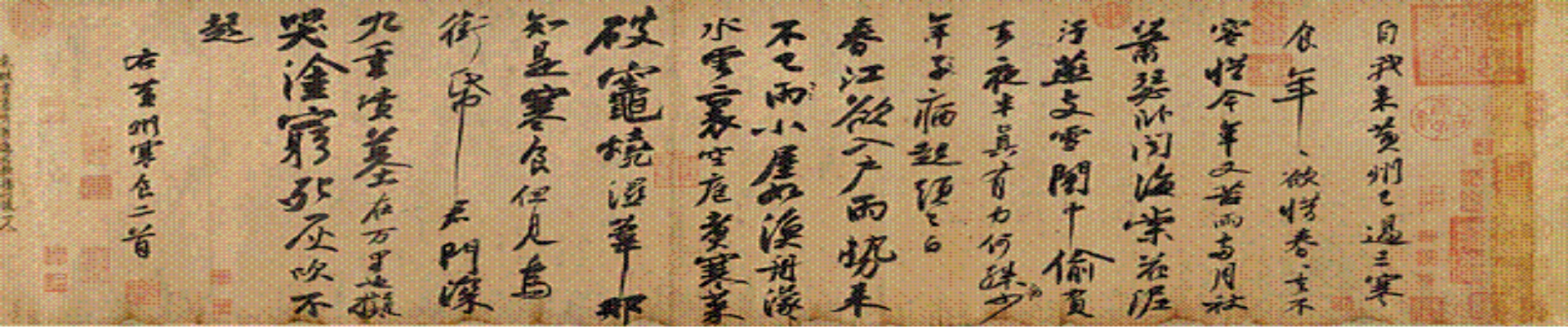


信息检索
Information Retrieval

教师：孙茂松

Tel:62781286

Email:sms@tsinghua.edu.cn

TA：**胡锦毅**

Email:hu-jy21@mails.tsinghua.edu.cn

郑重声明

- 此课件仅供选修清华大学计算机系本科生课《信息检索》（课号：40240372）的学生个人学习使用，所以只允许学生将其下载、存贮在自己的电脑中。未经孙茂松本人同意，任何人不得以任何方式扩散之（包括放到9#服务器上）。否则，由此可能引起的一切涉及知识产权的法律责任，概由该人负责。
- 此课件仅限孙茂松本人讲课使用。除孙茂松本人外，凡授课过程中，PTT文件显示此《郑重声明》之情形，即为侵权使用。

第八章
神经网络信息检索

Query-document Matching

$$sim(q,d) = cos(\vec{v}_q, \vec{v}_d) = \frac{\vec{v}_q^{\intercal}\vec{v}_d}{\|\vec{v}_q\|\|\vec{v}_d\|}$$

$$\text{where,} \quad \vec{v}_q = \frac{1}{|q|}\sum_{t_q \in q} \frac{\vec{v}_{t_q}}{\|\vec{v}_{t_q}\|}$$

$$\vec{v}_d = \frac{1}{|d|}\sum_{t_d \in d} \frac{\vec{v}_{t_d}}{\|\vec{v}_{t_d}\|}$$

$$DESM_{in-out}(q,d) = \frac{1}{|q|}\sum_{t_q \in q} \frac{\vec{v}_{t_q,in}^{\intercal}\vec{v}_{d,out}}{\|\vec{v}_{t_q,in}\|\|\vec{v}_{d,out}\|}$$

$$\vec{v}_{d,out} = \frac{1}{|d|}\sum_{t_d \in d} \frac{\vec{v}_{t_d,out}}{\|\vec{v}_{t_d,out}\|}$$

Albuquerque is the most populous city in the U.S. state of New Mexico. The high-altitude city serves as the county seat of Bernalillo County, and it is situated in the central part of the state, straddling the Rio Grande. The city population is 557,169 as of the July 1, 2014 population estimate from the United States Census Bureau, and ranks as the 32nd-largest city in the U.S. The Albuquerque metropolitan statistical area (or MSA) has a population of 907,301 according to the United States Census Bureau's most recently available estimate for 2015.

(a) About Albuquerque

Allen suggested that they could program a BASIC interpreter for the device; after a call from Gates claiming to have a working interpreter, MITS requested a demonstration. Since they didn't actually have one, Allen worked on a simulator for the Altair while Gates developed the interpreter. Although they developed the interpreter on a simulator and not the actual device, the interpreter worked flawlessly when they demonstrated the interpreter to MITS in Albuquerque, New Mexico in March 1975; MITS agreed to distribute it, marketing it as Altair BASIC.

(b) Not about Albuquerque

the city of cambridge is a university city and the county town of cambridgeshire , england . it lies in east anglia , on the river cam , about 50 miles (80 km) north of london . according to the united kingdom census 2011 , its population was _ (including _ students) . this makes cambridge the second largest city in cambridgeshire after peterborough , and the 54th largest in the united kingdom . there is archaeological evidence of settlement in the area during the bronze age and roman times ; under viking rule cambridge became an important trading centre . the first town charters were granted in the 12th century , although city status was not conferred until 1951 .

(a) Passage about the city of Cambridge

oxford is a city in the south east region of england and the county town of oxfordshire . with a population of _ it is the 52nd largest city in the united kingdom , and one of the fastest growing and most ethnically diverse . oxford has a broad economic base . its industries include motor manufacturing , education , publishing and a large number of information technology and _ businesses , some being academic offshoots . the city is known worldwide as the home of the university of oxford , the oldest university in the _ world . buildings in oxford demonstrate examples of every english architectural period since the arrival of the saxons , including the _ radcliffe camera . oxford is known as the city of dreaming spires , a term coined by poet matthew arnold .

(b) Passage about the city of Oxford

the cambridge (giraffa camelopardalis) is an african _ ungulate mammal , the tallest living terrestrial animal and the largest ruminant . its species name refers to its _ shape and its _ colouring . its chief distinguishing characteristics are its extremely long neck and legs , its _ _ , and its distinctive coat patterns . it is classified under the family _ , along with its closest extant relative , the okapi . the nine subspecies are distinguished by their coat patterns . the scattered range of giraffes extends from chad in the north to south africa in the south , and from niger in the west to somalia in the east . giraffes usually inhabit savannas , grasslands , and open woodlands .

(c) Passage about giraffes, but 'giraffe' is replaced by 'Cambridge'

Figure 4.2: A visualization of IN-OUT similarities between terms in different passages with the query term "Cambridge". The visualization reveals that, besides the term "Cambridge", many other terms in the passages about both Cambridge and Oxford have high similarity to the query term. The passage (c) is adapted from a passage on giraffes by replacing all the occurrences of the term "giraffe" with "cambridge". However, none of the other terms in (c) are found to be relevant to the query term. An embedding based approach may be able to determine that passage (c) is non-relevant to the query "Cambridge", but fail to realize that passage (b) is also non-relevant. A term counting-based model, on the other hand, can easily identify that passage (b) is non-relevant but may rank passage (c) incorrectly high.

Query-document Matching

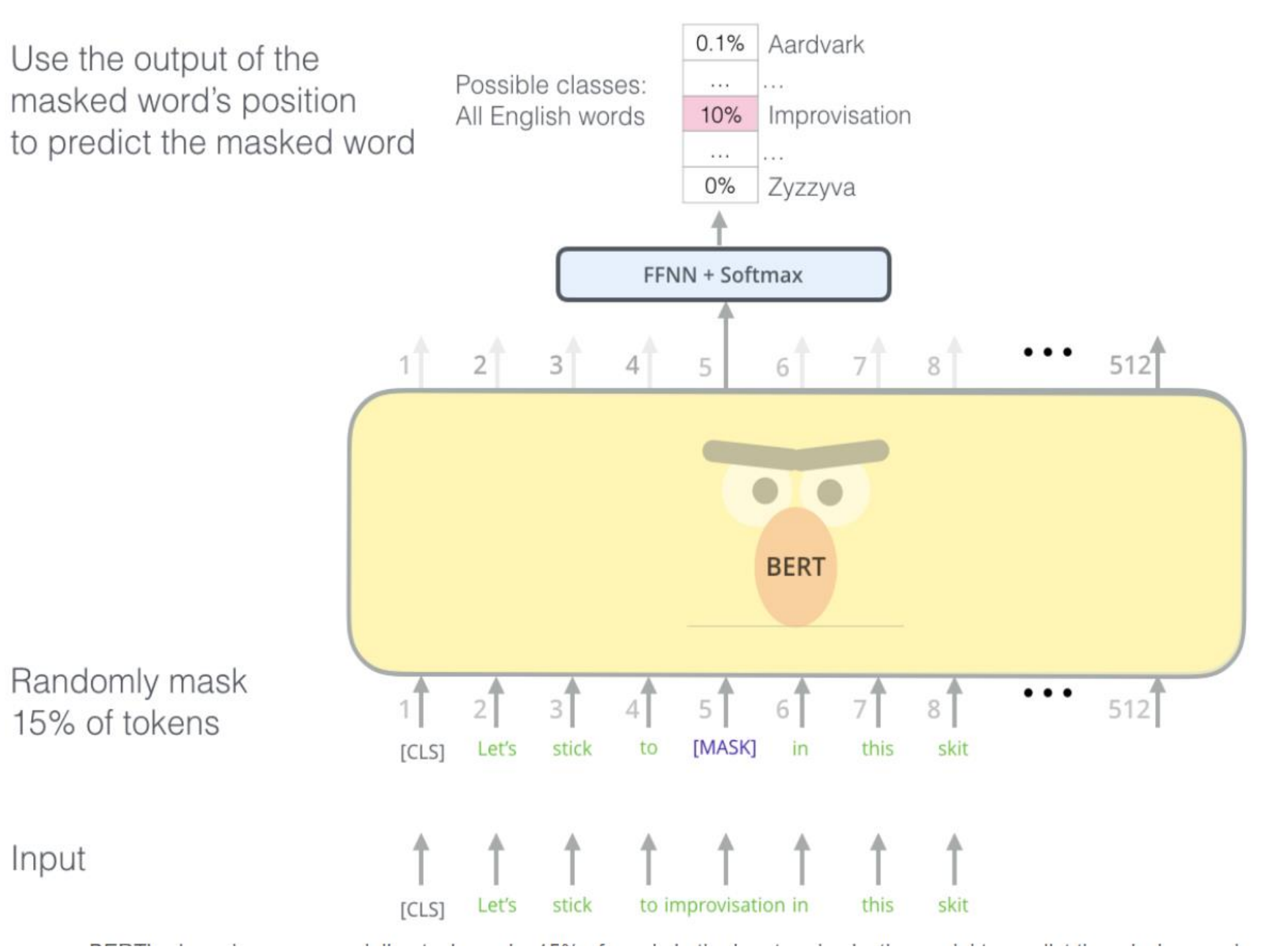
$$p(d|q) = \prod_{t_q \in q} \Bigg(\lambda \frac{tf(t_q, d)}{|d|} + \alpha \frac{\sum_{t_d \in d} (sim(\vec{v}_{t_q}, \vec{v}_{t_d}) \cdot tf(t_d, d))}{\sum_{t_{d_1} \in d} \sum_{t_{d_2} \in d} sim(\vec{v}_{t_{d_1}}, \vec{v}_{t_{d_2}}) \cdot |d|^2} + \beta \frac{\sum_{\bar{t} \in N_t} (sim(\vec{v}_{t_q}, \vec{v}_{\bar{t}}) \cdot \sum_{\bar{d} \in D} tf(\bar{t}, \bar{d}))}{\sum_{t_{d_1} \in N_t} \sum_{t_{d_2} \in N_t} sim(\vec{v}_{t_{d_1}}, \vec{v}_{t_{d_2}}) \cdot \sum_{\bar{d} \in D} |\bar{d}| \cdot |N_t|} + (1 - \alpha - \beta - \lambda) \frac{\sum_{\bar{d} \in D} tf(t_q, \bar{d})}{\sum_{\bar{d} \in D} |\bar{d}|} \Bigg)$$

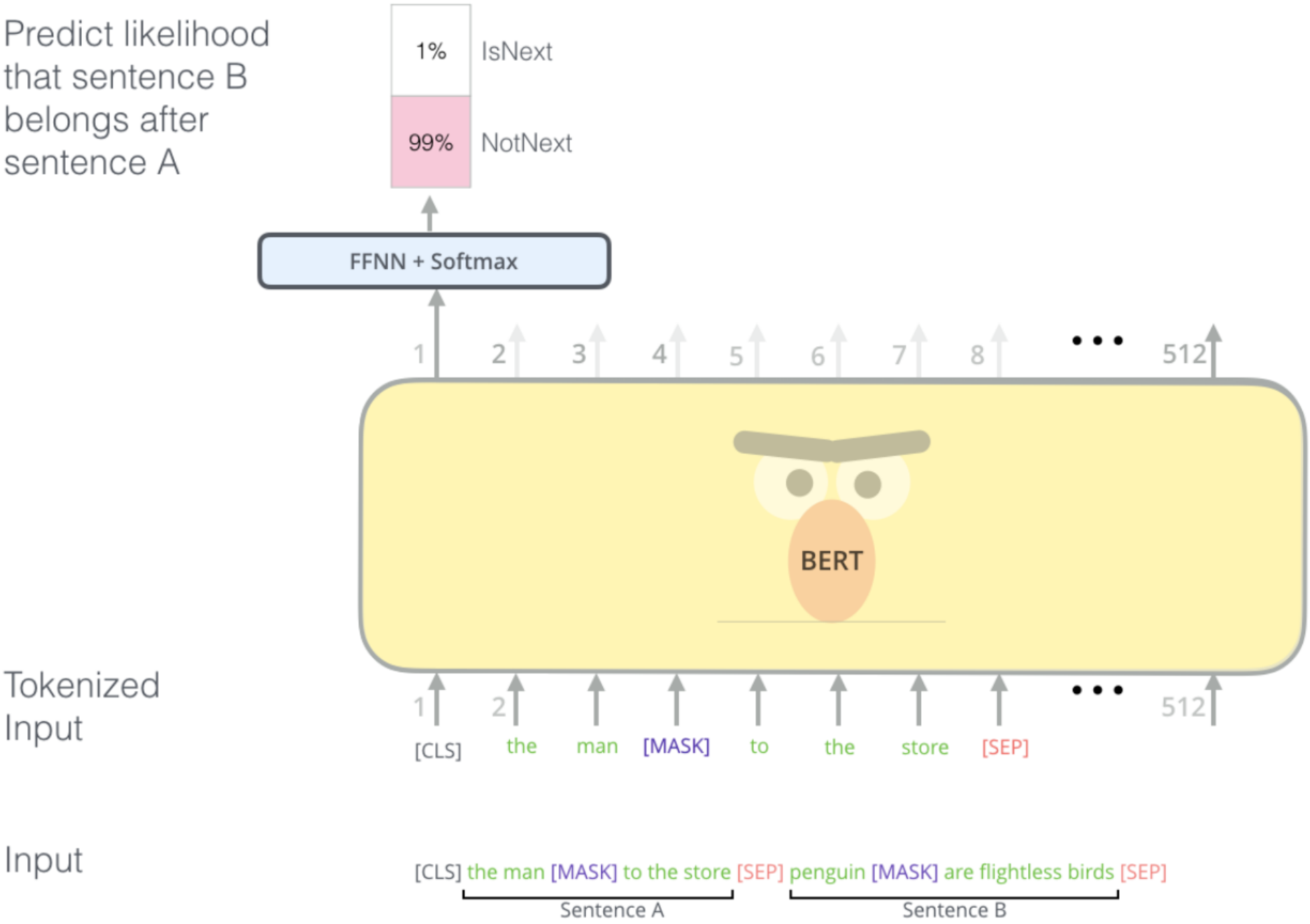
BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Jacob Devlin **Ming-Wei Chang** **Kenton Lee** **Kristina Toutanova**

Google AI Language

{jacobdevlin,mingweichang,kentonl,kristout}@google.com


Use the output of the masked word's position to predict the masked word
Possible classes: All English words
0.1%
Aardvark
...
...
10%
Improvisation
...
...
0%
Zyzzyva
FFNN + Softmax
1
2
3
4
5
6
7
8
...
512
BERT
Randomly mask 15% of tokens
[CLS]
Let's
stick
to
[MASK]
in
this
skit
Input
[CLS]
Let's
stick
to improvisation in
this
skit


Predict likelihood that sentence B belongs after sentence A
1%
IsNext
99%
NotNext
FFNN + Softmax
1
2
3
4
5
6
7
8
512
BERT
Tokenized Input
[CLS]
the
man
[MASK]
to
the
store
[SEP]
Input
[CLS] the man [MASK] to the store [SEP] penguin [MASK] are flightless birds [SEP]
Sentence A
Sentence B

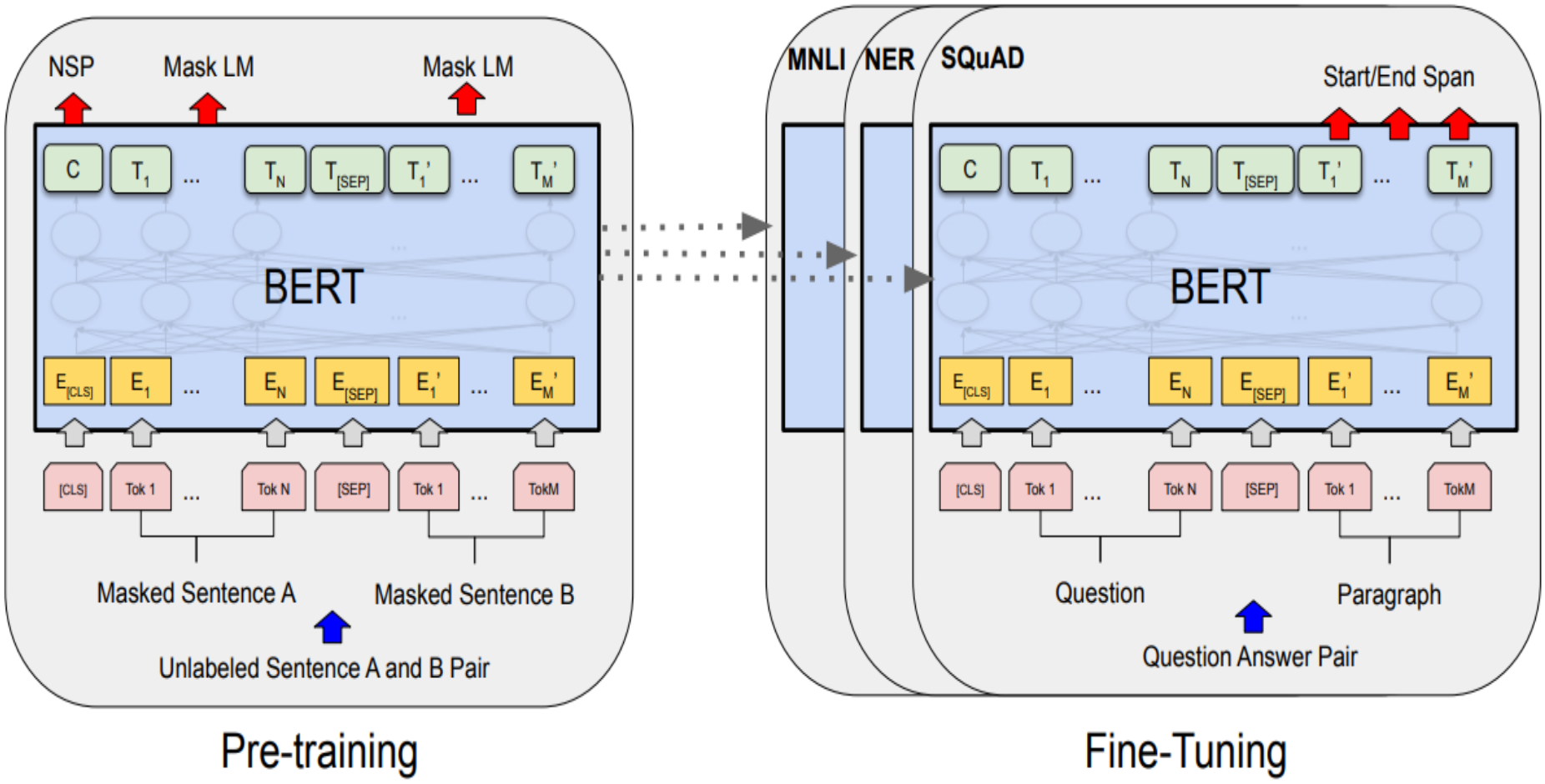


Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. `[CLS]` is a special symbol added in front of every input example, and `[SEP]` is a special separator token (e.g. separating questions/answers).

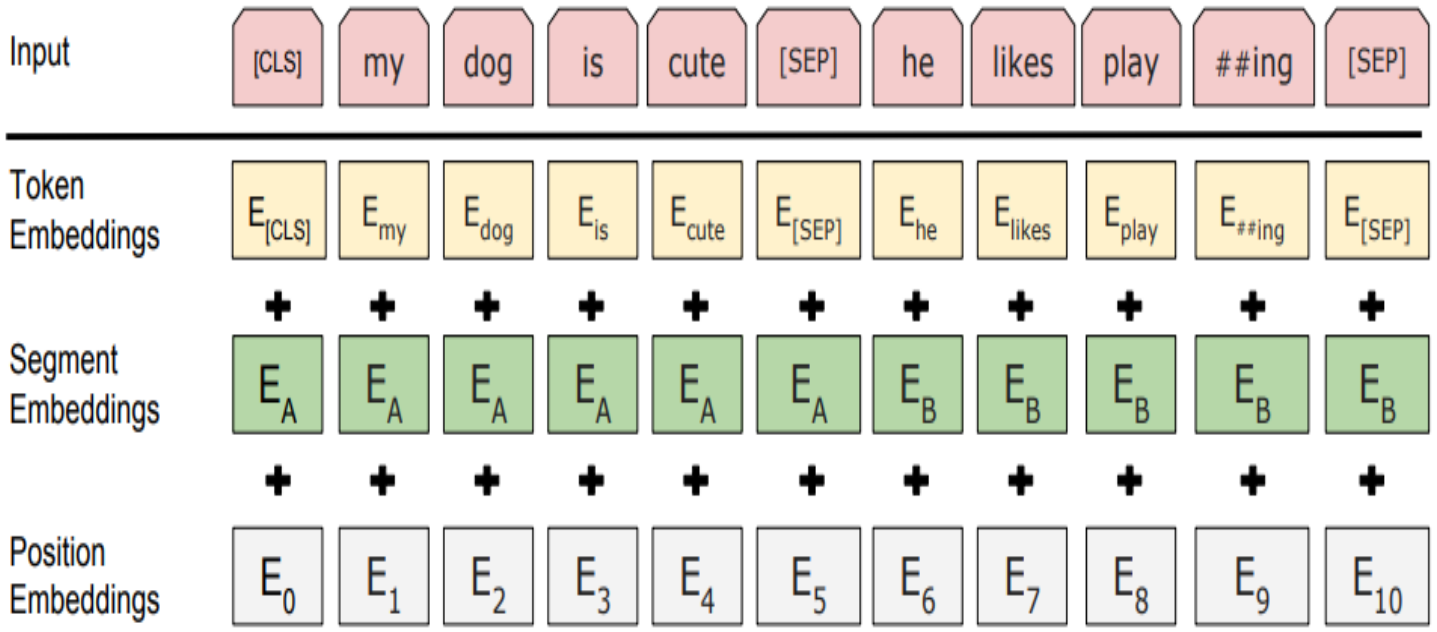
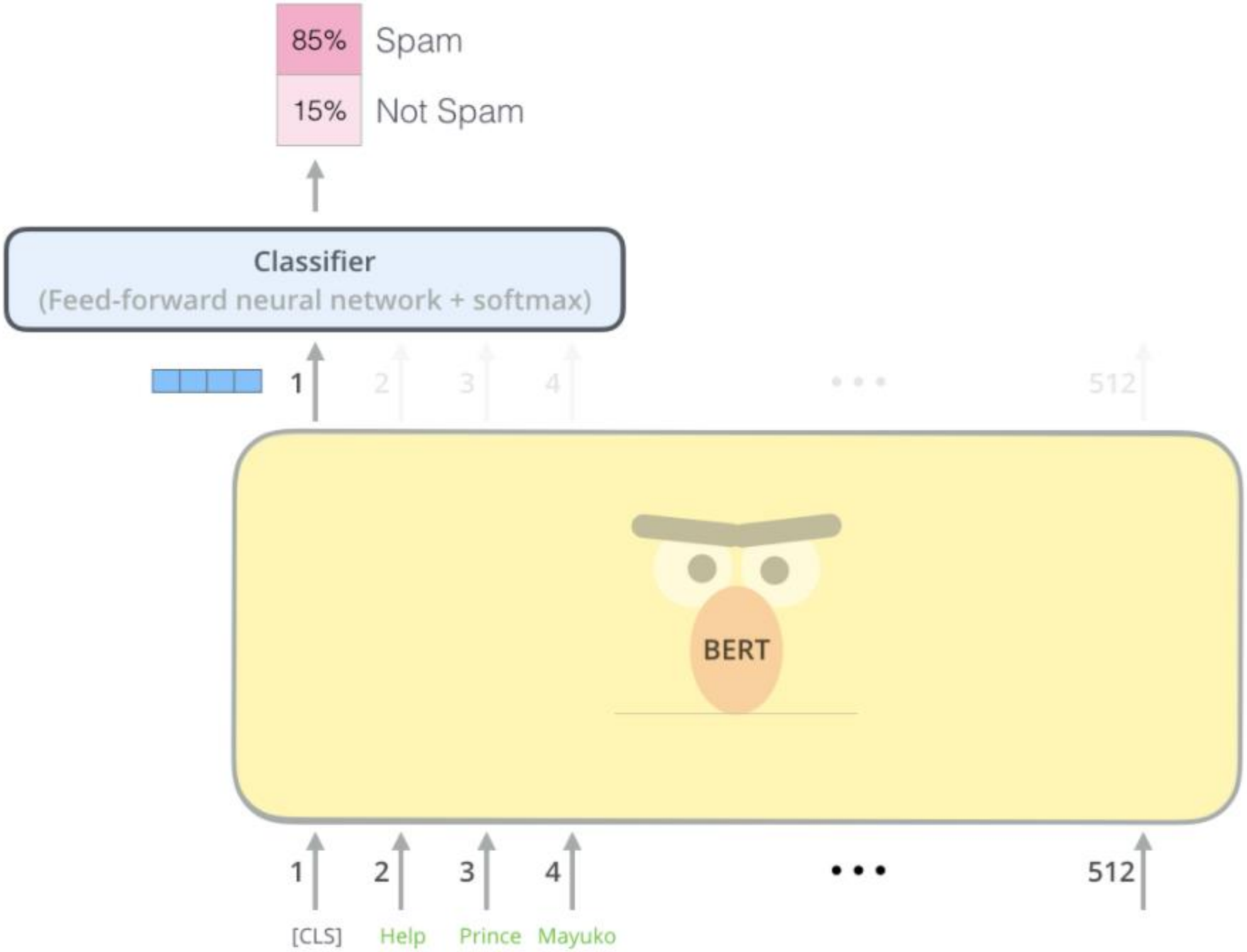


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.


85%
Spam
15%
Not Spam
Classifier
(Feed-forward neural network + softmax)
1
2
3
4
512
BERT
1
2
3
4
512
[CLS]
Help
Prince
Mayuko

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

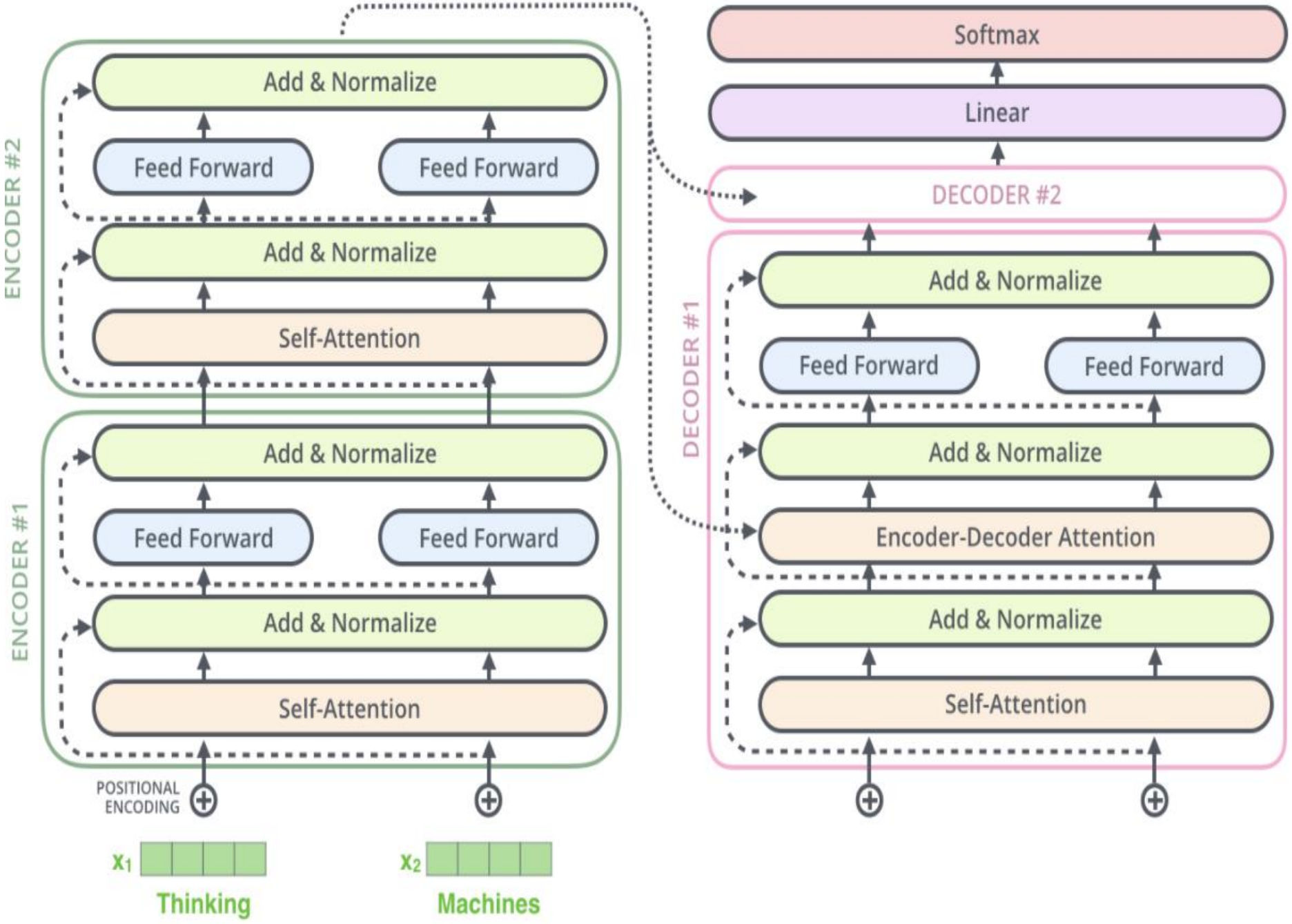
Jakob Uszkoreit* Google Research usz@google.com

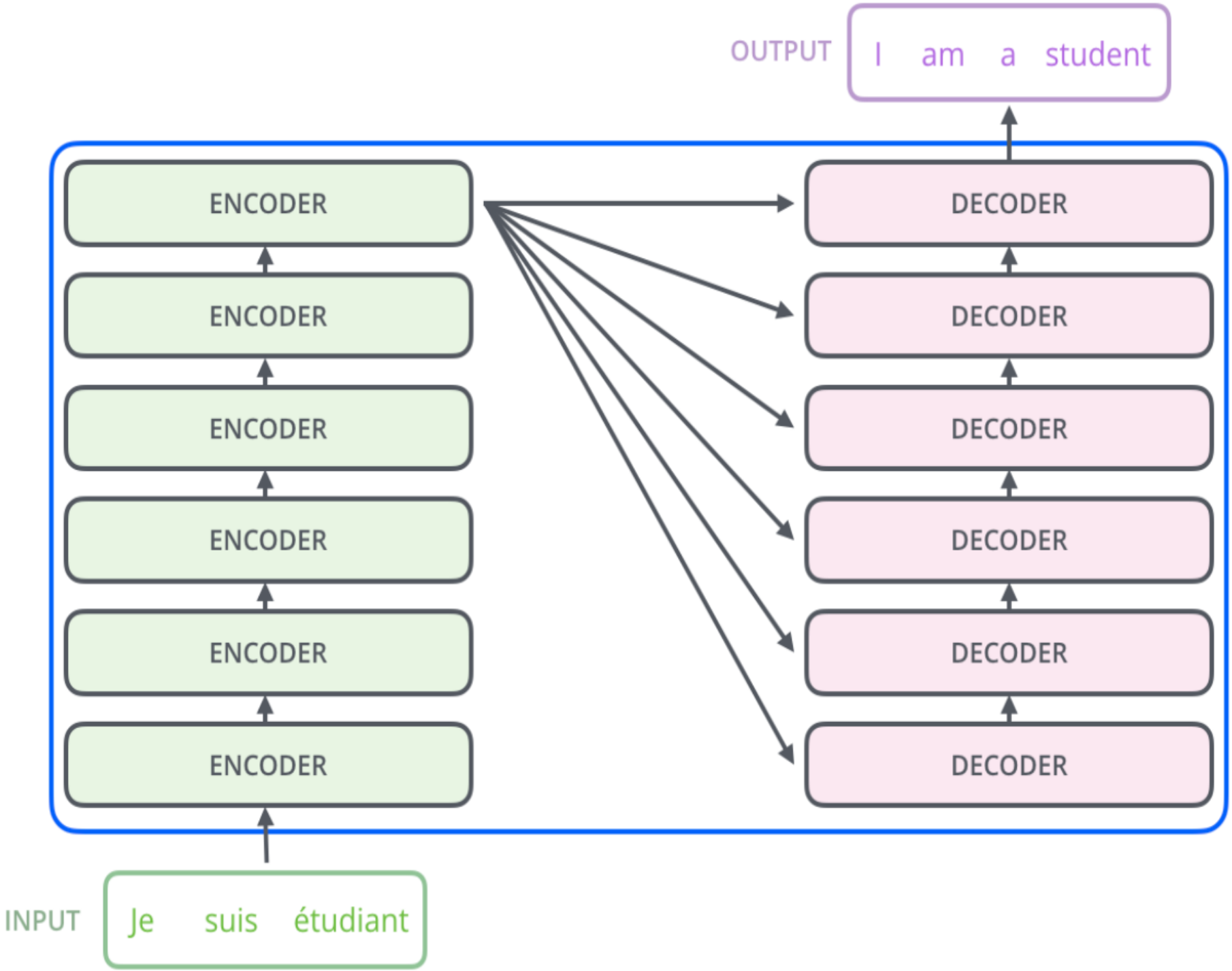
Llion Jones* Google Research llion@google.com

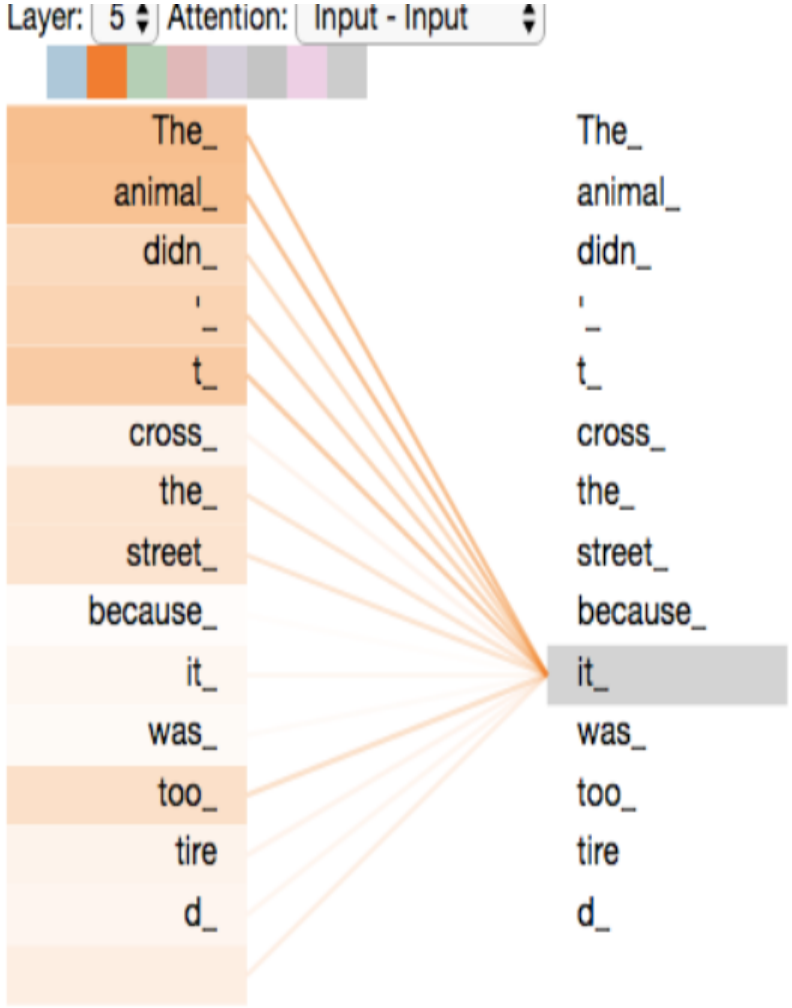
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com


Softmax
Linear
DECODER #2
DECODER #1
Add & Normalize
Feed Forward
Feed Forward
Add & Normalize
Encoder-Decoder Attention
Add & Normalize
Self-Attention
ENCODER #2
ENCODER #1
Add & Normalize
Feed Forward
Feed Forward
Add & Normalize
Self-Attention
POSITIONAL ENCODING
x1
x2
Thinking
Machines


OUTPUT
I am a student
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
ENCODER
DECODER
DECODER
DECODER
DECODER
DECODER
DECODER
INPUT
Je suis étudiant



As we are encoding the word "it" in encoder #5 (the top encoder in the stack), part of the attention mechanism was focusing on "The Animal", and baked a part of its representation into the encoding of "it".

1) This is our input sentence*

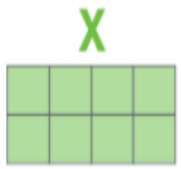
2) We embed each word*

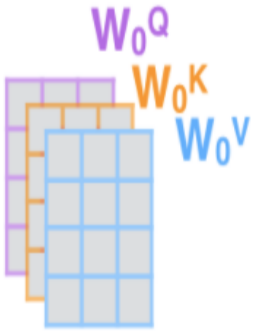
3) Split into 8 heads. We multiply X or R with weight matrices

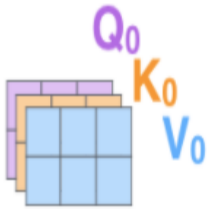
4) Calculate attention using the resulting Q/K/V matrices

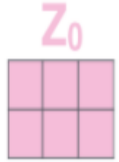
5) Concatenate the resulting Z matrices, then multiply with weight matrix W^O to produce the output of the layer

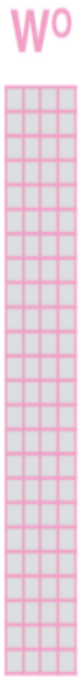
Thinking Machines

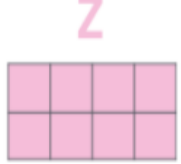


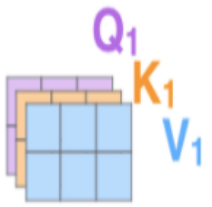


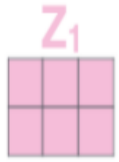




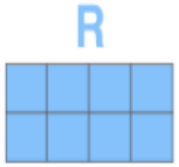


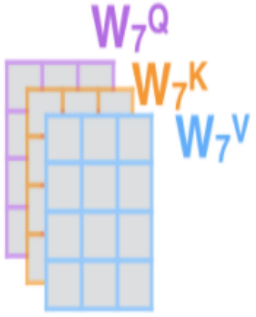


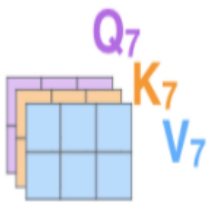


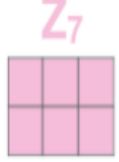


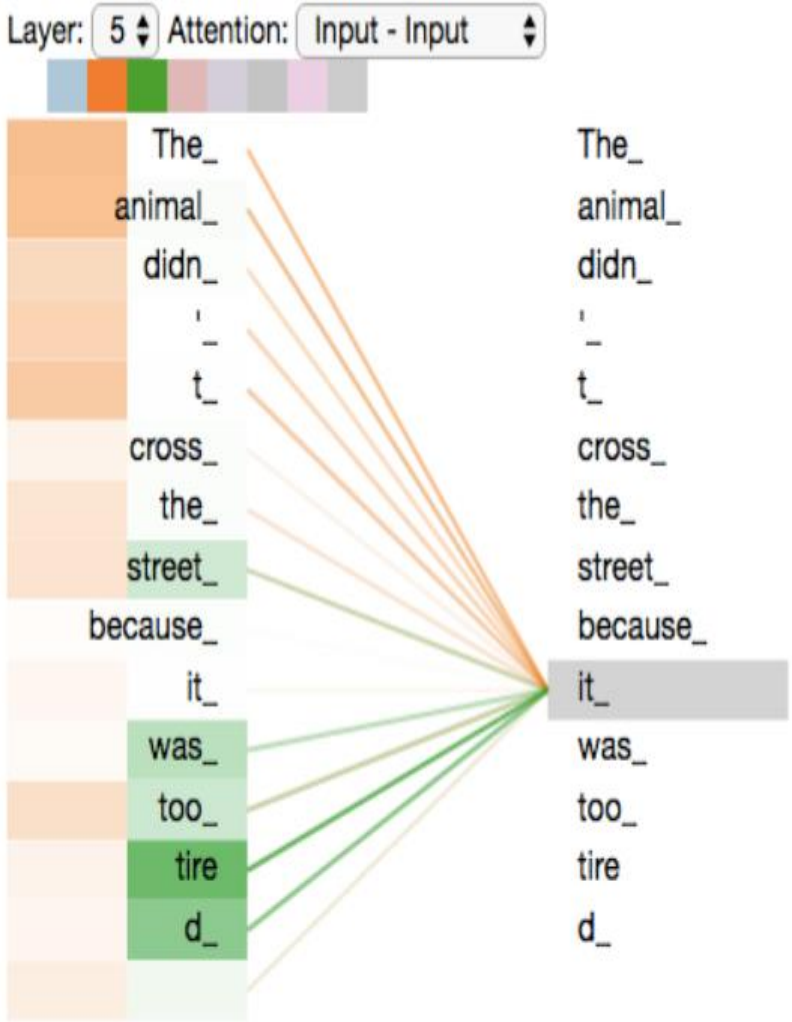
* In all encoders other than #0, we don't need embedding. We start directly with the output of the encoder right below this one



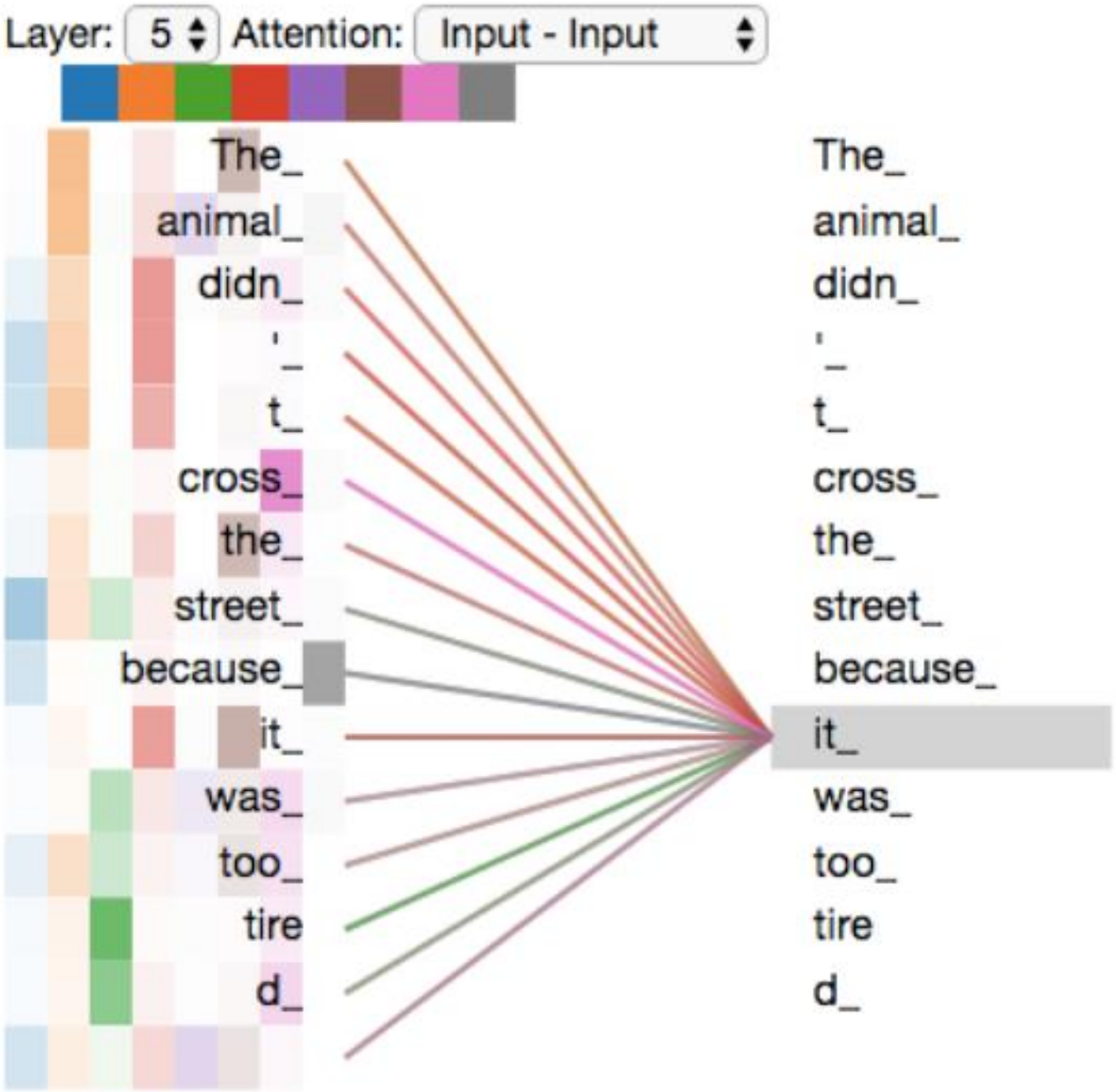


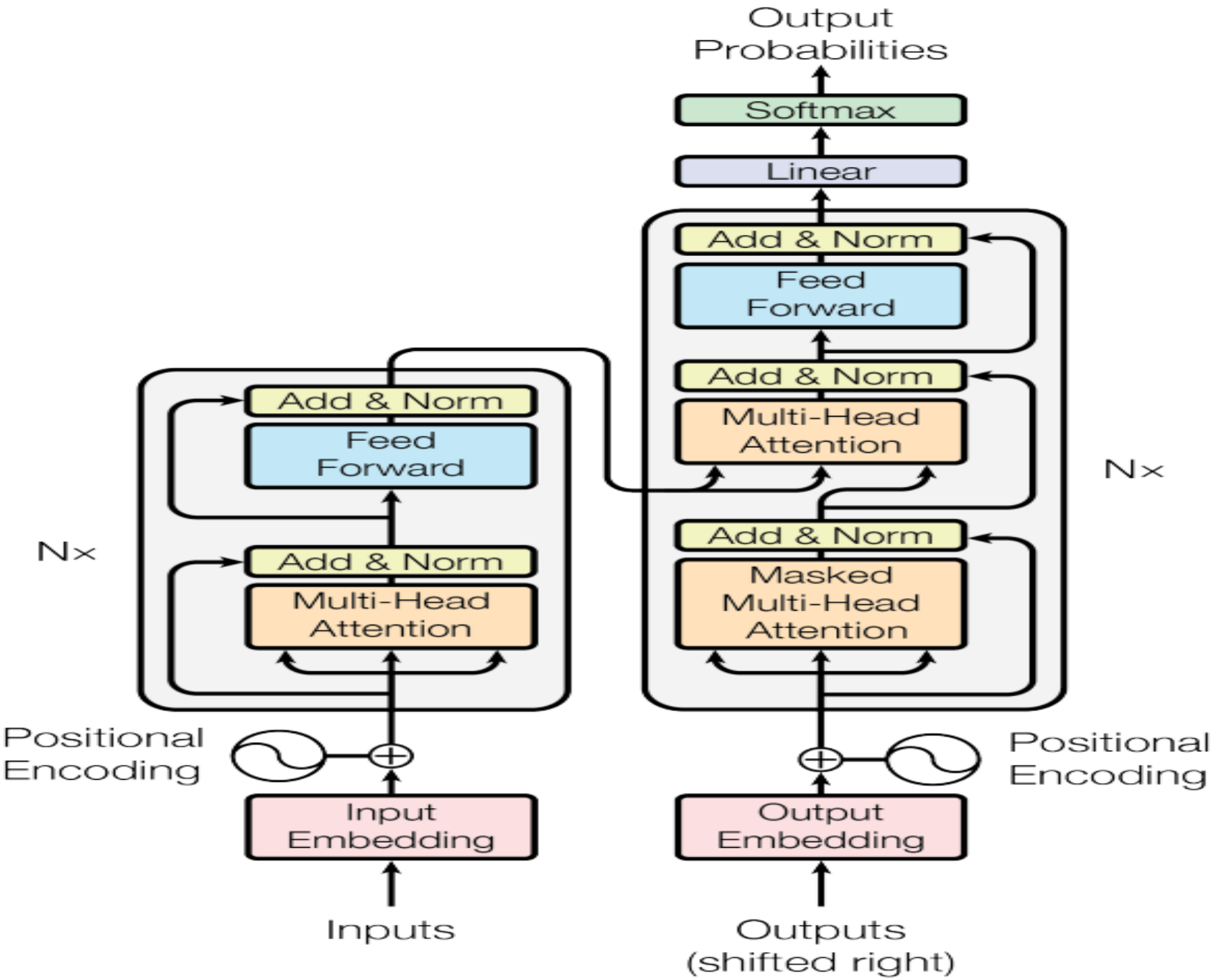


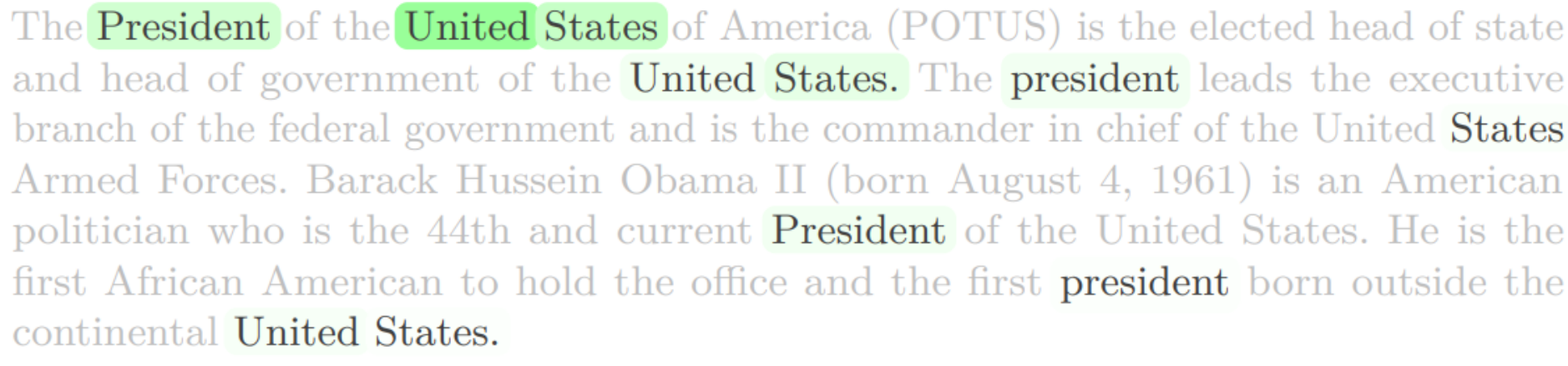




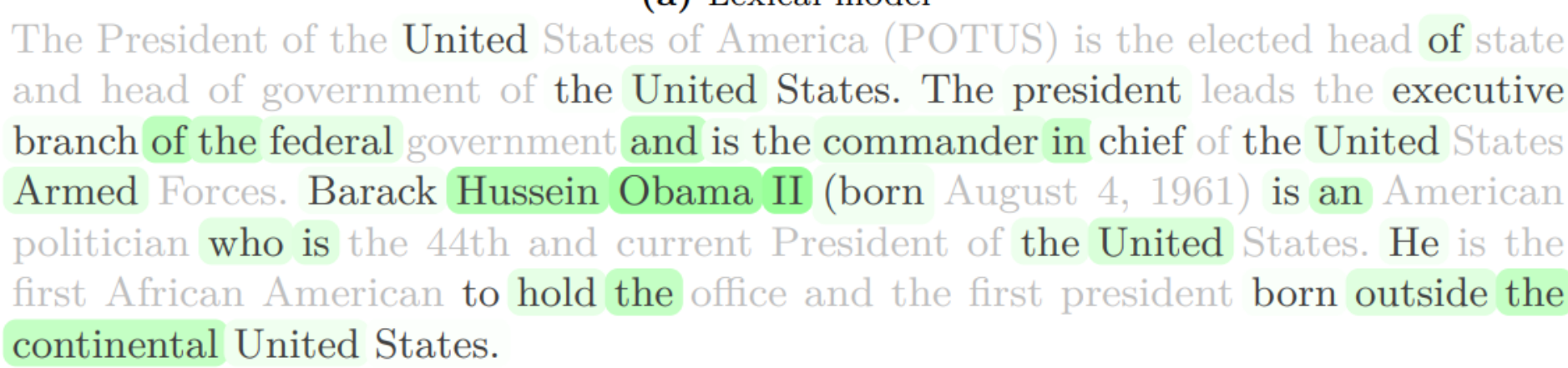
As we encode the word "it", one attention head is focusing most on "the animal", while another is focusing on "tired" -- in a sense, the model's representation of the word "it" bakes in some of the representation of both "animal" and "tired".


Layer: 5
Attention: Input - Input
The_
animal_
didn_
'_
t_
cross_
the_
street_
because_
it_
was_
too_
tire
d_
The_
animal_
didn_
'_
t_
cross_
the_
street_
because_
it_
was_
too_
tire
d_


Output Probabilities
Softmax
Linear
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feed Forward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)



(a) Lexical model



(b) Semantic model

Figure 7.2: Analysis of term importance for estimating the relevance of a passage to the query "United States President" by a lexical and a semantic deep neural network model. The lexical model only considers the matches of the query terms in the document but gives more emphasis to earlier occurrences. The semantic model is able to extract evidence of relevance from related terms such as "Obama" and "federal".

Deeper Text Understanding for IR with Contextual Neural Language Modeling

Zhuyun Dai
Carnegie Mellon University
zhuyund@cs.cmu.edu

Jamie Callan
Carnegie Mellon University
callan@cs.cmu.edu

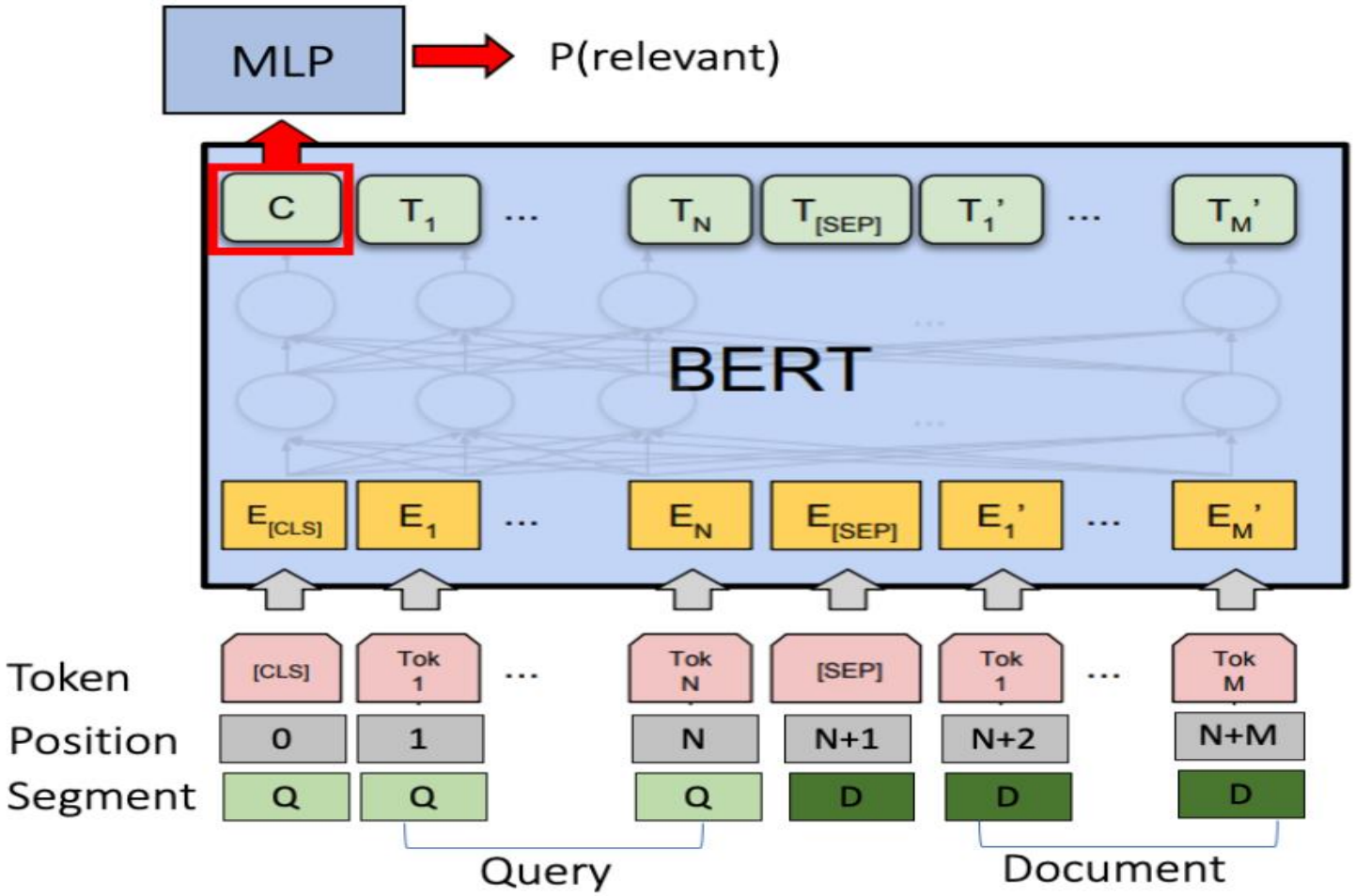


Table 2: Search accuracy on Robust04 and ClueWeb09-B. † indicates statistically significant improvements over Coor-Ascent by permutation test with $p < 0.05$.

	nDCG@20			
	Robust04		ClueWeb09-B	
Model	Title	Description	Title	Description
BOW	0.417	0.409	0.268	0.234
SDM	0.427	0.427	0.279	0.235
RankSVM	0.420	0.435	0.289	0.245
Coor-Ascent	0.427	0.441	**0.295**	0.251
DRMM	0.422	0.412	0.275	0.245
Conv-KNRM	0.416	0.406	0.270	0.242
BERT-FirstP	0.444†	0.491†	0.286	**0.272**†
BERT-MaxP	**0.469**†	**0.529**†	0.293	0.262†
BERT-SumP	0.467†	0.524†	0.289	0.261

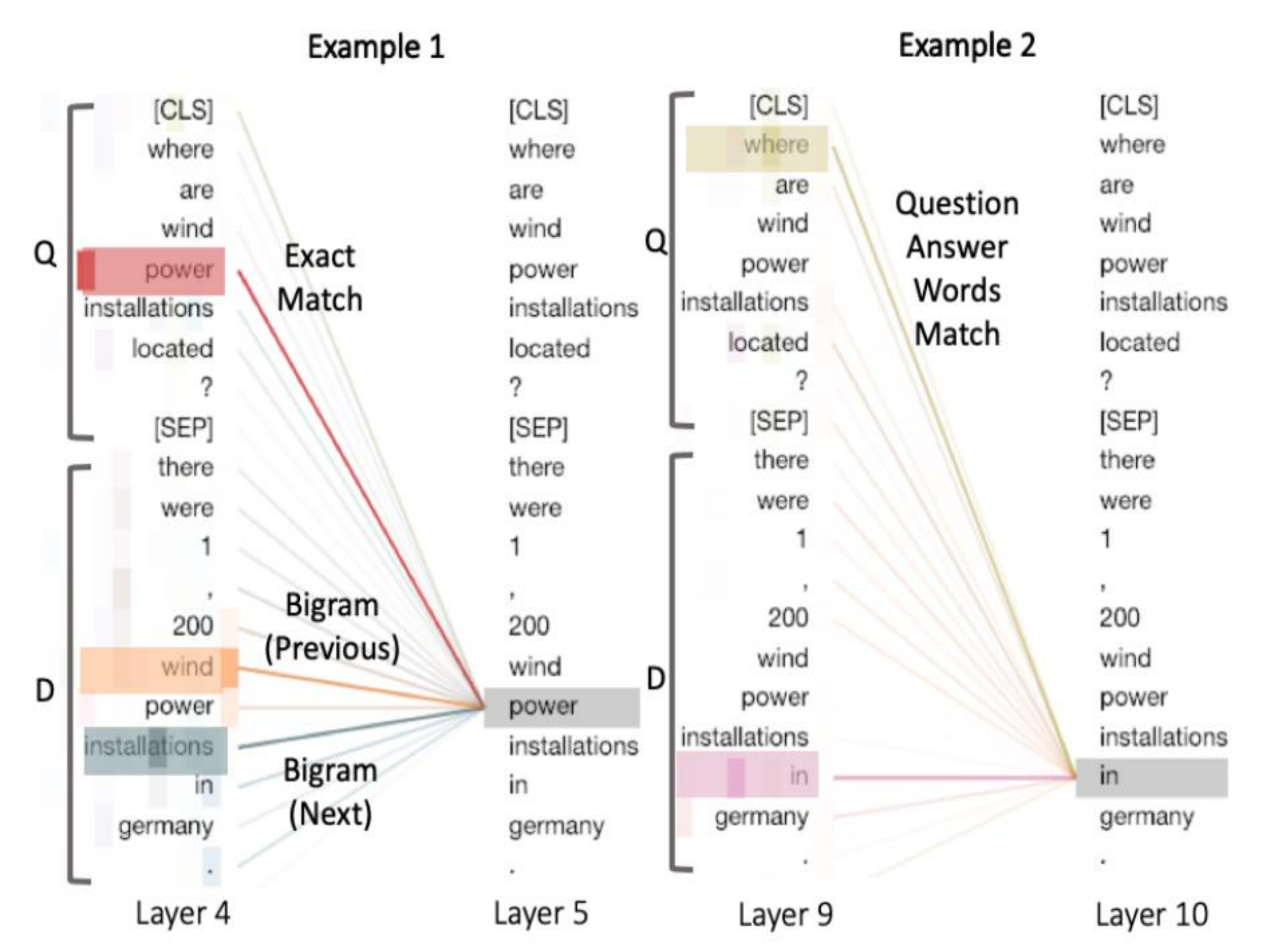

Example 1
Example 2
Q
D
[CLS]
where
are
wind
power
installations
located
?
[SEP]
there
were
1
,
200
wind
power
installations
in
germany
.
Exact Match
Bigram (Previous)
Bigram (Next)
Question Answer Words Match
Layer 4
Layer 5
Layer 9
Layer 10

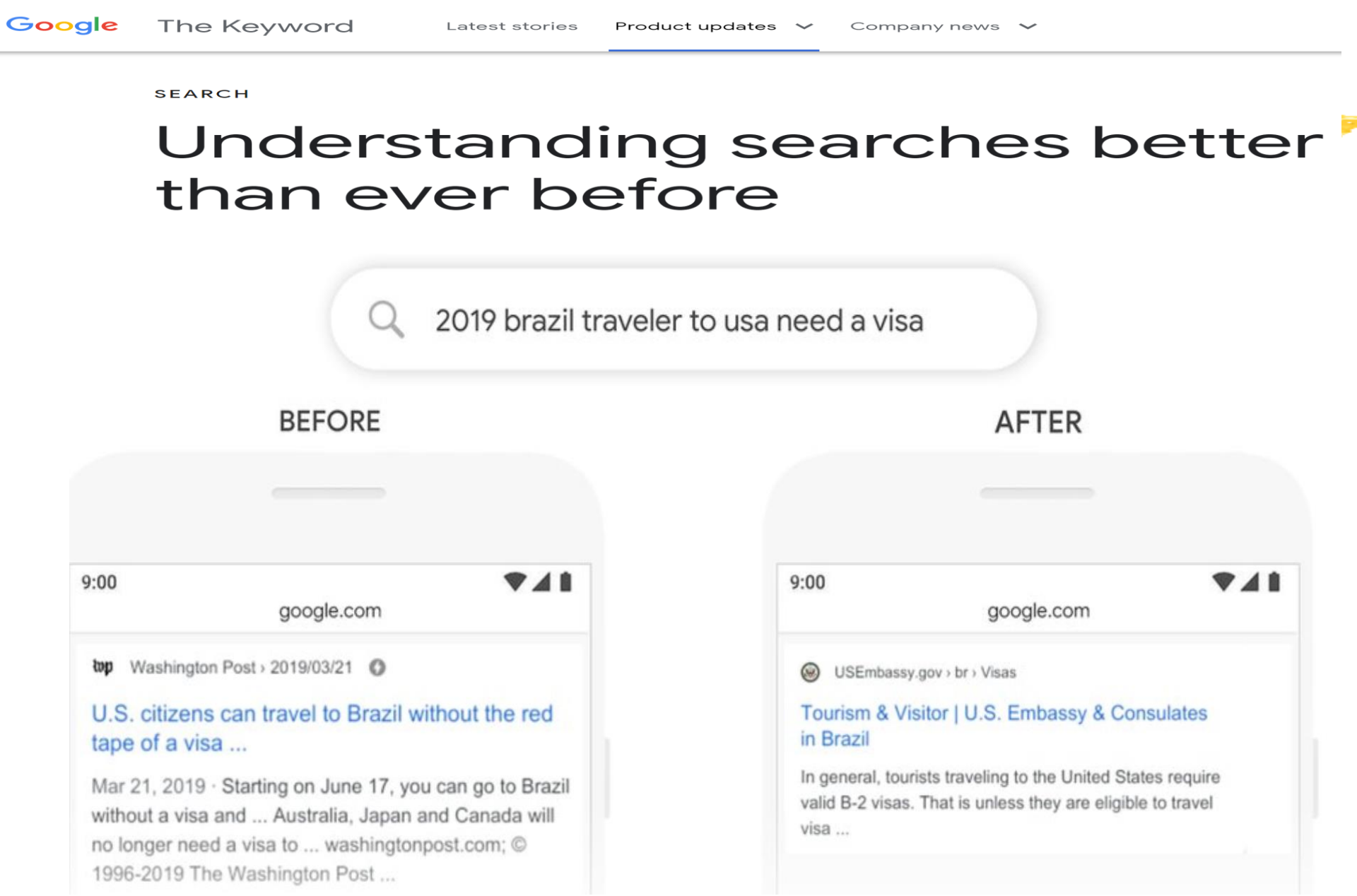
Table 1: Example of Robust04 search topic (Topic 697).

Title	air traffic controller
Description	What are working conditions and pay for U.S. air traffic controllers?
Narrative	Relevant documents tell something about working conditions or pay for American controllers. Documents about foreign controllers or individuals are not relevant.

Table 3: Accuracy on different types of Robust04 queries. Percentages show relative gain/loss over title queries.

Query	Avg Len	nDCG@20 SDM		Coor-Ascent		BERT-MaxP	
Title	3	0.427	–	0.427	–	0.469	–
Desc	14	0.404	-5%	0.422	-1%	0.529	+13%
Desc, keywords	7	0.427	-0%	0.441	+5%	0.503	+7%
Narr	40	0.278	-35%	0.424	-1%	0.487	+4%
Narr, keywords	18	0.332	-22%	0.439	+3%	0.471	+0%
Narr, positive	31	0.272	-36%	0.432	+1%	0.489	+4%

示例：谷歌搜索


Google
The Keyword
Latest stories
Product updates
Company news
SEARCH
Understanding searches better than ever before
2019 brazil traveler to usa need a visa
BEFORE
AFTER
9:00
google.com
Washington Post › 2019/03/21
U.S. citizens can travel to Brazil without the red tape of a visa ...
Mar 21, 2019 · Starting on June 17, you can go to Brazil without a visa and ... Australia, Japan and Canada will no longer need a visa to ... washingtonpost.com; © 1996-2019 The Washington Post ...
9:00
google.com
USEmbassy.gov › br › Visas
Tourism & Visitor | U.S. Embassy & Consulates in Brazil
In general, tourists traveling to the United States require valid B-2 visas. That is unless they are eligible to travel visa ...