MAI4CAREU

Επεξεργασία Φυσικής Γλώσσας

Distributed Contextual Embeddings

Demetris Paschalides

Department of Computer Science

University of Cyprus

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

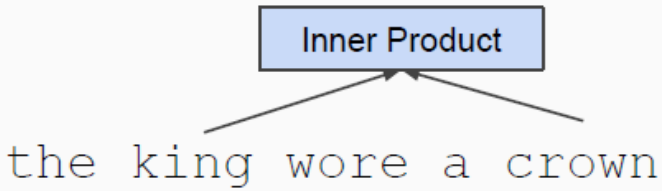
Word Embeddings

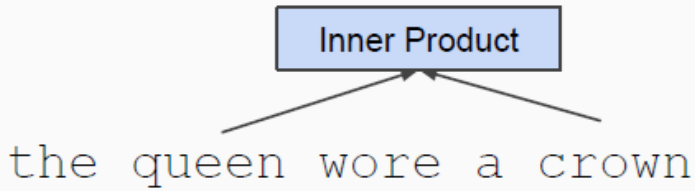
- Word embeddings αποτελούν τη βάση της βαθιάς μάθησης για NLP

king → [-0.5, -0.9, 1.4, …]

queen → [-0.6, -0.8, -0.2, …]

- Word embeddings (*word2vec*, *GloVe*) είναι συχνά προ-εκπαιδευμένοι σε σώματα κειμένων από στατιστικά στοιχεία





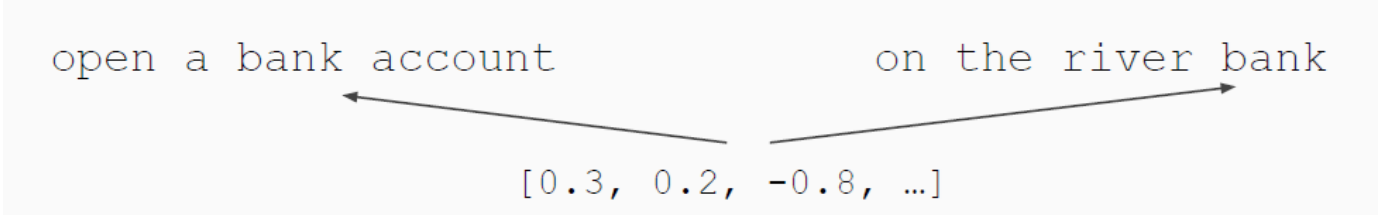
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ζητήματα Word2Vec Embeddings

- Word2Vec Embeddings έχουν δύο προβλήματα:
  - Ίδια αναπαράσταση για μια λέξη ανεξάρτητα από τα συμφραζόμενά της



  - Έχουμε μόνο μία αναπαράσταση για μια λέξη, αλλά οι λέξεις έχουν διαφορετικές πτυχές, συμπεριλαμβανομένης της σημασιολογίας, της συντακτικής συμπεριφοράς και του μητρώου/συνειρμών.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Παραστάσεις Word Embedding

- Count-based
  - TF-IDF, PPMI
- Class-based
  - Brown Clusters
- Distributed Prediction-based Embeddings
  - Word2Vec, FastText
- Distributed Contextual Embeddings from Language Models
  - Elmo, BERT
- Variants
  - Multi-lingual, Multi-sense etc.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Context είναι τα πάντα

- p(***play*** | *Elmo and Cookie Monster* ***play*** *a game.*)
- p(***play*** | *The Broadway* ***play*** *premiered yesterday.*)

≠

- Πώς να χειριστείτε διαφορετικές έννοιες λέξεων?
  - Μέχρι στιγμής, τα μοντέλα ενσωμάτωσης λέξεων παράγουν ένα διάνυσμα για τη λέξη "*play*".

- Εκπαιδεύστε ένα **neural language model** για να προβλέψετε την επόμενη λέξη που δίνεται προηγούμενες λέξεις στην πρόταση

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

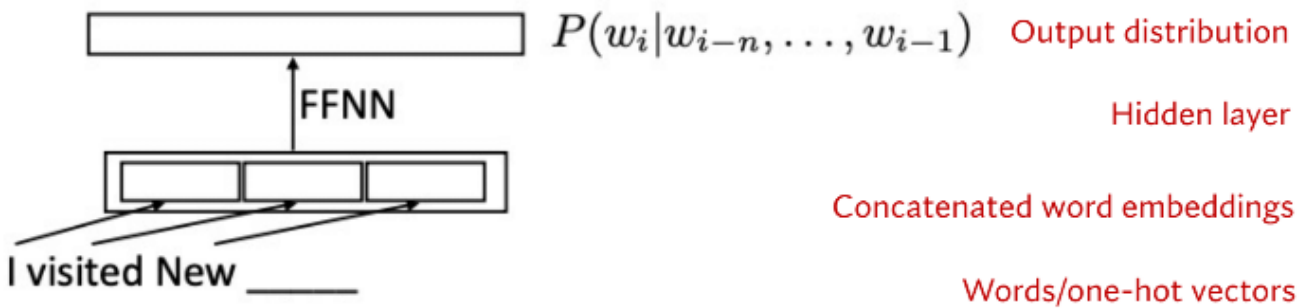
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Neural Language Models

$$P_{Add-1}(w_i \mid w_{i-1}) = \frac{c(w_{i-1}, w_i) + 1}{c(w_{i-1}) + V}$$

- Θυμηθείτε το Language Modeling task.
- **Input**: ακολουθία λέξεων.
- **Output**: πιθανότητα της επόμενης λέξης w.

- Αρχικές προσεγγίσεις: Feed Forward Neural Networks (FFNN) κοιτάζοντας context.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Neural Language Models

- Βελτιώσεις σε σχέση με N-Gram Language Model:
  - Δεν υπάρχει πρόβλημα αραιότητας.
  - Δεν χρειάζεται να αποθηκεύσετε όλα τα παρατηρούμενα N-Grams.
- Περιορισμούς:
  - Το σταθερό παράθυρο είναι πολύ μικρό.
  - Μεγέθυνση παραθύρου μεγεθύνει W.
  - Τα Windows δεν μπορούν ποτέ να είναι αρκετά μεγάλα!
  - Διαφορετικές λέξεις πολλαπλασιάζονται με εντελώς διαφορετικά βάρη. Δεν υπάρχει συμμετρία στον τρόπο επεξεργασίας των εισόδων.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Neural Language Models

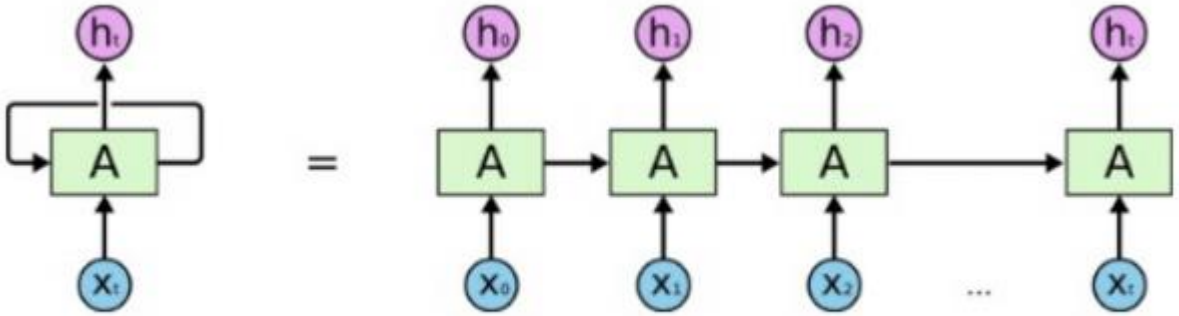
- Improvements over N-Gram Language Model:
  - No sparsity problem.
  - Don't need to store all observed N-Grams.

- Limitations:
  - Fixed window is too small.
  - Enlarging window enlarges *W*.
  - Windows can never be large enough!
  - Different words are multiplied by completely different weights. No symmetry in how the inputs are processed.

We need a neural architecture that can process any length.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Recurrent Neural Networks (RNN)



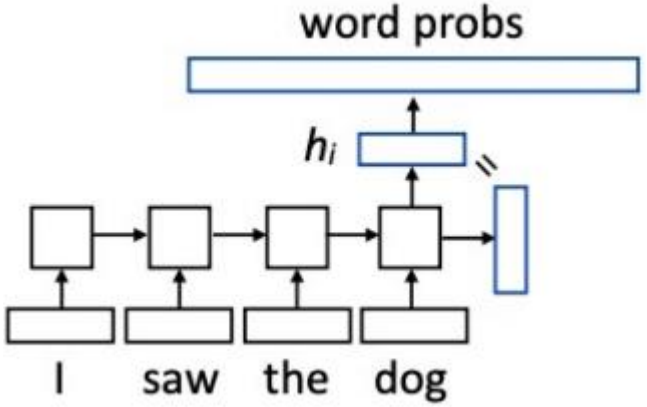
An unrolled recurrent neural network.

- Πάρτε διαδοχική είσοδο οποιουδήποτε μήκους.
- Εφαρμόστε τα ίδια βάρη σε κάθε βήμα.
- Προαιρετικά παράγετε έξοδο σε κάθε βήμα.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

RNN Language Modeling

- P(*w* | context) = softmax($W h_i$)
  - W είναι ένα matrix (μέγεθος λεξιλογίου) x (hidden size).

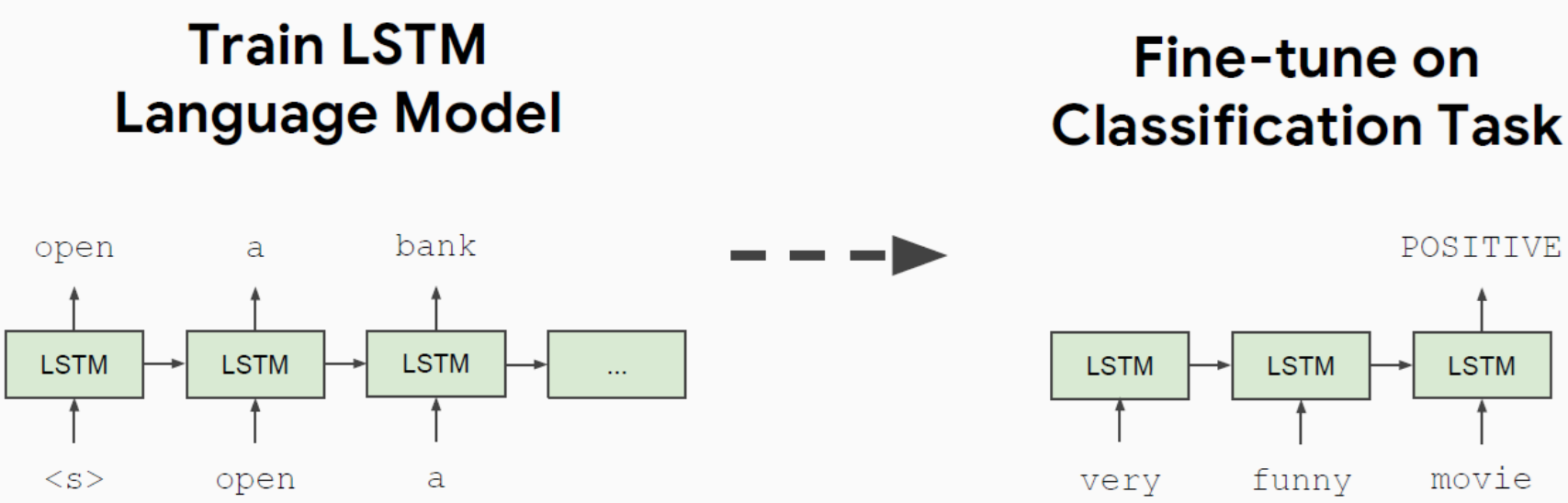


Training RNN Language Model

- Input = μια ακολουθία λέξεων
  Output = Οι λέξεις άλλαξαν μία θέση
  - Αυτό μας επιτρέπει να ομαδοποιούμε αποτελεσματικά την εκπαίδευση με την πάροδο του χρόνου.
  - Επίσης, παράγει χαρακτηριστικά από όλες τις προτάσεις / παραγράφους.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

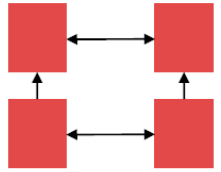
Ημι-εποπτευόμενη μάθηση ακολουθίας



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ζήτημα Linear Interaction Distance

- RNNs ξετυλίγονται "από αριστερά προς τα δεξιά".
  - Κωδικοποιεί γραμμική τοπικότητα: μια χρήσιμη ευρετική;



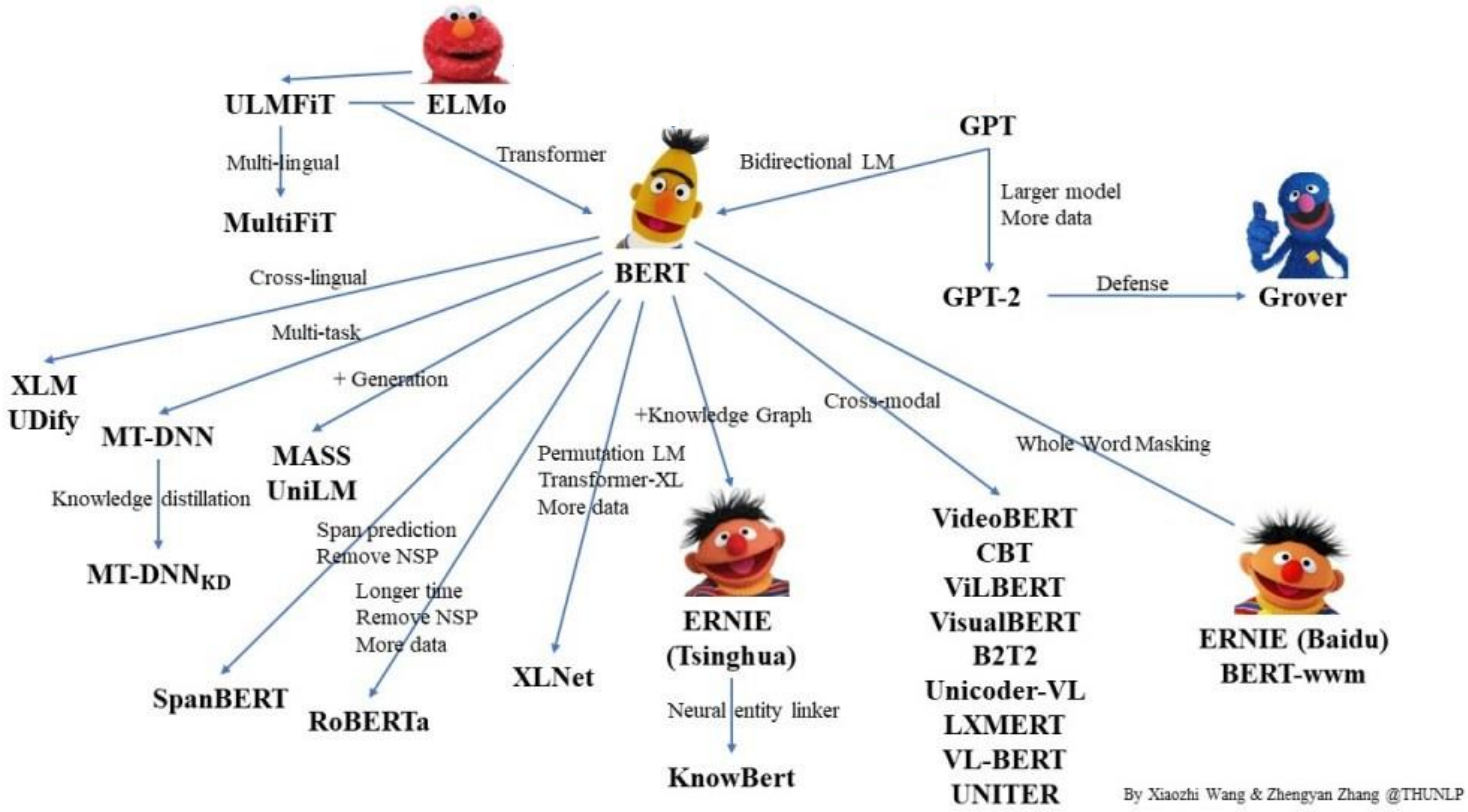
Linear Interaction Distance:

- O(sequence length) Βήματα για την αλληλεπίδραση μακρινών ζευγών λέξεων σημαίνει:
  - Δύσκολο να μάθουν εξαρτήσεις μεγάλων αποστάσεων (προβλήματα κλίσης);
  - Η γραμμική σειρά των λέξεων "ψήνεται";

- **Περιορισμούς**: Χρειάζεστε μηχανισμό κατάδειξης για να επαναλάβετε πρόσφατες λέξεις.
- **Solution**: Transformers

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ιστορία στις Transformers



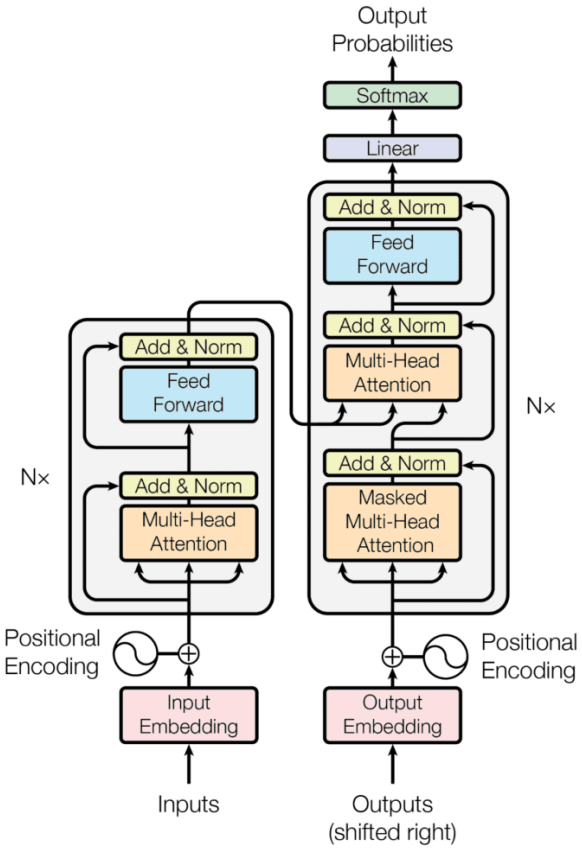
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Αρχιτεκτονική Transformer

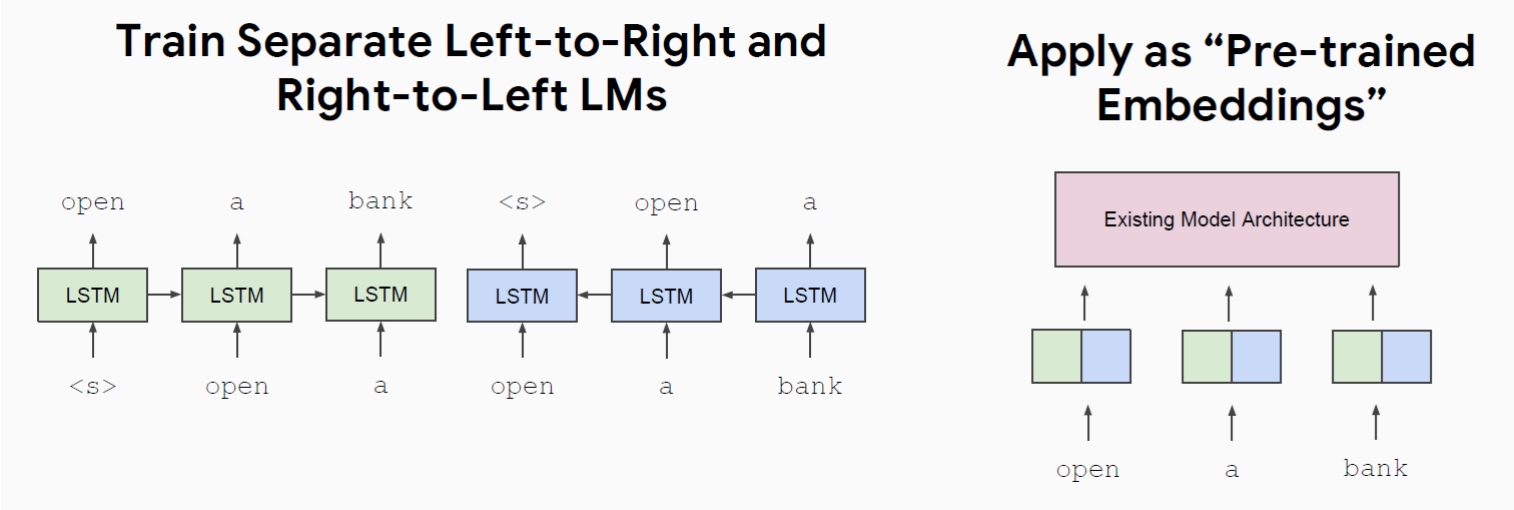
- Η αρχιτεκτονική μετασχηματιστών ακολουθεί μια δομή κωδικοποιητή-αποκωδικοποιητή.
- Ο κωδικοποιητής αντιστοιχίζει μια ακολουθία εισόδου σε μια ακολουθία συνεχών αναπαραστάσεων, η οποία στη συνέχεια τροφοδοτείται σε έναν αποκωδικοποιητή.
- Ο αποκωδικοποιητής λαμβάνει την έξοδο του κωδικοποιητή μαζί με την έξοδο του αποκωδικοποιητή στο προηγούμενο χρονικό βήμα για να δημιουργήσει μια ακολουθία εξόδου.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

ELMo: Deep Contextualized Word Representation

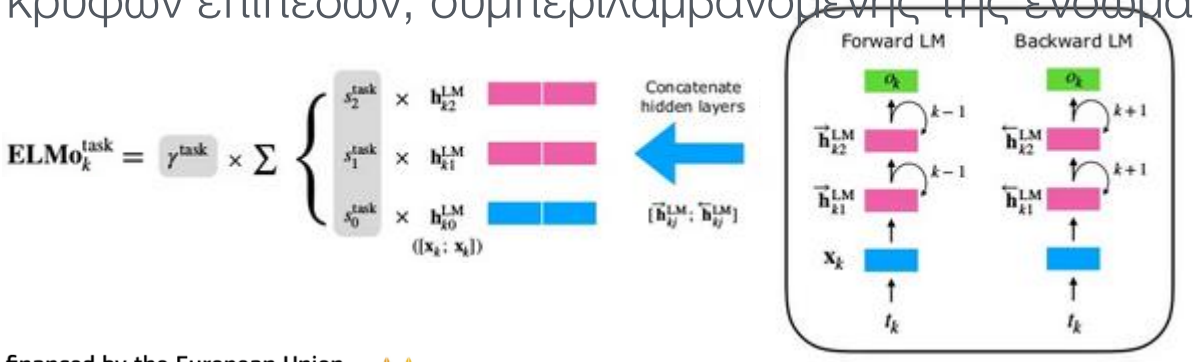
- Instead of using a fixed embedding for each word, ELMo looks at the entire sentence before assigning each word in it an embedding.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

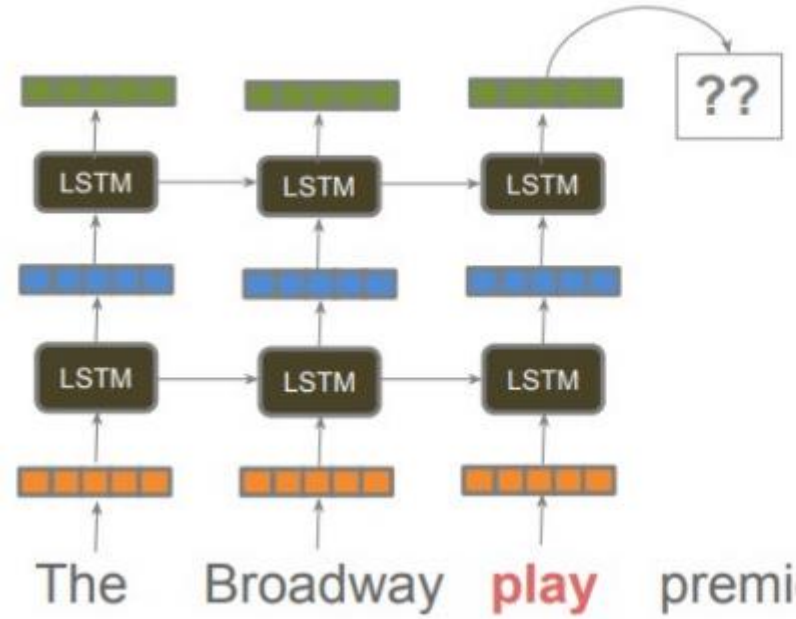
- Το ELMo μαθαίνει word embeddings μέσω της δημιουργίας αμφίδρομων LM (BiLMs).
  - BiLMs consist of forward and backward LMs.
  - Forward: $p(t_1, t_2, \ldots, t_N) = \prod_{k=1}^{N} p(t_k \mid t_1, t_2, \ldots, t_{k-1})$
  - Backward: $p(t_1, t_2, \ldots, t_N) = \prod_{k=1}^{N} p(t_k \mid t_{k+1}, t_{k+2}, \ldots, t_N)$
- ELMo αντιπροσωπεύει μια λέξη t_k ως γραμμικός συνδυασμός αντίστοιχων κρυφών επιπέδων, συμπεριλαμβανομένης της ενσωμάτωσής του.



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

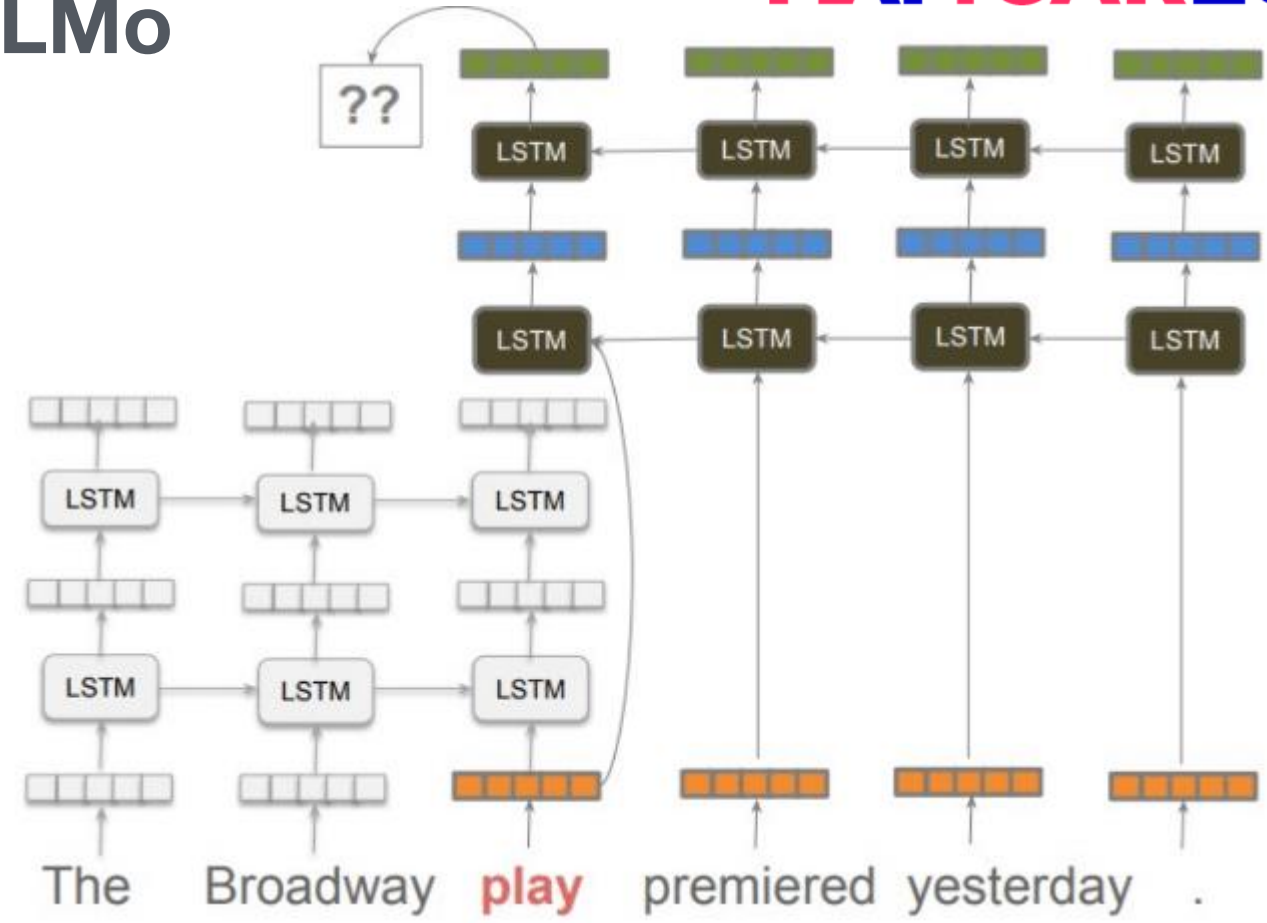
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

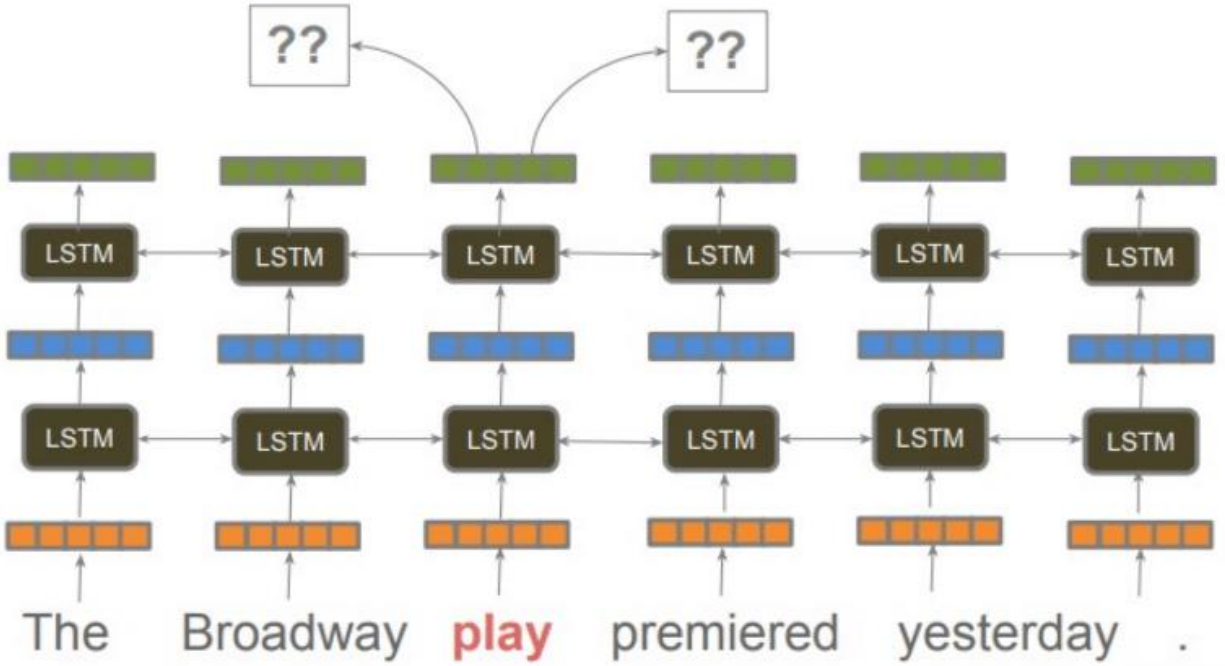
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

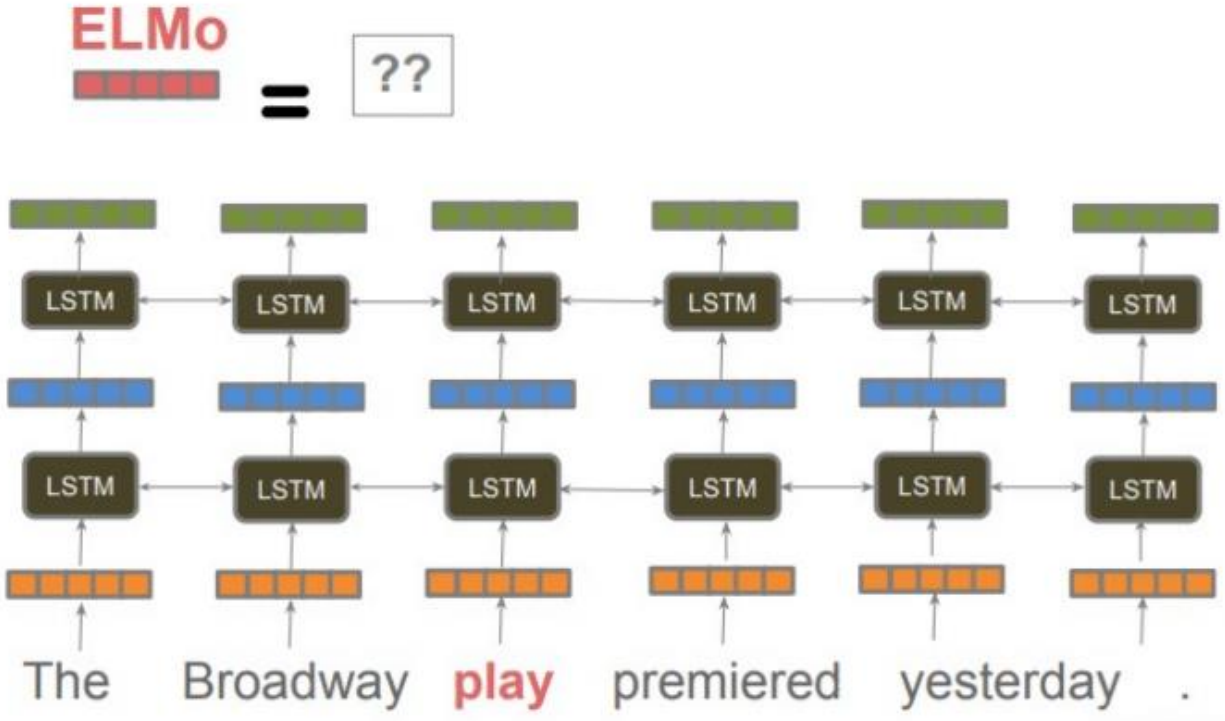
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

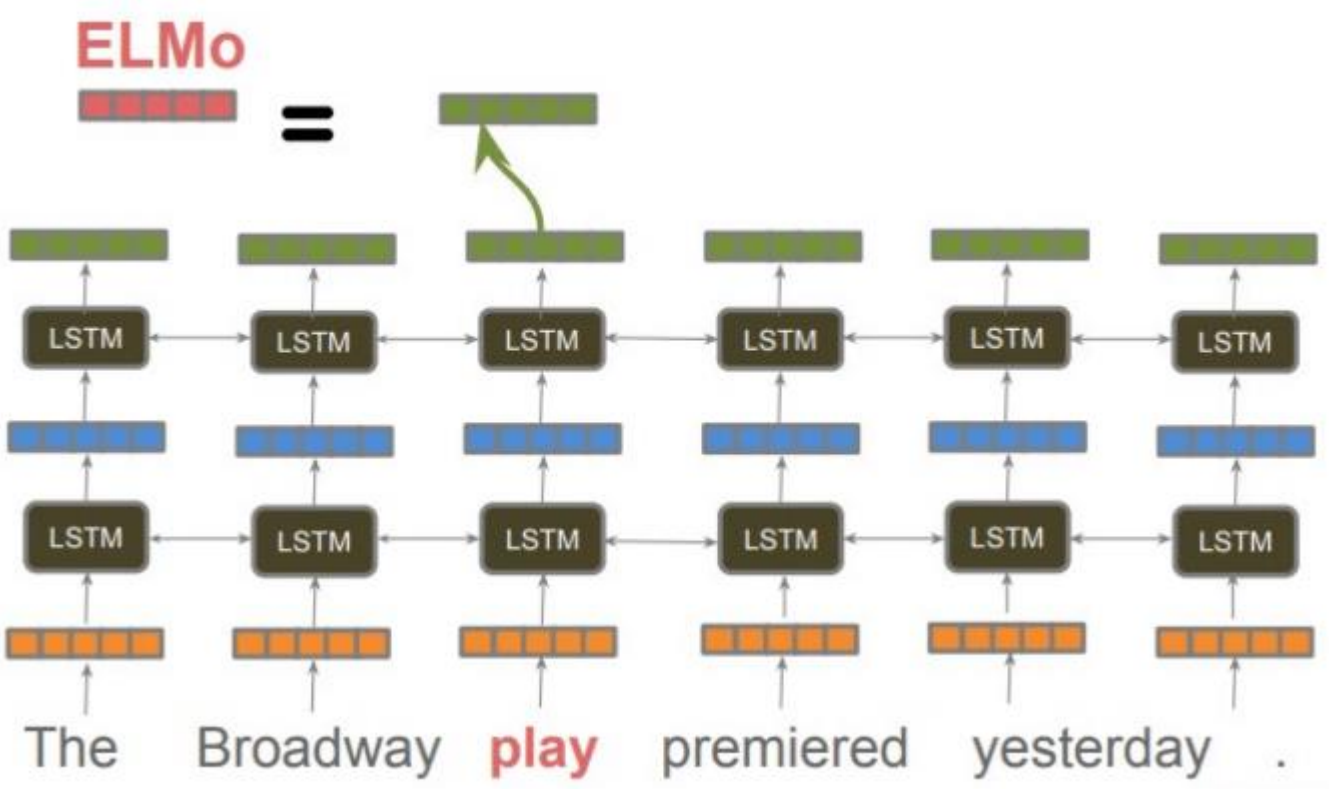
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

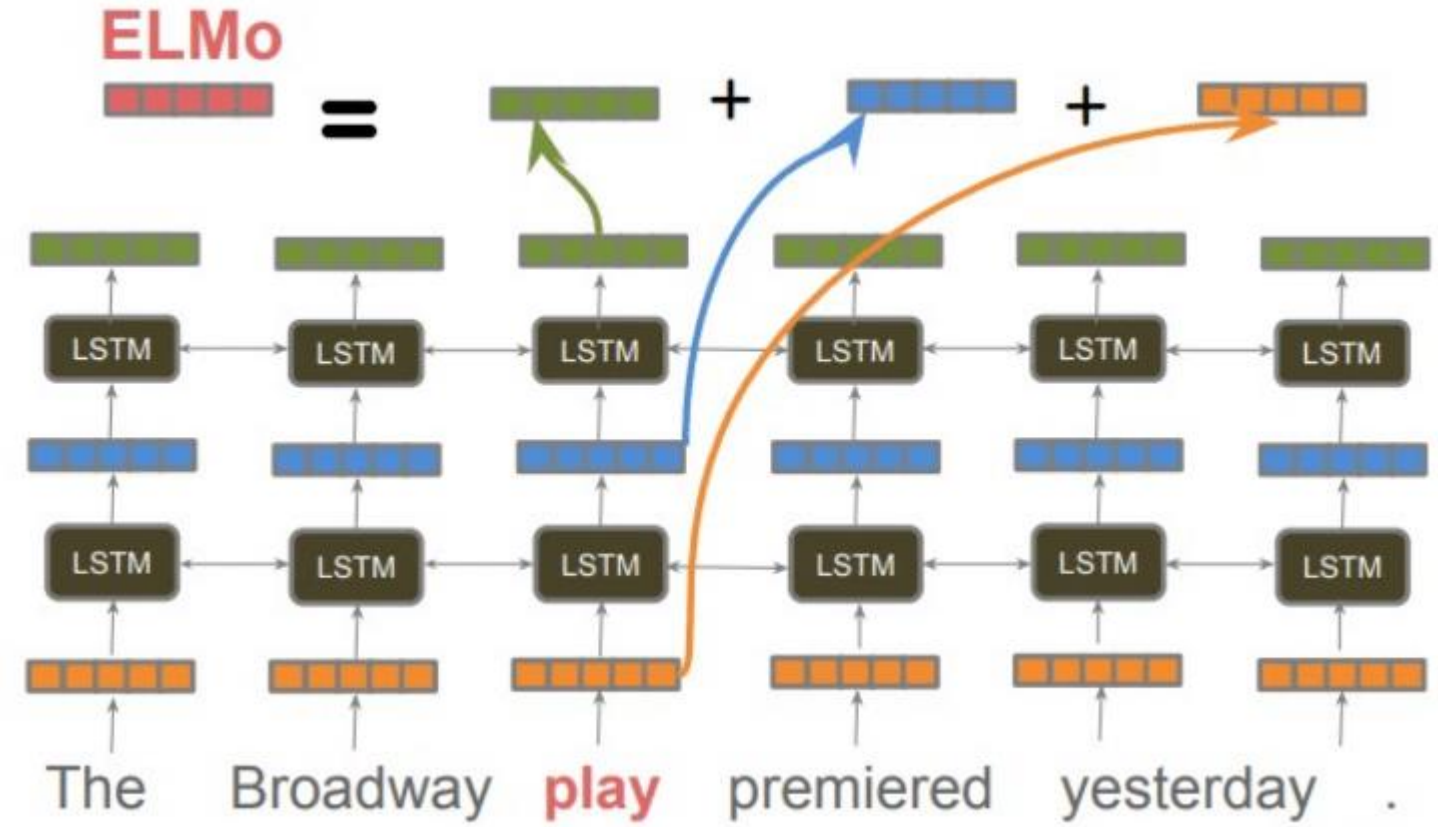
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

A Review of ELMo

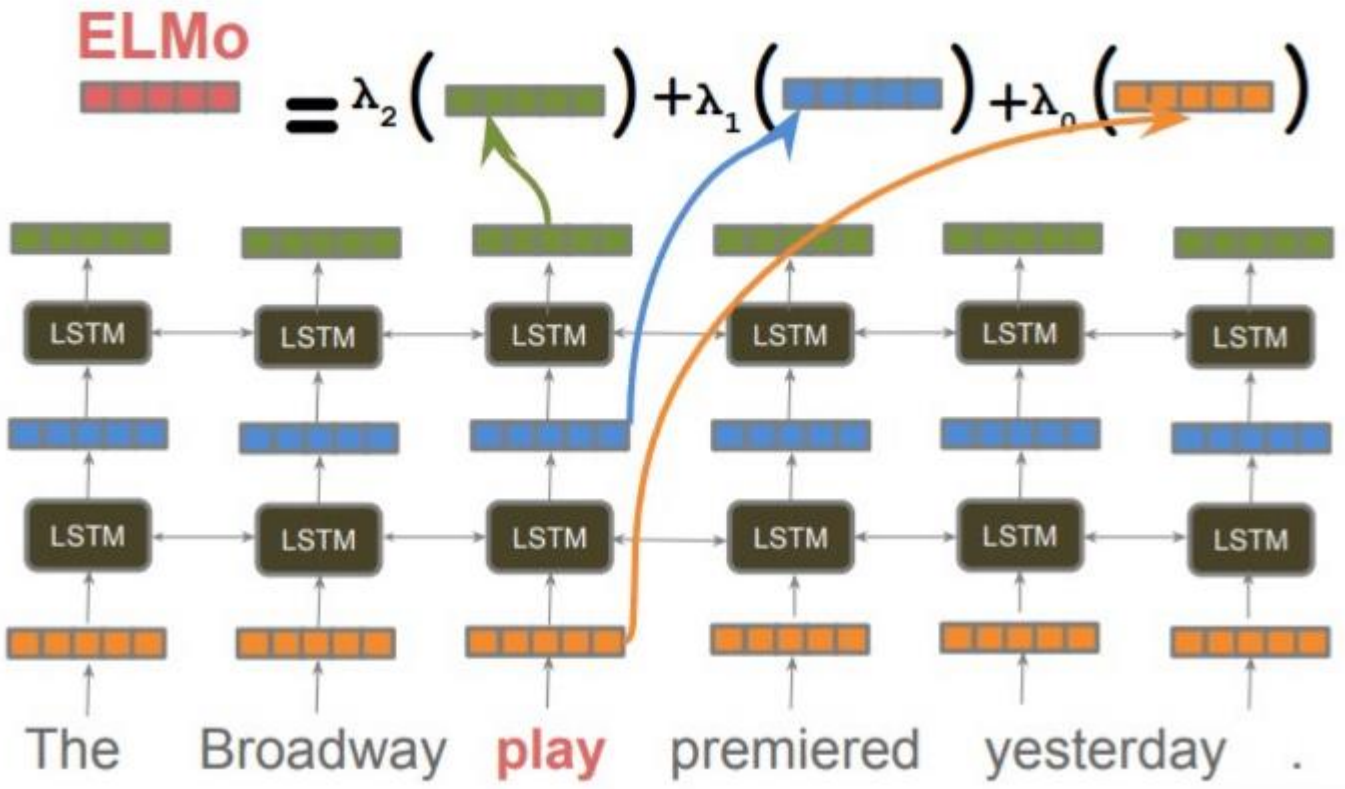
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

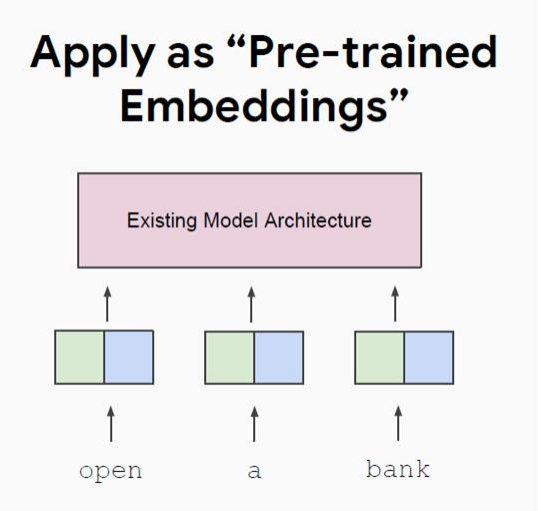
A Review of ELMo

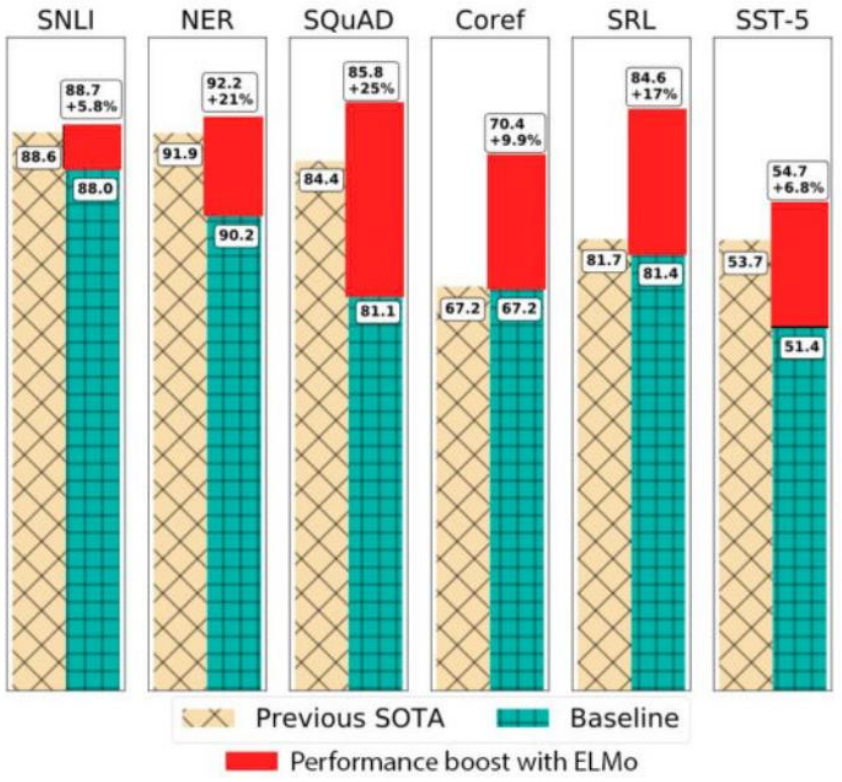
- An example of ELMo:



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

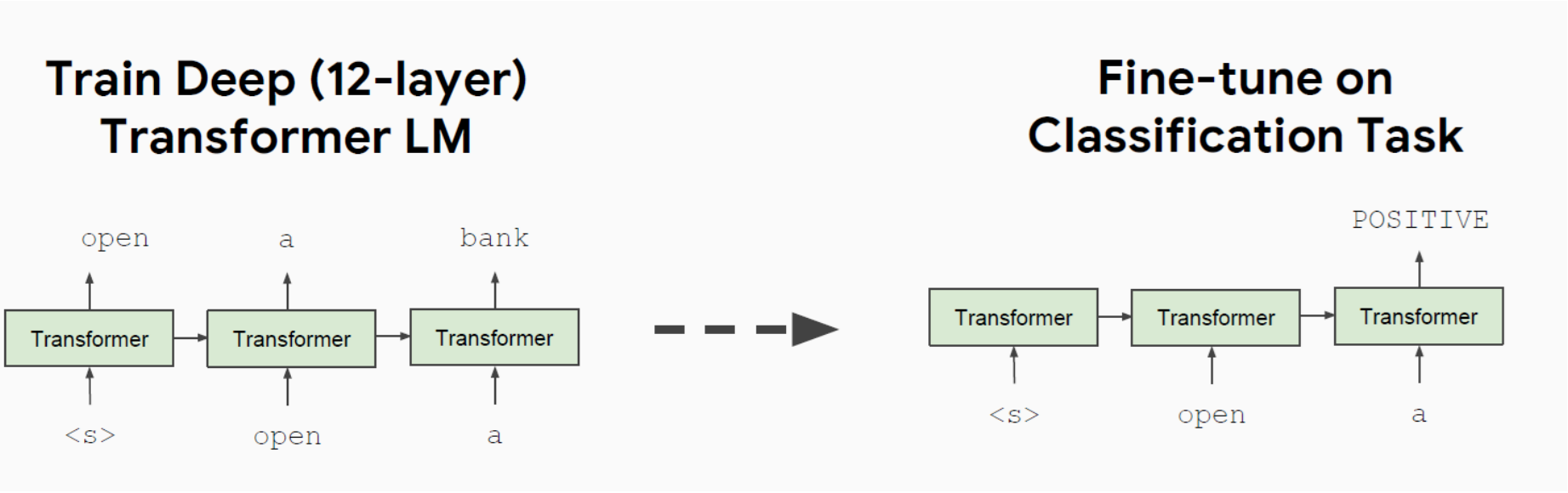
Αξιολόγηση με εξωγενή καθήκοντα





Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

OpenAI GPT: Generative Pre-trained Transformer



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

OpenAI GPT: Generative Pre-trained Transformer

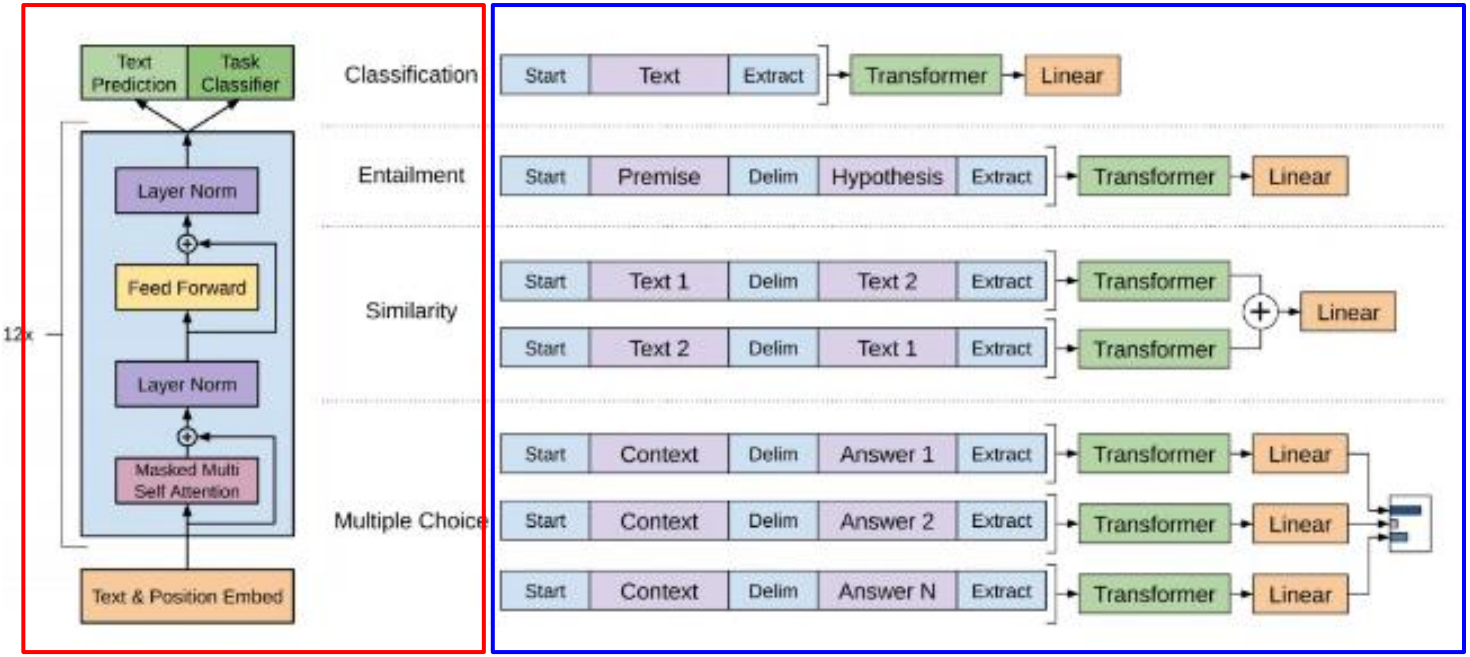
- Προ-εκπαίδευση χωρίς επίβλεψη για μεγιστοποίηση της log-likelihood :

$$L_1(\mathcal{U}) = \sum_i \log P(u_i | u_{i-k}, \ldots, u_{i-1}; \Theta)$$

Πού ***U*** = {u_1, …, u_n} is ένα μη εποπτευόμενο corpus από tokens, ***k*** είναι το μέγεθος του context window, ***P*** μοντελοποιείται ως νευρωνικό δίκτυο με παραμέτρους ***Θ*** (Radford et al., 2018).

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

OpenAI GPT: Generative Pre-trained Transformer



Μετασχηματισμοί εισόδου για τελειοποίηση διαφορετικών εργασιών.

Μετατρέψτε όλες τις δομημένες εισόδους σε ακολουθίες διακριτικών, ακολουθούμενες linear + softmax layer.

Αρχιτεκτονική μετασχηματιστών και στόχοι εκπαίδευσης

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Ζητήματα με προηγούμενες μεθόδους

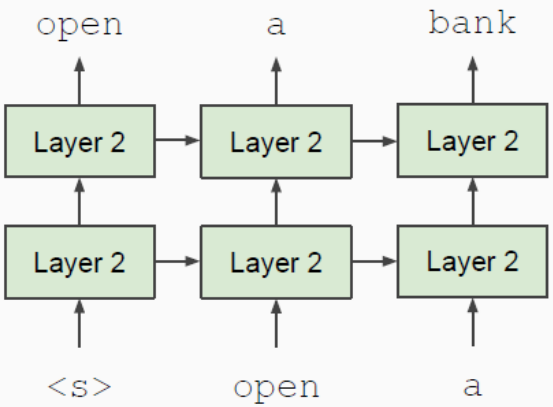
- **Πρόβλημα**: Language models Χρησιμοποιήστε μόνο αριστερό ή δεξί πλαίσιο, αλλά η κατανόηση της γλώσσας είναι αμφίδρομη.
- Γιατί τα LM είναι μονής κατεύθυνσης?
  - Λόγος 1: Η κατευθυντικότητα είναι απαραίτητη για τη δημιουργία μιας καλά διαμορφωμένης κατανομής πιθανότητας.
  - Λόγος 2: Οι λέξεις μπορούν να "δουν τον εαυτό τους" σε έναν αμφίδρομο κωδικοποιητή.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Unidirectional vs. Bidirectional Models

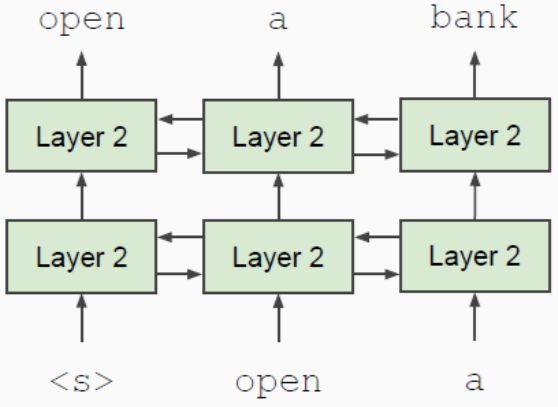
Unidirectional context

Build representation incrementally



Bidirectional context

Words can “see themselves”



Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Bidirectional Encoder Representations from Transformers (BERT)

Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Bidirectional Encoder Representations from Transformers (BERT)

- Το BERT εισήχθη από την Google Research το 2018 (Devlin et al., 2018)

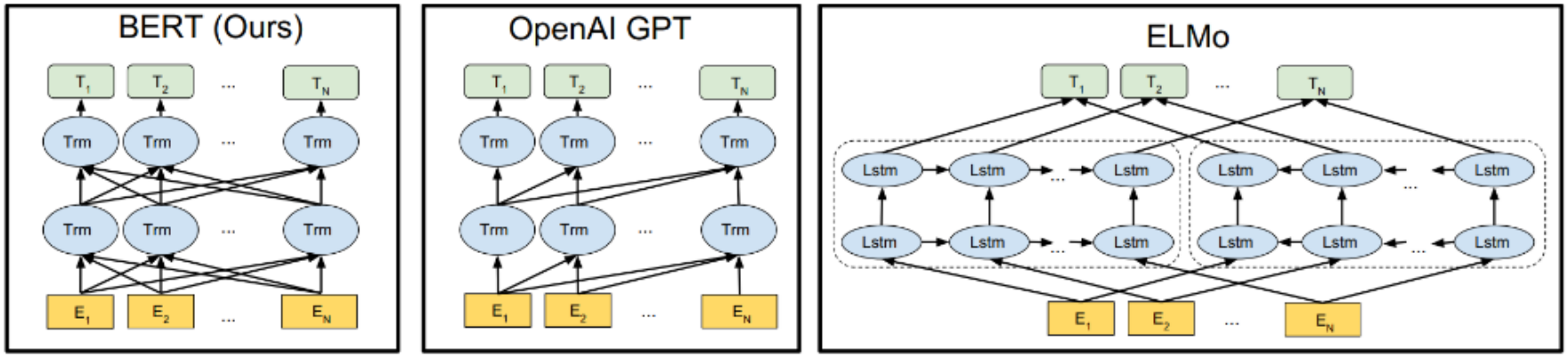
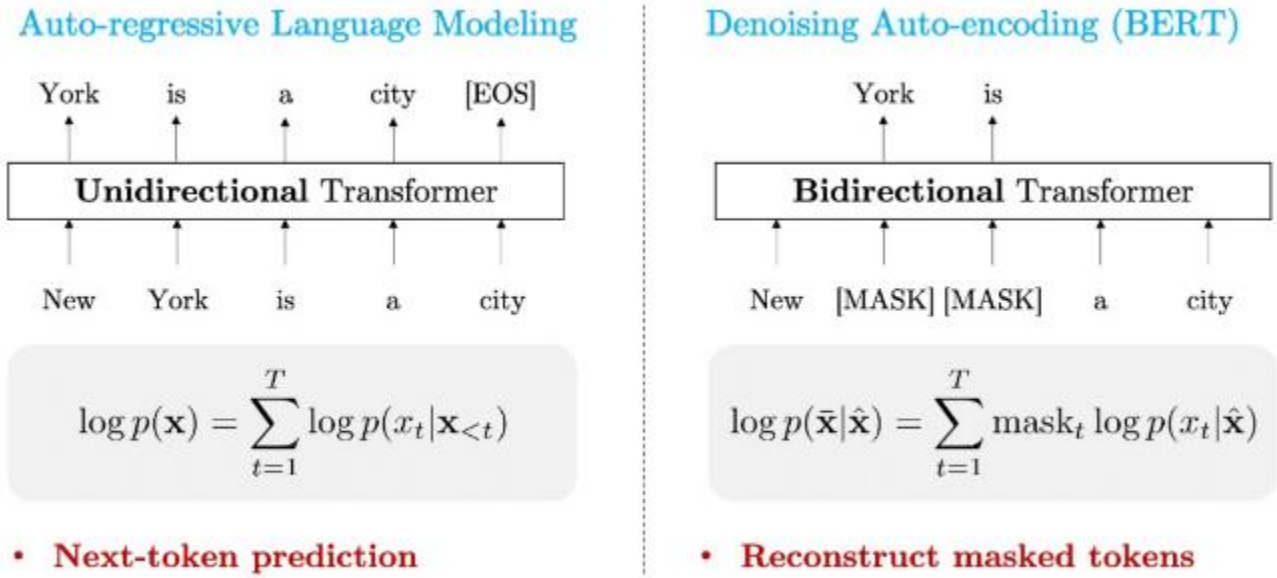


Figure 1: Differences in pre-training model architectures. BERT uses a bidirectional Transformer. OpenAI GPT uses a left-to-right Transformer. ELMo uses the concatenation of independently trained left-to-right and right-to-left LSTM to generate features for downstream tasks. Among three, only BERT representations are jointly conditioned on both left and right context in all layers.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Στόχοι για την κατάρτιση BERT

- Next-token prediction: Sequential / Unidirectional Transformer
- Masked-token prediction: Bidirectional Transformer



https://github.com/zihangdai/xlnet/blob/master/misc/slides.pdf

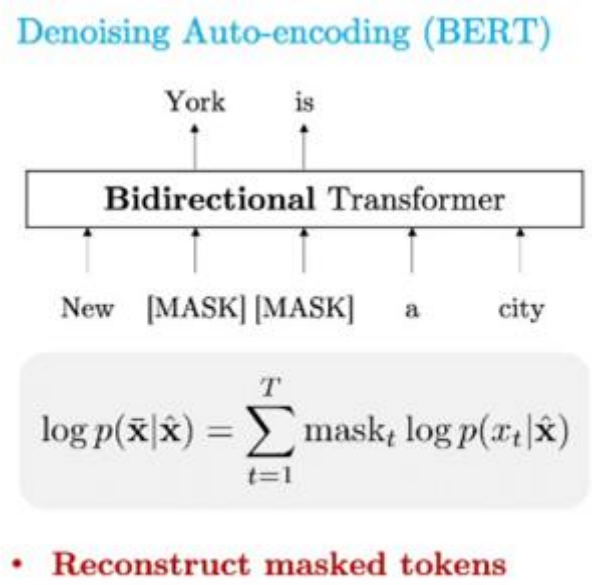
Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Στόχοι για την κατάρτιση BERT

- ❑ Next-token prediction: Sequential / Unidirectional Transformer
- ❑ Masked-token prediction: Bidirectional Transformer

- Mask out *k*% of the input words, and then predict the masked words.

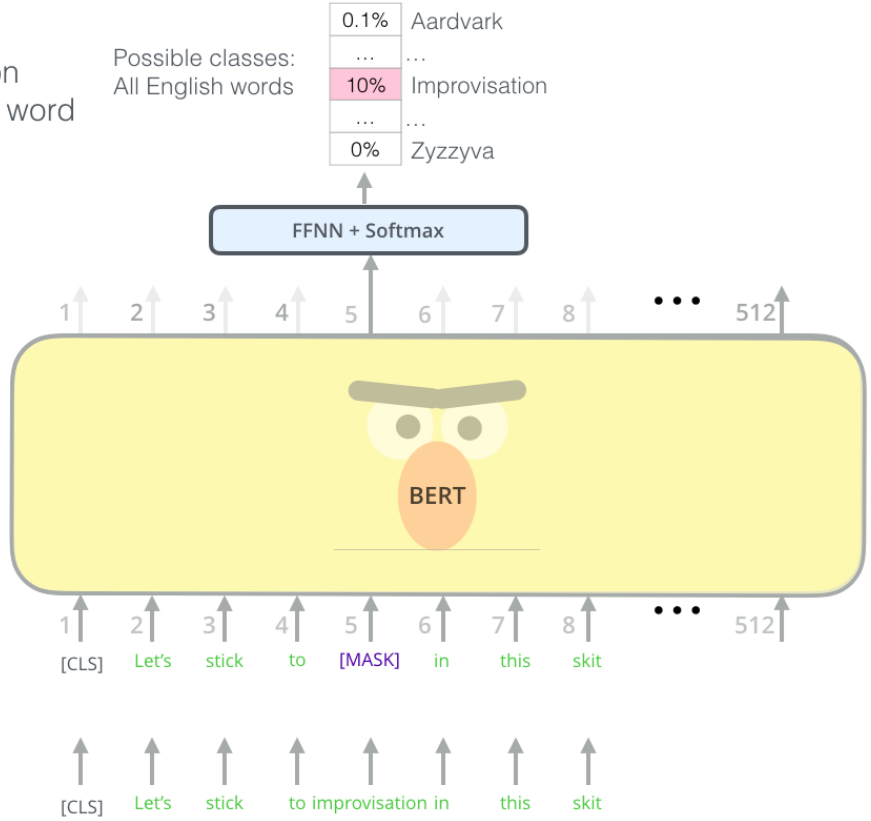
Input: The man went to the **[MASK]$_1$**.
He bought a **[MASK]$_2$** of milk.
Labels: **[MASK]$_1$** = *store*; **[MASK]$_2$** = *gallon*



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Masked Language Modeling

Use the output of the masked word's position to predict the masked word



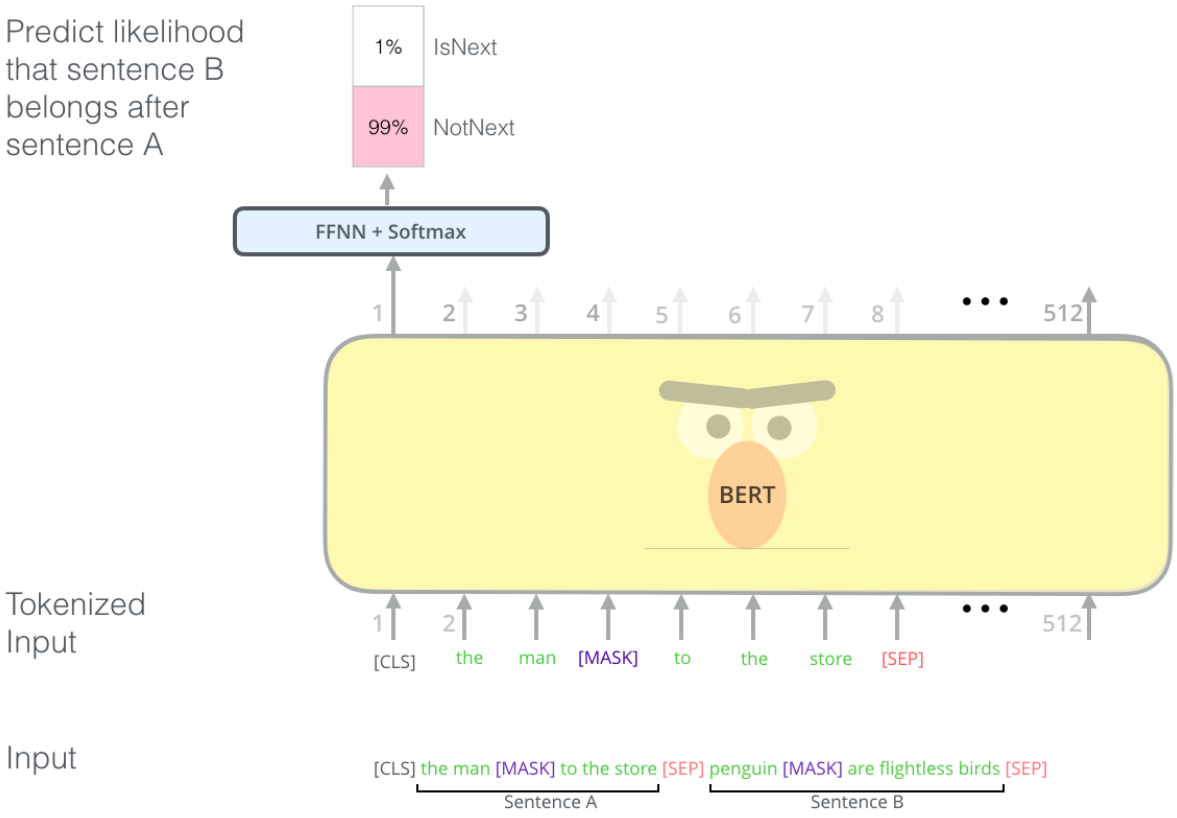
Randomly mask 15% of tokens

Input

Η έξυπνη εργασία γλωσσικής μοντελοποίησης του BERT καλύπτει το 15% των λέξεων στην είσοδο και ζητά από το μοντέλο να προβλέψει τη λέξη που λείπει.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

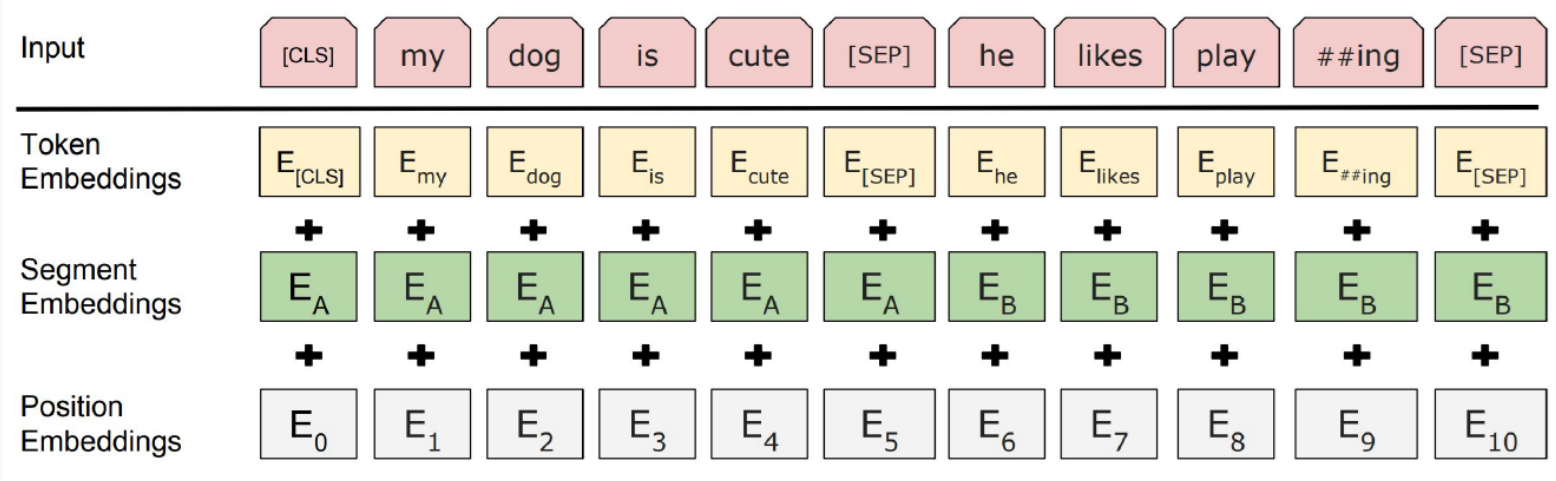
Next Sentence Prediction



Η δεύτερη εργασία στην οποία ο BERT είναι προ-εκπαιδευμένος είναι μια εργασία ταξινόμησης δύο προτάσεων. Το tokenization υπεραπλουστεύεται σε αυτό το γραφικό, καθώς ο BERT χρησιμοποιεί στην πραγματικότητα WordPieces ως μάρκες και όχι λέξεις --- έτσι ορισμένες λέξεις αναλύονται σε μικρότερα κομμάτια.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

BERT Input Representation



- Χρησιμοποιήστε 30.000 λεξιλόγιο WordPiece στην είσοδο.
- Κάθε διακριτικό είναι άθροισμα τριών ενσωματώσεων: Θέση, Τμήμα και Διακριτικό.
- Η ενιαία ακολουθία είναι πολύ πιο αποτελεσματική.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

BERT Performance Comparison

MAI4CAREU

GLUE Results

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	**Average** -
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	79.9	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	88.1	91.3	45.4	80.0	82.3	56.0	75.2
$BERT_{BASE}$	84.6/83.4	71.2	90.1	93.5	52.1	85.8	88.9	66.4	79.6
$BERT_{LARGE}$	**86.7/85.9**	**72.1**	**91.1**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**81.9**

MultiNLI

Premise: Hills and mountains are especially sanctified in Jainism.
Hypothesis: Jainism hates nature.
Label: Contradiction

CoLa

Sentence: The wagon rumbled down the road.
Label: Acceptable

Sentence: The car honked down the road.
Label: Unacceptable

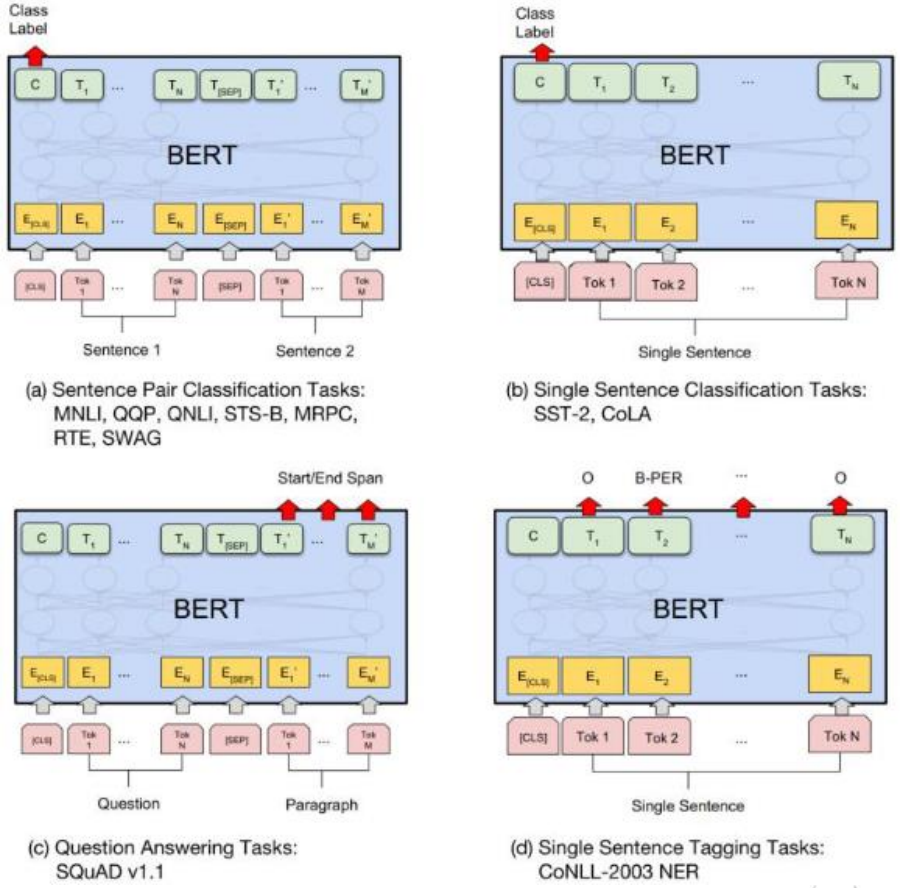
Co-financed by the European Union
Connecting Europe Facility

University of Cyprus

This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Fine-Tuning BERT

- **Context vector *C***: Πάρτε την τελική κρυφή κατάσταση που αντιστοιχεί στο πρώτο διακριτικό στην είσοδο [CLS].
- Μετασχηματισμός σε κατανομή πιθανότητας των ετικετών κλάσης: $P = \text{softmax}(CW^T)$.
- **Batch size**: 16, 32
- **Learning rate**: 5e-5, 3e-5, 2e-5
- **Number of epochs**: 3, 4



Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423

Περίληψη

- Εισαγωγή στις κατανεμημένες ενσωματώσεις με βάση τα συμφραζόμενα
- Ιστορία Μοντέλων Νευρωνικής Γλώσσας
  - Recurrent Neural Networks RNNs
  - ELMo
  - GPT
  - BERT
- Εσωτερική λειτουργία μετασχηματιστών και προσεγγίσεις τελειοποίησης
  - Masked Language Modeling and Next Sentence Prediction.

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action
No 2020-EU-IA-0087, co-financed by the EU CEF Telecom
under GA nr. INEA/CEF/ICT/A2020/2267423

Πόροι

- Jurafsky, D. and H. Martin Justin, Chapter 7. "Neural Networks and Neural Language Models" Speech and Language Processing
- Jurafsky, D. and H. Martin Justin, Chapter 11. "Transfer Learning with Contextual Embeddings and Pre-trained language models" Speech and Language Processing
- Illustrated BERT, ELMo by Jay Alammar: https://jalammar.github.io/illustrated-bert/
- Hugging Face's Transformers: https://huggingface.co/docs/transformers/index
- Transformer-based Models Hub: https://huggingface.co/models

Co-financed by the European Union
Connecting Europe Facility
University of Cyprus
This Master is run under the context of Action No 2020-EU-IA-0087, co-financed by the EU CEF Telecom under GA nr. INEA/CEF/ICT/A2020/2267423