are accessible from this public repository

- https://github.com/yazdipour/text-to-sparql
- https://github.com/yazdipour/text-to-sparql-development

Future work

- Study on DBPedia only Datasets
- Work on cleaner and bigger Dataset
- Create a clean DBPedia dataset from LC-QUAD
- Changing hyperparameters like different learning rate
- Adding a layer of lemmatization to the end
- or improve the finding strategy of properties in the WikiData dataset

Thank you for your attention

shahriar.yazdipour@tu-ilmenau.de

https://yazdipour.com