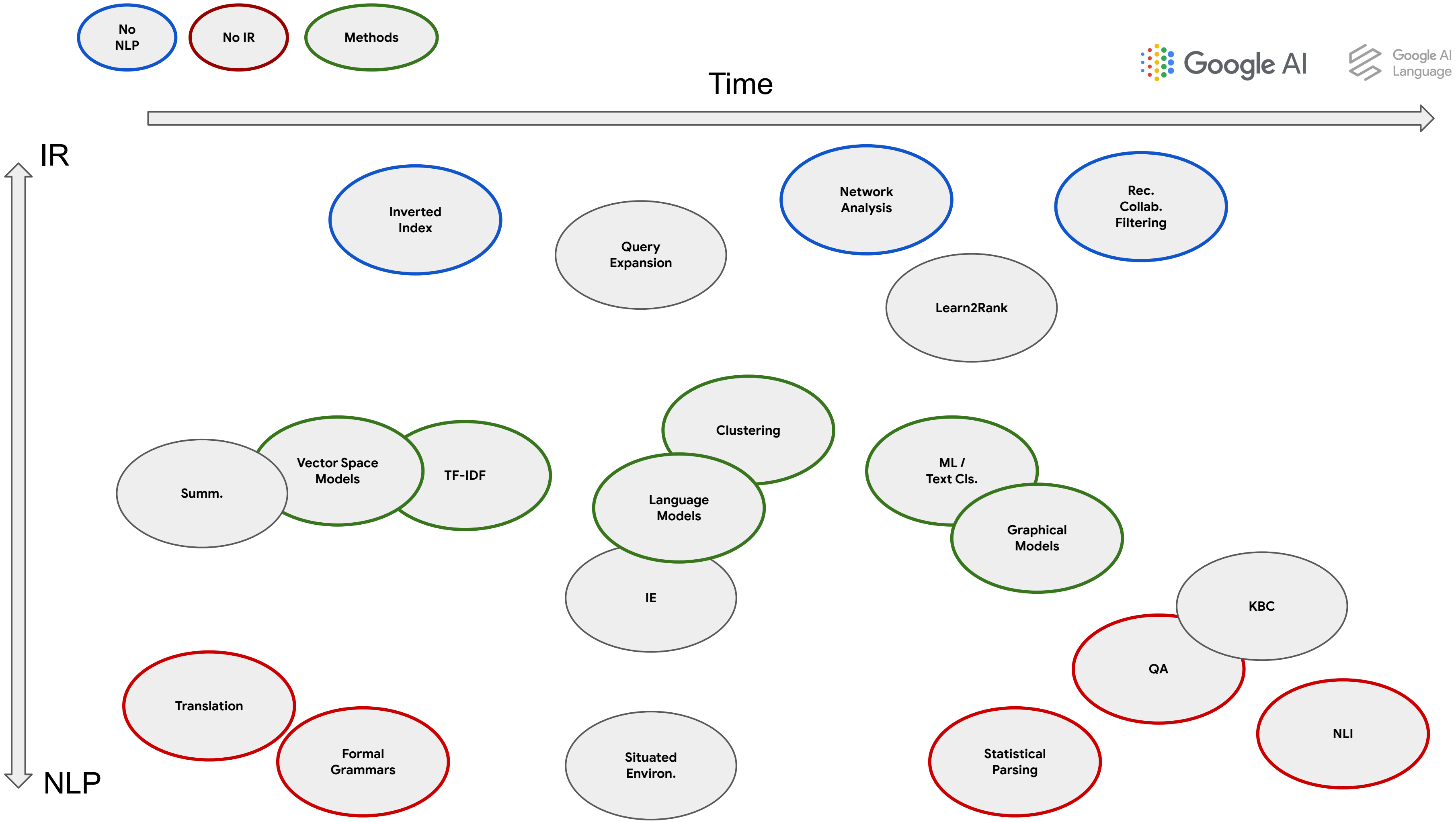
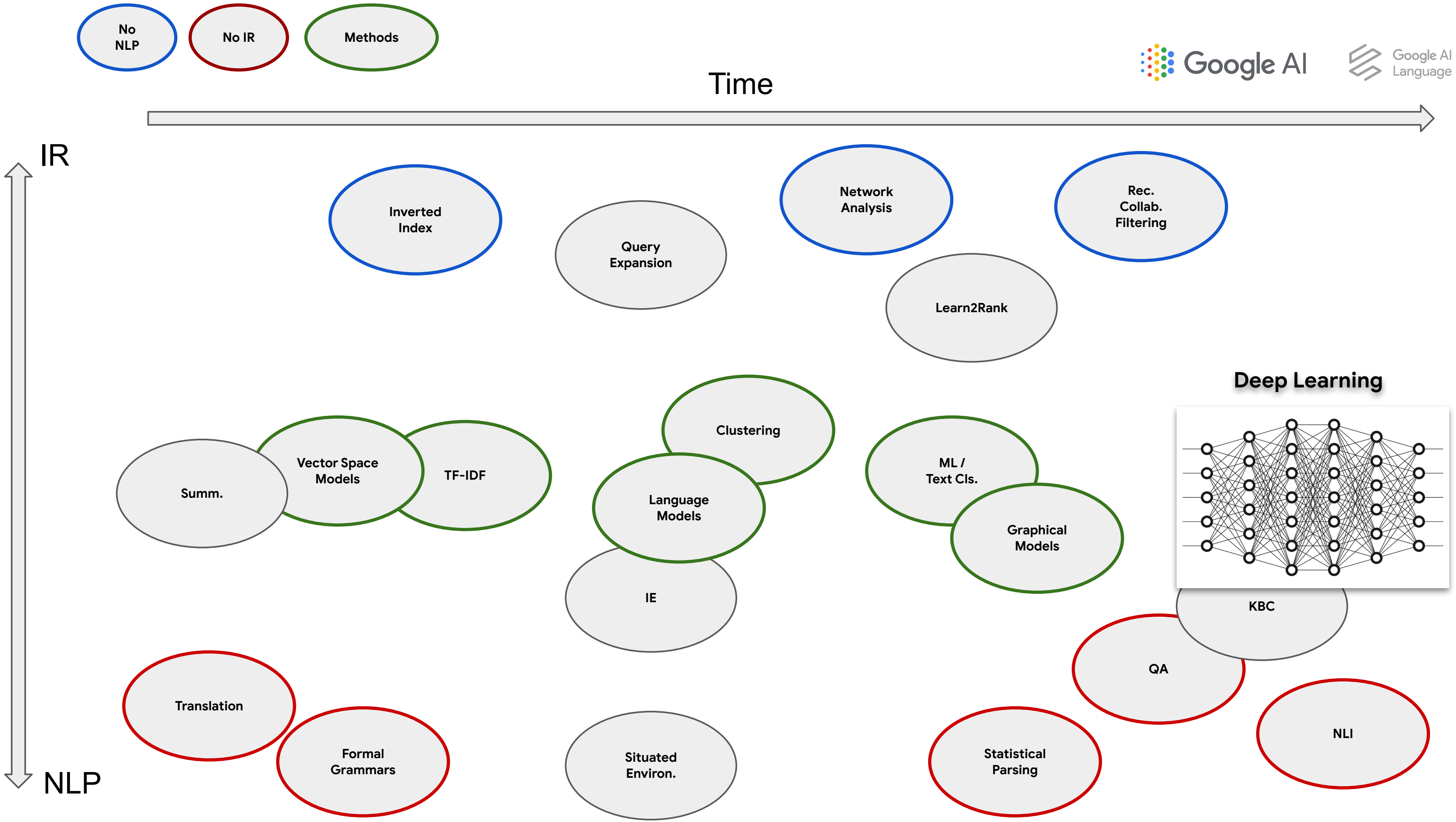
Google AI   Google AI Language

NLP & IR
How Deep Learning has Bridged the Gap

Ryan McDonald
Material from a number of Google research projects

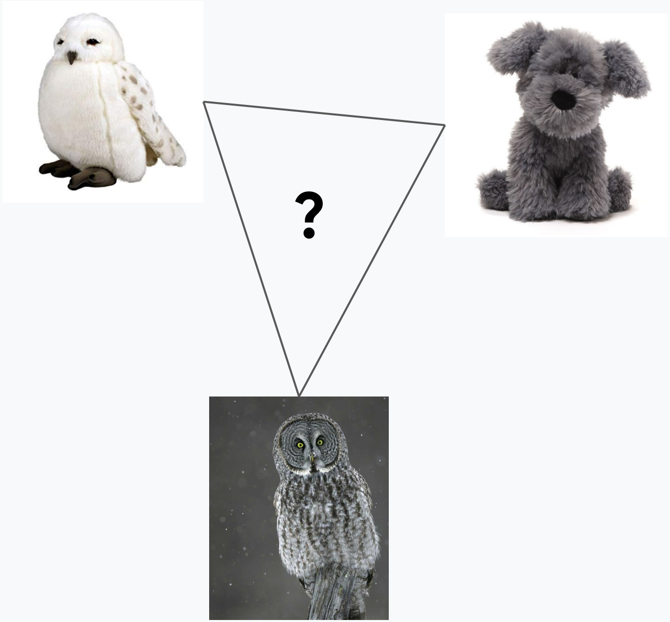


No NLP
No IR
Methods
Google AI
Google AI Language
Time
IR
NLP
Inverted Index
Query Expansion
Network Analysis
Rec. Collab. Filtering
Learn2Rank
Clustering
Vector Space Models
TF-IDF
Summ.
Language Models
ML / Text Cls.
Graphical Models
IE
KBC
QA
Translation
Formal Grammars
Situated Environ.
Statistical Parsing
NLI



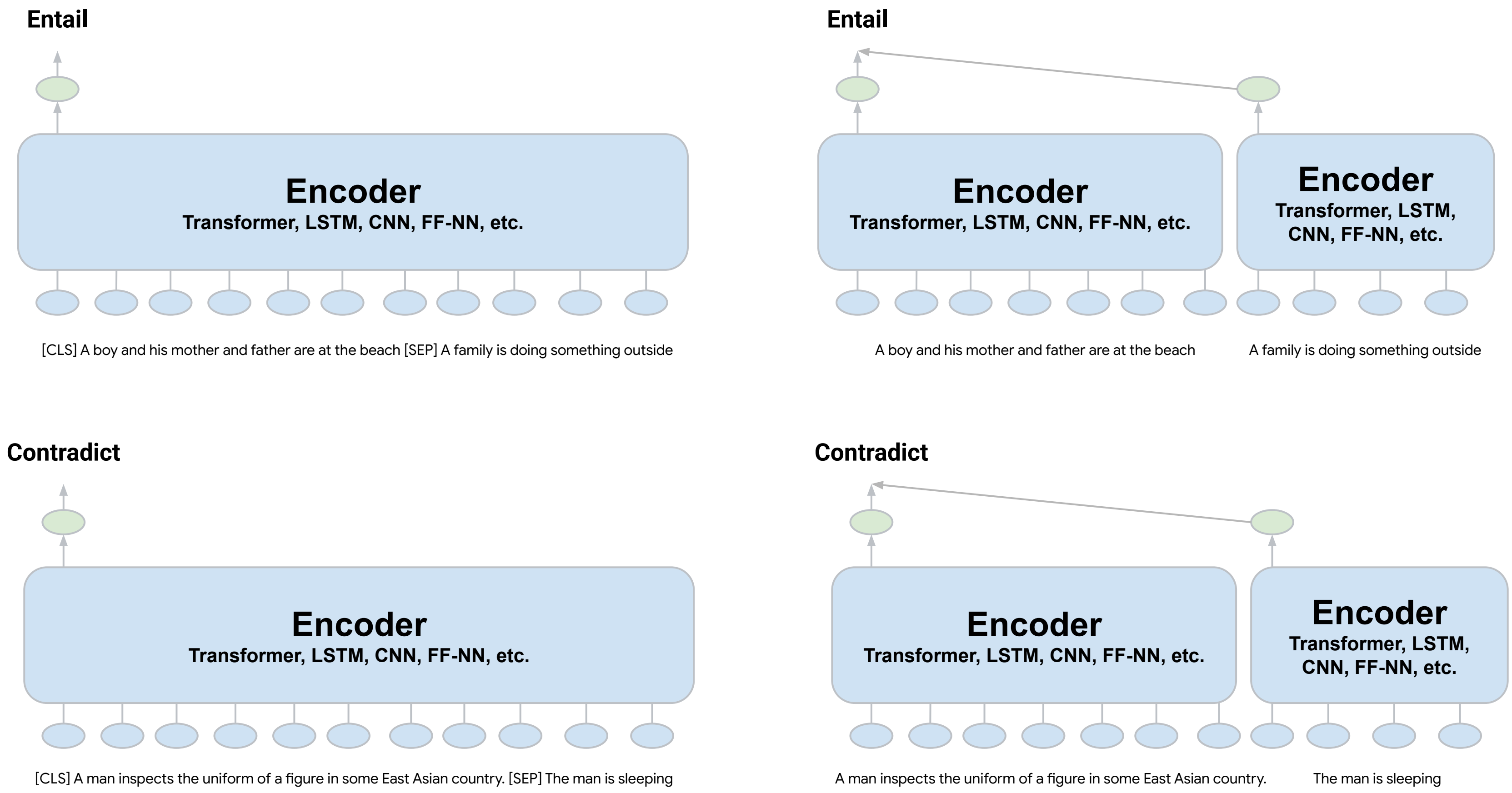
No NLP
No IR
Methods
Time
Google AI
Google AI Language
IR
NLP
Inverted Index
Query Expansion
Network Analysis
Rec. Collab. Filtering
Learn2Rank
Deep Learning
Summ.
Vector Space Models
TF-IDF
Clustering
Language Models
IE
ML / Text Cls.
Graphical Models
KBC
QA
NLI
Translation
Formal Grammars
Situated Environ.
Statistical Parsing

f(text, text)
f(text, entity)
f(text, image)
...
f(text, object)



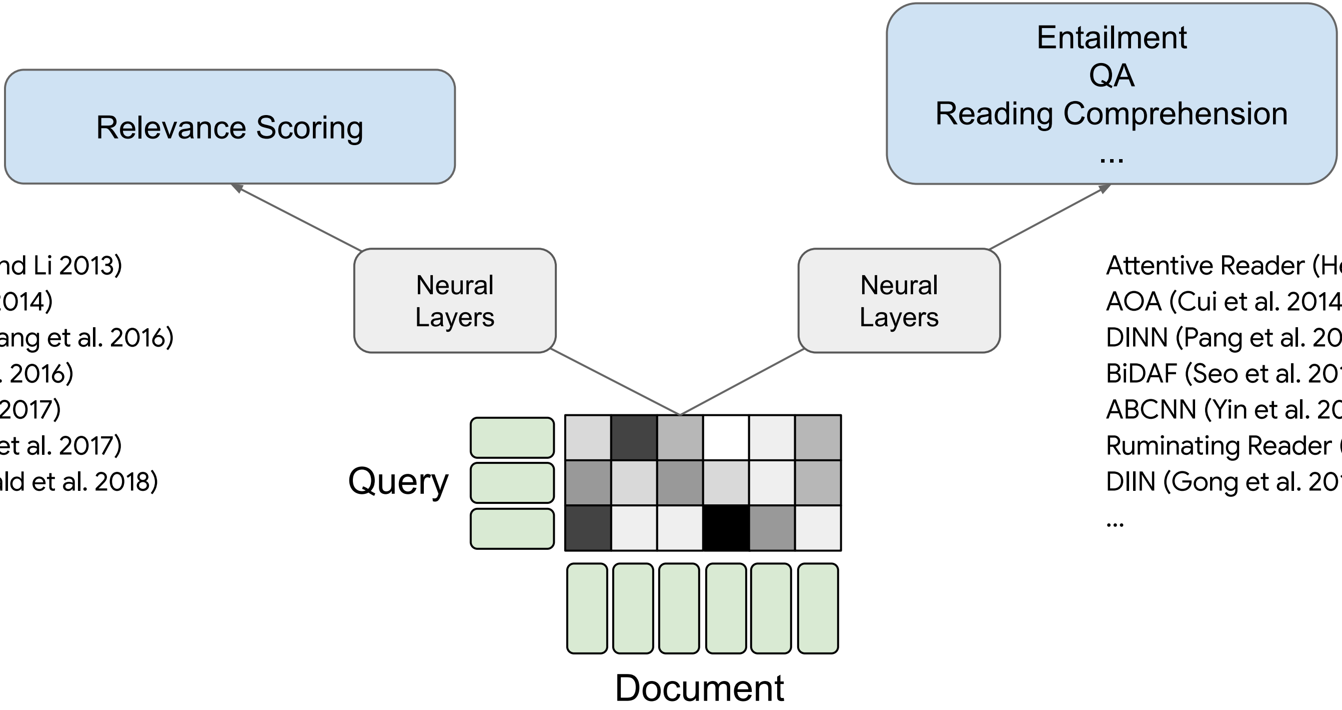
Google AI

f(text, text)


Entail
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
[CLS] A boy and his mother and father are at the beach [SEP] A family is doing something outside
Entail
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
A boy and his mother and father are at the beach
A family is doing something outside
Contradict
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
[CLS] A man inspects the uniform of a figure in some East Asian country. [SEP] The man is sleeping
Contradict
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
Encoder
Transformer, LSTM, CNN, FF-NN, etc.
A man inspects the uniform of a figure in some East Asian country.
The man is sleeping

Cross-attention (AKA one-tower; AKA interaction-based)

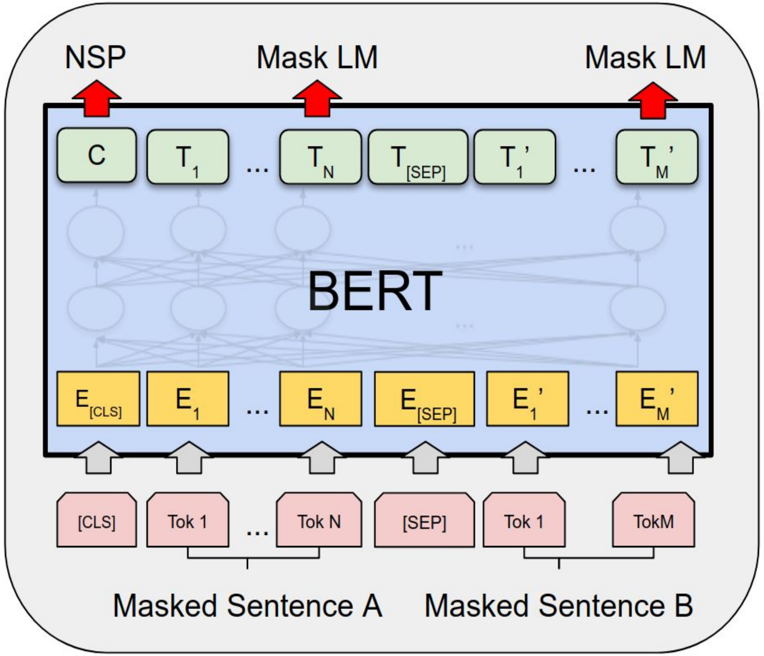
DeepMatch (Lu and Li 2013)
ARC-II (Hu et al. 2014)
MatchPyramid (Pang et al. 2016)
DRMM (Guo et al. 2016)
PACRR (Hui et al. 2017)
DeepRank (Pang et al. 2017)
PDRMM (McDonald et al. 2018)
...



Attentive Reader (Hermann et al. 2013)
AOA (Cui et al. 2014)
DINN (Pang et al. 2016)
BiDAF (Seo et al. 2016)
ABCNN (Yin et al. 2016)
Ruminating Reader (Gong et al. 2017)
DIIN (Gong et al. 2018)
...

BERT: Transformers + Pre-training + Fine-Tuning

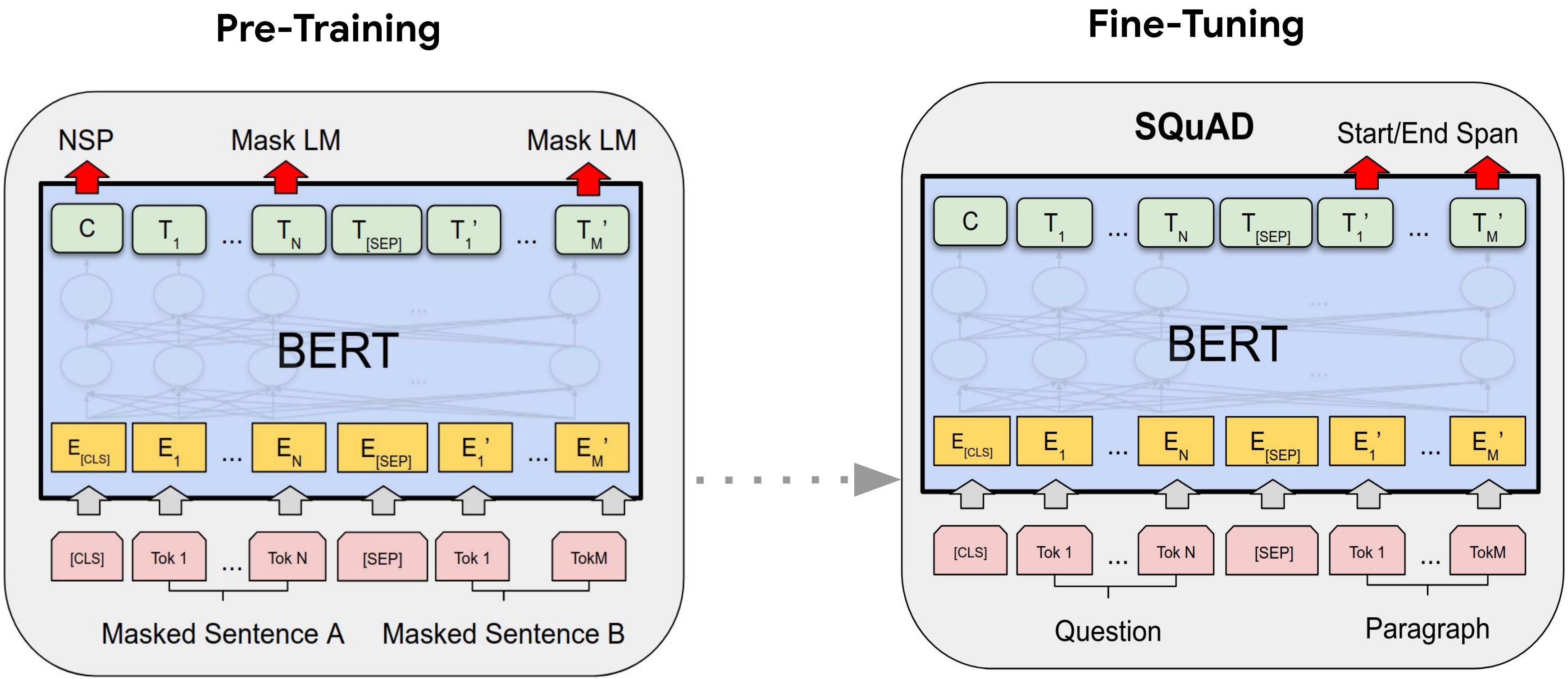
Pre-Training



Devlin et al. 2019

BERT: Transformers + Pre-training + Fine-Tuning

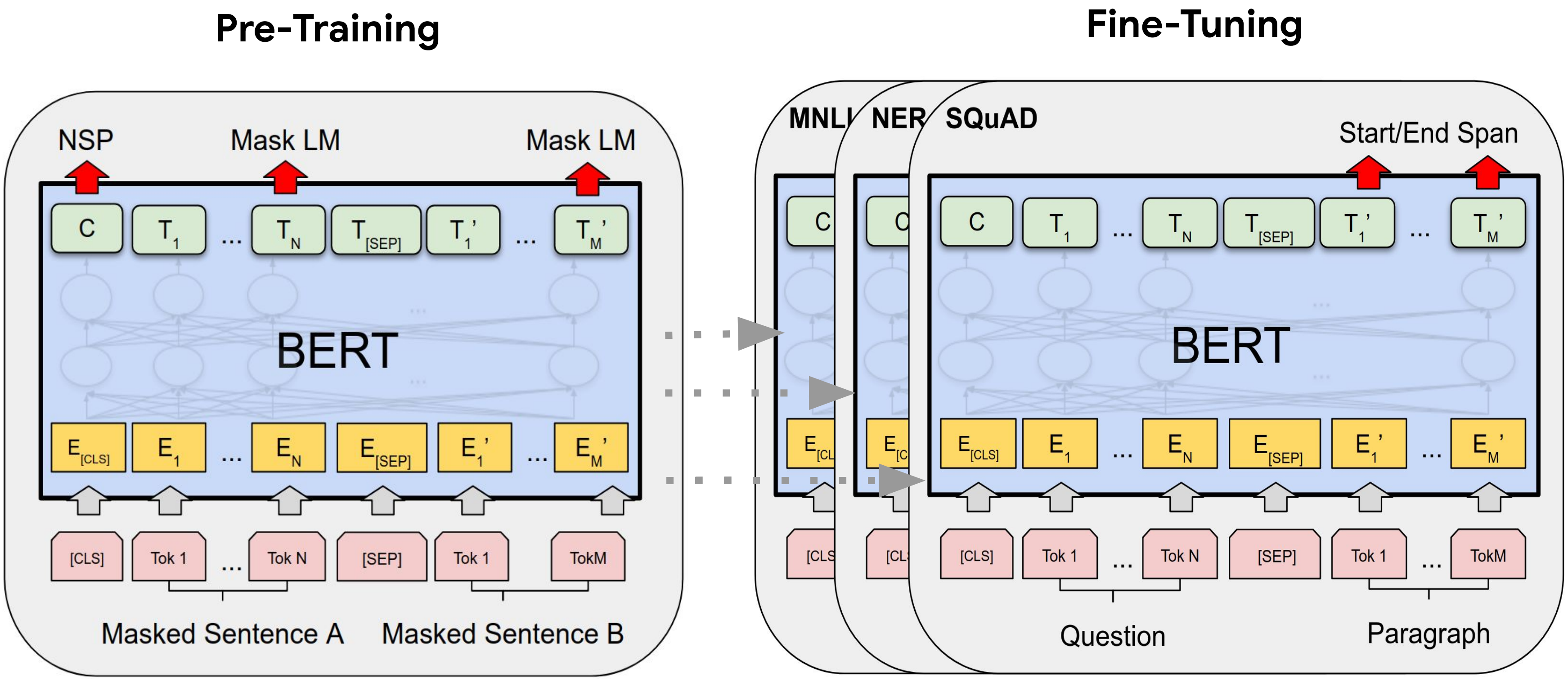
Google AI
Google AI Language



Devlin et al. 2019

BERT: Transformers + Pre-training + Fine-Tuning

Google AI
Google AI Language



Devlin et al. 2019

Transformers + Pre-training -- new dawn of NLP

Google AI

Google AI Language

System	MNLI-(m/mm)	QQP	QNLI	SST-2	CoLA	STS-B	MRPC	RTE	**Average**
	392k	363k	108k	67k	8.5k	5.7k	3.5k	2.5k	-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**92.7**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**82.1**

Devlin et al. 2019
Yang et al. 2019
Lan et al. 2019

Transformers + Pre-training -- new dawn of NLP

Google AI

Google AI Language

System	MNLI-(m/mm) 392k	QQP 363k	QNL 108k	erage -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	4.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	1.0
OpenAI GPT	82.1/81.4	70.3	87.4	5.1
$\text{BERT}_{\text{BASE}}$	84.6/83.4	71.2	90.5	9.6
$\text{BERT}_{\text{LARGE}}$	**86.7/85.9**	**72.1**	**92.7**	**2.1**

Dataset	XLNet-Large (as in paper)	XLNet-Large -wikibooks	BERT-Large -wikibooks best of 3 variants
SQuAD1.1 EM	89.0	88.2	86.7 (II)
SQuAD1.1 F1	94.5	94.0	92.8 (II)
SQuAD2.0 EM	86.1	85.1	82.8 (II)
SQuAD2.0 F1	88.8	87.8	85.5 (II)
RACE	81.8	77.4	75.1 (II)
MNLI	89.8	88.4	87.3 (II)
QNLI	93.9	93.9	93.0 (II)
QQP	91.8	91.8	91.4 (II)
RTE	83.8	81.2	74.0 (III)
SST-2	95.6	94.4	94.0 (II)
MRPC	89.2	90.0	88.7 (III)
CoLA	63.6	65.2	63.7 (II)
STS-B	91.8	91.1	90.2 (III)

Devlin et al. 2019
Yang et al. 2019
Lan et al. 2019

Transformers + Pre-training -- new dawn of NLP

System	MNLI-(m/mm) 392k	QQP 363k	QNL 108k	erage
				-
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	4.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	1.0
OpenAI GPT	82.1/81.4	70.3	87.4	5.1
BERT$_{\text{BASE}}$				9.6
BERT$_{\text{LARGE}}$				**2.1**

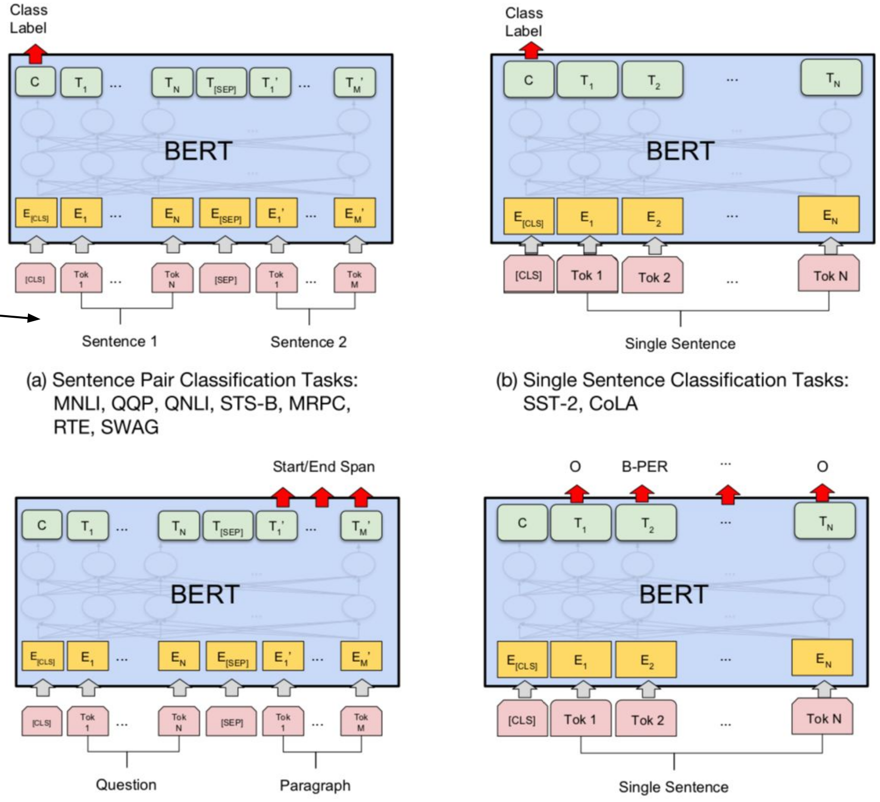
Dataset	**XLNet-Large (as in paper)**	**XLNet-Large -wikibooks**	**BERT-Large -wikibooks** best of 3 variants
SQuAD1.1 EM	89.0	88.2	86.7 (II)
SQuAD1.1 F1	94.5	94.0	92.8 (II)
SQuAD2.0 EM	86.1	85.1	82.8 (II)
SQuAD2.0 F1	88.8	87.8	85.5 (II)

Models	MNLI	QNLI	QQP	RTE	SST	MRPC	CoLA	STS	WNLI	Avg
Single-task single models on dev										
BERT-large	86.6	92.3	91.3	70.4	93.2	88.0	60.6	90.0	-	-
XLNet-large	89.8	93.9	91.8	83.8	95.6	89.2	63.6	91.8	-	-
RoBERTa-large	90.2	94.7	**92.2**	86.6	96.4	**90.9**	68.0	92.4	-	-
ALBERT (1M)	90.4	95.2	92.0	88.1	96.8	90.2	68.7	92.7	-	-
ALBERT (1.5M)	**90.8**	**95.3**	**92.2**	**89.2**	**96.9**	**90.9**	**71.4**	**93.0**	-	-
Ensembles on test (from leaderboard as of Sept. 16, 2019)										
ALICE	88.2	95.7	**90.7**	83.5	95.2	92.6	**69.2**	91.1	80.8	87.0
MT-DNN	87.9	96.0	89.9	86.3	96.5	92.7	68.4	91.1	89.0	87.6
XLNet	90.2	98.6	90.3	86.3	96.8	93.0	67.8	91.6	90.4	88.4
RoBERTa	90.8	98.9	90.2	88.2	96.7	92.3	67.8	92.2	89.0	88.5
Adv-RoBERTa	91.1	98.8	90.3	88.7	96.8	93.1	68.0	92.4	89.0	88.8
ALBERT	**91.3**	**99.2**	90.5	**89.2**	**97.1**	**93.4**	69.1	**92.5**	**91.8**	**89.4**

Devlin et al. 2019
Yang et al. 2019
Lan et al. 2019

BERT: Fine-tuning Paradigms

This can be a relevance scoring model



(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

(b) Single Sentence Classification Tasks: SST-2, CoLA

(c) Question Answering Tasks: SQuAD v1.1

(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

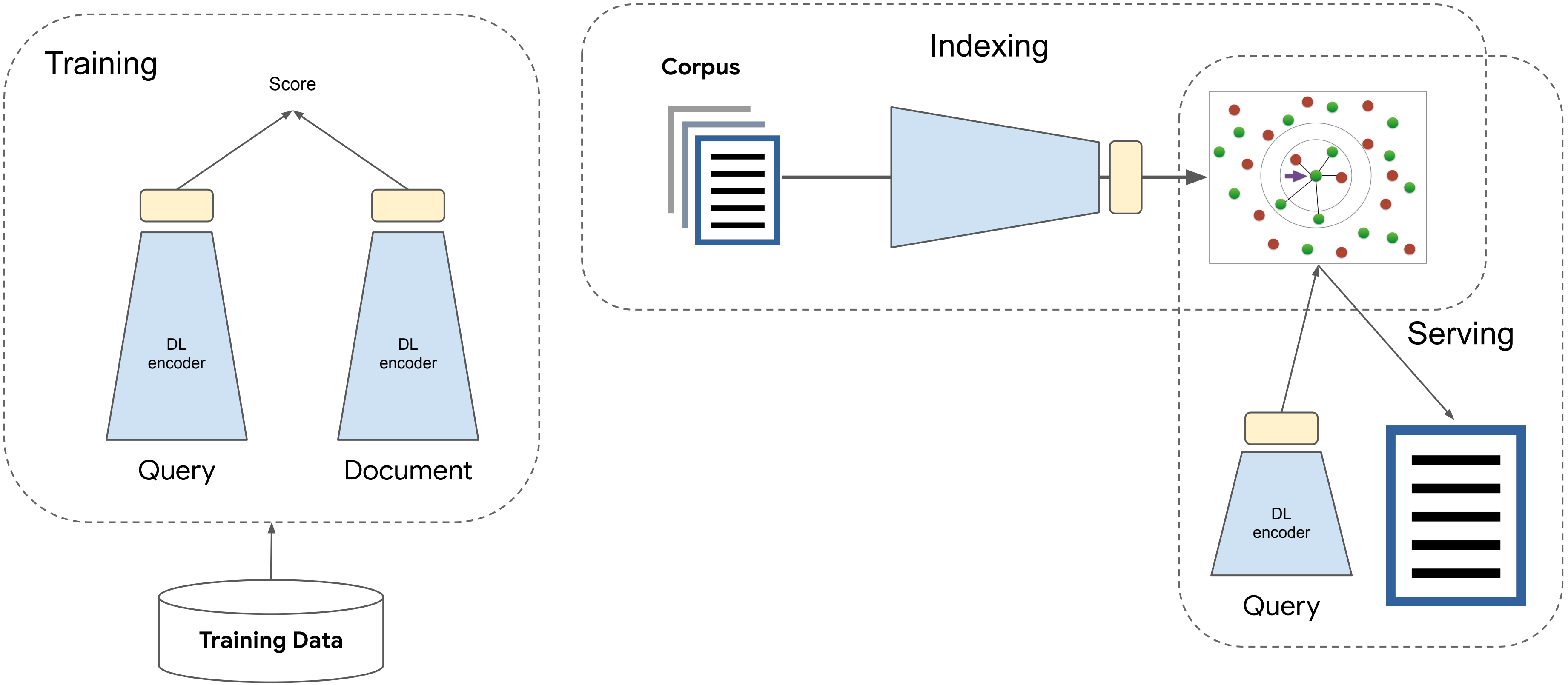
BERT 4 Document Relevance Scoring

Google AI
Google AI Language

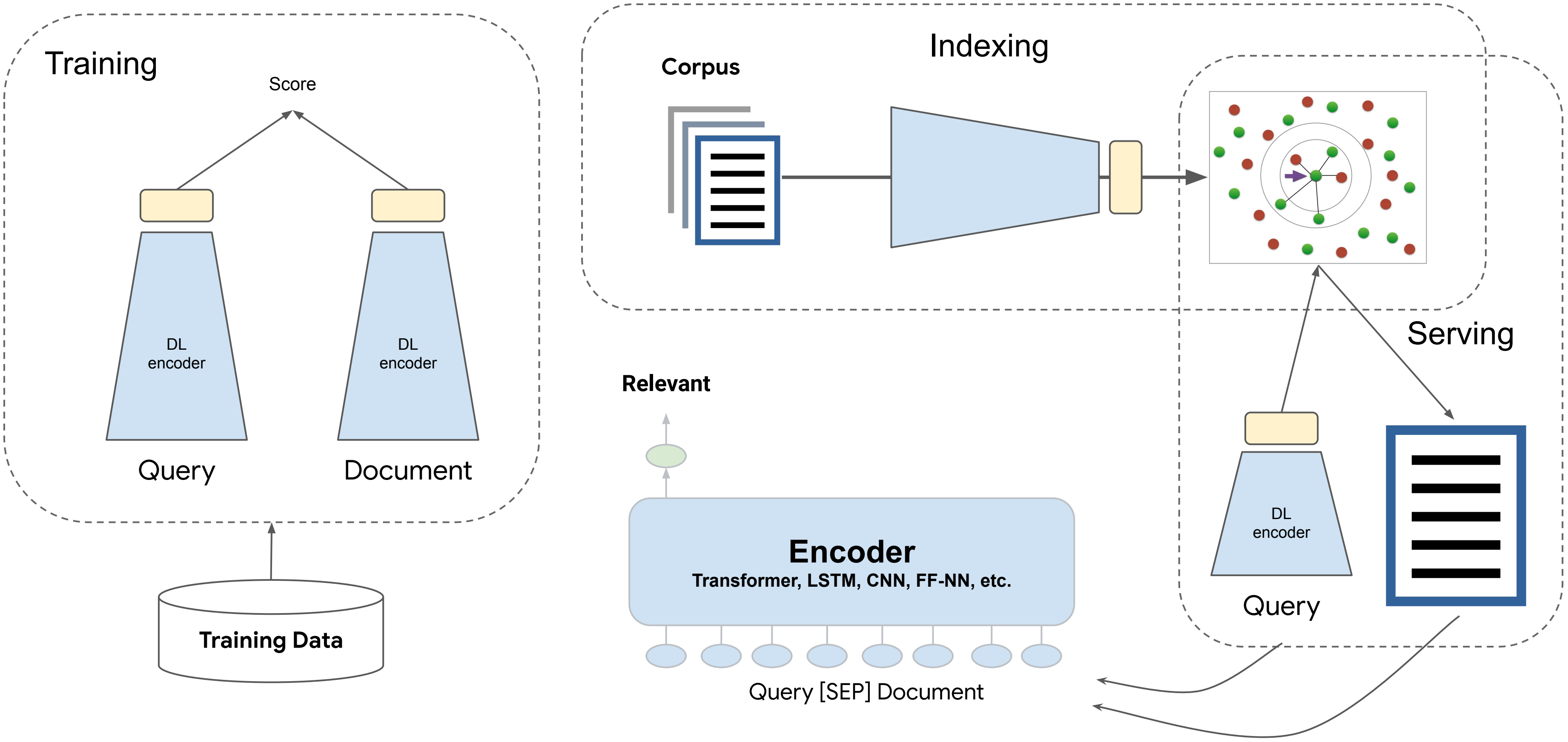
CEDR: Macaveney et al. 2019

		Robust04		WebTrack 2012–14	
Ranker	Input Representation	P@20	nDCG@20	nDCG@20	ERR@20
BM25	n/a	0.3123	0.4140	0.1970	0.1472
SDM [13]	n/a	0.3749	0.4353	-	-
TREC-Best	n/a	**0.4386**	**0.5030**	0.2855	**0.2530**
ConvKNRM	GloVe	0.3349	0.3806	[B] 0.2547	[B] 0.1833
Vanilla BERT	BERT (fine-tuned)	[BC] 0.4042	[BC] 0.4541	**[BC] 0.2895**	[BC] 0.2218
PACRR	GloVe	0.3535	[C] 0.4043	0.2101	0.1608
PACRR	ELMo	[C] 0.3554	[C] 0.4101	[BG] 0.2324	[BG] 0.1885
PACRR	BERT	[C] 0.3650	[C] 0.4200	0.2225	0.1817
PACRR	BERT (fine-tuned)	[BCVG] 0.4492	[BCVG] 0.5135	[BCG] 0.3080	[BCG] 0.2334
CEDR-PACRR	BERT (fine-tuned)	**[BCVG] 0.4559**	**[BCVG] 0.5150**	**[BCVGN] 0.3373**	**[BCVGN] 0.2656**
KNRM	GloVe	0.3408	0.3871	[B] 0.2448	0.1755
KNRM	ELMo	[C] 0.3517	[CG] 0.4089	0.2227	0.1689
KNRM	BERT	[BCG] 0.3817	[CG] 0.4318	[B] 0.2525	[B] 0.1944
KNRM	BERT (fine-tuned)	[BCG] 0.4221	[BCVG] 0.4858	[BCVG] 0.3287	[BCVG] 0.2557
CEDR-KNRM	BERT (fine-tuned)	**[BCVGN] 0.4667**	**[BCVGN] 0.5381**	**[BCVG] 0.3469**	**[BCVG] 0.2772**
DRMM	GloVe	0.2892	0.3040	0.2215	0.1603
DRMM	ELMo	0.2867	0.3137	[B] 0.2271	0.1762
DRMM	BERT	0.2878	0.3194	[BG] 0.2459	[BG] 0.1977
DRMM	BERT (fine-tuned)	[CG] 0.3641	[CG] 0.4135	[BG] 0.2598	[B] 0.1856
CEDR-DRMM	BERT (fine-tuned)	**[BCVGN] 0.4587**	**[BCVGN] 0.5259**	**[BCVGN] 0.3497**	**[BCVGN] 0.2621**

Dual Encoder Retrieval (AKA two-tower; AKA relevance-based)

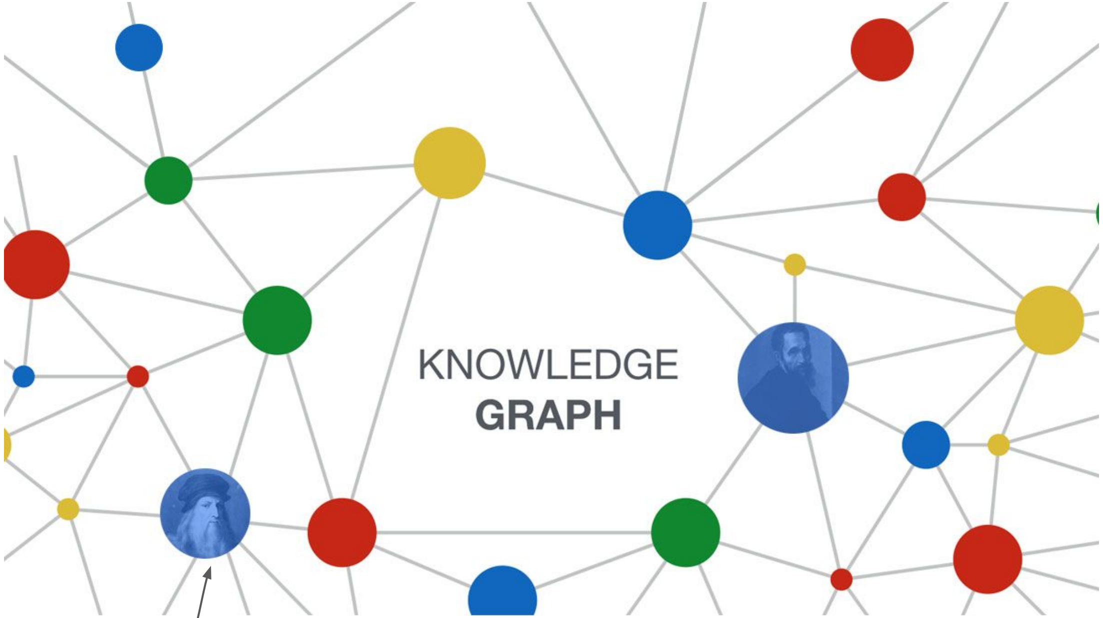


Dual Encoder Retrieval (AKA two-tower; AKA relevance-based)



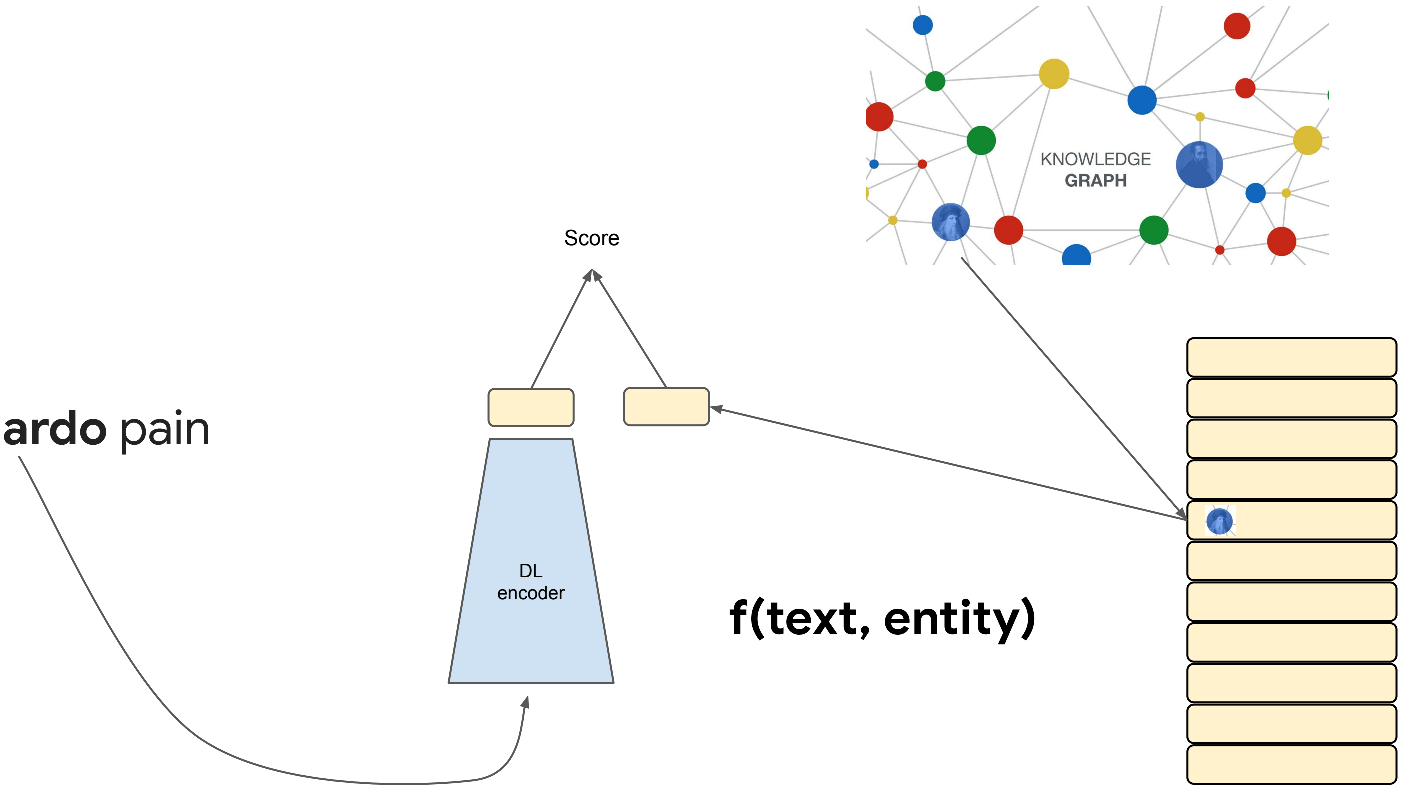
Classification with massive output spaces

When did **Leonardo** pain the Mona Lisa?



Embedding **Objects** vs. Descriptions

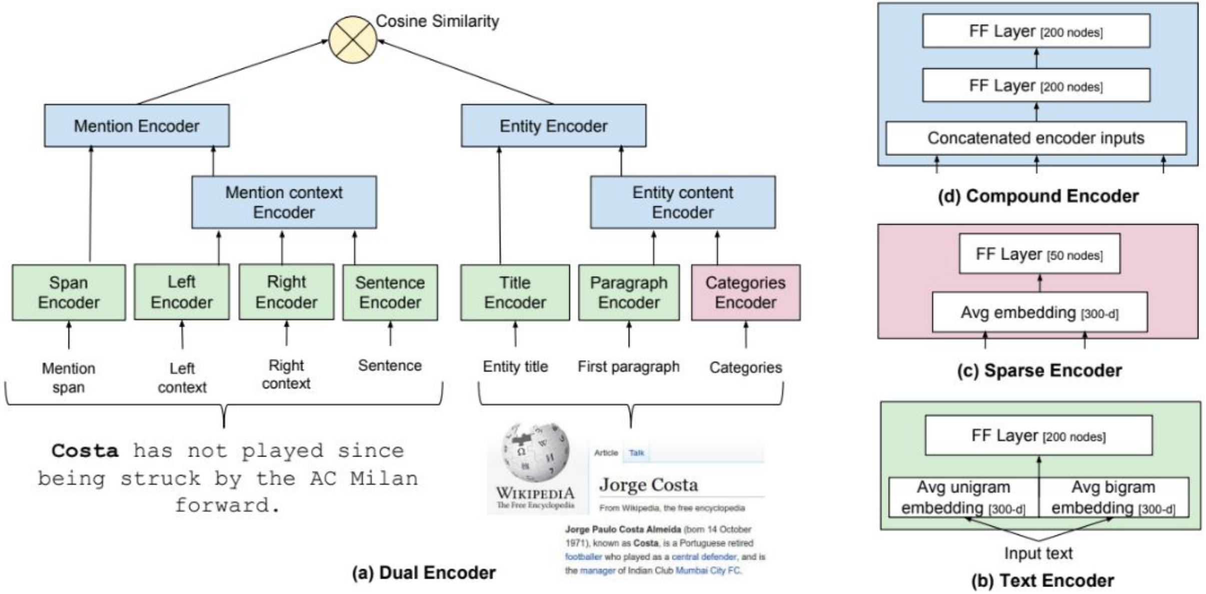
When did **Leonardo** pain the Mona Lisa?



f(text, entity)

Embedding Objects vs. Descriptions

Dataset	AT-Prior	R@1 DE-RN	DE-MN
CoNLL	65.71	40.87	**77.93**
ACE2004	80.93	55.64	**87.55**
AQUAINT	82.64	54.55	**86.78**
MSNBC	65.62	42.11	**75.30**
WikiSample	79.01	59.17	**84.06**
TACKBP 2009	69.40	51.34	**78.60**
TACKBP 2010	72.45	48.63	**87.35**
TACKBP 2011	55.52	35.85	**73.04**
TACKBP 2012	26.45	21.85	**49.91**
Wikinews	86.66	66.25	**91.56**
Average	68.44	47.63	**79.21**

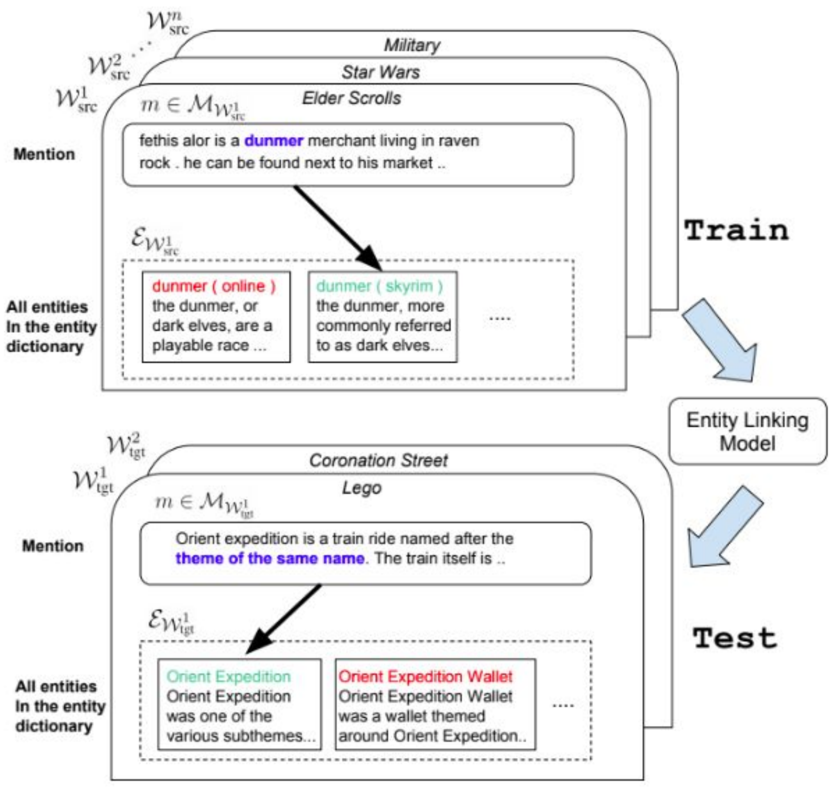


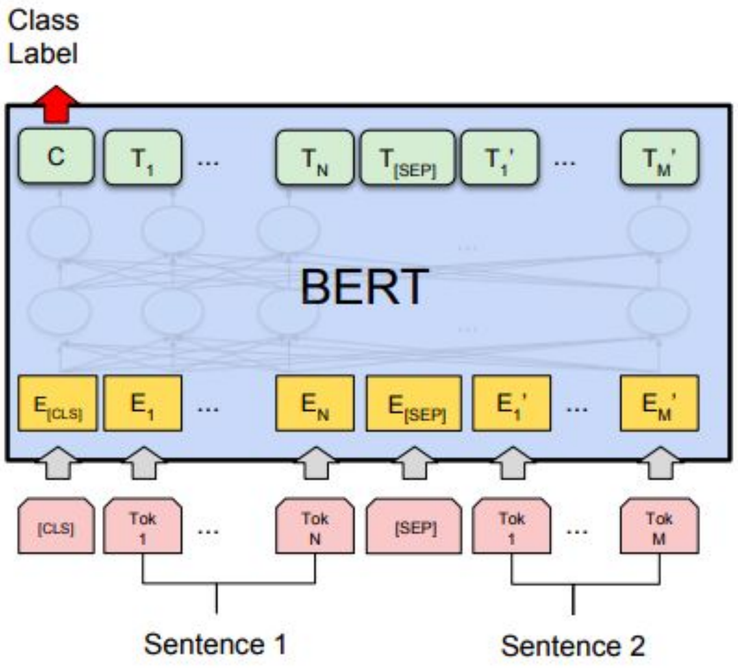
Gillick et al. 2019

f(text, text)

Zero-Shot Entity Linking

Lee et al. 2019





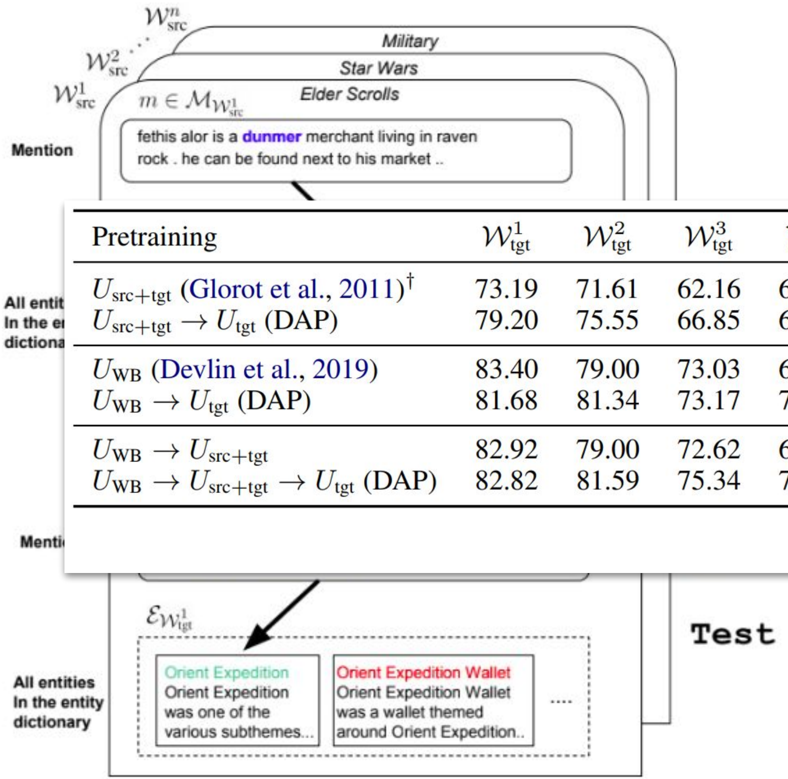
Fethis alor is a dunmer* merchant living in raven rock.

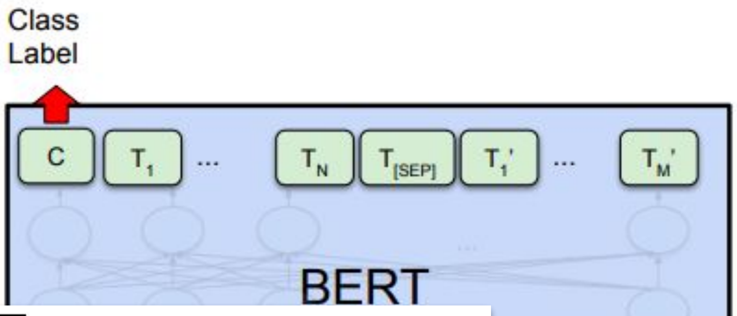
* *special mention vector added to indicate this is the focus mention*

dunmer (skyrim) the dunmer, more commonly referred to as dark elves

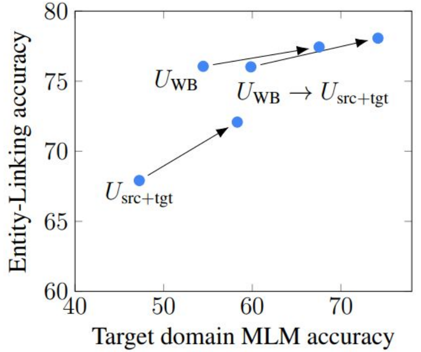
Zero-Shot Entity Linking

Lee et al. 2019





Pretraining	$\mathcal{W}^1_{\text{tgt}}$	$\mathcal{W}^2_{\text{tgt}}$	$\mathcal{W}^3_{\text{tgt}}$	$\mathcal{W}^4_{\text{tgt}}$	Avg
$U_{\text{src+tgt}}$ (Glorot et al., 2011)[†]	73.19	71.61	62.16	64.69	67.91
$U_{\text{src+tgt}} \rightarrow U_{\text{tgt}}$ (DAP)	79.20	75.55	66.85	66.72	**72.08**
U_{WB} (Devlin et al., 2019)	83.40	79.00	73.03	68.82	76.06
$U_{\text{WB}} \rightarrow U_{\text{tgt}}$ (DAP)	81.68	81.34	73.17	71.97	**77.04**
$U_{\text{WB}} \rightarrow U_{\text{src+tgt}}$	82.92	79.00	72.62	69.55	76.02
$U_{\text{WB}} \rightarrow U_{\text{src+tgt}} \rightarrow U_{\text{tgt}}$ (DAP)	82.82	81.59	75.34	72.52	**78.07**

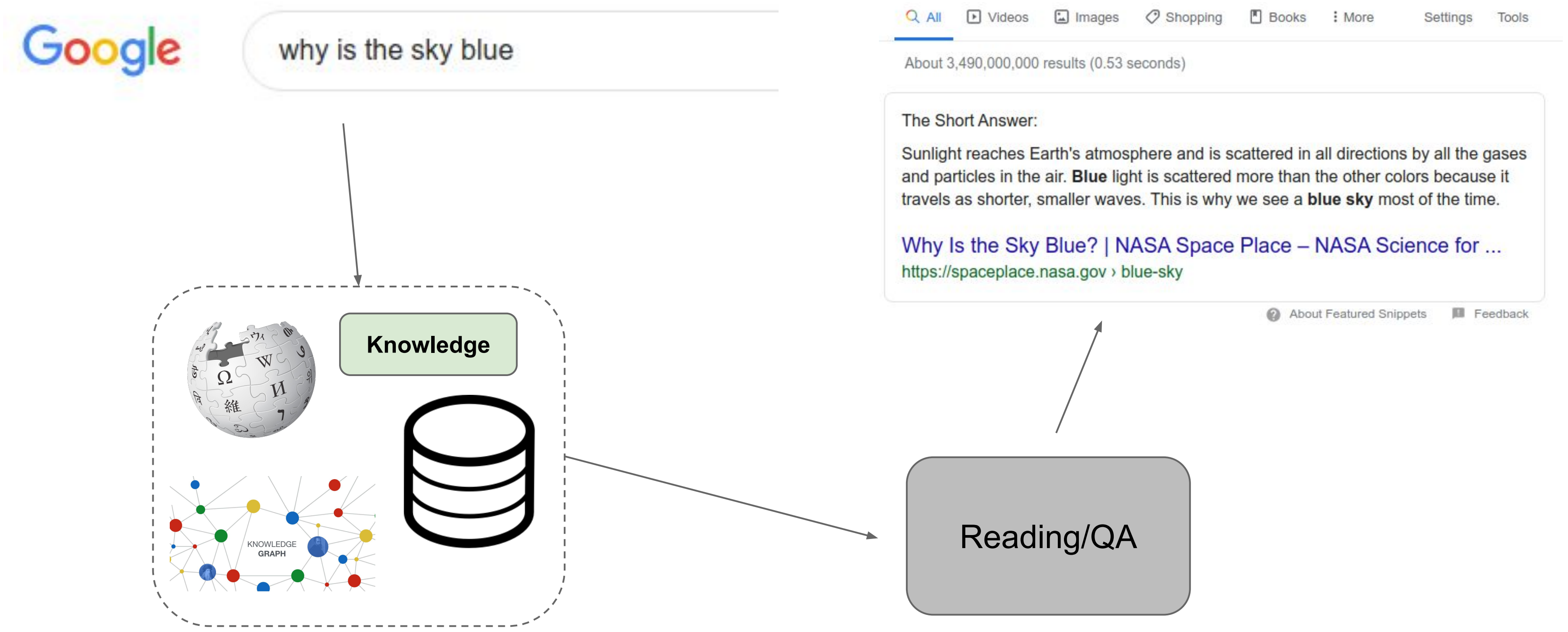


Fethis alor is a dunmer*
merchant living in raven rock.

* *special mention vector added to indicate this is the focus mention*

dunmer (skyrim) the dunmer, more commonly referred to as dark elves

Knowledge Retrieval ⟶ Comprehension



Pipelined Approach

Retrieval | **Reading Comprehension**

What does the zip in zip code stand for?

...The term ZIP is an acronym for Zone Improvement Plan...

Subject: /m/0dllr
Predicate: /common/topic/alias
Object: Zone Improvement Plan

Zone Improvement Plan

End-to-End

Latent Retrieval

$$P(a|q) = \frac{s(q,a)}{\sum_{a' \in \textsc{Wikipedia}} s(q,a')}$$

Input

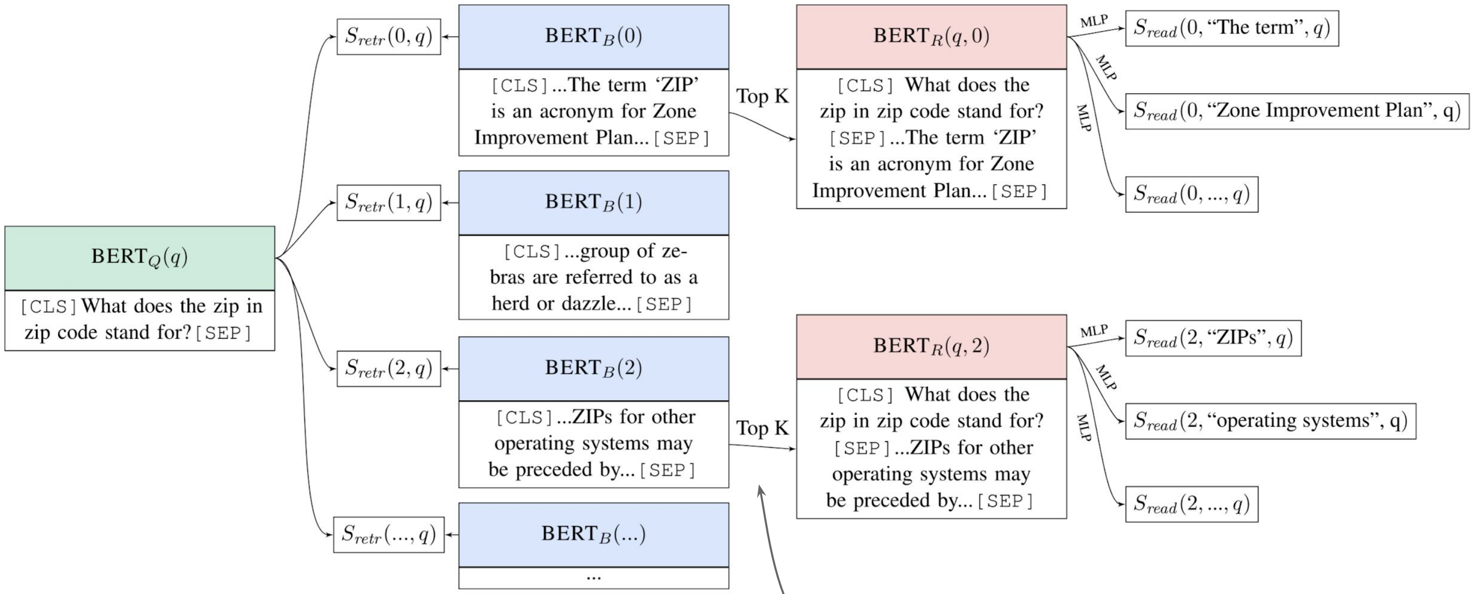
What does the zip in zip code stand for?

Knowledge

Output

Zone Improvement Plan

ORQA Overview



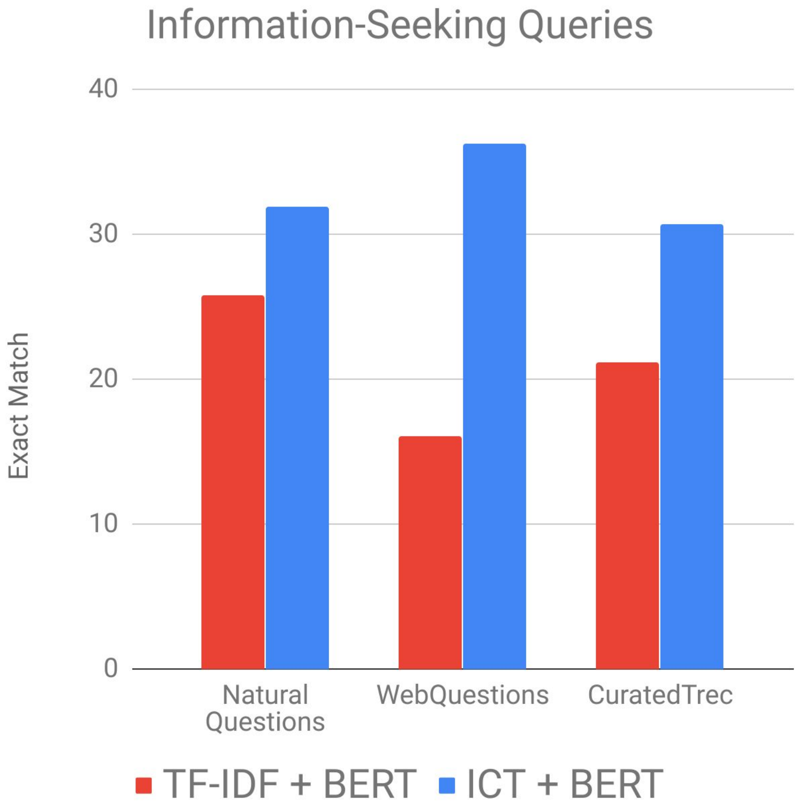
$$P(a|q) = \frac{s(q,a)}{\sum_{a' \in \text{WIKIPEDIA}} s(q,a')}$$

Approx w/ top-K

Lee et al. 2019

Results

End-to-end learning is crucial for information-seeking queries!



Lee et al. 2019

Sequential Question Answering

Building	City	Floors
First Canadian Place	Toronto	72
Commerce Court West	Toronto	57
Tour de la Bourse	Montreal	47
Place Ville-Marie	Montreal	44

What are the buildings in Toronto?
First Canadian Place, Commerce Court West

(Iyyer et al, 2017)

Sequential Question Answering

Building	City	Floors
First Canadian Place	Toronto	72
Commerce Court West	Toronto	57
Tour de la Bourse	Montreal	47
Place Ville-Marie	Montreal	44

What are the buildings in Toronto?
First Canadian Place, Commerce Court West
Of those, which buildings have more than 60 floors?
First Canadian Place

(Iyyer et al, 2017)

Sequential Question Answering

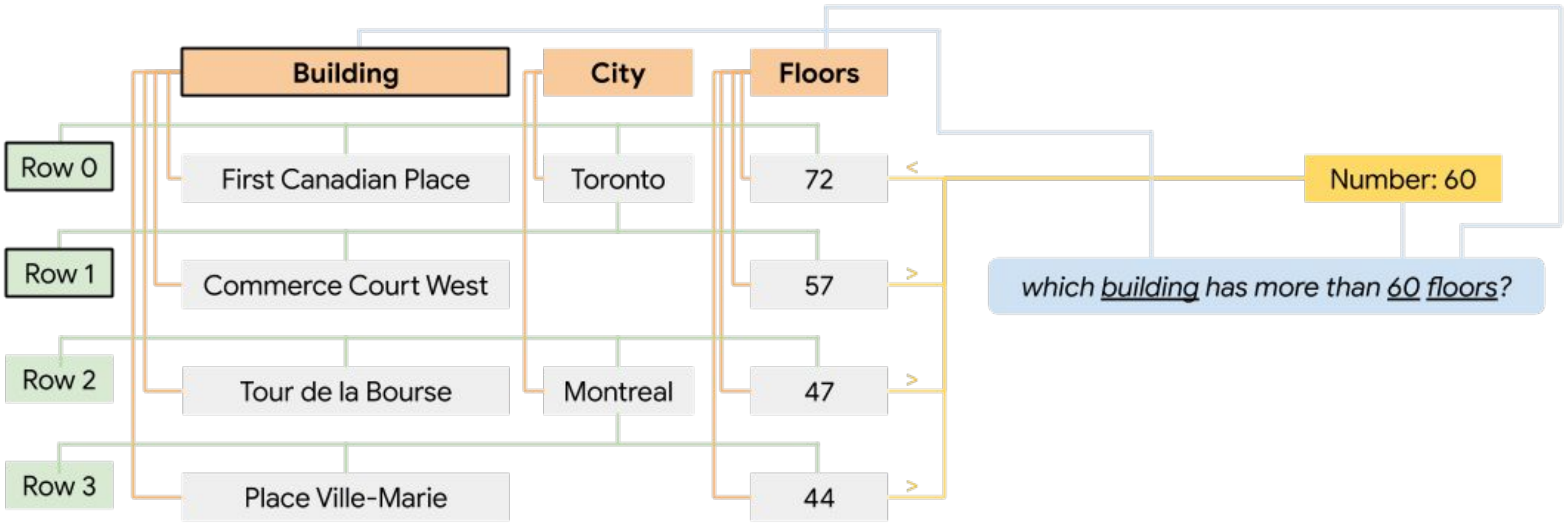
Building	City	Floors
First Canadian Place	Toronto	72
Commerce Court West	Toronto	57
Tour de la Bourse	Montreal	47
Place Ville-Marie	Montreal	44

What are the buildings in Toronto?
First Canadian Place, Commerce Court West
Of those, which buildings have more than 60 floors?
First Canadian Place
How many floors does it have?
72

(Iyyer et al, 2017)

Graph Transformer

Müller et al. 2019

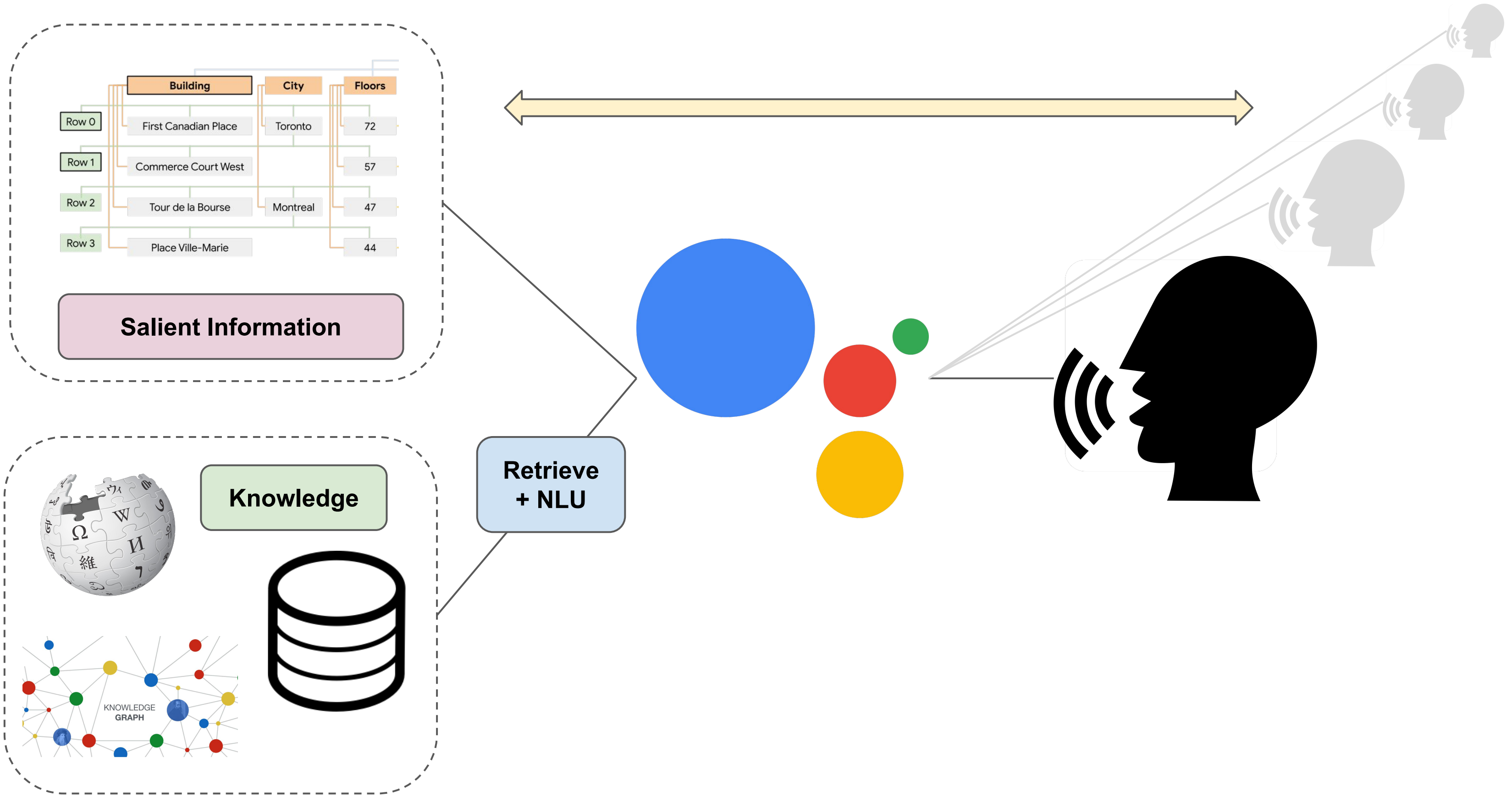


Results on SQA

Model	ALL	Q1	Q2	Q3
Iyyer et al. (2017) †	44.7	70.4	41.1	23.6
Sun et al. (2018) † *	45.6	70.3	42.6	24.8
Müller et al. (2019) † *	55.1	$\boxed{67.2}$	52.7	$\boxed{46.8}$

go/sqa-emnlp-2019

Retrieval + NLP

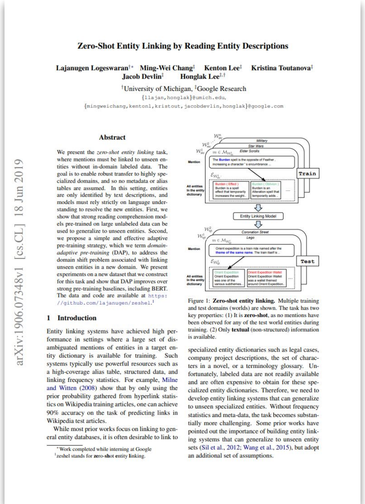


Thanks!


arXiv:1906.00300v3 [cs.CL] 27 Jun 2019

Latent Retrieval for Weakly Supervised Open Domain Question Answering

Kenton Lee Ming-Wei Chang Kristina Toutanova
Google Research
Seattle, WA
{kentonl,mingweichang,kristout}@google.com


arXiv:1906.07348v1 [cs.CL] 18 Jun 2019

Zero-Shot Entity Linking by Reading Entity Descriptions

Lajanugen Logeswaran Ming-Wei Chang Kenton Lee Kristina Toutanova Jacob Devlin Honglak Lee
University of Michigan, Google Research


arXiv:1810.04805v2 [cs.CL] 24 May 2019

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova
Google AI Language
{jacobdevlin,mingweichang,kentonl,kristout}@google.com


arXiv:1909.10506v1 [cs.CL] 23 Sep 2019

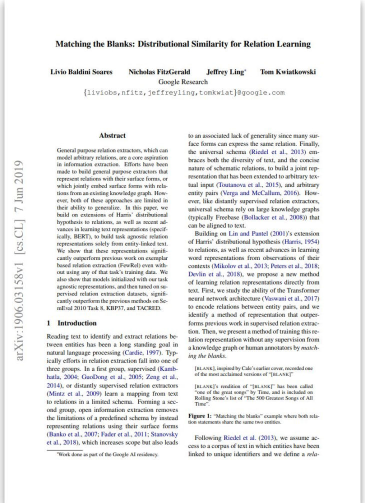
Learning Dense Representations for Entity Retrieval

Daniel Gillick Sayali Kulkarni Larry Lansing Alessandro Presta Jason Baldridge Eugene Ie Diego Garcia-Olano
Google Research, University of Texas at Austin


arXiv:1908.11787v1 [cs.CL] 30 Aug 2019

Answering Conversational Questions on Structured Data without Logical Forms

Thomas Müller, Francesco Piccinno, Massimo Nicosia, Peter Shaw, Yasemin Altun
Google Research
{thomasmueller,piccinno,massimon,petershaw,altun}@google.com


arXiv:1906.03158v1 [cs.CL] 7 Jun 2019

Matching the Blanks: Distributional Similarity for Relation Learning

Livio Baldini Soares Nicholas FitzGerald Jeffrey Ling Tom Kwiatkowski
Google Research
{liviobs,nfitz,jeffreyling,tomkwiat}@google.com