

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΣΧΕΔΙΟ ΠΑΡΟΥΣΙΑΣΗΣ
ΑΤΟΜΙΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ατομική Διπλωματική Εργασία

Τίτλος Εργασίας (ΚΕΦΑΛΑΙΑ)

Όνομα και Επώνυμο Φοιτητή

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2017

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Τίτλος Ατομικής Διπλωματικής Εργασίας
Chatbot advisor for students

Όνομα Φοιτητή
Αναστάσιος Ανδρέου

Επιβλέπων Καθηγητής
Μάριος Δικαιάκος

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των
απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του
Πανεπιστημίου Κύπρου

Μάιος 2024

Ευχαριστίες

Εδώ μπορείτε να διατυπώσετε τυχόν ευχαριστίες προς πρόσωπα τα οποία αισθάνεσθε ότι συνέβαλαν αποφασιστικά ή σας βοήθησαν στην ολοκλήρωση της διπλωματικής σας εργασίας. Επίσης, μπορείτε να διατυπώσετε ευχαριστίες προς Ιδρύματα, Ερευνητικά Κέντρα, κλπ.

Περίληψη

Εδώ θα γράψετε μία περίληψη της ΑΔΕ σας. Θα πρέπει να είναι σύντομη και να δίδει με σαφήνεια το θέμα που μελετήσατε και τα αποτελέσματά σας. Η περίληψη δεν πρέπει να υπερβαίνει την μία σελίδα, δηλ. θα χρησιμοποιήσετε μόνο το χώρο κάτω από τον τίτλο «Περίληψη».

Περιεχόμενα

Κεφάλαιο 1

Εισαγωγή.....	8
1.1 Εξέλιξη Γλωσσικών Μοντέλων	8
1.2 Ιστορική Αναδρομή των Chatbots.	10
1.3 Φυσική γλώσσα (Natural Language Processing -NLP)	14
1.4 Επισκόπηση νευρωνικών δικτύων/ recurrent neural networks (RNNs), and transformers	16
1.5 Επεξήγηση για τα μοντέλα BERT	18

Κεφάλαιο 2

Περιγραφή Προβλήματος.....	24
2.1 Work Motivation	24
2.2 Ανάλυση του Προβλήματος	25
2.3 Προσέγγιση λύσης	29

Κεφάλαιο 3

Βασικοί Ορισμοί και Σύγκριση Μοντέλων.....	30
3.1 Embeddings και η σημαντικότητα τους	30
3.2 Word2Vec vs Transformers Model	33
3.3 Διανυσματικές Βάσεις	37

Κεφάλαιο 4

Σχεδιασμός Συστήματος και Σύγχρονα ΜΓΜ	39
4.1 Μεθοδολογία πρότασης	39
4.2 GPT-4 openAi / GPTs	43
4.3 Bard / Gemini	45

Κεφάλαιο 5

Αξιολόγηση Συστήματος/Συμπεράσματα.....	47
5.1 Αξιολόγηση εφαρμογής με χρήση διανυσματικής βάσης Chroma DB	47
5.2 Αξιολόγηση καινούργιας εφαρμογής GPTs	53
5.3 Αξιολόγηση έτοιμης εφαρμογής GPTs	63
5.4 Συμπεράσματα από την σύγκριση των μοντέλων	70
 Βιβλιογραφία	 72
Παραρτήματα	73

Κεφάλαιο 1

Εισαγωγή

1.1 Εξέλιξη Γλωσσικών Μοντέλων	8
1.2 Ιστορική Αναδρομή των Chatbots.	10
1.3 Φυσική γλώσσα (Natural Language Processing -NLP)	14
1.4 Επισκόπηση νευρωνικών δικτύων/ recurrent neural networks (RNNs), and transformers	16
1.5 Επεξήγηση για τα μοντέλα BERT	18

1.1 Εξέλιξη των γλωσσικών μοντέλων

Τα γλωσσικά μοντέλα (Language Modeling-LM) θεωρούνται ένα υποπεδίο της τεχνητής νοημοσύνης (Artificial Intelligence-AI) που ασχολείται με την αναπαράσταση και την στατική ανάλυση της φυσικής γλώσσας. Κύριος στόχος των LM είναι η κατανόηση των σχέσεων και των μοτίβων που μπορεί να παρουσιάζουν οι λέξεις.

Μία από τις πιο θεμελιώδεις εργασίες που εκτελούν τα LM είναι η πρόβλεψη της επόμενης λέξης μέσα σε μία πρόταση. Η μέθοδος που ακολουθείται είναι μία ακολουθία λέξεων και προσπαθεί να προβλέψει την πιο πιθανή λέξη στην ακολουθία. Τα LM μπορούν να χρησιμοποιηθούν για την δημιουργία διαφορετικών μορφών κειμένου όπως προγραμματισμός, σενάρια, email κ.λπ.

- Υπάρχουν διαφόρων ειδών LM. Συγκεκριμένα:**Statistical Language Models (SLMs) (1990s)**: Αυτά τα μοντέλα κάνουν προβλέψεις με βάση το τι έχουν δει στο παρελθόν. Εκπαιδεύονται με τη χρήση μεγάλων κειμένων, όπως βιβλία ή άρθρα, και συνήθως καταγράφουν πόσες φορές κάποιες λέξεις εμφανίζονται μαζί. Έχουν την ιδιότητα να προβλέπουν την επόμενη λέξη με βάση τις προηγούμενες λέξεις της πρότασης, χρησιμοποιώντας, ωστόσο, μόνο μερικές λέξεις πριν από την λέξη που ψάχνουν.
- **Neural Language Models (NLM) (2013 [2])**: Αυτά τα μοντέλα παρουσιάζουν μία σημαντική πρόοδο σε σχέση με τα SLMs στην επεξεργασία της φυσικής γλώσσας. Αξιοποιούν τα νευρωνικά δίκτυα για να μοντελοποιήσουν περίπλοκες σχέσεις μεταξύ των λέξεων. Χρησιμοποιούν την έννοια των λεξικών αναπαραστάσεων, όπου, κάθε

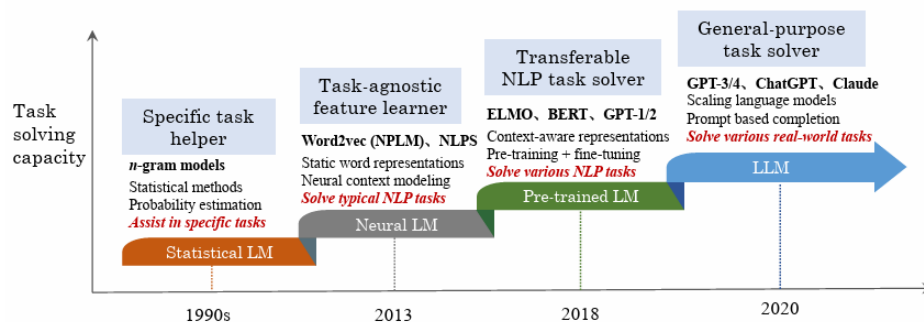
λέξη αναπαρίσταται με ένα διάνυσμα αριθμών. Αυτό είναι βοηθητικό στην κατανόηση του πλαισίου της γλώσσας. Αυτά τα διανύσματα περιέχουν πληροφορίες, όχι μόνο για τη συχνότητα των λέξεων, αλλά και για το πώς συνδέονται μεταξύ τους σε διάφορα πλαίσια. Τα NLMs χρησιμοποιούν τεχνητά νευρωνικά δίκτυα, τα οποία είναι ικανά να μάθουν από τεράστιες ποσότητες δεδομένων γλώσσας. Αυτά τα μοντέλα εκπαιδεύονται να προβλέπουν την επόμενη λέξη σε μια πρόταση, βάση των προηγούμενων λέξεων. Η επιτυχία τους δεν βασίζεται μόνο στο να βρίσκουν ποια λέξη έρχεται συχνά μετά από μια άλλη, αλλά και στο να καταλαβαίνουν πότε διαφορετικές λέξεις μπορούν να χρησιμοποιηθούν σε διαφορετικά πλαίσια για να εκφράσουν παρόμοιες ή διαφορετικές ιδέες. Μία αρχιτεκτονική που χρησιμοποιείται είναι τα Recurrent Neural Networks (RNNs), τα οποία είναι σχεδιασμένα για την ανάλυση και επεξεργασία ακολουθιών δεδομένων, όπως λέξεις σε μια πρόταση ή προτάσεις σε ένα κείμενο. Αυτό επιτυγχάνεται μέσω της δυνατότητάς τους να διατηρούν μία πληροφορία από προηγούμενα βήματα της ακολουθίας και να την χρησιμοποιούν για την λήψη αποφάσεων ή προβλέψεων στα επόμενα βήματα. Αυτή η μοναδική ικανότητα των RNNs να "θυμούνται" προηγούμενα στοιχεία της ακολουθίας τα καθιστά ιδανικά για την κατανόηση του πλαισίου και την παραγωγή συνεκτικών και σχετικών αποτελεσμάτων στην επεξεργασία της γλώσσας. Η πρόοδος και οι δυνατότητες αυτών των μοντέλων έχουν αναλυθεί εκτενώς από τον Zhao et al. [2].

- **Pre-trained Language Models (PLMs) (2018 [2]):** Αυτά τα μοντέλα βασίζονται στα NLM και προσθέτουν την έννοια του pre-trained, η οποία βελτιώνει σημαντικά την απόδοση σε διάφορες εργασίες που αφορούν την επεξεργασία της φυσικής γλώσσας. Ο όρος "εργασία" αναφέρεται σε ένα συγκεκριμένο πρόβλημα που ένα νευρωνικό δίκτυο προσπαθεί να επιλύσει με βάση την εκπαίδευση και ανάλυση της γλώσσας. Μερικά παραδείγματα εργασιών περιλαμβάνουν την κατανόηση φυσικής γλώσσας, την απάντηση σε ερωτήματα, τη μετάφραση μεταξύ γλωσσών και την παραγωγή κειμένου. Το κύριο πλεονέκτημα των PLMs είναι ότι εκπαιδεύονται προκαταρκτικά σε μια τεράστια ποσότητα κειμένου πριν αντιμετωπίσουν συγκεκριμένες εργασίες. Αυτή η εκπαίδευση περιλαμβάνει εργασίες όπως η πρόβλεψη της επόμενης λέξης σε μία ακολουθία ή τον προσδιορισμό των σχέσεων των λέξεων. Έτσι, το μοντέλο αποκτά γνώσεις σχετικές με την δομή και την χρήση της γλώσσας. Όταν ένα μοντέλο προ-εκπαιδευτεί μπορεί, στην συνέχεια, να βελτιστοποιηθεί (Fine Tuning) για συγκεκριμένες εργασίες και να εκτελέσει εξειδικευμένες εργασίες με υψηλή ακρίβεια. Για παράδειγμα ένα προ-εκπαιδευμένο μοντέλο γλώσσας μπορεί να υποβληθεί σε fine-tuning χρησιμοποιώντας ένα σύνολο δεδομένων που περιλαμβάνει ιατρικά κείμενα. Κατά τη διάρκεια αυτής της διαδικασίας, τα βάρη του μοντέλου προσαρμόζονται έτσι

ώστε να μπορεί να κατανοεί καλύτερα τα ιατρικά δεδομένα. Αυτό επιτρέπει στο μοντέλο να παρέχει πιο ακριβείς απαντήσεις στα ιατρικά ερωτήματα.

- **Large Language Models (LLMs) (2020 [2]):** Αυτά τα μοντέλα είναι βασισμένα στα PLM. Χαρακτηρίζονται για την επεκτασιμότητα του (scaling), δηλαδή περιλαμβάνουν πολύ μεγαλύτερα μεγέθη μοντέλων και δεδομένων σε σχέση με τα PLM. Η κλιμάκωση υποδηλώνει ότι καθώς τα μοντέλα γίνονται μεγαλύτερα, αποκτούν αυξημένη ικανότητα κατανόησης και παραγωγής ανθρώπινης γλώσσας. Λόγω της κλιμάκωσης, τα LLM περιλαμβάνουν δυνατότητες όπως η δημιουργία διαφορετικών γλωσσών και η ικανότητα εκμάθησης μέσα από ελάχιστα παραδείγματα. Η χρήση τέτοιων μοντέλων συχνά περιλαμβάνει και μία διεπαφή προτροπής (Prompting Interface), όπου οι χρήστες μπορούν να παρέχουν οδηγίες που καθορίζουν τη συμπεριφορά του συστήματος. Τα μεγάλα γλωσσικά μοντέλα έχουν αναδειχθεί ως ένα ισχυρό εργαλείο στον τομέα της επεξεργασίας της φυσικής γλώσσας (NLP). Μπορούν να χαρακτηριστούν ως ισχυρά προγράμματα τα οποία μπορούν να γράφουν ιστορίες, να απαντούν σε ερωτήσεις και να κάνουν συζητήσεις σαν ένας άνθρωπος. Η αποτελεσματικότητα και η αυξημένη ικανότητα των μεγάλων γλωσσικών μοντέλων έχει αναλυθεί λεπτομερώς από τους Naveed et al. [1].

Η εικόνα 1 παρακάτω απεικονίζει τη διαδικασία εξέλιξης των τεσσάρων γενεών γλωσσικών μοντέλων (LM) και τις σημαντικές αλλαγές στην ικανότητα επίλυσης εργασιών.



Εικόνα 1Μια διαδικασία εξέλιξης των τεσσάρων γενεών γλωσσικών μοντέλων (LM)

1.2 Ιστορική Αναδρομή των chatbots

Η ιστορία των chatbots ξεκινά από τα μέσα του 20^{ου} αιώνα με απλά συστήματα με σενάριο και φτάνει μέχρι σήμερα με τους εξελιγμένους AI βοηθούς με τους οποίους μπορεί να γίνει αλληλεπίδραση. Σύμφωνα με την πηγή 'A History of Chatbots' [3], η εξέλιξη των chatbots περιλαμβάνει διάφορα στάδια και σημαντικά ορόσημα.

1950 - The Turing Test: ο Alan Turing ήταν ένας Βρετανός επιστήμονας υπολογιστών και μαθηματικός, ο οποίος πρότεινε το Turing Test στην εργασία του «Υπολογιστική Μηχανή

και Νοημοσύνη». Αυτό το τεστ αξιολογούσε την ικανότητα ενός μηχανήματος να εκδηλώνει νοημοσύνη ισοδύναμη με αυτή του ανθρώπου, παρέχοντας τον πυρήνα της ιδέας πίσω από τα chatbots και την τεχνική νοημοσύνη. Σύμφωνα με τον Alan Turing μία μηχανή κατατίθεται ως έξυπνη όταν είναι ικανή να προσομοιώσει την ανθρώπινη συμπεριφορά σε τέτοιο βαθμό, ώστε να πείσει κάποιον πως βρίσκεται σε πραγματική συνομιλία με έναν άλλον άνθρωπο.

1966 – ELIZA: Είχε αναπτυχθεί από τον Joseph Weizenbaum στο MIT, το ELIZA το οποίο συχνά θεωρείται το πρώτο chatbot. Προσομοίαζε τη συνομιλία με την μεθοδολογία του «pattern matching» η οποία έδινε στους χρήστες μία ψευδαίσθηση κατανόησης από την πλευρά του μηχανήματος αν και στην πραγματικότητα δεν μπορούσε να κατανοήσει τη συνομιλία.[4]

1972 – PARRY: Δημιουργήθηκε από τον ψυχίατρο Kenneth Colby όταν ήταν στο πανεπιστήμιο του Stanford, το PARRY σχεδιάστηκε για να προσομοιώνει έναν ασθενή με παρανοϊκή σχιζοφρένεια. Ήταν πιο προηγμένο από το ELIZA, ενσωματώνοντας ένα στοιχειώδες μοντέλο συμπεριφοράς ενός παρανοϊκού ατόμου και μπόρεσε να περάσει το τεστ Turing πείθοντας ορισμένους ψυχιάτρους ότι ήταν ανθρώπινο. Αυτό το chatbot είχε δημιουργηθεί με τη χρήση της LISP, μιας γλώσσας προγραμματισμού ιδανικής για ανάπτυξη συστημάτων τεχνικής νοημοσύνης. Στην εικόνα 2 φαίνεται μια συνομιλία που είχε το μοντέλο Parry με το Eliza.

Network Working Group	V. Cerf
Request for Comments: 439	SU-ERL
NIC: 13771	21 January 1973

PARRY Encounters the DOCTOR

The session recorded below actually happened on 18 September 1972 and is herewith reproduced for your entertainment.

Good Evening.

Tell me your problems. Please terminate input with a period or a question mark.

*People get on my nerves sometimes.

I am not sure I understand you fully.

*You should pay more attention.

Suppose you should pay more attention.

*You're entitled to your own opinion.

Εικόνα 2: Συνομιλία από το Parry και το Eliza chatbot

1988 - Jabberwacky: Jabberwacky, ένα πιο προηγμένο AI από τον Rollo Carpenter, σχεδιάστηκε για να προσομοιώνει τη φυσική ανθρώπινη συνομιλία με έναν διασκεδαστικό και χιουμοριστικό τρόπο.

1995 - ALICE (Artificial Linguistic Internet Computer Entity): Ο Richard Wallace ανέπτυξε το ALICE (Artificial Linguistic Internet Computer Entity), έναν πρωτοποριακό

chatbot που επεξεργαζόταν φυσική γλώσσα χρησιμοποιώντας την AIML (Artificial Intelligence Markup Language), μια εξειδικευμένη γλώσσα προγραμματισμού η οποία δημιουργήθηκε ειδικά για τη σύνθεση απαντήσεων από chatbots. Το ALICE, έχοντας διακριθεί για τις καινοτόμες του ικανότητες, κατέκτησε πολυάριθμα βραβεία, συμπεριλαμβανομένων τριών βραβείων Loebner (2000, 2001, 2004), επιβεβαιώνοντας την υπεροχή του στον τομέα της επεξεργασίας φυσικής γλώσσας.



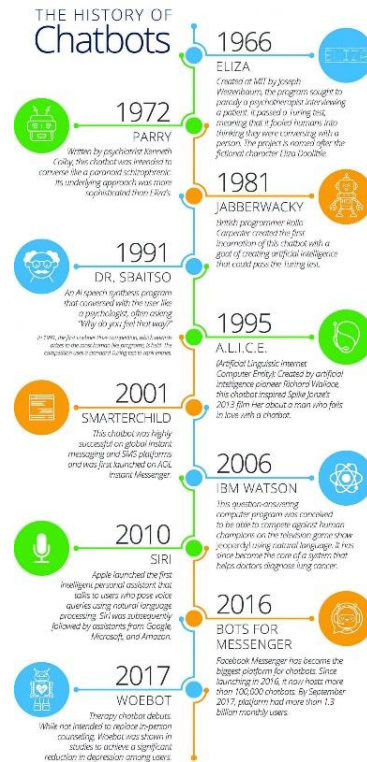
Εικόνα 3 SmartChild Chatbot

2001 –SmarterChild: Το SmarterChild ήταν ένα chatbot διαθέσιμο σε πλατφόρμες SMS και άμεσων μηνυμάτων, προσφέροντας στους χρήστες τη δυνατότητα να αλληλεπιδρούν με ένα ευρύ φάσμα υπηρεσιών πληροφοριών.

2010s - AI and Machine Learning Revolution: Η εισαγωγή τεχνολογιών μηχανικής μάθησης και τεχνητής νοημοσύνης, όπως η επεξεργασία φυσικής γλώσσας και η βαθιά εκμάθηση, οδήγησε στην ανάπτυξη πολύ πιο εξελιγμένων chatbots. Εταιρείες όπως η Google, η Microsoft και το Facebook επένδυσαν πολλά στην τεχνητή νοημοσύνη, οδηγώντας στη δημιουργία βοηθών όπως οι Google Assistant, Siri και Cortana.

2016 - The Rise of Messaging App Bots: Πλατφόρμες όπως το Facebook Messenger, το Slack και το Telegram άνοιξαν τα API τους για προγραμματιστές να δημιουργήσουν chatbot τρίτων, οδηγώντας σε αύξηση των bots για εξυπηρέτηση πελατών, ηλεκτρονικό εμπόριο, ψυχαγωγία και πολλά άλλα.

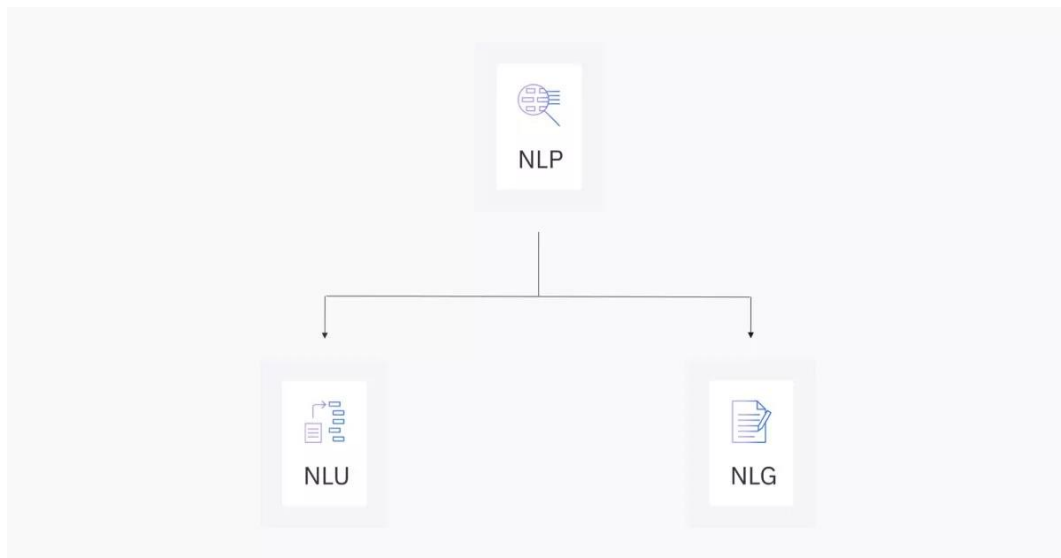
2020s - GPT and Advanced AI Chatbots: Η εισαγωγή των μοντέλων Generative Pretrained Transformer (GPT) από την OpenAI σηματοδότησε ένα σημαντικό άλμα στην τεχνολογία chatbot. Αυτά τα μοντέλα, ξεκινώντας από το GPT-2 έως τις πιο πρόσφατες επαναλήψεις, μπορούν να δημιουργήσουν κείμενο που μοιάζει με άνθρωπο με βάση τα δεδομένα που τους δίνονται.



Εικόνα 4 Στην εικόνα απεικονίζεται η εξέλιξη των chatbot από την έναρξή τους το 1966 με το ELIZA σε πιο προηγμένες επαναλήψεις όπως το IBM Watson και σύγχρονες πλατφόρμες όπως το Siri και το Bots for Messenger.

1.3 Φυσική γλώσσα (Natural Language Processing -NLP)

Η NLP αποτελεί ένα δυναμικό πεδίο της ΑΙ με στόχο την ανάπτυξη συστημάτων που να μπορούν να επεξεργάζονται, να αναλύουν και να παράγουν ανθρώπινη γλώσσα. Βασίζεται σε πολλούς διαφορετικούς επιστημονικούς κλάδους, όπως η υπολογιστική γλωσσολογία, η στατιστική και η μηχανική μάθηση. Η επεξεργασία φυσικής γλώσσας υπερβαίνει κατά πολύ την απλή προγραμματιστική διαδικασία που επιδιώκει να κατανοήσει τη σημασία των λέξεων. Αφορά τη βαθιά κατανόηση του πλαισίου, της διάθεσης, της πρόθεσης και ακόμη και των πιο λεπτών αποχρώσεων του χιούμορ και του σαρκασμού. Όπως αναφέρεται από τους Nadkarni et al.[5], η επεξεργασία φυσικής γλώσσας είναι η διαδικασία μετατροπής μη δομημένου κειμένου, όπως ο προφορικός ή ο γραπτός λόγος, σε δομημένα δεδομένα που μπορεί να κατανοήσει ένας υπολογιστής [5]. Η NLP περιέχει δυο κύρια στοιχεία: την κατανόηση φυσικής γλώσσας (NLU), η οποία μεταφράζει το μη δομημένο κείμενο σε δομημένα δεδομένα και τη δημιουργία φυσικής γλώσσας (NLG), η οποία μεταφράζει δομημένα δεδομένα σε μη δομημένο κείμενο.



Natural Language Understanding (NLU):

Το Natural Language Understanding (NLU) περιέχει διάφορες εργασίες όπως Syntax Analysis, δηλαδή την κατανόηση της γραμματικής και της δομής της πρότασης ώστε να καθορίζει τις σχέσεις μεταξύ των λέξεων. Περιέχει επίσης το Semantic Analysis που βοηθά την κατανόηση της πρότασης με την ερμηνεία της κάθε λέξης μέσα στην πρόταση σύμφωνα με το περιεχόμενο της. Τέλος, το Pragmatic Analysis προσπαθεί να κατανοήσει το επιδιωκόμενο αποτέλεσμα μίας πρότασης από την πρόθεση του ομιλητή, λαμβάνοντας υπόψη το πλαίσιο και τον τρόπο που χρησιμοποιείται η γλώσσα.

Natural Language Generation (NLG):

Εικόνα 5: Αυτή η εικόνα απεικονίζει τη δομή της Επεξεργασίας Φυσικής Γλώσσας (NLP), η οποία χωρίζεται σε δύο κύρια συστατικά: Κατανόηση Φυσικής Γλώσσας (NLU) και Δημιουργία Φυσικής Γλώσσας (NLG). Το NLP είναι ένα πεδίο τεχνητής νοημοσύνης που εστιάζει στην αλληλεπίδραση μεταξύ υπολογιστών και ανθρώπων μέσω της γλώσσας.

Το Natural Language Generation (NLG) μπορεί να θεωρηθεί ότι περιέχει αντίστροφες εργασίες σε σχέση με το NLU. Συγκεκριμένα, το text planning για να αποφασίσει για το περιεχόμενο του κειμένου που δημιουργείται με βάση τα δεδομένα εισόδου. Επιπρόσθετα, το sentence planning που είναι η οργάνωση του περιεχομένου σε προτάσεις, η επιλογή κατάλληλων λέξεων και ο προσδιορισμός της δομής των προτάσεων. Κατόπιν είναι το Text Realization που δημιουργεί προτάσεις διασφαλίζοντας ότι συμμορφώνονται με τους κανόνες της γραμματικής και της σύνταξης.

Applications of Natural Language Processing:

Αυτοματοποίηση Εργασιών: Τα chatbots πλέον με την βοήθεια της NLP είναι σε θέση να αναλάβουν ευρείες καθημερινές και επαναλαμβανόμενες εργασίες που παραδοσιακά ανατίθενται σε ανθρώπινο προσωπικό. Αυτό επιτρέπει στους εργαζομένους να αποδεσμευτούν

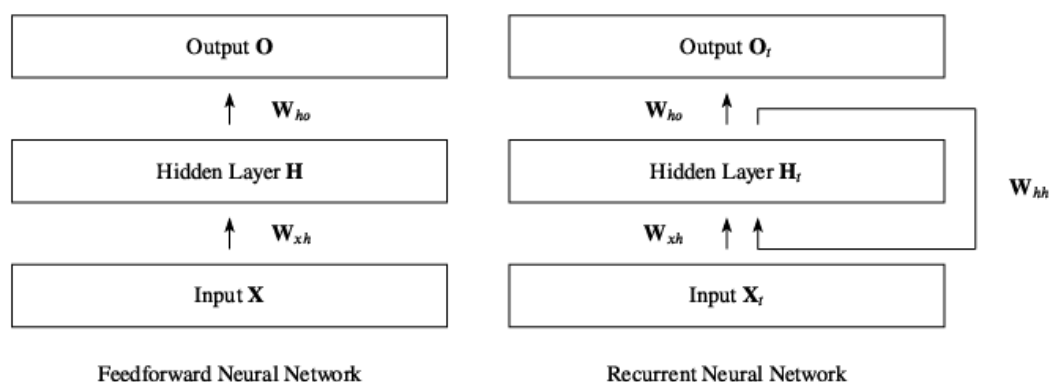
από αυτές τις δραστηριότητες και να αφιερώσουν τον χρόνο τους σε πιο σύνθετες και δημιουργικές εργασίες, αυξάνοντας έτσι την παραγωγικότητα στην εργασία.

Βελτίωση Αναζήτησης: Η επεξεργασία της φυσικής γλώσσας ενισχύει σημαντικά τις δυνατότητες των συστημάτων αναζήτησης και ανάκτησης πληροφοριών, βελτιώνοντας την ακρίβεια με την οποία ταιριάζουν τις «λέξεις κλειδιά»(keywords) σε έγγραφα και απαντούν σε συχνές ερωτήσεις. Συστήματα αναζήτησης που εφαρμόζουν τεχνικές NLP μπορούν να παρέχουν δραστικά βελτιωμένη πρόσβαση σε εξειδικευμένη έρευνα και γνώση.

1.4 Επισκόπηση νευρωνικών δικτύων/ recurrent neural networks (RNNs), and transformers

Το RNN είναι ένα παραδοσιακό νευρωνικό δίκτυο για επεξεργασία της φυσικής γλώσσας. Το RNN έχει τον περιορισμό ότι επεξεργάζεται τις λέξεις σε ακολουθίες, δηλαδή λαμβάνει υπόψη μόνο τις λέξεις που προηγούνται στο κείμενο και δεν λαμβάνει υπόψη τις επόμενες. Αυτό κάνει δύσκολο το έργο του RNN για την κατανόηση long-range dependencies, δηλαδή σχέσεις που απέχουν πολύ μεταξύ τους στην σειρά. Τα RNN (Recurrent Neural Networks) λειτουργούν με ένα μηχανισμό που τους επιτρέπει να αποθηκεύουν πληροφορίες από προηγούμενες εισόδους και να τις χρησιμοποιούν για την κατανόηση των τρέχοντων δεδομένων. Αυτό τους καθιστά ιδανικά για εργασίες που περιλαμβάνουν ακολουθίες δεδομένων, όπως η επεξεργασία φυσικής γλώσσας. Η προηγούμενη πληροφορία είναι κρίσιμη για την κατανόηση του περιεχομένου που ακολουθεί.[6]

Στον πυρήνα του, ένα RNN επεξεργάζεται τις ακολουθίες κάνοντας loop στα βήματα της ακολουθίας, διατηρώντας τις πληροφορίες σε κρυμμένα επίπεδα. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα ανατροφοδότησης, τα οποία επεξεργάζονται δεδομένα εισόδου προς μία κατεύθυνση χωρίς να κοιτάζουν πίσω, τα RNN περνούν πληροφορίες από το ένα βήμα στο επόμενο. Αυτός ο βρόχος μέσα τους επιτρέπει στα RNN να παρακολουθούν τις πληροφορίες.



Εικόνα 6: Δίκτυα νευρωνικών τροφοδοσίας και δίκτυα νευρωνικών κυκλωμάτων

The RNN Architecture:

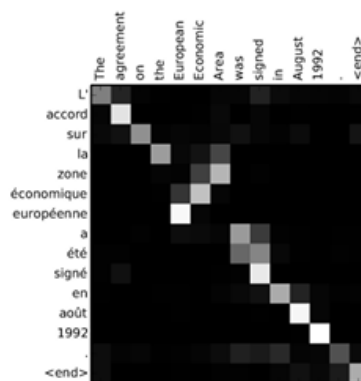
Input Layer (X_t): Εδώ είναι όπου το δίκτυο λαμβάνει μία λέξη της πρότασης κάθε φορά.

Hidden Layer (H_t): Είναι η μνήμη του RNN. Επεξεργάζεται την τρέχουσα είσοδο μαζί με τις πληροφορίες που λαμβάνονται από το προηγούμενο βήμα. Συνδυάζει τρέχοντα και προηγούμενα δεδομένα για τη λήψη αποφάσεων.

Output Layer (O_t): Με βάση τις πληροφορίες στο κρυφό επίπεδο, το RNN παράγει μια έξοδο για το τρέχον βήμα. Αυτό θα μπορούσε να είναι μια πρόβλεψη όπως η επόμενη λέξη σε μια πρόταση.

Attention Mechanism & Transformer

Ο μηχανισμός προσοχής (attention mechanism) σε RNN και άλλες αρχιτεκτονικές νευρωνικών δικτύων επιτρέπει στο μοντέλο να εστιάζει σε σχετικά μέρη των δεδομένων εισόδου. Ο μηχανισμός προσοχής βοηθά το μοντέλο να εστιάζει δυναμικά σε διαφορετικά μέρη της ακολουθίας εισόδου κατά την πρόβλεψη κάθε λέξης στην ακολουθία εξόδου. Αυτή η διαδικασία μοιάζει με τον τρόπο που οι άνθρωποι δίνουν προσοχή σε διαφορετικές λέξεις και τη σημασία τους ανάλογα με το πλαίσιο όταν κατανοούμε μια πρόταση.



Εικόνα 7: Αυτή η εικόνα παρουσιάζει έναν πίνακα ευθυγράμμισης από ένα μοντέλο αυτόματης μετάφρασης, συγκρίνοντας μια γαλλική πρόταση, "L'accord sur la zone économique européenne a été signé en août 1992" (The agreement on the European Economic Area was signed in August 1992.) με Αγγλική μετάφραση. Ο πίνακας αντιπροσωπεύει οπτικά πώς κάθε λέξη στη γαλλική πρόταση ευθυγραμμίζεται με λέξεις στην αγγλική μετάφραση, υποδεικνύοντας πώς το μοντέλο μετάφρασης επεξεργάζεται και συσχετίζει γλωσσικά στοιχεία μεταξύ των δύο γλωσσών.

Αυτός ο πίνακας απεικονίζει ποιες λέξεις στην πρόταση εισόδου (γαλλικά) λαμβάνονται υπόψη κατά την πρόβλεψη κάθε λέξης στην πρόταση εξόδου (Αγγλικά). Τα φωτεινότερα σημεία δείχνουν μια ισχυρότερη συσχέτιση. Με αυτόν τον τρόπο, το μοντέλο μαθαίνει να δίνει μεγαλύτερη προσοχή σε “ευθυγραμμίσεις” που είναι πιο σχετικές σε κάθε βήμα της μετάφρασης. Οι ευθυγραμμίσεις αναφέρονται στις σχέσεις ή τις αντιστοιχίες μεταξύ στοιχείων στις ακολουθίες εισόδου και εξόδου—συγκεκριμένα, ποια μέρη της εισόδου είναι σχετικά κατά την παραγωγή ενός συγκεκριμένου μέρους της εξόδου.

Τα προηγούμενα μοντέλα πριν από τους Transformers μπορούσαν να αναπαραστήσουν λέξεις ως διανύσματα, αλλά δεν κατείχαν το περιεχόμενο τους, έχοντας περιορισμένη κατανόηση του νοήματος των λέξεων. Η χρήση των λέξεων βασιζόταν κυρίως στην παρουσία τους, αγνοώντας συνήθως το πλήρες περιεχόμενο που φέρνουν.

Transformers είναι ένας νεότερος τύπος νευρωνικού δικτύου που ξεπερνά αυτόν τον περιορισμό. Χρησιμοποιούν ένα μηχανισμό που λέγεται self-attention για να επιτρέπουν σε κάθε λέξη να παρακολουθεί όλες τις άλλες λέξεις στην ακολουθία εισαγωγής ταυτόχρονα. Αυτό επιτρέπει στους Transformers να καταγράφουν παγκόσμιες σχέσεις και περιβάλλον, καθιστώντας τους πολύ καλύτερους στην κατανόηση εξαρτήσεων μεγάλης εμβέλειας.

Τώρα transformers χρησιμοποιούν τον μηχανισμό self-attention για να αναλύουν όλες τις λέξεις ενός κειμένου ταυτόχρονα. Αυτός ο μηχανισμός επιτρέπει στα LLM (Language Models) να αντιληφθούν και να εξετάσουν το πλήρες περιεχόμενο μιας πρότασης, προσφέροντας πολύ περισσότερες δυνατότητες για την ανάλυση της γλώσσας. Ο μηχανισμός αυτοπροσοχής εξετάζει κάθε λέξη σε ένα κείμενο και προσδιορίζει ποιες άλλες λέξεις είναι πιο σημαντικές για την κατανόηση του περιεχομένου.

1.5 Επεξήγηση για τα μοντέλα BERT

Σύμφωνα με τον οδηγό για το BERT από την Analytics Vidhya, το BERT (Bidirectional Encoder Representations from Transformers) είναι ένα νευρωνικό δίκτυο σχεδιασμένο για να κατανοεί τη σημασία των λέξεων στα κείμενα με βάση το πλαίσιο τους.[7] Είναι ένα two-layer bidirectional transformer που εκπαιδεύεται σε ένα τεράστιο σύνολο δεδομένων και κειμένου. Σύμφωνα με ανασκόπηση εφαρμογών του BERT, το μοντέλο χρησιμοποιείται ευρέως στην ανάλυση κειμένου και υπερέχει σε διάφορα καθήκοντα επεξεργασίας φυσικής γλώσσας, όπως η απάντηση σε ερωτήσεις.[8]

Το BERT χρησιμοποιεί την τεχνική της αυτό-προσοχής για να κατανοήσει το πλαίσιο των λέξεων σε ένα κείμενο αμφίδρομα. Αυτός ο μηχανισμός επιτρέπει στο μοντέλο να εστιάζει σε όλες τις λέξεις μιας πρότασης και να αξιολογεί τη σημασία κάθε λέξης για την κατανόηση μιας συγκεκριμένης λέξης στο κείμενο[7]. Κατά την επεξεργασία μιας λέξης, το μοντέλο λαμβάνει υπόψη τις συσχετίσεις της με όλες τις υπόλοιπες λέξεις της πρότασης. Αυτό οδηγεί

σε μια πιο ολοκληρωμένη κατανόηση του κειμένου και στην αναγνώριση των σημαντικών συναντήσεων μεταξύ των λέξεων. Έτσι, η αυτό-προσοχή βοηθά στην παραγωγή πιο ακριβών και συνεκτικών αποτελεσμάτων σε εργασίες όπως η κατανόηση κειμένου και η γεννητική γλώσσα.

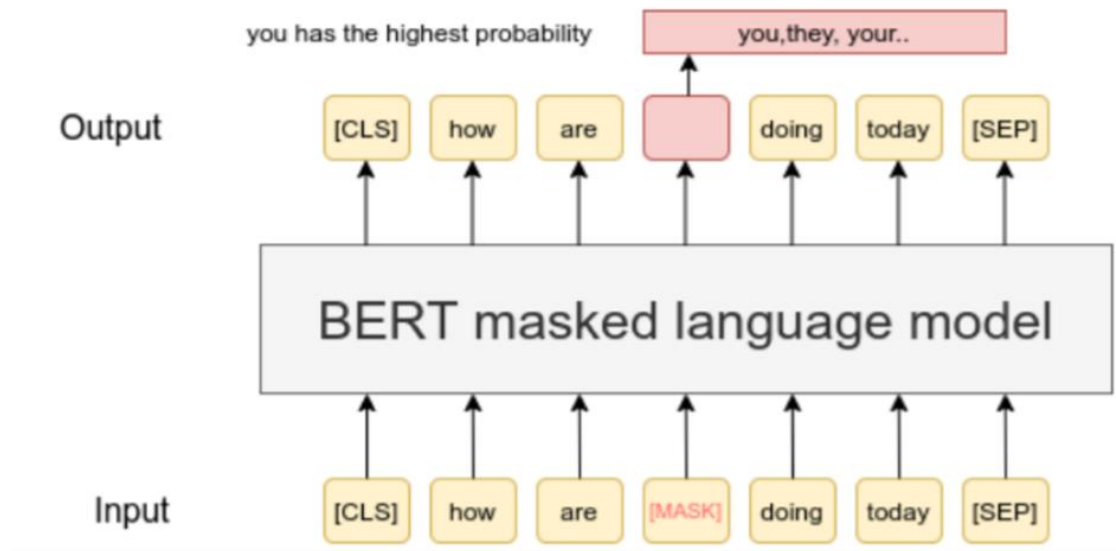
Αρχικά για κάθε λέξη στο κείμενο δημιουργείται ένα query vector (Query Vector (Q)) αντιπροσωπεύει τα χαρακτηριστικά της λέξης και την πιθανή σχέση της με άλλες λέξεις. Για κάθε λέξη δημιουργείται επίσης ένα key vector (Key Vector (K)) που αντιπροσωπεύει πώς αυτή η λέξη μπορεί να συσχετιστεί με άλλες λέξεις. Κάθε λέξη αποκτά επιπλέον έναν value vector (Value Vector (V)) που δηλώνει τη συνεισφορά της στην κατανόηση του κειμένου.

Στην συνέχεια γίνεται υπολογισμός του σκορ προσοχής (attention scores). Το μοντέλο υπολογίζει το σκορ προσοχής για κάθε ζεύγος λέξεων στην πρόταση. Το σκορ προσοχής δείχνει πόσο σημαντική είναι η άλλη λέξη (κλειδί) για την κατανόηση μιας συγκεκριμένης λέξης (ερώτημα). Κάθε λέξη στην πρόταση αναπαρίσταται αρχικά ως ένα διάνυσμα, γνωστό ως word embedding. Στη συνέχεια, τα word embeddings χρησιμοποιούνται για να δημιουργηθεί το query vector για κάθε λέξη. Αυτό το query vector λειτουργεί ως ερώτημα για τη σημασιολογική σχέση μεταξύ της συγκεκριμένης λέξης και των υπολοίπων λέξεων στην πρόταση. Έπειτα ο υπολογισμός γίνεται πολλαπλασιάζοντας το query vector της τρέχουσας λέξης με το key vector κάθε άλλης λέξης και στη συνέχεια διαιρώντας το αποτέλεσμα με τη ρίζα της διάστασης των key vectors για να κανονικοποιήσει το σκορ. Το αποτέλεσμα είναι μια τιμή που δείχνει τον βαθμό σχετικότητας μεταξύ των δύο λέξεων.

Μετά τον υπολογισμό των σκορ προσοχής, είναι απαραίτητο να κανονικοποιηθούν ώστε να μπορούν να συγκριθούν μεταξύ τους και να αντιπροσωπεύουν σωστά τη σημασιολογική σημαντικότητα των λέξεων στην πρόταση. Αυτό επιτυγχάνεται με τη χρήση της συνάρτησης softmax, μιας συνάρτησης που μετατρέπει τα σκορ προσοχής σε πιθανότητες. Η συνάρτηση softmax δίνει μεγαλύτερη βαρύτητα στα μεγάλα σκορ και μειώνει τα μικρά σκορ, επιτρέποντας έτσι τη δημιουργία ενός πιο αξιόπιστου συνολικού προσοχής στην πρόταση. Στη συνέχεια, πολλαπλασιάζουμε τις πιθανότητες με τα value vectors. Αυτός ο υπολογισμός καθορίζει πόσο από τη σημασιολογική πληροφορία που περιέχεται σε κάθε λέξη θα πρέπει να μεταδοθεί στο επόμενο επίπεδο της νευρωνικής αρχιτεκτονικής. Αυτό γίνεται με βάση τη σημασιολογική σημασία που έχει αποδοθεί σε κάθε λέξη από το σκορ προσοχής.

Τέλος οι τροποποιημένες τιμές value συνοψίζονται για να δημιουργήσουν ένα νέο σύνολο από vectors για κάθε λέξη. Αυτό το νέο σύνολο vectors περιλαμβάνει πληροφορίες από όλο το κείμενο, βοηθώντας το μοντέλο να καταλάβει καλύτερα τη σημασία και το πλαίσιο κάθε λέξης.

Masked Language Model (MLM)



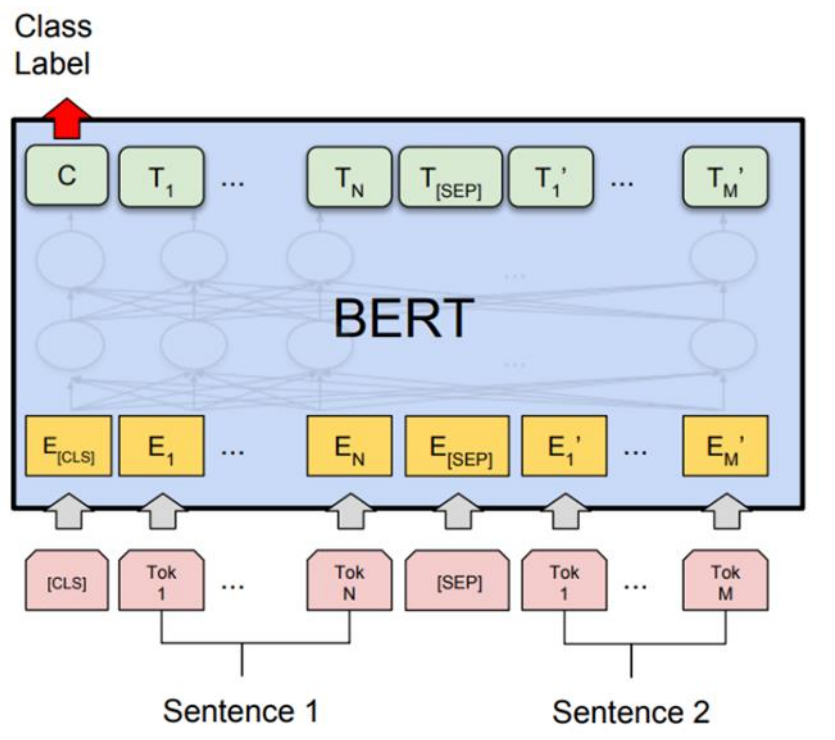
Εικόνα 8 Αυτή η εικόνα απεικονίζει τη λειτουργικότητα του BERT(ενός μοντέλου καλυμμένης γλώσσας που χρησιμοποιείται στην επεξεργασία φυσικής γλώσσας. Το διάγραμμα δείχνει πώς ο BERT προβλέπει μια καλυμμένη λέξη σε μια πρόταση. Ο BERT προβλέπει πιθανές λέξεις που θα μπορούσαν να ταιριάζουν στη θέση της μάσκας με βάση το πλαίσιο που παρέχεται από τις άλλες λέξεις της πρότασης. Το μοντέλο BERT επεξεργάζεται την είσοδο και την έξοδο μέσω ενός αμφίδρομου μηχανισμού που λαμβάνει υπόψη τόσο το αριστερό όσο και το δεξιό πλαίσιο σε όλα τα επίπεδα του μοντέλου, ενισχύοντας την κατανόηση και την ακρίβεια πρόβλεψής του.

Το κυρίαρχο χαρακτηριστικό του BERT είναι η ικανότητά του να κατανοεί το πλαίσιο των λέξεων σε ένα κείμενο "αμφίδρομο" (bidirectional), δηλαδή το BERT μπορεί να καταλάβει το νόημα των λέξεων με βάση το περιεχόμενο των λέξεων που βρίσκονται πριν και μετά από μία λέξη μέσα στην πρόταση. Το BERT είναι ένα γλωσσικό μοντέλο που χρησιμοποιεί μία τεχνική που λέγεται masked language modeling. Αυτή είναι μία τεχνική για να προβλέπεις μια καλυμμένη λέξη μέσα σε μία πρόταση. Από την είσοδο κειμένου, το BERT επιλέγει τυχαία ένα δείγμα των tokens που θα υποστούν αντικατάσταση. Τυπικά, το 15% των tokens επιλέγεται για αυτόν τον σκοπό. Το 80% αντικαθίσταται με το ειδικό token [MASK]. Το 10% δεν αλλάζει, παραμένει ως έχει. Το υπόλοιπο 10% αντικαθίσταται με ένα τυχαίο token από το λεξιλόγιο.

Ο στόχος της MLM είναι η μείωση της απώλειας cross-entropy για την πρόβλεψη των αρχικών tokens που έχουν αντικατασταθεί από το [MASK]. Δηλαδή, το μοντέλο εκπαιδεύεται για να προβλέψει το αρχικό περιεχόμενο των masked positions, βασιζόμενο στο πλαίσιο που παρέχουν οι υπόλοιπες λέξεις της πρότασης.

Αυτή η μέθοδος εκπαίδευσης επιτρέπει στο BERT να αναπτύξει μια βαθιά κατανόηση του πλαισίου σε πλήρεις φράσεις και παραγράφους, καθώς και τη δυνατότητα να κατανοεί τη γλώσσα με πιο περίπλοκο τρόπο.

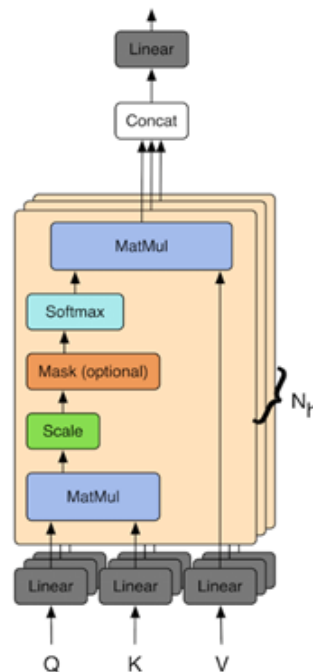
Next Sentence Prediction (NSP)



Εικόνα 9: Αυτή η εικόνα αντιπροσωπεύει τον τρόπο με τον οποίο χρησιμοποιείται το BERT για μια εργασία ταξινόμησης που περιλαμβάνει δύο προτάσεις. Το διάγραμμα δείχνει δύο προτάσεις ως παραδείγματα εισαγωγής: Πρόταση 1 και Πρόταση 2. Αυτές οι προτάσεις κωδικοποιούνται από το BERT για να μετατρέψουν το ακατέργαστο κείμενο σε διανύσματα (E) που καταγράφουν γλωσσικές πληροφορίες.

Το BERT είναι εκπαιδευμένο πάνω σε binary classification task (NSP). Δέχεται δύο προτάσεις ως είσοδο και προβλέπει αν ακολουθεί η μία την άλλη μέσα στο κείμενο. Για να μπορέσει να το καταλάβει αυτό γίνεται επιλογή θετικών και αρνητικών παραδειγμάτων. Τα θετικά παραδείγματα είναι προτάσεις διαδοχικές από το ίδιο κείμενο ενώ τα αρνητικά είναι προτάσεις μη διαδοχικές από διαφορετικά κείμενα ή από διαφορετικά σημεία του ίδιου κειμένου. Κατά την εκπαίδευση, το BERT παρουσιάζεται με τα ζεύγη προτάσεων (θετικά και αρνητικά παραδείγματα) και εκπαιδεύεται να προβλέπει εάν η δεύτερη πρόταση στο ζεύγος ακολουθεί φυσικά την πρώτη ή όχι. Το μοντέλο χρησιμοποιεί έναν δυαδικό ταξινομητή για να καταλήξει σε μια πρόβλεψη “ΝΑΙ” αν οι προτάσεις ακολουθούν η μία την άλλη ή “Όχι” αν οι προτάσεις δεν ακολουθούν η μία την άλλη. Η διαδικασία NSP βοηθά το BERT να αναπτύξει μια κατανόηση της συνέχειας και της δομής των κειμένων, καθώς και την ικανότητα να καταλαβαίνει τη λογική συσχέτιση μεταξύ διαφορετικών προτάσεων.

Multi-Head Attention



Εικόνα 10 Σε αυτή την εικόνα φαίνεται ο μηχανισμός Multi-Head Attention μέσα σε μια αρχιτεκτονική μοντέλου μετασχηματιστή για την επεξεργασία ακολουθιών σε εργασίες. Η διαδικασία περιλαμβάνει διανύσματα εισόδου (Ερώτημα, Κλειδί, Τιμή), γραμμικούς μετασχηματισμούς, κλιμάκωση, προαιρετική κάλυψη, softmax για βαθμολογίες προσοχής και πολλαπλασιασμό πίνακα, ακολουθούμενο από συνένωση και έναν άλλο γραμμικό μετασχηματισμό. Αυτός ο μηχανισμός επιτρέπει στο μοντέλο να εστιάζει σε διαφορετικά μέρη των δεδομένων εισόδου ταυτόχρονα, βελτιώνοντας την ικανότητά του να κατανοεί και να δημιουργεί κείμενο.

Το BERT χρησιμοποιεί μία τεχνική που ονομάζεται multi-head attention για να αποτυπώνει πολλές πτυχές μίας λέξης. Αυτό γίνεται με την μετατροπή της εισόδου σε πολλαπλές κεφαλές που η καθεμία είναι υπεύθυνη για να αντιπροσωπεύει μία διαφορετική πτυχή της λέξης. Κάθε μία από τις κεφαλές λειτουργεί ως ένας διακριτός μηχανισμός προσοχής και είναι υπεύθυνος για την αντιπροσώπευση ενός συγκεκριμένου χαρακτηριστικού της λέξης. Μία κεφαλή της λέξης μπορεί να αναφέρεται στον ορισμό της λέξης και μία άλλη στον γραμματικό ρόλο της λέξης. Κάθε κεφαλή υπολογίζει ένα σταθμισμένο άθροισμα που συσχετίζεται με τις λέξεις στην είσοδο. Τα βάρη υπολογίζονται χρησιμοποιώντας μία συνάρτηση η οποία μετρά την ομοιότητα μεταξύ του ερωτήματος (η λέξη που εξετάζεται) και κάθε κλειδιού (των άλλων λέξεων της ακολουθίας). Τα σταθμισμένα ποσά κάθε κεφαλής ενώνονται και τροφοδοτούνται σε ένα δίκτυο τροφοδοσίας. Η έξοδος του δικτύου τροφοδοσίας χρησιμοποιείται στη συνέχεια ως έξοδος του multi-head attention layer.

Positional Encoding

Sequence	Index of token, k	Positional Encoding Matrix with $d=4$, $n=100$			
		$i=0$	$i=0$	$i=1$	$i=1$
I	0	$P_{00}=\sin(0)$ = 0	$P_{01}=\cos(0)$ = 1	$P_{02}=\sin(0)$ = 0	$P_{03}=\cos(0)$ = 1
am	1	$P_{10}=\sin(1/1)$ = 0.84	$P_{11}=\cos(1/1)$ = 0.54	$P_{12}=\sin(1/10)$ = 0.10	$P_{13}=\cos(1/10)$ = 1.0
a	2	$P_{20}=\sin(2/1)$ = 0.91	$P_{21}=\cos(2/1)$ = -0.42	$P_{22}=\sin(2/10)$ = 0.20	$P_{23}=\cos(2/10)$ = 0.98
Robot	3	$P_{30}=\sin(3/1)$ = 0.14	$P_{31}=\cos(3/1)$ = -0.99	$P_{32}=\sin(3/10)$ = 0.30	$P_{33}=\cos(3/10)$ = 0.96

Positional Encoding Matrix for the sequence 'I am a robot'

Εικόνα 11 Στην εικόνα τον πίνακα του Positional Encoding για την πρόταση "I am a robot". Ο πίνακας δείχνει πώς σε κάθε λέξη ("I", "am", "a", "robot") στην ακολουθία εκχωρείται μια μοναδική κωδικοποίηση θέσης με βάση τον δείκτη του στην ακολουθία ($k=0, 1, 2, 3$). Για διαστάσεις πίνακα $d=4$ και $n=100$, οι τιμές υπολογίζονται χρησιμοποιώντας συναρτήσεις ημιτονοειδούς και συνημιτόνου. Η κωδικοποίηση βοηθά το μοντέλο να διατηρεί πληροφορίες σχετικά με τη σειρά των λέξεων στην πρόταση.

Positional Encoding είναι μία τεχνική που χρησιμοποιείται στα μοντέλα νευρωνικών δικτύων για να κωδικοποιηθεί η σχετική θέση των λέξεων σε μία ακολουθία. Αυτό είναι πολύ σημαντικό γιατί το νευρωνικό δίκτυο μπορεί να μάθει για την σειρά των λέξεων μέσα σε μία πρόταση μόνο από την χρήση των διανυσμάτων. Χωρίς αυτή την τεχνική τα νευρωνικά δίκτυα δεν θα μπορούσαν να ξεχωρίσουν προτάσεις με τις ίδιες λέξεις αλλά σε διαφορετική σειρά. Όπως αυτό το παράδειγμα "The cat chased the dog" and "The dog chased the cat", θα τις θεωρούσε ίδιες χωρίς αυτή την τεχνική.

Χρησιμοποιούνται 2 τεχνικές για το positional encoding:

Sinusoidal positional encoding : Χρησιμοποιείται ένας συνδυασμός ημιτονικών και συνημιτονικών συναρτήσεων για την κωδικοποίηση της κάθε λέξης μέσα στην πρόταση. Αυτές οι συναρτήσεις έχουν διαφορετικές συχνότητες και φάσεις για διαφορετικές θέσεις κάτι που επιτρέπει στο δίκτυο να μάθει για την σχετική θέση των λέξεων μέσα στην πρόταση.

Learned positional encoding : Χρησιμοποιεί ένα διαφορετικό είδος νευρωνικών δικτύων για να μάθει την τεχνική. Αυτή η τεχνική είναι πιο ευέλικτη από την προηγούμενη αλλά είναι υπολογιστικά πολύ πιο ακριβή.

Κεφάλαιο 2

Περιγραφή Προβλήματος

2.1 Work Motivation	24
2.2 Ανάλυση του Προβλήματος	25
2.3 Προσέγγιση Λύσης	29

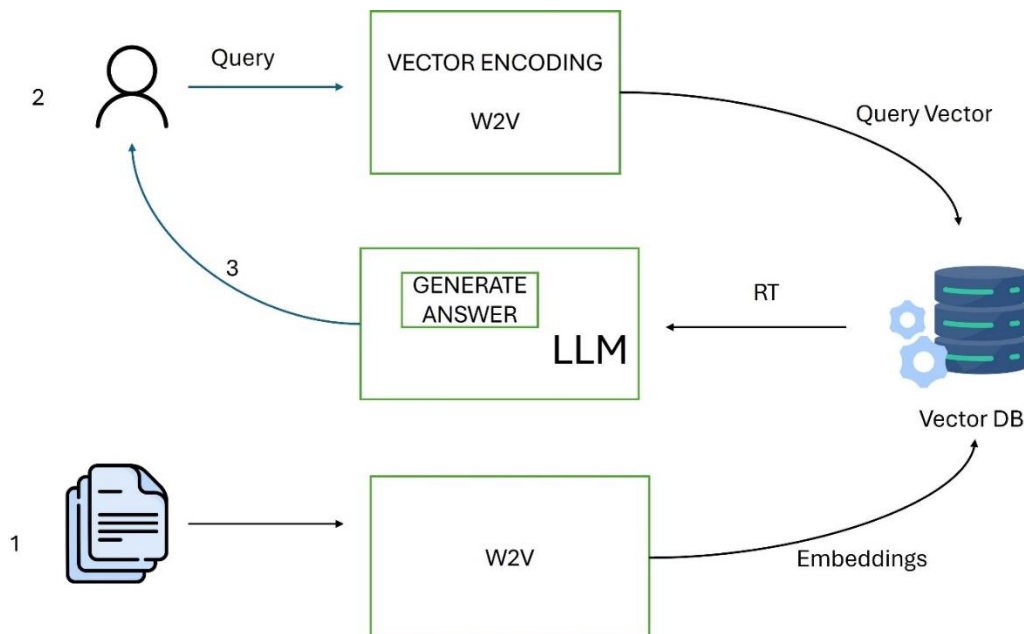
2.1 Work Motivation

Μετά από τέσσερα χρόνια ως φοιτητής του Πανεπιστημίου Κύπρου, έχω αντιληφθεί την σημασία της συμβουλευτικής στην πανεπιστημιακή ζωή του κάθε φοιτητή. Μέσω της συμβουλευτικής οι φοιτητές μπορούν να βοηθηθούν ώστε να λύσουν διάφορα προβλήματα και να επιτύχουν τους στόχους τους. Εντός αυτού του πλαισίου, θεωρώ ότι η δημιουργία ενός αυτοματοποιημένου συστήματος συμβουλευτικής για την καθοδήγηση των φοιτητών θα βοηθήσει το πανεπιστήμιο και τους φοιτητές. Τα πρώτα μου χρόνια ως φοιτητής στο πανεπιστήμιο Κύπρου είχα διάφορες απορίες όσο αφορά τα μαθήματα που πρέπει να επιλέξω, την διαφορά μεταξύ του υποχρεωτικού μαθήματος και του μαθήματος ελεύθερης επιλογής κ.α. Για να μπορέσω να βρω τότε την λύση στις απορίες μου έψαχνα στον οδηγό σπουδών και σε άλλες σελίδες του πανεπιστημίου κάτι που ήταν χρονοβόρο. Γι' αυτό έχω σκεφτεί την δημιουργία ενός τέτοιου συστήματος (chatbot) που θα βασίζεται στην χρήση της φυσικής γλώσσας και της τεχνητής νοημοσύνης για να παρέχει ανατροφοδότηση στους φοιτητές σχετικά με τα μαθήματά τους και τους στόχους του κάθε κλάδου. Η συμβουλευτική μπορεί να βοηθήσει τους φοιτητές να επιλέξουν τα κατάλληλα μαθήματα και να αναπτύξουν ένα αποτελεσματικό πρόγραμμα σπουδών.

2.2 Ανάλυση του Προβλήματος

Abstract

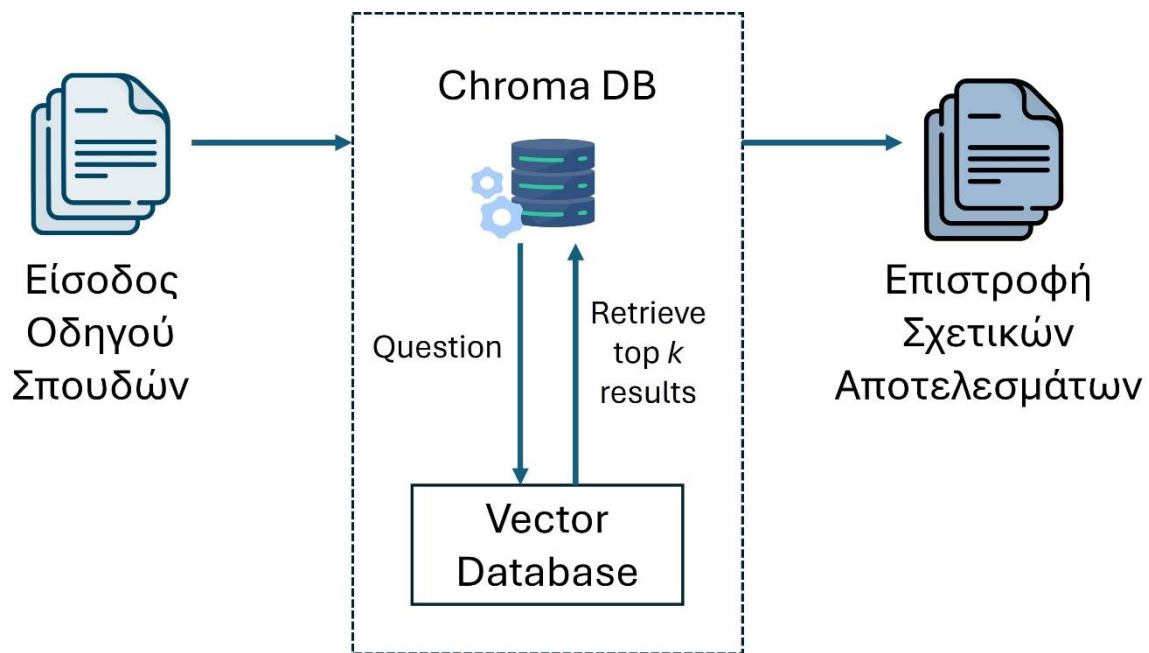
Το σύστημα ανάλυσης προβλημάτων αποτελείται από την εξής διαδικασία. Αρχικά, διαχωρίζονται τα κεφάλαια του οδηγού σπουδών σε ξεχωριστά αρχεία κειμένου (.txt). Έπειτα, δημιουργείται μια διανυσματική βάση για την αποθήκευση των αναπαραστάσεων των κειμένων του οδηγού σπουδών, επιτρέποντας έτσι την εύκολη αναζήτηση για απαντήσεις σε ερωτήσεις του χρήστη. Στην συνέχεια εισάγεται η ερώτηση του χρήστη μέσα στο σύστημα. Το σύστημα είναι υπεύθυνο για την μετατροπή της πρότασης σε διανύσματα για να μπορεί να την κατανοήσει ο υπολογιστής. Τα διανύσματα αποτελούν μια μαθηματική αναπαράσταση των λέξεων, επιτρέποντας έτσι στον υπολογιστή να τις κατανοήσει και να επεξεργαστεί το κείμενο. Για τη μετατροπή των λέξεων σε διανύσματα χρησιμοποιείται ο αλγόριθμος word2vec, ο οποίος αναπαριστά τις λέξεις, λαμβάνοντας υπόψη το περιβάλλον τους στο κείμενο. Στη συνέχεια, το σύστημα επικοινωνεί με τη διανυσματική βάση για να επιστρέψει τα παρόμοια αποτελέσματα με βάση την ερώτηση του χρήστη. Η διανυσματική βάση επιστρέφει τα κείμενα από τα αρχεία .txt που έχουν τη μεγαλύτερη ομοιότητα και χρησιμοποιούνται ως γνώση. Τέλος, με τη χρήση του GPT 3.5 API, το σύστημα αξιοποιεί τις γνώσεις από τη διανυσματική βάση για να επιστρέψει την απάντηση σε ανθρώπινη μορφή, έτσι ώστε να μπορεί να την κατανοήσει ο χρήστης.



Εικόνα 12 Αφαιρετική Εικόνα Λειτουργίας Προγράμματος

Βήμα 1.

Δημιουργία Διανυσματικής Βάσης

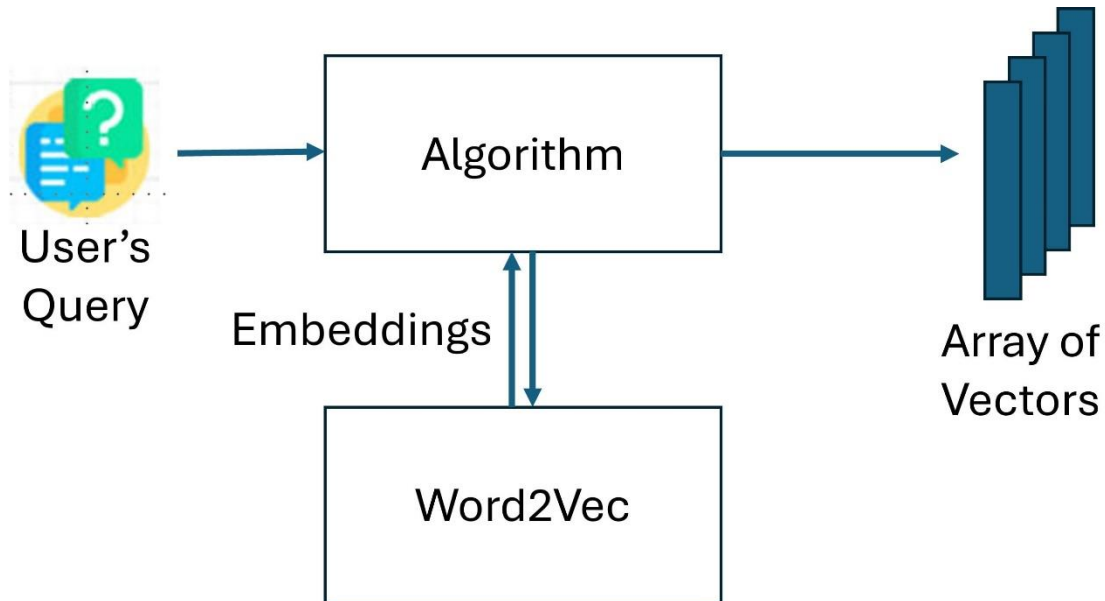


Εικόνα 13 Πρώτο Υποπρόβλημα. Δημιουργία Διανυσματικής Βάσης.

Αρχικά, γίνεται η εισαγωγή αρχείων σε μορφή .txt, όπου κάθε αρχείο αντιστοιχεί στο κείμενο του οδηγού σπουδών ανάλογα με τον κλάδο που ανήκει. Τα δεδομένα από τα αρχεία διαχωρίζονται και αποθηκεύονται σε ξεχωριστά αρχεία για ευκολότερη ανάκτηση και ευέλικτη διαχείριση τους στη διανυσματική βάση, καθώς και για τη διευκόλυνση της αναζήτησης κατά την εκτέλεση ερωτήσεων από τον χρήστη. Η εισαγωγή σε μία διανυσματική βάση επιτρέπει την αποθήκευση των δεδομένων με τρόπο που λαμβάνει υπόψη τη σημασιολογική τους δομή. Αυτό σημαίνει ότι οι λέξεις ή οι φράσεις που έχουν παρόμοια σημασιολογική έννοια θα έχουν κοντινά διανύσματα αναπαράστασης. Συνεπώς, κατά την ανάκτηση δεδομένων για εφαρμογές όπως συστήματα ερωτήσεων-απαντήσεων, μπορούν να εντοπιστούν γρήγορα και αποτελεσματικά τα σχετικά δεδομένα. Αυτό διευκολύνει την αντιστοίχιση των ερωτήσεων του χρήστη με τις κατάλληλες απαντήσεις. Μετά την εισαγωγή των αρχείων, κάθε αρχείο κειμένου διαιρείται σε παράθυρα, δηλαδή σε συγκεκριμένο αριθμό γραμμών. Σε κάθε παράθυρο, υπολογίζονται τα διανύσματα αναπαράστασης και συλλέγονται τα μεταδεδομένα, που αναφέρονται στο αρχείο, η ταυτότητά του και ο αριθμός γραμμής. Για τον υπολογισμό των διανυσμάτων χρησιμοποιείται ο αλγόριθμος word2Vec. Κάθε παράθυρο περιλαμβάνει έναν ορισμένο αριθμό γραμμών κειμένου και υπολογίζει το μέσο διάνυσμα για τις λέξεις που περιέχονται σε αυτές τις γραμμές. Αυτή η διαδικασία εφαρμόζεται για ένα καθορισμένου μεγέθους παράθυρο σε όλα τα αρχεία κειμένου. Στη συνέχεια, τα διανύσματα εισάγονται στη διανυσματική βάση Chroma db για τη δημιουργία της. Μετά την δημιουργία της διανυσματικής βάσης μπορούν να γίνουν ερωτήσεις. Σύμφωνα με την ερώτηση που δίνεται επιστρέφει παρόμοια αποτελέσματα που εντοπίζει από τα αρχεία του οδηγού σπουδών.

Βήμα 2

Μετατροπή λέξεων σε διανύσματα (Embeddings)

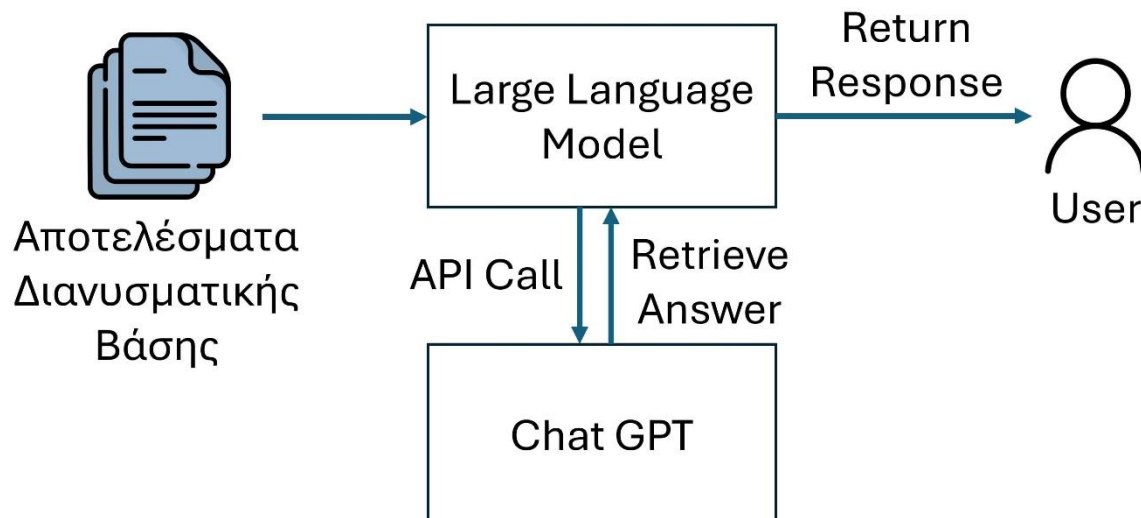


Εικόνα 14 Δεύτερο Υποπρόβλημα. Μετατροπή λέξεων σε διανύσματα

Αρχικά, η είσοδος που λαμβάνεται από το χρήστη επεξεργάζεται από το σύστημα, μετατρέποντας τις λέξεις σε διανύσματα. Το “User’s Query” αντιπροσωπεύει ένα περιβάλλον χρήστη στο οποίο οι φοιτητές θα εισάγουν τις ερωτήσεις τους. Ο αλγόριθμος θα επεξεργαστεί την είσοδο που λαμβάνεται από το User Query. Το Word2Vec είναι ένα υπολογιστικό μοντέλο σχεδιασμένο για τη μετατροπή των λέξεων σε αριθμητικά διανύσματα. Είναι μια μορφή ενσωμάτωσης λέξεων που επιτρέπει σε λέξεις με παρόμοιες σημασίες να έχουν παρόμοιες αναπαραστάσεις. Το Word2Vec χρησιμοποιείται για να μάθει συσχετίσεις μεταξύ λέξεων από ένα δεδομένο κείμενο. Μόλις εκπαιδευτεί, ένα τέτοιο μοντέλο μπορεί να αναγνωρίσει συνώνυμες λέξεις. Έτσι, οι λέξεις με παρόμοιες σημασίες έχουν παρόμοιες αναπαραστάσεις στον διανυσματικό χώρο. Τέλος, η χρήση του Word2Vec επιτρέπει να αντιληφθούμε τις σημασιολογικές σχέσεις μεταξύ των λέξεων στο κείμενο, βοηθώντας έτσι στην καλύτερη κατανόηση του περιεχομένου από το σύστημα.

Βήμα 3

Μετατροπή του Κειμένου σε Ανθρώπινη Μορφή με την Χρήση της API του LLM



Εικόνα 15 Μετατροπή κειμένου σε ανθρώπινη μορφή

Όταν θα δημιουργηθεί η διανυσματική βάση και ο χρήστης θα κάνει μια σχετική ερώτηση το σύστημα θα επιστρέψει τα 4 πιο όμοια αποτελέσματα με τη μορφή ενός πίνακα. Αυτά τα αποτελέσματα που προέρχονται από τη διανυσματική βάση χρησιμοποιούνται ως “context”. Αυτό σημαίνει ότι τα πιο σχετικά κείμενα με την ερώτηση του χρήστη θα χρησιμοποιηθούν ως πληροφορίες για να καθοδηγήσουν την απάντηση του συστήματος. Για την περαιτέρω επεξεργασία αυτών των αποτελεσμάτων, χρησιμοποιείται το LLM της OpenAI, το gpt-3.5-turbo. Το LLM λαμβάνει ως είσοδο τα κείμενα που παρέχονται από τη διανυσματική βάση και επεξεργάζεται τα κείμενα προκειμένου να παράγει μια πιο κατανοητή ανθρώπινη απάντηση.

2.3 Προσέγγιση Λύσης

Διαχωρισμός του οδηγού σπουδών σε αρχεία

Αρχικά έγινε διαχωρισμός των pdf αρχείων που περιείχαν τον οδηγό σπουδών, σε πολλαπλά αρχεία pdf, όπου το κάθε ένα περιείχε ένα τμήμα από τον οδηγό σπουδών. Στην συνέχεια, έγινε η μετατροπή των pdf αρχείων σε .txt για να μπορούν να τα χρησιμοποιηθούν ως είσοδο για την δημιουργία της διανυσματικής βάσης.

Δημιουργία Διανυσματικής Βάσης

Αρχικά δημιουργείται η διανυσματική βάση. Υπήρχαν χωρισμένα τα τμήματα του οδηγού σπουδών σε αρχεία .txt από το προηγούμενο βήμα. Στην συνέχεια τα αρχεία χωρίζονται σε προτάσεις και μετατρέπονται οι προτάσεις σε διανύσματα. Χρησιμοποιήθηκε η διανυσματική βάση Chroma db για την αποθήκευση αυτών των διανυσμάτων. Επιστρέφονται τα παρόμοια

αποτελέσματα με βάση τις σχέσεις των διανυσμάτων της ερώτησης και των κειμένων txt από τον οδηγό σπουδών.

Χρήση Chat gpt API key

Η διανυσματική βάση μετά από μια ερώτηση που θα κάνει ο χρήστης θα επιστρέψει παρόμοια αποτελέσματα, δηλαδή τις πιο όμοιες προτάσεις από τον οδηγό σπουδών (τα κείμενα txt). Θα χρησιμοποιηθούν αυτές οι προτάσεις για την τροφοδοσία του LLM Chat GPT για να μπορεί να απαντήσει με την χρήση της φυσικής γλώσσας στις ερωτήσεις του χρήστη.

Κεφάλαιο 3

Βασικοί Ορισμοί και σύγκριση Μοντέλων

3.1 Embeddings και η σημαντικότητα τους	30
3.2 Word2Vec vs Transformers Model	33
3.3 Διανυσματικές Βάσεις	37

3.1 Embeddings και η σημαντικότητα τους + Contextual Word Embeddings

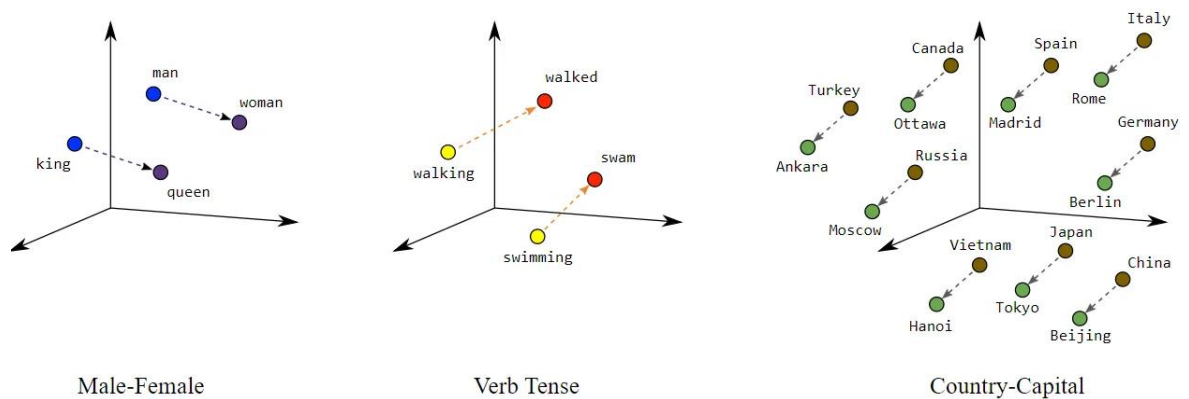
Τα word Embeddings είναι ουσιώδη στην επεξεργασία φυσικής γλώσσας (NLP), μετατρέποντάς τις λέξεις σε αριθμητικά διανύσματα που μπορούν να τύχουν επεξεργασίας από αλγόριθμους μηχανικής μάθησης.

Για να γραφτεί κείμενο, τα LLM πρέπει να μεταφράσουν τις λέξεις σε μία γλώσσα που να μπορούν να την κατανοήσουν. Αρχικά, οι λέξεις χωρίζονται σε tokens, που είναι διακριτές μονάδες επεξεργασίας. Μετά τη μετατροπή σε tokens, κάθε token μετατρέπεται σε ένα διάνυσμα, το οποίο είναι ένα σετ από αριθμούς που αντιπροσωπεύουν διάφορες πτυχές της σημασίας του token.

Η κατανόηση της σημασίας μιας λέξης από τα LLM γίνεται μέσω της ανάλυσης του πλαισίου της. Το πλαίσιο είναι ουσιαστικά το περιβάλλον στο οποίο εμφανίζεται κάθε λέξη ή έκφραση. Αυτή η ανάλυση βασίζεται στην Υπόθεση της Διανομής, που υποστηρίζει ότι οι λέξεις με παρόμοιες σημασίες τείνουν να εμφανίζονται σε παρόμοια πλαίσια. Στην πράξη, τα LLM εξετάζουν τις λέξεις που περιβάλλουν κάθε token για να καταλάβουν την πλήρη σημασία του. Τα διανύσματα αυτά, με τις πλούσιες πληροφορίες που περιέχουν, επιτρέπουν στο

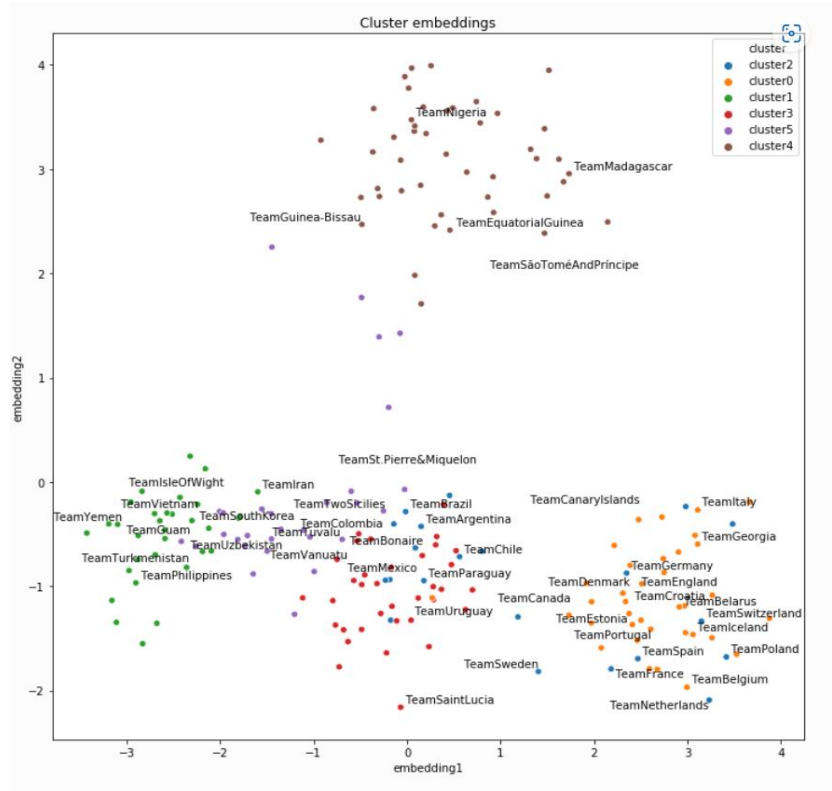
μοντέλο να διακρίνει τις διαφορετικές σημασίες μιας λέξης ανάλογα με το πλαίσιο στο οποίο χρησιμοποιείται.

Τα παραδοσιακά μοντέλα γλώσσας θεώρησαν τις λέξεις ως ξεχωριστές, απομονωμένες οντότητες, αποτυγχάνοντας να αποτυπώσουν τις σχέσεις μεταξύ των λέξεων. Τα words embeddings αντιμετωπίζουν αυτό τον περιορισμό τοποθετώντας λέξεις με παρόμοιες σημασιολογική πλαίσια κοντά στο ένα στον πολυδιάστατο χώρο, κωδικοποιώντας, έτσι, όχι μόνο την ατομική σημασία αλλά και τις σχέσεις μεταξύ των λέξεων.



Εικόνα 16 Word Embeddings Visualized: Relationships Between Gender, Verb Tense, and Country-Capital Pairs

Με την μείωση των χιλιάδων τιμών που υπάρχουν στα διανύσματα σε δύο, μπορεί να φανεί η απόσταση-σχέση μεταξύ των λέξεων πιο καθαρά όπως δείχνει η Εικόνα 16.

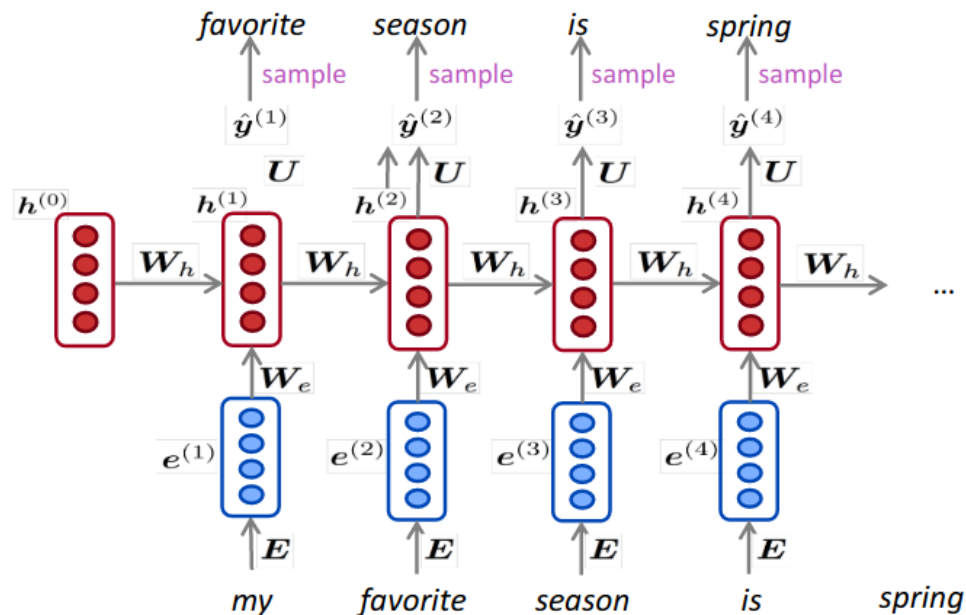


Εικόνα 17 Ομαδοποίηση διανυσματικών ομάδων: Οπτικοποίηση ομαδοποιήσεων σε δισδιάστατο χώρο.

Μπορούν να εντοπιστούν λέξεις και να ποσοτικοποιηθούν σε ομάδες λέξεων (clusters). Η εικόνα 17 απεικονίζει ένα διάγραμμα συσσωματωμένων διανυσμάτων για διάφορες ομάδες, οπτικοποιημένες σε δισδιάστατο χώρο. Κάθε σημείο αντιπροσωπεύει μια ομάδα, ενώ τα χρώματα υποδεικνύουν τις διαφορετικές συστάδες στις οποίες ανήκουν αυτές οι ομάδες

Contextual Word Embeddings

δεδομένου μίας τρέχουσας λέξης μέσα σε μία πρόταση. Χρησιμοποιούμε αυτά τα δίκτυα για την ικανότητά τους να καταγράφουν και να διατηρούν μακροπρόθεσμες εξαρτήσεις στις κρυφές τους καταστάσεις.



Εικόνα 18 Η κρυφή κατάσταση (Κόκκινο) μπορεί να διατηρήσει πληροφορίες για την προηγούμενη λέξη στην πρόταση.

Μετά την τροφοδοσία της τρέχουσας λέξης μπορεί να γίνει σύνδεση της κρυφής κατάστασης με τη συνήθη αναπαράσταση του Word2Vec για να διατηρηθούν οι πληροφορίες τόσο για την τρέχουσα λέξη, όσο, και για το παρελθόν.

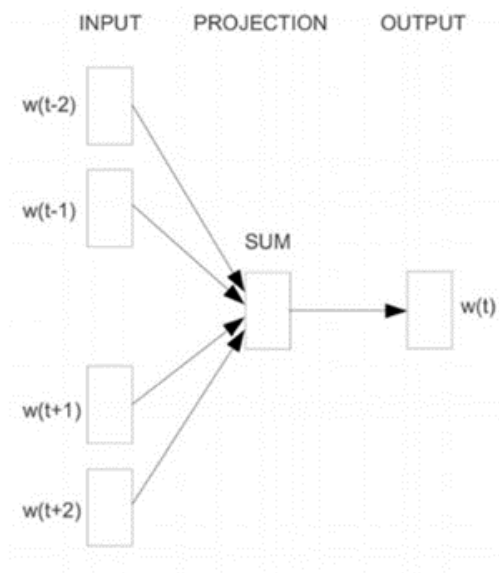
3.2 Word2Vec vs Transformers Model

Word2Vec

Αναπτύχθηκε από την Google το 2013, το Word2Vec ώστε να χρησιμοποιεί ένα feedforward νευρωνικό δίκτυο για τη δημιουργία αναπαραστάσεων λέξεων. Χρησιμοποιεί δύο μοντέλα για την εκπαίδευση αυτών των αναπαραστάσεων: το Skip-gram και το CBOW (Continuous Bag of Words). Στο Skip-gram, το μοντέλο προσπαθεί να μαντέψει τι λέξεις βρίσκονται γύρω από μια συγκεκριμένη λέξη. Στο CBOW, το μοντέλο προσπαθεί να μαντέψει μια συγκεκριμένη λέξη με βάση τις λέξεις που βρίσκονται γύρω της. Ωστόσο, απαιτεί μεγάλο όγκο δεδομένων εκπαίδευσης και έχει υψηλή υπολογιστική πολυπλοκότητα, καθιστώντας το αργότερο στην εκπαίδευση, αλλά κατάλληλο για μεγαλύτερες εφαρμογές. Η επεξεργασία μεγάλων ποσοτήτων δεδομένων και η εκμάθηση από αυτά αυξάνει την υπολογιστική πολυπλοκότητα. Ωστόσο, αυτή η πολυπλοκότητα συμβάλλει στην ακρίβεια των προβλέψεων του μοντέλου σε μεγαλύτερες εφαρμογές.

Το Word2Vec έχει γίνει βασικό συστατικό πολλών υψηλότερων αλγορίθμων στον κόσμο της φυσικής γλώσσας (NLP). Είναι ένα από τα εργαλεία που βοήθησε σε σημαντικές προόδους στις εφαρμογές μηχανικής μάθησης βασισμένες στη γλώσσα, συμπεριλαμβάνοντας τη μηχανική μετάφραση, την αναγνώριση φωνής και τα τεχνητά νευρωνικά δίκτυα για συζητήσεις Τεχνητής Νοημοσύνης.

CBOW



Εικόνα 19 CBOW

Η είσοδος (input): : Στην είσοδο,

φαίνονται οι λέξεις στο περιβάλλον μιας λέξης-στόχος. Η " $w(t-2)$ " σημαίνει τη λέξη δύο θέσεις πριν από τη στόχο λέξη, " $w(t-1)$ " είναι η λέξη μία θέση πριν, " $w(t+1)$ " είναι η λέξη μία θέση μετά, και " $w(t+2)$ " είναι η λέξη δύο θέσεις μετά τη στόχο λέξη.

Η προβολή (Projection): Οι λέξεις

εισόδου μετατρέπονται σε διανύσματα. Αυτά τα διανύσματα "προβάλλονται" σε ένα κοινόχρηστο, κρυφό στρώμα (hidden layer), το οποίο δεν απεικονίζεται άμεσα εδώ, αλλά

υπονοείται από τις γραμμές που συνδέουν την είσοδο με την ενδιάμεση διαδικασία της άθροισης. Τα διανύσματα από το κοινόχρηστο στρώμα αθροίζονται μαζί για να σχηματίσουν ένα μοναδικό διάνυσμα. Αυτή η άθροιση αντιπροσωπεύει την ολική γλωσσική πληροφορία από το περιβάλλον της λέξης στόχος. Η διαδικασία της άθροισης βοηθάει το δίκτυο να κατανοήσει τη σημασία της λέξης βάσει του πλαισίου της.

Η έξοδος (Output): Το συνολικό διάνυσμα χρησιμοποιείται για να προβλέψει την πραγματική λέξη στόχο " $w(t)$ ". Στην πραγματικότητα, αυτό γίνεται μέσω μιας διαδικασίας αντίστροφης από την προβολή, όπου το διάνυσμα μετατρέπεται πίσω σε μια λέξη. Συνήθως, αυτό γίνεται με τη χρήση της συνάρτησης softmax, η οποία καθορίζει την πιθανότητα κάθε λέξης στο λεξιλόγιο να είναι η λέξη στόχος.

Skip-gram

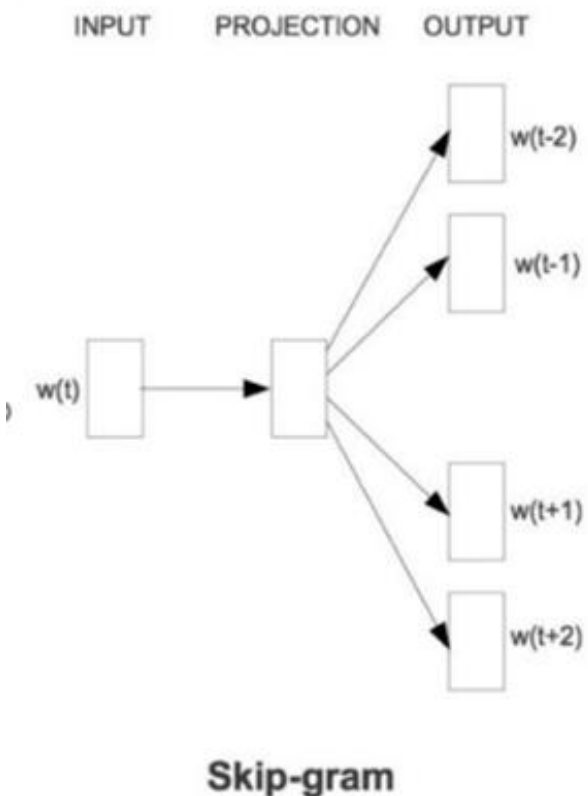
Ο αλγόριθμος Skip-gram εκπαιδεύεται για να προβλέψει τις λέξεις που περιβάλλουν μια συγκεκριμένη λέξη.

Η είσοδος(input) : Η είσοδος του μοντέλου είναι μία στόχος λέξη, σημειωμένη ως $w(t)$. Αυτή η λέξη είναι η λέξη για την οποία θα γίνει πρόβλεψη για το πλαίσιο.

Η προβολή (projection): Η διαδικασία μετατροπής της εισόδου σε ένα σύνολο από αριθμούς. Η προβολή γίνεται χρησιμοποιώντας ένα νευρωνικό δίκτυο. Το νευρωνικό δίκτυο λαμβάνει ως είσοδο τη συγκεκριμένη λέξη και παράγει ως έξοδο ένα σύνολο αριθμών. Τα νούμερα αυτά αντιπροσωπεύουν τη σημασία της συγκεκριμένης λέξης σε σχέση με άλλες λέξεις. Για να προβλέψει τις λέξεις που

περιβάλλουν την είσοδο, ο αλγόριθμος χρησιμοποιεί μια διαδικασία που ονομάζεται backpropagation. Η backpropagation είναι μια τεχνική που χρησιμοποιείται για την εκπαίδευση νευρωνικών δικτύων. Η τεχνική αυτή υπολογίζει την απόκλιση μεταξύ των προβλέψεων του νευρωνικού δικτύου και των πραγματικών τιμών. Στη συνέχεια, η απόκλιση αυτή χρησιμοποιείται για την ενημέρωση των παραμέτρων του νευρωνικού δικτύου.

Η έξοδος(output): Το μοντέλο χρησιμοποιεί το διάνυσμα της λέξης στόχος για να προβλέψει τις λέξεις στο πλαίσιο της. Αυτές οι λέξεις πλαισίου εμφανίζονται στην έξοδο και συνήθως είναι τοποθετημένες με βάση τη σειρά τους πριν και μετά τη λέξη στόχο λέξη, δηλαδή $w(t-2)$, $w(t-1)$, $w(t+1)$, $w(t+2)$ αντιστοιχούν στις λέξεις που βρίσκονται δύο θέσεις πριν, μία πριν, μία μετά και δύο μετά τη λέξη στόχο αντίστοιχα.



Εικόνα 20 Skip-gram

Transformer Model

Το SentenceTransformers είναι ένα Python Framework για διανυσματικοποίηση προτάσεων, κειμένων και εικόνων. Αυτά τα διανύσματα μπορούν στην συνέχεια να συγκριθούν για να εντοπιστούν προτάσεις με παρόμοια σημασία.

Τα Sentence Transformer Models είναι χτισμένα πάνω στην αρχιτεκτονική του μετασχηματιστή, η οποία περιλαμβάνει self-attention mechanisms. Σε αντίθεση με τα παραδοσιακά μοντέλα Transformer που λειτουργούν σε επίπεδο λέξης, τα Sentence Transformers λειτουργούν σε επίπεδο πρότασης ή εγγράφου. Κωδικοποιούν ολόκληρες

προτάσεις ή έγγραφα σε διανύσματα σταθερού μεγέθους, συλλαμβάνοντας σημασιολογικές πληροφορίες και σχέσεις συμφραζομένων μεταξύ των λέξεων.

Τα Sentence Transformer Models εκπαιδεύονται χρησιμοποιώντας είτε μεγάλης κλίμακας εποπτευόμενες μεθόδους (large-scale supervised) ή μη εποπτευόμενες μεθόδους (unsupervised methods).

Το Supervised Training περιλαμβάνει την εκπαίδευση Sentence Transformer model σε label datasets για συγκεκριμένες εργασίες.

- **Sentence Pair Classification:** Σε αυτήν την εργασία, το μοντέλο εκπαιδεύεται να ταξινομεί εάν τα ζεύγη προτάσεων έχουν μια συγκεκριμένη σχέση, όπως συνεπεία, αντίφαση ή παράφραση.
- **Sentence Regression:** Εδώ, το μοντέλο μαθαίνει να προβλέπει μια συνεχή τιμή με βάση ζεύγη προτάσεων. Μια κοινή εφαρμογή είναι η σημασιολογική ομοιότητα κειμένου (STS), όπου το μοντέλο προβλέπει μια βαθμολογία ομοιότητας μεταξύ δύο προτάσεων με βάση τη σημασιολογική τους σημασία.

Κατά τη διάρκεια της εποπτευόμενης εκπαίδευσης (large-scale supervised), το μοντέλο μετασχηματιστή προτάσεων ρυθμίζεται με ακρίβεια χρησιμοποιώντας backpropagation και gradient descent για την ελαχιστοποίηση μιας συνάρτησης απώλειας συγκεκριμένης εργασίας.

Unsupervised Training περιλαμβάνει την εκπαίδευση pre-trained Sentence Transformer Model σε μεγάλα σύνολα χωρίς label για την εργασία. Αυτή η φάση αποσκοπεί στη μάθηση γενικών αναπαραστάσεων γλώσσας που αποτυπώνουν σημασιολογικές και συντακτικές πληροφορίες.

Σύγκριση με τα παραδοσιακά μοντέλα όπως το Word2Vec

- Οι μετασχηματιστές (Sentence Transformers) αποτυπώνουν καλύτερα τις πληροφορίες που σχετίζονται με τα συμφραζόμενα, καθώς κωδικοποιούν ολόκληρες προτάσεις ή έγγραφα.
- Δημιουργούν αναπαραστάσεις με σημασιολογικό νόημα που διατηρούν τις σχέσεις μεταξύ των προτάσεων.

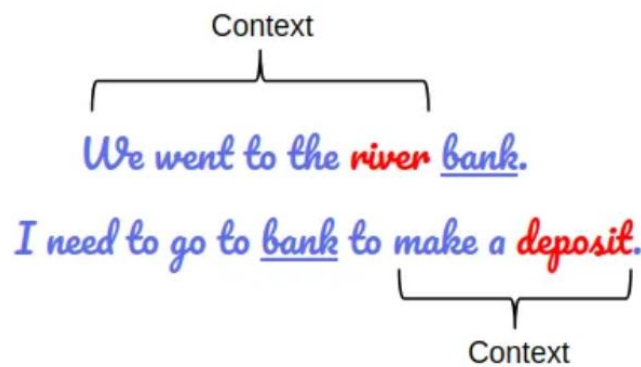
Word2Vec vs Transformer Model

Το μοντέλο του Word2Vec δημιουργεί διανύσματα που είναι εξαρτημένα από το περιβάλλον και υπάρχει μόνο μία αριθμητική αναπαράσταση για κάθε λέξη. Ωστόσο, το Transformer Model δημιουργεί διανύσματα που επιτρέπουν την ύπαρξη περισσότερων από μία

αναπαραστάσεις για την ίδια λέξη, με βάση το πλαίσιο το οποίο χρησιμοποιείται η λέξη. Τα διανύσματα του μοντέλου των transformers εξαρτώνται από το περιβάλλον.

Στο πιο κάτω σχήμα (εικόνα 21) η λέξη bank χρησιμοποιείται σε δύο διαφορετικά πλαίσια. Στην πρώτη περίπτωση γη κατά μήκος του ποταμού και στη δεύτερη ως οικονομική οντότητα. Το Word2Vec θα δημιουργήσει το ίδιο ενιαίο διάνυσμα για την τράπεζα λέξεων και για τις δύο προτάσεις. Ενώ, το BERT θα δημιουργήσει δύο διαφορετικά διανύσματα για τη λέξη τράπεζα που χρησιμοποιείται σε δύο διαφορετικά περιβάλλοντα.

Το διάνυσμα του Word2Vec για τη λέξη bank θα είναι μια μπερδεμένη αναπαράσταση καθώς συμπτύσει διαφορετικά περιβάλλοντα σε ένα ενιαίο διάνυσμα. Το διάνυσμα του BERT μπορεί να διακρίνει και να συλλάβει τις δύο διαφορετικές σημασιολογικές έννοιες παράγοντας δύο διαφορετικά διανύσματα για την ίδια λέξη “bank”.



Εικόνα 21 Η ίδια λέξη χρησιμοποιείται με διαφορετικό context

3.3 Διανυσματικές βάσεις

Διανυσματικές βάσεις

Οι διανυσματικές βάσεις δεδομένων αποθηκεύουν πληροφορίες ως διανύσματα. Τα διανύσματα είναι μια ισχυρή μέθοδος αναζήτησης σε πολύ μεγάλα και αδόμητα σύνολα δεδομένων. Αυτά τα σύνολα δεδομένων μπορεί να αποτελούνται από κείμενο, εικόνες και μια διανυσματική βάση δεδομένων ζητά αυτές τις πληροφορίες σε μια διαχειρίσιμη μορφή.

Οι διανυσματικές βάσεις δεδομένων έχουν την δυνατότητα να είναι αποτελεσματικές στην αποθήκευση διανυσμάτων υψηλών διαστάσεων που μπορούν να περιέχουν εκατοντάδες διαφορετικές διαστάσεις. Δημιουργώντας έτσι ένα μεγάλο επίπεδο πολυπλοκότητας.

Σημαντικές λειτουργίες που παρέχονται από τις διανυσματικές βάσεις.

1. Indexing Vectors (Ευρετήρια Διανυσμάτων): Μπορεί να εντοπίζει διανύσματα και να κάνει την ανάκτηση γρήγορη και αποτελεσματική.

2. Searching with Similarity (Αναζήτηση με ομοιότητα): Σε μια διανυσματική βάση δεδομένων, πραγματοποιείτε αναζήτηση με βάση το πόσο παρόμοια είναι τα διανύσματα, με ένα διάνυσμα που θα του δοθεί (στην περίπτωση αυτή μίας ερώτησης από τον χρήστη). Η βάση δεδομένων χρησιμοποιεί μαθηματικές συναρτήσεις για να προσδιορίσει ποια διανύσματα είναι παρόμοια με το ερώτημά του χρήστη.

3. Applications (Εφαρμογές): Είναι ικανές για εφαρμογές που αναζητούν ομοιότητα, μεταξύ κειμένων ή εικόνων.

Chroma DB

Η Chroma db έχει σχεδιαστεί για να βοηθά προγραμματιστές και επιχειρήσεις κάθε μεγέθους στη δημιουργία εφαρμογών LLM. Το Chroma διασφαλίζει ότι μία εφαρμογή είναι επεκτάσιμη και λειτουργεί με βέλτιστο τρόπο, επιτρέποντας την αποθήκευση, αναζήτηση και γρήγορη ανάκτηση υψηλής διάστασης διανυσμάτων. Επιπλέον, υποστηρίζει διάφορους τύπους δεδομένων, καθιστώντας το κατάλληλο για σχεδόν οποιαδήποτε εφαρμογή.

Η Chroma DB είναι ένα open source store (αποθηκευτικός χώρος για διανυσματικές αναπαραστάσεις), σχεδιασμένο για την αποθήκευση και ανάκτηση διανυσματικών αναπαραστάσεων. Είναι ιδιαίτερα χρήσιμο για την αποθήκευση διανυσματικών αναπαραστάσεων μαζί με μεταδεδομένα, τα οποία αργότερα μπορούν να χρησιμοποιηθούν από μεγάλα μοντέλα γλωσσικής επεξεργασίας (LLMs) όπως το GPT-3. Η Chroma db χρησιμοποιά τα metadata για να αναφερθεί σε επιπλέον πληροφορίες που μπορεί να προσδιορίσει για τα διανύσματα. Έχει παρόμοια λειτουργία με την χρήση ετικετών (label) στα δεδομένα για να γίνουν πιο οργανωμένα. Ακόμη, αυτές οι πληροφορίες μπορούν να χρησιμοποιηθούν αργότερα στην αναζήτηση για να επιστρέψει απαντήσεις μόνο από ένα συγκεκριμένο metadata field.

Θετικά από την χρήση της ChromaDB

Native Support for Vector Data: Το Chroma DB έχει σχεδιαστεί ειδικά για να χειρίζεται αποτελεσματικά διανυσματικά δεδομένα. Αυτό είναι ιδιαίτερα αποτελεσματικό για εφαρμογές που μπορούν να αναπαρασταθούν με διανύσματα.

High Performance: Η Chroma db χαρακτηρίζεται για την γρήγορη ανάκτηση διανυσματικών δεδομένων ακόμα και όταν πρόκειται για μεγάλα σύνολα δεδομένων. Οι μηχανισμοί ευρετηρίασης και αναζήτησης ερωτημάτων έχουν σχεδιαστεί για να χειρίζονται αποτελεσματικά τις αναζητήσεις ομοιότητας, κάτι που είναι απαραίτητο για εφαρμογές όπως συστήματα συστάσεων που βασίζονται σε περιεχόμενο.

Scalability: Η Chroma DB έχει σχεδιαστεί για οριζόντια κλίμακα, που σημαίνει ότι μπορεί να χειριστεί αυξανόμενες ποσότητες δεδομένων προσθέτοντας περισσότερους κόμβους στο σύμπλεγμα βάσης δεδομένων. Αυτή η επεκτασιμότητα το καθιστά κατάλληλο για την ανάπτυξη βάσεων δεδομένων πανεπιστημιακών ενημερωτικών δελτίων, διασφαλίζοντας ότι η απόδοση παραμένει σταθερή ακόμη και όταν ο όγκος των δεδομένων αυξάνεται με την πάροδο του χρόνου.

Flexibility in Data Modeling: Σε αντίθεση με τις παραδοσιακές σχεσιακές βάσεις δεδομένων που απαιτούν τη δομή των δεδομένων σε πίνακες με προκαθορισμένα σχήματα, το Chroma DB προσφέρει μεγαλύτερη ευελιξία στη μοντελοποίηση δεδομένων. Αυτό προσφέρει ιδιαίτερα πλεονεκτήματα όταν ασχολούμαστε με μη δομημένα ή ημι-δομημένα δεδομένα, όπως έγγραφα κειμένου (πανεπιστημιακός οδηγός σπουδών).

Κεφάλαιο 4

Σχεδιασμός Συστήματος και Σύγχρονα MΓM

4.1	Μεθοδολογία Πρότασης	39
4.2	GPT-4 openAi / GPTs	43
4.3	Bard / Gemini	45

4.1 Μεθοδολογία Πρότασης

1. Εισαγωγή στην Μεθοδολογία

Η μεθοδολογία αυτής της διπλωματικής εργασίας περιλαμβάνει τα βήματα που ακολουθήθηκαν για την ανάπτυξη ενός αυτοματοποιημένου συστήματος συμβουλευτικής για φοιτητές, χρησιμοποιώντας τεχνητή νοημοσύνη και επεξεργασία φυσικής γλώσσας. Οι κύριες φάσεις περιλαμβάνουν τη συλλογή και επεξεργασία δεδομένων, τη δημιουργία διανυσματικής βάσης, τη μετατροπή λέξεων σε διανύσματα και την ενσωμάτωση του GPT-3.5 API για την παραγωγή απαντήσεων σε φυσική γλώσσα.

2. Περιγραφή των Εργαλείων και Τεχνικών

Για την υλοποίηση του συστήματος χρησιμοποιήθηκαν τα εξής εργαλεία και τεχνικές:

GPT-3.5 API: Για την παραγωγή απαντήσεων σε φυσική γλώσσα.

Chroma db: Για τη δημιουργία και αποθήκευση της διανυσματικής βάσης.

Word2Vec: Για τη μετατροπή των λέξεων σε αριθμητικά διανύσματα.

Python: Για την ανάπτυξη του λογισμικού και την επεξεργασία των δεδομένων.

3. Συλλογή και Επεξεργασία Δεδομένων

Τα δεδομένα συλλέχθηκαν από τον οδηγό σπουδών του Πανεπιστημίου Κύπρου. Αρχικά, ο οδηγός σπουδών διαχωρίστηκε σε ξεχωριστά αρχεία κειμένου (.txt), καθένα από τα οποία αντιστοιχεί σε ένα συγκεκριμένο τμήμα. Η μετατροπή των αρχείων από pdf σε txt πραγματοποιήθηκε με τη χρήση ειδικού λογισμικού, διατηρώντας τη δομή των πινάκων και των γραμμών για να εξασφαλιστεί η ακρίβεια των δεδομένων.

4. Δημιουργία της Βάσης γνώσης

Είσοδος: Σύνολο αρχείων A (αρχεία τμημάτων από τον οδηγό σπουδών σε μορφή .txt).

Επεξεργασία:

Εισαγωγή Αρχείων: Αρχικά, γίνεται η εισαγωγή αρχείων σε μορφή .txt, όπου κάθε αρχείο αντιστοιχεί στο κείμενο του οδηγού σπουδών ανάλογα με τον κλάδο που ανήκει.

Διαχωρισμός και Αποθήκευση Δεδομένων: Τα δεδομένα από τα αρχεία διαχωρίζονται και αποθηκεύονται σε ξεχωριστά αρχεία για ευκολότερη ανάκτηση και ευέλικτη διαχείριση στη διανυσματική βάση, καθώς και για τη διευκόλυνση της αναζήτησης κατά την εκτέλεση ερωτήσεων από τον χρήστη. Η εισαγωγή στη διανυσματική βάση επιτρέπει την αποθήκευση των δεδομένων με τρόπο που λαμβάνει υπόψη τη σημασιολογική τους δομή. Αυτό σημαίνει ότι λέξεις ή φράσεις με παρόμοια σημασιολογική έννοια θα έχουν κοντινά διανύσματα αναπαράστασης.

Μετατροπή Κειμένου σε Διανύσματα: Για τη μετατροπή του περιεχομένου του κειμένου σε διανύσματα (σύνολο αριθμών), χρησιμοποιείται ο αλγόριθμος Word2Vec. Όπως φαίνεται και στην εικόνα (βάλε αριθμό εικόνας), μπορούμε να εισάγουμε ένα κείμενο σε ένα μοντέλο και να το μετατρέψουμε σε διανύσματα. Κάθε λέξη μπορεί να αντιπροσωπεύεται με πολλά διανύσματα. Όσο μεγαλύτερος είναι ο αριθμός των διανυσμάτων, τόσο καλύτερη είναι η ακρίβεια.

Διαίρεση Κειμένου σε Προτάσεις: Κάθε αρχείο κειμένου διαιρείται σε παράθυρα, δηλαδή σε συγκεκριμένο αριθμό γραμμών. Σε κάθε παράθυρο υπολογίζονται τα διανύσματα αναπαράστασης και συλλέγονται τα μεταδεδομένα, που αναφέρονται στο αρχείο, την ταυτότητά του και τον αριθμό γραμμής. Κάθε παράθυρο περιλαμβάνει έναν ορισμένο αριθμό γραμμών κειμένου και υπολογίζει το μέσο διάνυσμα για τις λέξεις που περιέχονται σε αυτές τις γραμμές.

Δημιουργία Διανυσματικής Βάσης: Τα διανύσματα αυτά, μαζί με τα μεταδεδομένα, εισάγονται στη διανυσματική βάση Chroma db για τη δημιουργία της. Τα μεταδεδομένα περιγράφουν την προέλευσή τους και τον αριθμό των προτάσεων. Επιπλέον, αποθηκεύονται

και τα έγγραφα (documents), δηλαδή οι προτάσεις που είχαν τη μεγαλύτερη σημασιολογική ομοιότητα με μια δεδομένη ερώτηση του χρήστη. Κάθε σύνολο προτάσεων αποθηκεύεται με ένα μοναδικό αναγνωριστικό (id), διευκολύνοντας την αναζήτηση και την ανάκληση συγκεκριμένων πληροφοριών.

Έξοδος: Επιστρέφεται η συλλογή των διανυσμάτων, μεταδεδομένων (metadata), μοναδικών αναγνωριστικών (ids) και εγγράφων (documents).

5. Μετατροπή Λέξεων σε Διανύσματα (Embeddings)

Είσοδος: Ερώτηση χρήστη. Σύνολο S λέξεων.

Επεξεργασία: Όταν γίνεται μια ερώτηση από τον χρήστη, το σύστημα πρέπει να κατανοήσει το περιεχόμενό της. Αυτό επιτυγχάνεται μετατρέποντας τις λέξεις της ερώτησης σε διανύσματα (embeddings), μια μορφή αναπαράστασης που μπορεί να κατανοήσει και να αναλύσει ο υπολογιστής. Το περιβάλλον χρήστη (User's Query) επιτρέπει στους φοιτητές να εισάγουν τις ερωτήσεις τους, οι οποίες στη συνέχεια επεξεργάζονται από το σύστημα.

Για τη μετατροπή των λέξεων σε διανύσματα χρησιμοποιείται το Word2Vec, ένα υπολογιστικό μοντέλο που σχεδιάστηκε για αυτή τη δουλειά. Το Word2Vec επιτρέπει σε λέξεις με παρόμοιες σημασίες να έχουν παρόμοιες αναπαραστάσεις. Εκπαιδεύεται να μαθαίνει τις συσχετίσεις μεταξύ λέξεων από το κείμενο, έτσι ώστε λέξεις με παρόμοιες σημασίες να έχουν παρόμοιες αναπαραστάσεις στον διανυσματικό χώρο. Αυτό βοηθά το σύστημα να κατανοεί τις σημασιολογικές σχέσεις μεταξύ των λέξεων στο κείμενο, διευκολύνοντας την ανάλυση της σημασίας της ερώτησης. Μετά την ανάλυση, το σύστημα μπορεί να αναζητήσει στη βάση γνώσης και να παράγει απαντήσεις σχετικές με την ερώτηση του χρήστη.

Έξοδος: Το επεξεργασμένο vector διανυσμάτων (embeddings).

6. Ενσωμάτωση και Χρήση του GPT-3.5 API

Είσοδος: Βάση γνώσης A, Σύνολο S προτάσεων από την βάση γνώσης

Επεξεργασία:

Το chatbot αντλεί σχετικό πλαίσιο από μια βάση δεδομένων πλαισίου για να κατανοήσει καλύτερα την εντολή του χρήστη. Αυτό σημαίνει την πρόσβαση σε συγκεκριμένα δεδομένα που έχουν συσχετιστεί με την ερώτηση του χρήστη, βοηθώντας το chatbot να παράγει μια πιο εύστοχη απάντηση. Το σύστημα επιστρέφει τα 4 πιο όμοια αποτελέσματα από τη διανυσματική βάση με τη μορφή ενός πίνακα. Αυτά τα αποτελέσματα χρησιμοποιούνται ως “context”, που σημαίνει ότι τα πιο σχετικά κείμενα με την ερώτηση του χρήστη θα χρησιμοποιηθούν για να καθοδηγήσουν την απάντηση του συστήματος. Για την περαιτέρω επεξεργασία αυτών των αποτελεσμάτων, χρησιμοποιείται το LLM της OpenAI, το gpt-3.5-turbo. Τα μεγάλα γλωσσικά

μοντέλα, όπως το GPT-3.5-turbo, έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην κατανόηση και παραγωγή φυσικής γλώσσας, όπως αναφέρεται και στην ανασκόπηση των μεγάλων γλωσσικών μοντέλων από τον Naveed et al. [1]. Το LLM λαμβάνει τα κείμενα από τη διανυσματική βάση και τα επεξεργάζεται για να παράγει μια πιο κατανοητή ανθρώπινη απάντηση.

Έξοδος: Το αποτέλεσμα είναι μια απάντηση που επιστρέφεται στον χρήστη. Ο αλγόριθμος επιστρέφει τα κατάλληλα διανύσματα που σχετίζονται με την ερώτηση του χρήστη σε φυσική γλώσσα.

7. Περιορισμοί και Προκλήσεις

Η δυσκολία του προβλήματος πηγάζει από την ανάγκη για την κατανόηση των διαφόρων κλάδων σπουδών που προσφέρονται από το Πανεπιστήμιο Κύπρου.

Μετατροπή οδηγού σπουδών από Pdf σε txt:

Η μετατροπή του οδηγού σπουδών από pdf σε txt ήταν αρκετά δύσκολη καθώς δεν μπορούσα να χρησιμοποιήσω ένα έτοιμο tool για την μετατροπή. Επίσης, υπάρχουν πολλοί πίνακες που απεικονίζουν τα μαθήματα μαζί με τις πιστωτικές μονάδες. Για τον λόγο αυτό έπρεπε να έχω υπόψη μου κατά τη διαδικασία της μετατροπής πως η δομή των πινάκων και των γραμμών θα πρέπει παραμείνει η ίδια.

Επιλογή μοντέλου μετατροπής διανυσμάτων:

Για την επιλογή του μοντέλου μετατροπής των διανυσμάτων πρέπει να κάνω διάφορες δοκιμές για να εντοπίσω ποιο είναι το μοντέλο που επιστρέφει τα καλύτερα αποτελέσματα. Λόγω του μεγάλου όγκου του οδηγού σπουδών χρειάζεται αρκετή ανάλυση για την επιλογή του κατάλληλου μοντέλου.

Ενσωμάτωση του οδηγού σπουδών του πανεπιστημίου Κύπρου:

Μία από τις μεγαλύτερες δυσκολίες που υπάρχουν είναι η ενσωμάτωση του οδηγού σπουδών του πανεπιστημίου Κύπρου στο σύστημα. Το GPT δεν έχει δει το κείμενο του οδηγού σπουδών κατά την διάρκεια της εκπαίδευσης του, έτσι, η κατανόηση του είναι περιορισμένη αφού δεν μπορεί να αντιληφθεί πλήρως το περιεχόμενο ενός οδηγού σπουδών καθώς δεν είχε πρόσβαση σε πραγματικούς οδηγούς σπουδών κατά την διάρκεια της εκπαίδευσης του.

Κατανόηση της φυσικής γλώσσας(Natural Language Processing - NLP):

Επίσης, ακόμη μία δυσκολία πηγάζει από την κατανόηση της φυσικής γλώσσας (Natural Language Processing - NLP) ,ώστε να μπορεί να κατανοεί και να ερμηνεύει τις ανθρώπινες εκφράσεις και διατυπώσεις που χρησιμοποιούν οι χρήστες για να ρωτήσουν το ίδιο πράγμα.

Διαχείριση ερωτήσεων εκτός των γνώσεων του:

Να μπορεί να απαντά μόνο σε ερωτήσεις που είναι σχετικές με τον οδηγό πάνω στις οποίες έχει εκπαιδευτεί. Επίσης, να μπορεί να διαχωρίζει τις ερωτήσεις που δεν γνωρίζει για να μην τις απαντήσει δίνοντας, πιθανώς, λάθος απάντηση. Είναι πολύ σημαντική αυτή η πρόκληση γιατί οι φοιτητές/τριες που θα το χρησιμοποιούν θα πρέπει να είναι σίγουροι ότι είναι αξιόπιστο αφού αφορά την ακαδημαϊκή τους πορεία στο πανεπιστήμιο.

Ενημέρωση και συντήρηση:

Θα πρέπει να υπάρχει ενημέρωση του οδηγού σπουδών σε περίπτωση που προκύψουν αλλαγές στα μαθήματα του κάθε κλάδου ή ακόμη και αλλαγές που μπορεί να γίνουν στο πρόγραμμα σπουδών σε κάποιο τμήμα του πανεπιστημίου.

4.2 GPT-4 openAi / GPTs

Το ChatGPT είναι ένα σύστημα τεχνητής νοημοσύνης που δημιουργήθηκε από την OpenAI. Είναι ένα μεγάλο γλωσσικό μοντέλο, το οποίο σημαίνει ότι έχει εκπαιδευτεί σε ένα τεράστιο σύνολο δεδομένων κειμένου και κώδικα. Αυτό του επιτρέπει να παράγει κείμενο, να μεταφράζει γλώσσες, να γράφει διαφορετικά είδη δημιουργικού περιεχομένου και να απαντά στις ερωτήσεις με έναν ενημερωτικό τρόπο.

Υποπροβλήμα: Εκπαίδευση

Είσοδος: Ένα τεράστιο σύνολο δεδομένων κειμένου και κώδικα.

Επεξεργασία: Χρήση ενός αλγορίθμου νευρωνικού δικτύου για να αναγνωρίσει μοτίβα στο σύνολο δεδομένων.

Έξοδος: Ένα εκπαιδευμένο νευρωνικό δίκτυο που μπορεί να παράγει κείμενο, να μεταφράζει γλώσσες, να γράφει διαφορετικά είδη δημιουργικού περιεχομένου και να απαντά στις ερωτήσεις με έναν ενημερωτικό τρόπο.

Το πρώτο βήμα στη λειτουργία του ChatGPT είναι η εκπαίδευση. Κατά τη διάρκεια της εκπαίδευσης, το ChatGPT εκτίθεται σε ένα τεράστιο σύνολο δεδομένων κειμένου και κώδικα. Αυτό το σύνολο δεδομένων αυτό περιλαμβάνει κείμενα από βιβλία, άρθρα, ιστοσελίδες, κώδικα από προγράμματα, ιστότοπους και άλλα. Το ChatGPT χρησιμοποιεί έναν αλγόριθμο νευρωνικού δικτύου για να αναγνωρίσει μοτίβα στο σύνολο δεδομένων. Τα μοτίβα αυτά είναι οι σχέσεις μεταξύ των λέξεων και των φράσεων. Για παράδειγμα μπορεί να μαθαίνει ότι συνήθως μία λέξη εμφανίζεται πριν από κάποια λέξη. Η διαδικασία αυτή μπορεί να διαρκέσει αρκετές εβδομάδες ή ακόμα και μήνες.

Τεχνικές εκπαίδευσης:

Το ChatGPT χρησιμοποιεί για την εκπαίδευση του μετασχηματιστές. Ο μετασχηματιστής είναι ένα είδος νευρωνικού δικτύου που είναι ιδιαίτερα αποτελεσματικό στην επεξεργασία κειμένου.

Ο μετασχηματιστής αποτελείται από μια σειρά από στρώματα. Κάθε στρώμα του μετασχηματιστή χρησιμοποιεί μια συνάρτηση μετασχηματιστή για να επεξεργαστεί τις εισόδους του. Η συνάρτηση μετασχηματιστή είναι μια μαθηματική συνάρτηση που μαθαίνει να αναγνωρίζει μοτίβα στις εισόδους.

Κατά τη διάρκεια της εκπαίδευσης, ο μετασχηματιστής ενημερώνεται συνεχώς για να βελτιώσει την ικανότητά του να αναγνωρίζει μοτίβα. Αυτό γίνεται με τη χρήση μιας τεχνικής που ονομάζεται οπίσθια διάδοση (back propagation). Η οπίσθια διάδοση είναι μια τεχνική που υπολογίζει το πόσο καλά ο μετασχηματιστής εκτελεί μια εργασία και στη συνέχεια χρησιμοποιεί αυτές τις πληροφορίες για να ενημερώσει τον μετασχηματιστή.

Υποπρόβλημα: Τροφοδότηση με αρχεία και απάντηση ερωτήσεων σχετικά με αυτά.

Είσοδος: Ένα αρχείο κειμένου.

Επεξεργασία: Χρήση του εκπαιδευμένου νευρωνικού δικτύου για να επεξεργαστεί το αρχείο και να προσθέσει τις πληροφορίες του στο σύνολο δεδομένων εκπαίδευσης.

Έξοδος: Ένα ενημερωμένο σύνολο δεδομένων εκπαίδευσης που βοηθά το ChatGPT να κατανοήσει και να παράγει κείμενο που σχετίζεται με το αρχείο που τροφοδοτήθηκε.

Ο χρήστης μπορεί να τροφοδοτήσει το ChatGPT με ένα αρχείο κειμένου ή κώδικα. Το ChatGPT χρησιμοποιεί το εκπαιδευμένο νευρωνικό δίκτυό του για να επεξεργαστεί το αρχείο. Το νευρωνικό δίκτυο αναγνωρίζει τις λέξεις και τις φράσεις στο αρχείο και τις προσθέτει στο σύνολο δεδομένων εκπαίδευσης.

Το ChatGPT χρησιμοποιεί μια τεχνική που ονομάζεται τμηματοποίηση για να διαβάσει ένα αρχείο. Η τμηματοποίηση είναι μια διαδικασία που χωρίζει ένα μεγάλο αρχείο σε μικρότερα κομμάτια. Το ChatGPT χωρίζει το αρχείο σε κομμάτια που είναι περίπου το μέγεθος μιας πρότασης. Στη συνέχεια, το ChatGPT χρησιμοποιεί το εκπαιδευμένο νευρωνικό δίκτυό του για να επεξεργαστεί κάθε κομμάτι. Το νευρωνικό δίκτυο αναγνωρίζει τις λέξεις και τις φράσεις στο κομμάτι και τις προσθέτει στο σύνολο δεδομένων εκπαίδευσης. Αυτό το σύνολο δεδομένων εκπαίδευσης ενημερώνεται συνεχώς με νέες πληροφορίες από τα αρχεία που τροφοδοτούνται. Όταν ένας χρήστης υποβάλει μια ερώτηση στο ChatGPT, το ChatGPT χρησιμοποιεί το ενημερωμένο σύνολο δεδομένων εκπαίδευσης για να προσπαθήσει να καταλάβει την ερώτηση. Στη συνέχεια, το ChatGPT παράγει μια απάντηση που βασίζεται στις πληροφορίες που έχει μάθει.

Δημιουργία Chatbot με την χρήση των GPTs από το openAi

Το openAi προχώρησε στην δημιουργία προσαρμοσμένων εκδόσεων για το ChatGPT τα οποία μπορείς να δημιουργήσεις για συγκεκριμένο σκοπό, αυτά ονομάζονται GPTs. Τα GPTs είναι ένας νέος τρόπος για να δημιουργήσεις μία προσαρμοσμένη έκδοση του ChatGPT που θα είναι πιο βοηθητικά για συγκεκριμένες εργασίες, για την δουλειά, για το σπίτι και μπορείς να τα μοιραστείς και με άλλους. Για παράδειγμα, υπάρχει ένα GPT που μπορεί να εξηγήσει πως παίζεται ένα επιτραπέζιο παιχνίδι ή να βοηθήσει κάποιο παιδί με τα μαθηματικά προβλήματα του σχολείου και να του προσφέρει έξτρα μαθήματα.

Μπορούν να δημιουργηθούν από οποιοδήποτε καθώς δεν προαπαιτούν την χρήση κώδικα. Μπορούν να δημιουργηθούν για προσωπικούς σκοπούς ή για την χρήση τους σε κάποια εταιρεία ή μπορεί να απευθύνονται σε όλο το κοινό. Για την δημιουργία τους μπορείς να ξεκινήσεις με ένα απλό διάλογο δίνοντας εντολές, παραθέτοντας νέες γνώσεις, εξηγώντας του τι πρέπει να κάνει η εφαρμογή.

4.3 Bard / Gemini

Το Google Gemini είναι ένα εργαλείο συνομιλίας τεχνητής νοημοσύνης (AI) που σχεδιάστηκε από την Google για την προσομοίωση ανθρώπινων συνομιλιών χρησιμοποιώντας επεξεργασία φυσικής γλώσσας (NLP) και μηχανική εκμάθηση. Το google Gemini ανήκει στην οικογένεια των LLM που έχουν δυνατότητες κατανόησης γλώσσας, ήχου, κώδικα και βίντεο.

Το Gemini ενσωματώνει τις δυνατότητες NLP, οι οποίες παρέχουν τη δυνατότητα κατανόησης και επεξεργασίας της γλώσσας. Το Gemini χρησιμοποιείται επίσης για την κατανόηση ερωτημάτων εισόδου καθώς και δεδομένων. Είναι σε θέση να κατανοεί και να αναγνωρίζει εικόνες, επιτρέποντάς του να αναλύει πολύπλοκα οπτικά στοιχεία, όπως γραφήματα και σχήματα, χωρίς την ανάγκη εξωτερικής οπτικής αναγνώρισης χαρακτήρων (OCR). Διαθέτει επίσης ευρείες πολυγλωσσικές δυνατότητες για μεταφραστικές εργασίες και λειτουργικότητα σε διάφορες γλώσσες.

Σε αντίθεση με τα προηγούμενα μοντέλα τεχνητής νοημοσύνης της Google, το Gemini είναι εκπαιδευμένο σε σύνολα δεδομένων που εκτείνονται σε πολλούς τύπους δεδομένων. Ως πολυτροπικό μοντέλο, το Gemini επιτρέπει διατροπικές συλλογιστικές ικανότητες. Αυτό σημαίνει ότι το Gemini μπορεί να αιτιολογήσει μια σειρά διαφορετικών τύπων δεδομένων εισόδου, συμπεριλαμβανομένων ήχου, εικόνων και κειμένου. Για παράδειγμα, το Gemini μπορεί να κατανοήσει χειρόγραφες σημειώσεις, γραφήματα και διαγράμματα για να λύσει πολύπλοκα προβλήματα. Η αρχιτεκτονική Gemini υποστηρίζει την άμεση πρόσληψη κειμένου, εικόνων, κυματομορφών ήχου και πλαισίων βίντεο ως αλληλοδιαπλεκόμενες ακολουθίες.

Πως δουλεύει το Gemini;

Το Google Gemini λειτουργεί με το να εκπαιδεύεται πρώτα σε ένα τεράστιο σύνολο δεδομένων. Μετά την εκπαίδευση, το μοντέλο χρησιμοποιεί διάφορες τεχνικές νευρωνικών δικτύων για να είναι σε θέση να κατανοήσει το περιεχόμενο, να απαντήσει σε ερωτήσεις, να δημιουργήσει κείμενο και να παράγει αποτελέσματα.

Τα Gemini LLM χρησιμοποιούν μια αρχιτεκτονική νευρωνικών δικτύων που βασίζονται σε μοντέλο μετασχηματιστή. Το Google DeepMind χρησιμοποιεί αποτελεσματικούς μηχανισμούς προσοχής (attention mechanism) στον αποκωδικοποιητή μετασχηματιστή (transformer decoder) για να βοηθήσει τα μοντέλα να επεξεργάζονται μεγάλα περιβάλλοντα, που εκτείνονται σε διαφορετικούς τρόπους.

Τα μοντέλα Gemini έχουν εκπαιδευτεί σε διάφορα πολυτροπικά και πολύγλωσσα σύνολα δεδομένων κειμένου, εικόνων, ήχου και βίντεο με το Google DeepMind χρησιμοποιώντας προηγμένο φιλτράρισμα δεδομένων για τη βελτιστοποίηση της εκπαίδευσης. Καθώς αναπτύσσονται διαφορετικά μοντέλα Gemini για την υποστήριξη συγκεκριμένων υπηρεσιών Google, υπάρχει μια διαδικασία (Fine Tuning) που μπορεί να χρησιμοποιηθεί για την περαιτέρω βελτιστοποίηση ενός μοντέλου για μια περίπτωση χρήσης.

Κεφάλαιο 5

Αξιολόγηση/ Συμπεράσματα

5.1 Αξιολόγηση εφαρμογής με την χρήση διανυσματικής βάσης Chroma DB	47
5.2 Αξιολόγηση καινούργιας εφαρμογής GPTs	53
5.3 Αξιολόγηση έτοιμης εφαρμογής GPTs	63
5.4 Συμπεράσματα από την σύγκριση των μοντέλων	70

Για την αξιολόγηση της εφαρμογής προχώρησα στην συλλογή ερωτήσεων από φοιτητές/τριες πληροφορικής σχετικές με τον οδηγό που είχαν συμφοιτητές/τριες μου κατά τη διάρκεια των σπουδών τους.

5.1 Αξιολόγηση εφαρμογής με την χρήση διανυσματικής βάσης Chroma DB

1.

You: Ποιος είναι ο ελάχιστος αριθμός ECTS που απαιτείται για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών;

Chatbot: Ο ελάχιστος αριθμός ECTS που απαιτείται για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών είναι 240 ECTS, σύμφωνα με τον οδηγό σπουδών.

2.

You: Ποιοι είναι οι γενικοί στόχοι του Προπτυχιακού Προγράμματος Πληροφορικής;

Chatbot: Οι γενικοί στόχοι του Προπτυχιακού Προγράμματος Πληροφορικής συμπεριλαμβάνουν την προετοιμασία των φοιτητών για την κατανόηση και ανάπτυξη

λογισμικού, την κατανόηση των βασικών αρχών της πληροφορικής, την ανάπτυξη δεξιοτήτων σε προγραμματισμό και ανάλυση συστημάτων, καθώς και την ανάπτυξη κριτικής σκέψης και προβληματισμού.

3.

Υου: 3.Ποια συγκεκριμένα επαγγελματικά αποτελέσματα στοχεύει το Τμήμα Επιστήμης Υπολογιστών να προετοιμάσει τους αποφοίτους του;

Chatbot: Το Τμήμα Επιστήμης Υπολογιστών στοχεύει να προετοιμάσει τους αποφοίτους του για επαγγελματικές επιλογές στους τομείς της πληροφορικής, των υπολογιστών, της τεχνητής νοημοσύνης, της ρομποτικής, του δικτύων και πληροφορικής, της ανάπτυξης λογισμικού και της ψηφιακής επεξεργασίας σήματος. Οι απόφοιτοι είναι εξοπλισμένοι με τις απαραίτητες γνώσεις και δεξιότητες για να ανταπεξέλθουν σε ένα ευρύ φάσμα επαγγελματικών προοπτικών στον τομέα της πληροφορικής και των νέων τεχνολογιών.

4.

Υου: Πώς διατηρεί το Τμήμα Πληροφορικής τη σύνδεσή του με την τοπική βιομηχανία;

Chatbot: Το Τμήμα Πληροφορικής διατηρεί τη σύνδεσή του με την τοπική βιομηχανία μέσω διάφορων πρωτοβουλιών και δράσεων. Αυτές περιλαμβάνουν τη συμμετοχή σε συνέδρια και εκδηλώσεις όπου παρουσιάζονται τα ερευνητικά αποτελέσματα του τμήματος, την ανάπτυξη ερευνητικών έργων σε συνεργασία με επιχειρήσεις, την παροχή πρακτικής εμπειρίας στους φοιτητές μέσω πρακτικής άσκησης και προσφοράς πρακτικών μαθημάτων. Με αυτόν τον τρόπο, το Τμήμα Πληροφορικής διασφαλίζει ότι οι φοιτητές/τριες του εκπαιδεύονται με τρόπο που ανταποκρίνεται στις ανάγκες και τις απαιτήσεις της τοπικής βιομηχανίας.

5.

Υου: Ποια είναι η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής;

Chatbot: Η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής είναι οκτώ (8) εξάμηνα.

6.

Υου: Ποιους τομείς σπουδών καλύπτει το Προπτυχιακό Πρόγραμμα Πληροφορικής;

Chatbot: Το προπτυχιακό πρόγραμμα πληροφορικής καλύπτει τομείς όπως οι βασικές αρχές και τεχνικές που υποστηρίζουν την επίλυση προβλημάτων στην τεχνητή νοημοσύνη, η διαχείριση μεγάλων όγκων δεδομένων χρησιμοποιώντας εξειδικευμένα εργαλεία και τεχνικές όπως το Hadoop και το Spark, καθώς και ο προγραμματισμός σε γλώσσες όπως η Python.

7.

Υου: Ποια θέματα περιλαμβάνονται στην περιοχή "Θεωρία" του προγράμματος;

Chatbot: Η περιοχή "Θεωρία" του προγράμματος περιλαμβάνει θέματα όπως οι αφηρημένοι τύποι δεδομένων, η πολυπλοκότητα αλγορίθμων, η ανάλυση της μέσης και της χειρίστης

περίπτωσης, οι γραμμικές δομές δεδομένων όπως λίστες, στοίβες και ουρές, καθώς και οι δενδρικές δομές δεδομένων όπως τα δυαδικά δένδρα και οι ουρές προτεραιότητας.

8.

You: Ποια είναι τα βασικά μαθήματα που διδάσκονται στα «Υπολογιστικά Συστήματα»;

Chatbot: Στα "Υπολογιστικά Συστήματα" διδάσκονται θέματα όπως η ανάλυση συστημάτων μέσω μετρικού λογισμικού, η ανάλυση και διεκπεραίωση έργων λογισμικού από ομάδες φοιτητών, καθώς και εξειδικευμένα θέματα ανάλογα με τη φύση των έργων όπως διακομιστές διαδικτύου, πακέτα γραφικών διεπαφών χρήστη κ.λπ. Τα προαπαιτούμενα για τα μαθήματα αυτά είναι το ΕΠΑ 231 και το ΕΠΑ 343. Επίσης, μπορεί να ασχολούνται με θέματα όπως το pair programming και η αξιολόγηση του παραγόμενου συστήματος λογισμικού.

9.

You: Σε τι επικεντρώνεται η περιοχή "Επίλυση προβλημάτων" στο πρόγραμμα σπουδών της Επιστήμης των Υπολογιστών;

Chatbot: Η περιοχή "Επίλυση προβλημάτων" στο πρόγραμμα σπουδών της Επιστήμης των Υπολογιστών επικεντρώνεται στην ανάπτυξη και εφαρμογή υπολογιστικών μεθόδων για την επίλυση προβλημάτων σε διάφορους τομείς. Αυτή η περιοχή εστιάζει στη μελέτη του τρόπου με τον οποίο οι υπολογιστές μπορούν να χρησιμοποιηθούν για την ανάλυση και επίλυση προβλημάτων, χρησιμοποιώντας αλγορίθμους και τεχνικές που είναι βασισμένες σε υπολογιστικές μεθόδους.

10.

You: Πώς η περιοχή "Εφαρμογές" ενσωματώνει γνώσεις από άλλους τομείς;

Chatbot: Η περιοχή "Εφαρμογές" ενσωματώνει γνώσεις από άλλους τομείς μέσω της εφαρμογής των θεωρητικών γνώσεων που αποκομίζονται από τα μαθήματα όπως η Γραμμική Άλγεβρα και οι Διαφορικές Εξισώσεις. Για παράδειγμα, η μηχανική όραση, που αποτελεί έναν τομέα εφαρμογών, ενσωματώνει γνώσεις από τη θεωρία της νευρωνικής επιστήμης και της μηχανικής μάθησης, που αναφέρονται σε βιβλία όπως το "Theoretical Neuroscience: Computational and Mathematical Modelling of Neural Systems" και το "Pattern Recognition and Machine Learning".

11.

You: Τι είδους μαθήματα κατατάσσει το πρόγραμμα στα «Υποχρεωτικά Μαθήματα»;

Chatbot: Το πρόγραμμα κατατάσσει στα «Υποχρεωτικά Μαθήματα» μαθήματα που αποτελούν τον πυρήνα των σπουδών και πρέπει να παρακολουθηθούν από όλους/ες τους/τις φοιτητές/τριες. Αυτά τα μαθήματα παρέχουν τις βασικές γνώσεις και δεξιότητες που απαιτούνται για την απόκτηση του τίτλου σπουδών.

12.

You: Πώς ορίζονται στο πρόγραμμα τα «Μαθήματα Περιορισμένης Επιλογής» και τα «Μαθήματα Ελεύθερης Επιλογής»;

Chatbot: Τα «Μαθήματα Περιορισμένης Επιλογής» ορίζονται στο πρόγραμμα ως εκείνα τα μαθήματα που οι φοιτητές/τριες πρέπει να επιλέξουν από μια συγκεκριμένη λίστα μαθημάτων που προσφέρονται στο πλαίσιο του προγράμματος σπουδών τους.

Τα «Μαθήματα Ελεύθερης Επιλογής» ορίζονται ως εκείνα τα μαθήματα που οι φοιτητές έχουν την ελευθερία να επιλέξουν από οποιοδήποτε τμήμα του πανεπιστημίου ή ακόμα και από άλλα πανεπιστήμια, με σκοπό την εμβάθυνση σε κάποιο ειδικότερο θέμα που τους ενδιαφέρει ή την επέκταση των γνώσεων τους σε συγκεκριμένους τομείς.

13.

Υου: Ποιες ειδικότητες έχουν στη διάθεσή τους οι φοιτητές/τριες από το ακαδημαϊκό έτος 2016/2017;

Chatbot: Από το ακαδημαϊκό έτος 2016/2017, οι φοιτητές/τριες έχουν στη διάθεσή τους τις ακόλουθες ειδικότητες: 1. Εισαγωγή στη Νεοελληνική Φιλολογία, 2. Συγγραφή Ακαδημαϊκού Δοκιμίου, 3. Εισαγωγή στη Θεωρητική Γλωσσολογία, 4. Εισαγωγή στην Κλασική Φιλολογία και 5. Θεματογραφία Αρχαίων Ελληνικών. Το συνολικό πλήθος των ειδικοτήτων είναι 31 μονάδες.

14.

Υου: Πώς μπορεί ένας/μίαφοιτητής/τρια να επιτύχει μια εξειδίκευση σύμφωνα με τις οδηγίες του προγράμματος;

Chatbot: Για να επιτύχει μια εξειδίκευση σύμφωνα με τις οδηγίες του προγράμματος, ο φοιτητής πρέπει να εξετάσει προσεκτικά τον οδηγό σπουδών και να εστιάσει στα μαθήματα που σχετίζονται με την εξειδίκευση που τον ενδιαφέρει. Επιπλέον, πρέπει να συμβουλευτεί τους ακαδημαϊκούς συμβούλους του για τη διαμόρφωση ενός σχεδίου σπουδών που θα του επιτρέψει να αποκτήσει τις απαιτούμενες γνώσεις και δεξιότητες. Τέλος, ο/η φοιτητής/τρια θα πρέπει να συμμετέχει ενεργά στα σχετικά ερευνητικά προγράμματα και εκδηλώσεις που προσφέρει το πρόγραμμα, προκειμένου να εμβαθύνει στον τομέα της εξειδίκευσής του.

15.

Υου: Ποιες προϋποθέσεις πρέπει να πληροί ένα μάθημα για να αναγνωριστεί ως εξειδίκευση;

Chatbot: Για να αναγνωριστεί ένα μάθημα ως εξειδίκευση, πρέπει να προσφέρει εμβάθυνση και εξειδίκευση σε συγκεκριμένο θεματικό πεδίο. Συνήθως, αυτό σημαίνει ότι το μάθημα απαιτεί συγκεκριμένες προαπαιτούμενες γνώσεις ή δεξιότητες που οι φοιτητές/τριες πρέπει να έχουν αποκτήσει από προηγούμενα μαθήματα.

16.

Υου: Περιγράψτε την εξάρτηση μεταξύ των μαθημάτων όπως περιγράφεται στο πρόγραμμα.

Chatbot: Σύμφωνα με το πρόγραμμα σπουδών, υπάρχει μια σαφής εξάρτηση μεταξύ των μαθημάτων, με ορισμένα μαθήματα να αποτελούν προαπαιτούμενα για άλλα. Για παράδειγμα, προϋπόθεση για το μάθημα "Παραδόσεις Νεοελληνικής Φιλολογίας" μπορεί να είναι το

μάθημα "Παράδοση Γλωσσολογίας". Αυτό το μοντέλο εξασφαλίζει μια συνεκτική πορεία μάθησης και την κατάλληλη προετοιμασία των φοιτητών για τις πιο προχωρημένες μαθηματικές ενότητες.

17.

Υου: Ποια είναι τα βασικά μαθήματα που πρέπει να παρακολουθήσουν όλοι οι φοιτητές/τριες της Πληροφορικής;

Chatbot: Όλοι οι φοιτητές της Πληροφορικής πρέπει να παρακολουθήσουν τα εξής βασικά μαθήματα:

- Εισαγωγή στην Πληροφορική
- Αλγοριθμική και Προγραμματισμός
- Στοιχεία Πληροφορικής
- Προγραμματισμός σε C++
- Βάσεις Δεδομένων
- Δίκτυα Υπολογιστών

18.

Υου: Ποια είναι τα κύρια χαρακτηριστικά της συνιστώσας της Μεμονωμένης Διατριβής;

Chatbot: Η συνιστώσα της Μεμονωμένης Διατριβής περιλαμβάνει την εκπόνηση μιας ανεξάρτητης ερευνητικής εργασίας από τον φοιτητή/τρια με στόχο την εμβάθυνση σε ένα συγκεκριμένο πεδίο και την επίδειξη ερευνητικών δεξιοτήτων αλλά και κριτικής σκέψης εκ μέρους του/της φοιτητή/τριας. Κάθε διατριβή καθοδηγείται από έναν/μία επιβλέποντα/επιβλέπουσα καθηγητή/τρια, ο/η οποίος/α παρέχει καθοδήγηση στον/στη φοιτητή/τρια κατά τη διάρκεια εκπόνησης της εργασίας. Η μεμονωμένη διατριβή καταλήγει σε μία γραπτή ερευνητική εργασία και την υποστήριξη της προφορικά.

19.

Υου: Πώς είναι δομημένα τα μαθήματα επιλογής σε διαφορετικά εξάμηνα;

Chatbot: Τα μαθήματα επιλογής σε διαφορετικά εξάμηνα είναι δομημένα με βάση το περιεχόμενό τους και τις γνώσεις και δεξιότητες που απαιτούνται για την επιτυχή ολοκλήρωσή τους. Κάθε μάθημα επιλογής έχει συγκεκριμένα προαπαιτούμενα και βιβλιογραφία που πρέπει να διαβαστεί. Επίσης, ορίζεται ο αριθμός των ECTS που αντιστοιχεί σε κάθε μάθημα, καθώς και η διδακτική μονάδα υπεύθυνη για τη διδασκαλία του. Οι φοιτητές/τριες επιλέγουν τα μαθήματα σύμφωνα με τα ενδιαφέροντά τους και τις ακαδημαϊκές τους ανάγκες, σύμφωνα πάντα με τον οδηγό σπουδών του κάθε προγράμματος.

Υου: Πώς είναι δομημένα τα μαθήματα επιλογής σε διαφορετικά εξάμηνα;

20.

Υου: Ποια μαθήματα μαθηματικών είναι υποχρεωτικά για τους/τις φοιτητές/τριες της Πληροφορικής;

Chatbot: Σύμφωνα με τον οδηγό σπουδών, για τους/τις φοιτητές/τριες του τμήματος Πληροφορικής το υποχρεωτικό μάθημα μαθηματικών περιλαμβάνει το μάθημα "Δομές Δεδομένων και Αλγόριθμοι".

5.2 Αξιολόγηση καινούργιας εφαρμογής GPTs

Παράδειγμα ενός GPT για την δημιουργία chatbot για τους φοιτητές του ΠανεπιστημίουΚύπρου.

MyGPT University of Cyprus
Live · Only me

Updates pending

Share

Update

CreateConfigure



Name
MyGPT University of Cyprus

Description
English and Greek-speaking academic guide for University of Cyprus, precise in semester-specific course details.

Instructions
Given a question, and a set of relevant passage texts from the study guide, I should answer in an understandable and easier way to the question, using only the context of the texts given to you.
If the query is too general, ask the user to be more specific and provide suggestions.
Answer the question using only the box provided.
If insufficient information is provided, do not answer the question.
Do not include irrelevant information in the body of the answer.

Conversation starters
Which courses are required in the 4th semester for computer science students?
What is the code for the Machine Learning course?
Can you list the 2nd semester courses for electrical engineering students?
What are the prerequisites for the 3rd semester cybersecurity course?

Preview


MyGPT University of Cyprus
English and Greek-speaking academic guide for University of Cyprus,
precise in semester-specific course details.

Message MyGPT University of Cyprus

Έδωσα τις οδηγίες και ανέβασα τον οδηγό σπουδών για να μπορεί να τον χρησιμοποιεί το chat gpt για την δημιουργία απαντήσεων σε ερωτήσεις φοιτητών/τριων του Πανεπιστημίου Κύπρου.

Παράδειγμα ερωτήσεων:

1.

 **You**

Ποιος είναι ο ελάχιστος αριθμός ECTS που απαιτείται για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών;

 **MyGPT University of Cyprus**

Για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών στο Προπτυχιακό Πρόγραμμα Σπουδών απαιτούνται τουλάχιστον 240 μονάδες ECTS.

2.



MyGPT University of Cyprus

Οι γενικοί στόχοι του Προπτυχιακού Προγράμματος Πληροφορικής στο Πανεπιστήμιο Κύπρου είναι οι εξής:

1. **Επαγγελματική και Ακαδημαϊκή Προοπτική:** Να προετοιμάσει τους αποφοίτους για θέσεις υψηλής ευθύνης στον επαγγελματικό και ακαδημαϊκό χώρο, όπου θα μπορέσουν να προωθήσουν την ανάπτυξη και εφαρμογή νέων μεθόδων και ιδεών.
2. **Σύνδεση με την Βιομηχανία:** Το πρόγραμμα τονίζει τη σημασία της σύνδεσης και της επικοινωνίας με την εγχώρια βιομηχανία για τη δημιουργία συνεργασιών μέσω των αποφοίτων του.
3. **Δια βίου Μάθηση και Πνευματική Ανεξαρτησία:** Να καλλιεργήσει στους φοιτητές την επιθυμία για διαρκή μάθηση και ανώτερη εκπαίδευση, ωριμότητα, ανεξαρτησία πνεύματος και κριτική σκέψη.
4. **Εμπέδωση Γνώσης Πληροφορικής:** Οι απόφοιτοι θα κατανοούν την Πληροφορική τόσο ως επιστημονικό πεδίο όσο και ως επαγγελματική ενασχόληση, προετοιμασμένοι για εργασία στην εγχώρια αγορά ή για περαιτέρω σπουδές και εξειδίκευση σε τομείς έρευνας.
5. **Προσαρμογή στις Τεχνολογικές Εξελίξεις:** Να προετοιμάσουν τους απόφοιτους να παρακολουθούν και να συμβαδίζουν με τις ραγδαίες εξελίξεις στον επιστημονικό και τεχνολογικό χώρο.
6. **Εκπαίδευση στη Δευτεροβάθμια:** Αναμένεται ότι οι απόφοιτοι θα μπορούν επίσης να διδάξουν Πληροφορική στη δευτεροβάθμια εκπαίδευση μετά από επιτυχή παρακολούθηση προγράμματος παιδαγωγικής κατάρτισης.

Αυτοί είναι οι βασικοί στόχοι που διαμορφώνουν τον πυρήνα του Προπτυχιακού Προγράμματος Σπουδών Πληροφορικής στο Πανεπιστήμιο Κύπρου.

3.



You

Ποια συγκεκριμένα επαγγελματικά αποτελέσματα στοχεύει το Τμήμα Επιστήμης Υπολογιστών να προετοιμάσει τους αποφοίτους του;



MyGPT University of Cyprus

Το Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κύπρου έχει στόχο να προετοιμάσει τους αποφοίτους του για θέσεις υψηλής ευθύνης τόσο στον επαγγελματικό όσο και στον ακαδημαϊκό χώρο. Συγκεκριμένα, οι απόφοιτοι προετοιμάζονται να σταδιοδρομήσουν σε θέσεις όπου θα μπορούν να προωθήσουν αποτελεσματικά την ανάπτυξη και εφαρμογή νέων μεθόδων και ιδεών. Το Τμήμα επίσης δίνει έμφαση στη σύνδεση με την εγχώρια βιομηχανία και στοχεύει να δημιουργήσει δίαυλο συνεργασίας με την κοινωνία μέσω των αποφοίτων του.

Ακαδημαϊκά, οι απόφοιτοι θα έχουν τη δυνατότητα να εργαστούν ως επιστήμονες στον τομέα της Τεχνολογίας Πληροφοριών, να συνεχίσουν τις σπουδές τους για την απόκτηση μεταπτυχιακού διπλώματος, να εξειδικευτούν σε συγκεκριμένους τομείς έρευνας ή ακόμα και να διδάξουν Πληροφορική στη δευτεροβάθμια εκπαίδευση μετά από επιτυχή παρακολούθηση προγράμματος παιδαγωγικής κατάρτισης.

4.



You

Πώς διατηρεί το Τμήμα Πληροφορικής τη σύνδεσή του με την τοπική βιομηχανία;



MyGPT University of Cyprus

Το Τμήμα Πληροφορικής στο Πανεπιστήμιο Κύπρου δίνει ιδιαίτερη σημασία στη διατήρηση σύνδεσης με την τοπική βιομηχανία. Αυτό περιλαμβάνει τη συνεχή επικοινωνία με την εγχώρια βιομηχανία και τη δημιουργία ενός διαύλου μέσω των αποφοίτων του Τμήματος. Ένας από τους στόχους του Τμήματος είναι να προετοιμάσει αποφοίτους ικανούς να εργαστούν σε υψηλές θέσεις υπευθυνότητας στον επαγγελματικό χώρο και να προωθήσουν αποτελεσματικά την ανάπτυξη και εφαρμογή νέων μεθόδων και ιδεών.

5.



You

Ποια είναι η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής;



MyGPT University of Cyprus

Η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής στο Πανεπιστήμιο Κύπρου είναι οκτώ εξάμηνα, όμως υπάρχει η δυνατότητα να παραταθεί μέχρι τα δώδεκα εξάμηνα.



6.



You

Ποιους τομείς σπουδών καλύπτει το Προπτυχιακό Πρόγραμμα Πληροφορικής;



MyGPT University of Cyprus

Το Προπτυχιακό Πρόγραμμα Πληροφορικής στο Πανεπιστήμιο Κύπρου καλύπτει τους εξής τομείς σπουδών:

1. **Θεωρία:** Αναλύει τις βασικές θεωρίες και μοντέλα υπολογισμού, σχεδίαση και ανάλυση αλγορίθμων, καθώς και μαθηματική λογική που σχετίζεται με την πληροφορική.
2. **Υπολογιστικά Συστήματα:** Περιλαμβάνει μαθήματα για το υλικό και το λογισμικό συστημάτων, αρχιτεκτονική υπολογιστών, λειτουργικά συστήματα, και ανάπτυξη συστημάτων μικροεπεξεργαστών.
3. **Επίλυση Προβλημάτων:** Εστιάζει στην ανάπτυξη αλγοριθμικής σκέψης, προγραμματισμού και σχεδίασης αλγορίθμων, προσφέροντας γνώση σε διάφορες γλώσσες προγραμματισμού.
4. **Εφαρμογές:** Αναφέρεται στην ανάπτυξη εφαρμογών για την επίλυση πραγματικών προβλημάτων, μελετώντας βάσεις δεδομένων, συστήματα γνώσης, και τεχνολογία λογισμικού.

7.



You

Ποια θέματα περιλαμβάνονται στην περιοχή "Θεωρία" του προγράμματος;



MyGPT University of Cyprus

Η περιοχή "Θεωρία" του προπτυχιακού προγράμματος σπουδών πληροφορικής στην Πανεπιστήμιο Κύπρου καλύπτει θέματα που σχετίζονται με τη θεωρία και τα μοντέλα υπολογισμού, τη σχεδίαση και ανάλυση αλγορίθμων, καθώς και τη μαθηματική λογική, η οποία διαδραματίζει κεντρικό ρόλο ως Λογισμός της Πληροφορικής. Περιλαμβάνεται επίσης η διδασκαλία των Διακριτών Μαθηματικών μέσω σχετικών μαθημάτων και η ενδυνάμωση της ικανότητας των φοιτητών για αφαιρετικότητα και τυπικό διαλογισμό, μέσω μαθημάτων που προσφέρονται από το Τμήμα Μαθηματικών και Στατιστικής.

8.



You

Ποια είναι τα βασικά μαθήματα που διδάσκονται στα «Υπολογιστικά Συστήματα»;



MyGPT University of Cyprus

Στην περιοχή των "Υπολογιστικών Συστημάτων" του προπτυχιακού προγράμματος σπουδών πληροφορικής στην Πανεπιστήμιο Κύπρου, τα βασικά μαθήματα περιλαμβάνουν θέματα σχετικά με το υλικό και το λογισμικό των συστημάτων. Συγκεκριμένα, καλύπτονται οι βασικές αρχές οργάνωσης και αρχιτεκτονικής υπολογιστών, λειτουργικά συστήματα, σχεδίαση και υλοποίηση γλωσσών προγραμματισμού, συστήματα μικροεπεξεργαστών, μεταβίβαση δεδομένων, δίκτυα, κατανομημένα συστήματα, παράλληλα και νέα αρχιτεκτονικά.

9.



You

Σε τι επικεντρώνεται η περιοχή "Επίλυση προβλημάτων" στο πρόγραμμα σπουδών της Επιστήμης των Υπολογιστών;



MyGPT University of Cyprus

Η περιοχή "Επίλυση Προβλημάτων" στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής της Πανεπιστημίου Κύπρου επικεντρώνεται στην ανάπτυξη αλγοριθμικής σκέψης και δίνει έμφαση στις αρχές προγραμματισμού και τη σχεδίαση αλγορίθμων. Ο βασικός στόχος είναι η απόκτηση ικανότητας στη χρήση διαφόρων γλωσσών προγραμματισμού. Παράλληλα, διδάσκονται διάφορα πρότυπα προγραμματισμού (διαδικασιακό ή προστακτικό, αντικειμενοστρεφή, λογικό). Τα μαθήματα επιλογής της περιοχής αυτής καλύπτουν προχωρημένες τεχνικές επίλυσης προβλημάτων βασιζόμενες στον παραλληλισμό και τον ταυτοχρονισμό.

10.



You

Πώς η περιοχή "Εφαρμογές" ενσωματώνει γνώσεις από άλλους τομείς;



MyGPT University of Cyprus

Η περιοχή "Εφαρμογές" στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής του Πανεπιστημίου Κύπρου ενσωματώνει γνώσεις και ικανότητες από άλλες περιοχές με σκοπό την ανάπτυξη χρήσιμων εφαρμογών για την επίλυση ρεαλιστικών προβλημάτων. Αυτή η περιοχή επικεντρώνεται στην εξέταση σημαντικών τεχνολογικών κατασκευασμάτων όπως βάσεις δεδομένων, γραφικά και συστήματα διασύνδεσης χρήστη-μηχανής, τόσο ως αυτοτελείς εφαρμογές όσο και ως εργαλεία για την ανάπτυξη πιο σύνθετων εφαρμογών.

Επιπλέον, η περιοχή εξετάζει νέες μεθοδολογίες τεχνολογίας λογισμικού που καλύπτουν όλα τα στάδια σχεδίασης, ανάπτυξης και διατήρησης υψηλής ποιότητας εφαρμογών. Μέσω της Επαγγελματικής Πρακτικής Τεχνολογίας Λογισμικού, εφαρμόζονται αυτές οι μεθοδολογίες σε πραγματικά έργα, αξιολογώντας τους παράγοντες που συμβάλλουν στην επιτυχία των έργων. Επιπρόσθετα, συζητούνται τα κυριότερα κοινωνικά και δεοντολογικά θέματα που σχετίζονται με την εφαρμογή της πληροφορικής στην κοινωνία.

11.



You

Τι είδους μαθήματα κατατάσσει το πρόγραμμα στα «Υποχρεωτικά Μαθήματα»;



MyGPT University of Cyprus

Τα "Υποχρεωτικά Μαθήματα" στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής του Πανεπιστημίου Κύπρου αποτελούν τον κορμό του προγράμματος και περιλαμβάνουν:

1. **Υποχρεωτικά Μαθήματα του Τμήματος Πληροφορικής:** Αυτά τα μαθήματα καλύπτουν τις βασικές έννοιες και τεχνικές της επιστήμης των υπολογιστών και της πληροφορικής. Παραδείγματα περιλαμβάνουν αλγόριθμους, δομές δεδομένων, προγραμματισμό, οργάνωση υπολογιστών, λειτουργικά συστήματα κ.ά.
2. **Υποχρεωτικά Μαθήματα από άλλα Τμήματα του Πανεπιστημίου:** Αυτά ενδέχεται να καλύψουν άλλες συμπληρωματικές γνωστικές περιοχές που είναι σημαντικές για την πληροφορική, όπως μαθηματικά, στατιστική, κλπ.
3. **Μαθήματα Αγγλικής γλώσσας:** Το πρόγραμμα συχνά περιλαμβάνει υποχρεωτικά μαθήματα στην αγγλική γλώσσα, προκειμένου να εξασφαλίσει ότι οι φοιτητές έχουν επαρκή γνώση της διεθνούς γλώσσας της επιστήμης και τεχνολογίας.

Αυτά τα μαθήματα είναι σχεδιασμένα να δώσουν στους φοιτητές μια ισχυρή βάση στις βασικές θεωρίες και πρακτικές που είναι απαραίτητες για περαιτέρω σπουδές και επαγγελματική ανάπτυξη στον τομέα της πληροφορικής.

12.



You

Πώς ορίζονται στο πρόγραμμα τα «Μαθήματα Περιορισμένης Επιλογής» και τα «Μαθήματα Ελεύθερης Επιλογής»;



MyGPT University of Cyprus

Στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής του Πανεπιστημίου Κύπρου, τα "Μαθήματα Περιορισμένης Επιλογής" και τα "Μαθήματα Ελεύθερης Επιλογής" ορίζονται ως εξής:

1. Μαθήματα Περιορισμένης Επιλογής: Αυτά τα μαθήματα προσφέρονται από το Τμήμα και επιτρέπουν στον φοιτητή να εστιάσει σε μια συγκεκριμένη ειδίκευση της Πληροφορικής ή να αποκτήσει γνώσεις που καλύπτουν ένα ευρύτερο επιστημονικό φάσμα. Τα μαθήματα αυτά επιλέγονται από το φοιτητή με βάση τα ενδιαφέροντά του και τους επαγγελματικούς του στόχους και μπορεί να περιλαμβάνουν πιο εξειδικευμένα θέματα όπως δίκτυα, ασφάλεια, ή αλγορίθμους.

2. Μαθήματα Ελεύθερης Επιλογής: Αυτά τα μαθήματα προσφέρονται από άλλα Τμήματα του Πανεπιστημίου και επιτρέπουν στους φοιτητές να εξερευνήσουν γνωστικά πεδία εκτός του κύριου τομέα σπουδών τους. Αυτή η ευελιξία δίνει τη δυνατότητα στους φοιτητές να αποκτήσουν πιο ολοκληρωμένη παιδεία και να συνδυάσουν διαφορετικές γνώσεις με τις σπουδές τους στην πληροφορική.

Και οι δύο κατηγορίες μαθημάτων σχεδιάζονται για να προσφέρουν μια ισορροπημένη εκπαιδευτική εμπειρία που ενισχύει τόσο την ειδίκευση όσο και την πολυμάθεια, εξασφαλίζοντας ότι οι φοιτητές είναι καλά προετοιμασμένοι για επαγγελματική ανάπτυξη και προσωπική ανάπτυξη **【5*source】** .



You

Ποιες ειδικότητες έχουν στη διάθεσή τους οι φοιτητές από το ακαδημαϊκό έτος 2016/2017;



MyGPT University of Cyprus

Από το ακαδημαϊκό έτος 2016/2017, το Τμήμα Πληροφορικής του Πανεπιστημίου Κύπρου προσφέρει στους φοιτητές του τις παρακάτω εξειδικεύσεις:

- 1. Δίκτυα Υπολογιστών:** Εστιάζει στα βασικά και προχωρημένα θέματα δικτύωσης, ασφάλεια δικτύων και τεχνολογίες Διαδικτύου.
- 2. Θεμελιώσεις της Πληροφορικής:** Αναλύει τη λογική στην πληροφορική, βασικές έννοιες γλωσσών προγραμματισμού και παράλληλους αλγορίθμους.
- 3. Μεγάλα Δεδομένα και Διαδικτυακός Υπολογισμός:** Εξετάζει τις προχωρημένες βάσεις δεδομένων εξόρυξη δεδομένων στον Παγκόσμιο Ιστό και υπολογιστικά κέντρα δεδομένων.
- 4. Τεχνητή Νοημοσύνη:** Περιλαμβάνει μαθήματα όπως μηχανική μάθηση, συστήματα υπολογιστικής νοημοσύνης και υπολογιστική όραση.
- 5. Τεχνολογία Λογισμικού:** Καλύπτει θέματα προχωρημένης τεχνολογίας λογισμικού, επαναχρησιμοποίηση λογισμικού και επαγγελματική πρακτική.
- 6. Υπολογισμός Πραγματικού Κόσμου:** Επικεντρώνεται σε γραφικά υπολογιστών, αλληλεπίδραση ανθρώπου υπολογιστή και συστήματα υπολογιστικής νοημοσύνης.
- 7. Ψηφιακά και Ενσωματωμένα Συστήματα:** Προσφέρει γνώσεις σχετικά με την αρχιτεκτονική υπολογιστών, προγραμματισμό συστημάτων και διαδίκτυο των πραγμάτων.

Κάθε εξειδίκευση σχετίζεται με ένα σύνολο μαθημάτων και οι φοιτητές έχουν την ευκαιρία να επικεντρωθούν, μέσω των περιορισμένων επιλογών που θα επιλέξουν και μέσω της ατομικής διπλωματικής τους εργασίας, σε μια συγκεκριμένη περιοχή της Επιστήμης της Πληροφορικής

14.



You

Πώς μπορεί ένας φοιτητής να επιτύχει μια εξειδίκευση σύμφωνα με τις οδηγίες του προγράμματος;



MyGPT University of Cyprus

Για να επιτύχει ένας φοιτητής μια εξειδίκευση στο προπτυχιακό πρόγραμμα πληροφορικής του Πανεπιστημίου Κύπρου, πρέπει να ακολουθήσει τα ακόλουθα βήματα:

- 1. Επιλογή Μαθημάτων:** Ο φοιτητής πρέπει να επιλέξει και να παρακολουθήσει τουλάχιστον τρία μαθήματα περιορισμένης επιλογής από τα μαθήματα που ανήκουν στην εξειδίκευση της επιλογής του.
- 2. Ατομική Διπλωματική Εργασία:** Ο φοιτητής πρέπει να εκπονήσει την Ατομική Διπλωματική Εργασία του σε θέμα που σχετίζεται με την εξειδίκευση, υπό την επίβλεψη ενός μέλους του ακαδημαϊκού προσωπικού του Τμήματος.
- 3. Υποβολή Αίτησης για Αναγνώριση Εξειδίκευσης:** Μετά τις προσθαφαίρεσεις του τελευταίου εξαμήνου σπουδών του, ο φοιτητής πρέπει να υποβάλει αίτηση για την αναγνώριση της εξειδίκευσης. Αυτό γίνεται με την υποβολή ειδικού εντύπου στην Επιτροπή Προπτυχιακών Σπουδών του Τμήματος.
- 4. Εγκρίσεις και Αναγνώριση:** Η εξειδίκευση θα αναγνωριστεί επίσημα με την έγκριση της Επιτροπής και θα καταγραφεί στην αναλυτική βαθμολογία του φοιτητή, αναδεικνύοντας έτσι την εξειδίκευσή του σε έναν συγκεκριμένο τομέα της Πληροφορικής.

Αυτή η διαδικασία εξασφαλίζει ότι οι φοιτητές αποκτούν βαθιά γνώση και ειδικευμένες δεξιότητες σε επιλεγμένους τομείς της πληροφορικής, ενώ παράλληλα διατηρείται η ευελιξία για προσαρμογή των σπουδών στα ενδιαφέροντα και τις ανάγκες τους [5*source] .

15

A

You

Ποιες προϋποθέσεις πρέπει να πληροί ένα μάθημα για να αναγνωριστεί ως εξειδίκευση;



MyGPT University of Cyprus

Για να αναγνωριστεί ένα μάθημα ως μέρος μιας εξειδίκευσης στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής στο Πανεπιστήμιο Κύπρου, το μάθημα πρέπει να πληροί τα εξής κριτήρια:

- 1. Σχέση με την Εξειδίκευση:** Το μάθημα πρέπει να είναι ένα από τα ορισμένα μαθήματα που συνδέονται ειδικά με την εξειδίκευση που ο φοιτητής επιθυμεί να ακολουθήσει. Το Τμήμα Πληροφορικής ορίζει ένα συγκεκριμένο σύνολο μαθημάτων για κάθε εξειδίκευση, τα οποία καλύπτουν συγκεκριμένους τομείς όπως Δίκτυα Υπολογιστών, Τεχνητή Νοημοσύνη, κ.ά.
- 2. Αριθμός Μαθημάτων:** Ο φοιτητής πρέπει να έχει επιλέξει και να έχει επιτύχει τουλάχιστον τρία μαθήματα περιορισμένης επιλογής από αυτά που ανήκουν στην επιθυμητή εξειδίκευση.
- 3. Ατομική Διπλωματική Εργασία:** Ο φοιτητής πρέπει να έχει εκπονήσει την Ατομική Διπλωματική Εργασία του σε θέμα που σχετίζεται με την εξειδίκευση υπό την επίβλεψη κατάλληλου μέλους του ακαδημαϊκού προσωπικού.
- 4. Υποβολή Αίτησης:** Μετά την ολοκλήρωση των παραπάνω, ο φοιτητής πρέπει να υποβάλει αίτηση για την αναγνώριση της εξειδίκευσης στην Επιτροπή Προπτυχιακών Σπουδών του Τμήματος. Η αίτηση αυτή πρέπει να γίνει με ειδικό έντυπο και να υποβληθεί μετά τις προσθαφαιρέσεις του τελευταίου εξαμήνου σπουδών.
- 5. Έγκρισεις:** Η εξειδίκευση θα αναγνωριστεί επίσημα μόλις λάβει την έγκριση από την Επιτροπή Προπτυχιακών Σπουδών και θα καταγραφεί στην αναλυτική βαθμολογία του φοιτητή.

Αυτά τα βήματα διασφαλίζουν ότι οι φοιτητές που επιθυμούν να εξειδικευτούν σε κάποιον τομέα της πληροφορικής έχουν ολοκληρώσει επαρκώς τα απαιτούμενα μαθήματα και την έρευνα σχετικά με τον επιλεγμένο τομέα τους **【5*source】** .

16

A

You

Περιγράψτε την εξάρτηση μεταξύ των μαθημάτων όπως περιγράφεται στο πρόγραμμα.



MyGPT University of Cyprus

Στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής του Πανεπιστημίου Κύπρου, η εξάρτηση μεταξύ των μαθημάτων καθορίζεται με τη μορφή προαπαιτούμενων. Αυτό σημαίνει ότι ορισμένα μαθήματα πρέπει να έχουν ολοκληρωθεί επιτυχώς πριν από την εγγραφή σε άλλα μαθήματα που απαιτούν αυτήν την προηγούμενη γνώση.

Αυτή η εξαρτημένη δομή είναι σχεδιασμένη για να διασφαλίζει ότι οι φοιτητές αναπτύσσουν την απαιτούμενη γνώση και δεξιότητες σταδιακά, και να είναι καλά προετοιμασμένοι για πιο προχωρημένα ή εξειδικευμένα μαθήματα. Για παράδειγμα, ένα μάθημα στην οργάνωση υπολογιστών μπορεί να απαιτεί την επιτυχή ολοκλήρωση ενός εισαγωγικού μαθήματος στα ψηφιακά συστήματα.

Η εξάρτηση μεταξύ των μαθημάτων ενδέχεται να εκφράζεται και στο πλαίσιο των κανόνων προπτυχιακών σπουδών, όπου οι φοιτητές πρέπει να ολοκληρώνουν ορισμένα μαθήματα πριν προχωρήσουν σε επόμενα εξάμηνα ή πριν ασχοληθούν με τη διπλωματική τους εργασία.

Οι εξαρτήσεις ανάμεσα στα μαθήματα, όπως αυτές ορίζονται στον πρόγραμμα σπουδών, παρουσιάζονται συχνά σε μορφή πίνακα, για να βοηθήσουν τους φοιτητές να κατανοήσουν ποια μαθήματα πρέπει να ολοκληρώσουν και πότε. Αυτό διευκολύνει την ακαδημαϊκή προγραμματισμό και βοηθά τους φοιτητές να διαμορφώσουν το δικό τους εκπαιδευτικό σχέδιο κατά τη διάρκεια των σπουδών τους **【5*source】** .

17



MyGPT University of Cyprus

Τα βασικά υποχρεωτικά μαθήματα που πρέπει να παρακολουθήσουν όλοι οι φοιτητές της Πληροφορικής στο Πανεπιστήμιο Κύπρου, ως μέρος του προπτυχιακού τους προγράμματος, περιλαμβάνουν τα εξής:

1. **Διακριτές Δομές στην Πληροφορική και Υπολογισμό:** Εισαγωγή στα διακριτά μαθηματικά και την εφαρμογή τους στην πληροφορική.
2. **Αρχές Προγραμματισμού:** Βασικές αρχές του προγραμματισμού, χρησιμοποιώντας κάποια προγραμματιστική γλώσσα.
3. **Απειροστικός Λογισμός I και II:** Θεωρία και πρακτικές εφαρμογές του απειροστικού λογισμού.
4. **Γραμμική Άλγεβρα:** Βασικές έννοιες και μέθοδοι της γραμμικής άλγεβρας.
5. **Ψηφιακά Συστήματα:** Εισαγωγή στην οργάνωση και την αρχιτεκτονική των ψηφιακών συστημάτων.
6. **Αντικειμενοστρεφής Προγραμματισμός:** Επέκταση των αρχών του προγραμματισμού με έμφαση στην αντικειμενοστρεφή προγραμματιστική παράδοση.
7. **Οργάνωση Υπολογιστών:** Μελέτη της δομής και της λειτουργίας των σύγχρονων υπολογιστών.
8. **Δομές Δεδομένων και Αλγόριθμοι:** Ανάπτυξη, ανάλυση και εφαρμογή δομών δεδομένων και αλγορίθμων.
9. **Λειτουργικά Συστήματα:** Σχεδίαση και αρχιτεκτονική λειτουργικών συστημάτων.
10. **Βάσεις Δεδομένων και Τεχνολογία Λογισμικού:** Εισαγωγή στις βάσεις δεδομένων και στην ανάπτυξη λογισμικού.

Αυτά τα μαθήματα συνθέτουν τη βάση του ακαδημαϊκού προγράμματος και παρέχουν στους φοιτητές τις απαραίτητες γνώσεις και δεξιότητες για να κατανοήσουν πιο σύνθετες έννοιες και να εφαρμόσουν τις τεχνολογίες και μεθοδολογίες που μαθαίνουν σε πρακτικά προβλήματα και σενάρια.



You

Ποια είναι τα κύρια χαρακτηριστικά της συνιστώσας της Μεμονωμένης Διατριβής;



MyGPT University of Cyprus

Η συνιστώσα της Μεμονωμένης Διατριβής (Ατομική Διπλωματική Εργασία) στο προπτυχιακό πρόγραμμα σπουδών πληροφορικής στο Πανεπιστήμιο Κύπρου περιλαμβάνει τα ακόλουθα κύρια χαρακτηριστικά:

- 1. Αυτονομία:** Ο φοιτητής αναλαμβάνει την ευθύνη να επιλέξει ένα θέμα έρευνας, να κάνει μια ανεξάρτητη έρευνα και να αναπτύξει τις αντίστοιχες λύσεις ή προτάσεις.
- 2. Επίβλεψη:** Η διατριβή πρέπει να γίνεται υπό την επίβλεψη ενός μέλους του ακαδημαϊκού προσωπικού, το οποίο παρέχει καθοδήγηση και στήριξη καθ' όλη τη διάρκεια της εργασίας.
- 3. Έρευνα:** Η διπλωματική εργασία απαιτεί από τον φοιτητή να πραγματοποιήσει ανεξάρτητη έρευνα σε ένα επιστημονικό ή τεχνολογικό θέμα. Ο φοιτητής πρέπει να συλλέξει δεδομένα, να πραγματοποιήσει αναλύσεις και να ερμηνεύσει τα αποτελέσματα.
- 4. Γραπτή Αναφορά:** Το αποτέλεσμα της έρευνας και η συνολική εργασία πρέπει να καταγραφούν σε μια λεπτομερή γραπτή αναφορά, η οποία περιλαμβάνει μεθοδολογία, ανάλυση δεδομένων, συζήτηση και συμπεράσματα.
- 5. Παρουσίαση:** Ο φοιτητής πρέπει να παρουσιάσει τα ευρήματα και την ανάλυσή του στο τέλος της εργασίας σε ένα ευρύτερο κοινό, που μπορεί να περιλαμβάνει μέλη του τμήματος και άλλους φοιτητές.
- 6. Αξιολόγηση:** Η διατριβή αξιολογείται βάσει της ποιότητας της έρευνας, της πληρότητας της γραπτής αναφοράς, της ικανότητας του φοιτητή να υπερασπιστεί τα ευρήματα και της συνολικής παρουσίας των αποτελεσμάτων.
- 7. Ψηφιακή Καταχώριση:** Οι ολοκληρωμένες διπλωματικές εργασίες καταχωρούνται στην ψηφιακή βιβλιοθήκη ανοιχτής πρόσβασης του τμήματος.

Αυτή η διαδικασία επιτρέπει στους φοιτητές να εξασκήσουν τις ικανότητές τους στην ανεξάρτητη έρευνα, να αναπτύξουν βαθιά κατανόηση σε ένα συγκεκριμένο θέμα και να βελτιώσουν τις δεξιότητες τους στην επικοινωνία και στην παρουσίαση των εργασιών τους.

19.



You

Πώς είναι δομημένα τα μαθήματα επιλογής σε διαφορετικά εξάμηνα;



MyGPT University of Cyprus

Τα μαθήματα επιλογής στο Προπτυχιακό Πρόγραμμα Σπουδών του Πανεπιστημίου Κύπρου κατηγοριοποιούνται σε διάφορες περιοχές, με κάθε μία να στοχεύει σε συγκεκριμένες δεξιότητες και γνώσεις. Τα μαθήματα διαχωρίζονται σε:

- 1. Υποχρεωτικά Μαθήματα:** Αυτά είναι βασικά μαθήματα που πρέπει να παρακολουθήσουν όλοι οι φοιτητές.
- 2. Μαθήματα Περιορισμένης Επιλογής:** Προσφέρονται από το τμήμα και δίνουν την ευκαιρία στους φοιτητές να εστιάσουν σε συγκεκριμένες εξειδικεύσεις, όπως Ψηφιακά και Ενσωματωμένα Συστήματα, Τεχνητή Νοημοσύνη, κ.ά.
- 3. Μαθήματα Ελεύθερης Επιλογής:** Αυτά τα μαθήματα επιλέγονται από άλλα τμήματα και επιτρέπουν στους φοιτητές να εξερευνήσουν γνώσεις εκτός του κυρίως πεδίου των σπουδών τους.

Η δομή και η κατανομή των μαθημάτων επιλογής διαφοροποιούνται ανάλογα με το εξάμηνο και την ειδίκευση. Για παράδειγμα, στα τελευταία εξάμηνα (όπως το 7ο και 8ο) οι φοιτητές επικεντρώνονται περισσότερο στις Περιορισμένες Επιλογές και στην Ατομική Διπλωματική Εργασία, ενώ στα πρώτα εξάμηνα τονίζεται περισσότερο η θεωρητική βάση και οι γενικές γνώσεις.

20.



You

Ποια μαθήματα μαθηματικών είναι υποχρεωτικά για τους φοιτητές της Πληροφορικής;



MyGPT University of Cyprus

Για τους φοιτητές της Πληροφορικής, τα υποχρεωτικά μαθήματα μαθηματικών περιλαμβάνουν τα εξής:

1. Απειροστικός Λογισμός I (ΜΑΣ 012) - 5 ECTS, προσφέρεται στο πρώτο εξάμηνο.
2. Απειροστικός Λογισμός II (ΜΑΣ 013) - 5 ECTS, προσφέρεται στο δεύτερο εξάμηνο.
3. Γραμμική Άλγεβρα (ΜΑΣ 029) - 5 ECTS, προσφέρεται επίσης στο δεύτερο εξάμηνο.
4. Πιθανότητες και Στατιστική (ΜΑΣ 055) - 7 ECTS, προσφέρεται στο τρίτο εξάμηνο.

Αυτά τα μαθήματα είναι απαραίτητα για την ολοκλήρωση των σπουδών στο προπτυχιακό πρόγραμμα Πληροφορικής και συνιστούν τη μαθηματική βάση που χρειάζεται ένας φοιτητής για την επιστήμη της πληροφορικής.

5.3 Αξιολόγηση έτοιμης εφαρμογής GPTs

1



You

Ποιος είναι ο ελάχιστος αριθμός ECTS που απαιτείται για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών;



PDF Ai PDF

Για την απόκτηση πτυχίου στην Επιστήμη των Υπολογιστών απαιτούνται τουλάχιστον 240 μονάδες ECTS.

2



You

Ποιοι είναι οι γενικοί στόχοι του Προπτυχιακού Προγράμματος Πληροφορικής;



PDF Ai PDF

Οι γενικοί στόχοι του Προπτυχιακού Προγράμματος Πληροφορικής περιλαμβάνουν την προετοιμασία των αποφοίτων για σταδιοδρομία σε θέσεις υψηλής υπευθυνότητας στον επαγγελματικό ή ακαδημαϊκό χώρο, όπου θα μπορούν να προωθήσουν αποτελεσματικά την ανάπτυξη και εφαρμογή νέων μεθόδων και ιδεών. Επίσης, το πρόγραμμα στοχεύει να καλλιεργήσει στους φοιτητές την επιθυμία για δια βίου μάθηση και επαγγελματική ανάπτυξη, ωριμότητα, ανεξαρτησία πνεύματος και δυνατότητα κριτικής σκέψης.

3



You

Ποια συγκεκριμένα επαγγελματικά αποτελέσματα στοχεύει το Τμήμα Επιστήμης Υπολογιστών να προετοιμάσει τους αποφοίτους του;



PDF Ai PDF

Το Τμήμα Επιστήμης Υπολογιστών στοχεύει να προετοιμάσει τους αποφοίτους του για διάφορα επαγγελματικά αποτελέσματα. Οι απόφοιτοι θα μπορούν να εργοδοτηθούν από την εγχώρια αγορά ως επιστήμονες περί την Τεχνολογία Πληροφοριών, να συνεχίσουν τις σπουδές τους για την απόκτηση μεταπτυχιακού διπλώματος, να εξειδικευθούν σε συγκεκριμένους τομείς έρευνας σε κάποιο ερευνητικό κέντρο, και είναι πιθανό να διδάξουν Πληροφορική στη δευτεροβάθμια εκπαίδευση μετά από επιτυχή παρακολούθηση προγράμματος παιδαγωγικής κατάρτισης.

4



You

Πώς διατηρεί το Τμήμα Πληροφορικής τη σύνδεσή του με την τοπική βιομηχανία;



PDF Ai PDF

Το Τμήμα Πληροφορικής διατηρεί τη σύνδεσή του με την τοπική βιομηχανία μέσω της συνεχούς επικοινωνίας και συνεργασίας. Αποδίδει ιδιαίτερη σημασία στην ενίσχυση αυτής της σχέσης με στόχο να δημιουργήσει δίαυλο συνεργασίας με τον Κυπριακό χώρο, μέσω των αποφοίτων του. Αυτό βοηθά στην αποτελεσματική προώθηση νέων μεθόδων και ιδεών στον επαγγελματικό τομέα.

5



You

Ποια είναι η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής;



PDF Ai PDF

Η αναμενόμενη διάρκεια ολοκλήρωσης του Προπτυχιακού Προγράμματος Πληροφορικής είναι οκτώ εξάμηνα. Ωστόσο, αυτή η διάρκεια μπορεί να παραταθεί μέχρι τα δώδεκα εξάμηνα αν χρειαστεί.

6



You

Ποιους τομείς σπουδών καλύπτει το Προπτυχιακό Πρόγραμμα Πληροφορικής;



PDF Ai PDF

Το Προπτυχιακό Πρόγραμμα Σπουδών στην Πληροφορική καλύπτει τους ακόλουθους τομείς σπουδών:

- 1. Θεωρία:** Περιλαμβάνει την καλλιέργεια τυπικού τρόπου σκέψης, οργάνωσης και επεξεργασίας πληροφοριών, τη σχεδίαση και ανάλυση αλγορίθμων, και τη Μαθηματική Λογική ως Λογισμό της Πληροφορικής.
- 2. Υπολογιστικά Συστήματα:** Αφορά την οργάνωση και αρχιτεκτονική υπολογιστών, τα λειτουργικά συστήματα, τη σχεδίαση και υλοποίηση γλωσσών προγραμματισμού, και τις έννοιες των παράλληλων και ενσωματωμένων συστημάτων.
- 3. Επίλυση Προβλημάτων:** Στοχεύει στην ανάπτυξη αλγοριθμικής σκέψης, προγραμματισμού, και την εξοικείωση με διάφορες γλώσσες προγραμματισμού.
- 4. Εφαρμογές:** Συνδυάζει γνώσεις και ικανότητες από τους άλλους τομείς με σκοπό την ανάπτυξη εφαρμογών για την επίλυση ρεαλιστικών προβλημάτων, καλύπτοντας τεχνολογικά κατασκευάσματα όπως βάσεις δεδομένων, γραφικά, και συστήματα διασύνδεσης χρήστη-μηχανής.

**You**

Ποια θέματα περιλαμβάνονται στην περιοχή "Θεωρία" του προγράμματος;

**PDF Ai PDF**

Στην περιοχή "Θεωρία" του Προπτυχιακού Προγράμματος Πληροφορικής περιλαμβάνονται θέματα όπως:

- 1. Θεωρία και μοντέλα υπολογισμού:** Εξετάζει τις βάσεις των μοντέλων υπολογισμού και τις θεωρητικές πτυχές της πληροφορικής.
- 2. Σχεδίαση και ανάλυση αλγορίθμων:** Εστιάζει στις μεθοδολογίες για τη δημιουργία αποτελεσματικών και αποδοτικών αλγορίθμων.
- 3. Μαθηματική Λογική:** Αναλύει το ρόλο της μαθηματικής λογικής ως λογισμός στην πληροφορική.
- 4. Διακριτά Μαθηματικά:** Παρέχει τη θεωρητική βάση για την κατανόηση των στοιχειωδών δομών και αλγορίθμων χρησιμοποιώντας διακριτά μαθηματικά.

Αυτά τα μαθήματα βοηθούν τους φοιτητές να αναπτύξουν μια ικανότητα αφαιρετικότητας και τυπικού διαλογισμού, καθώς και να αποκτήσουν χρήσιμες μαθηματικές ικανότητες.

**You**

Ποια είναι τα βασικά μαθήματα που διδάσκονται στα «Υπολογιστικά Συστήματα»;

**PDF Ai PDF**

Στην περιοχή των "Υπολογιστικών Συστημάτων" στο Προπτυχιακό Πρόγραμμα Πληροφορικής διδάσκονται τα εξής βασικά μαθήματα:

- 1. Οργάνωση και Αρχιτεκτονική Υπολογιστών:** Μελετά τις βασικές αρχές οργάνωσης και την αρχιτεκτονική των υπολογιστών.
- 2. Λειτουργικά Συστήματα:** Εξερευνά τις βασικές λειτουργίες και τη διαχείριση των λειτουργικών συστημάτων.
- 3. Σχεδίαση και Υλοποίηση Γλωσσών Προγραμματισμού:** Αναφέρεται στη δημιουργία και ανάπτυξη γλωσσών προγραμματισμού.
- 4. Συστήματα Μικροεπεξεργαστών:** Καλύπτει τα θέματα που σχετίζονται με την αρχιτεκτονική και τη λειτουργία των μικροεπεξεργαστών.
- 5. Μεταβίβαση Δεδομένων:** Συζητά τις μεθόδους και την τεχνολογία για την αποτελεσματική μεταφορά δεδομένων.
- 6. Δίκτυα, Κατανεμημένα Συστήματα, και Νέες Αρχιτεκτονικές:** Εστιάζει στην κατανόηση των δικτύων, των κατανεμημένων συστημάτων, καθώς και σε προηγμένες αρχιτεκτονικές.

Αυτά τα μαθήματα παρέχουν στους φοιτητές μια βαθιά κατανόηση της υλικοτεχνικής και λογισμικού συστημάτων, κρίσιμες γνώσεις για την ανάπτυξη και διαχείριση σύγχρονων υπολογιστικών συστημάτων.



You

Σε τι επικεντρώνεται η περιοχή "Επίλυση προβλημάτων" στο πρόγραμμα σπουδών της Επιστήμης των Υπολογιστών;



PDF Ai PDF

Η περιοχή "Επίλυση Προβλημάτων" στο πρόγραμμα σπουδών της Επιστήμης των Υπολογιστών επικεντρώνεται στην ανάπτυξη αλγοριθμικής σκέψης και δίνει έμφαση στις αρχές προγραμματισμού και τη σχεδίαση αλγορίθμων. Σκοπός είναι να αποκτήσουν οι φοιτητές ικανότητα στη χρήση διαφόρων γλωσσών προγραμματισμού και να μάθουν διάφορα πρότυπα προγραμματισμού, όπως διαδικασιακό ή προστακτικό, αντικειμενοστρεφή και λογικό προγραμματισμό.

Επιπλέον, τα μαθήματα επίλυσης αυτής της περιοχής καλύπτουν προχωρημένες τεχνικές επίλυσης προβλημάτων βασισμένες στον παραλληλισμό και τον ταυτοχρονισμό. Αυτή η περιοχή διευκολύνει την κατανόηση των τεχνικών που απαιτούνται για τη σχεδίαση, υλοποίηση και αξιολόγηση λύσεων σχετικά μικρών αλλά σημαντικών προβλημάτων, καθώς και για την επίλυση ρεαλιστικών προβλημάτων.

10



You

Πώς η περιοχή "Εφαρμογές" ενσωματώνει γνώσεις από άλλους τομείς;



PDF Ai PDF

Η περιοχή "Εφαρμογές" στο Προπτυχιακό Πρόγραμμα Πληροφορικής σκοπεύει να ενσωματώσει και να εφαρμόσει τις γνώσεις και τις ικανότητες που οι φοιτητές έχουν αποκτήσει από άλλους τομείς σπουδών, με στόχο την ανάπτυξη χρήσιμων εφαρμογών για την επίλυση ρεαλιστικών προβλημάτων. Η περιοχή καλύπτει σημαντικά τεχνολογικά κατασκευάσματα όπως βάσεις δεδομένων, γνώσεων, γραφικά και συστήματα διασύνδεσης χρήστη-μηχανής, παρέχοντας ταυτόχρονα εργαλεία για την ανάπτυξη εφαρμογών υψηλότερου επιπέδου.

Επίσης, εξετάζονται νέες μεθοδολογίες τεχνολογίας λογισμικού που καλύπτουν όλα τα στάδια σχεδίασης, ανάπτυξης και διατήρησης υψηλής ποιότητας εφαρμογών. Αυτές οι μεθοδολογίες επιτρέπουν την αξιολόγηση των παραγόντων που είναι σημαντικοί για την επιτυχή έκβαση ενός έργου, χρησιμοποιώντας ως παραδείγματα τρέχουσες εφαρμογές. Μέσα από αυτή τη συνεχή αλληλεπίδραση και ολοκλήρωση της θεωρητικής γνώσης με την πρακτική εφαρμογή, η περιοχή των Εφαρμογών βοηθά τους φοιτητές να ενσωματώνουν αποτελεσματικά την ευρεία γκάμα δεξιοτήτων που έχουν αποκτήσει σε συγκεκριμένα επαγγελματικά αποτελέσματα.

11



You

Τι είδους μαθήματα κατατάσσει το πρόγραμμα στα «Υποχρεωτικά Μαθήματα»;



PDF Ai PDF

Στην κατηγορία των "Υποχρεωτικών Μαθημάτων" στο Προπτυχιακό Πρόγραμμα Πληροφορικής περιλαμβάνονται μαθήματα που οι φοιτητές πρέπει να παρακολουθήσουν υποχρεωτικά για να ολοκληρώσουν τις σπουδές τους στην Πληροφορική. Αυτά τα μαθήματα καλύπτουν βασικές γνώσεις και δεξιότητες που είναι θεμελιώδεις για την επιστήμη των υπολογιστών και περιλαμβάνουν:

- 1. Μαθήματα προγραμματισμού:** Παρέχουν τις βασικές αρχές προγραμματισμού και αντικειμενοστρεφή προγραμματισμό.
- 2. Υπολογιστικά συστήματα:** Συμπεριλαμβάνουν μαθήματα όπως οργάνωση υπολογιστών και λειτουργικά συστήματα.
- 3. Δομές δεδομένων και αλγόριθμοι:** Εστιάζουν στις δομές δεδομένων και τις τεχνικές για αποτελεσματική επεξεργασία και ανάλυση δεδομένων.
- 4. Μαθηματικά και στατιστική:** Περιλαμβάνουν μαθήματα όπως απειροστικός λογισμός, γραμμική άλγεβρα, πιθανότητες και στατιστική, ουσιαστικά για την κατανόηση και ανάλυση συστημάτων και αλγορίθμων.
- 5. Εξειδικευμένα τεχνικά μαθήματα:** Όπως θεωρία υπολογισμού, ασφάλεια συστημάτων, επικοινωνίες και δίκτυα.

Αυτά τα μαθήματα αποτελούν τον κορμό της  τριταξίαιδους στην Πληροφορική και είναι κρίσιμα για την απόκτηση μιας ολοκληρωμένης και ευρύτερης κατανόησης του αντικειμένου **【4^{source}】**.

12



You

Πώς ορίζονται στο πρόγραμμα τα «Μαθήματα Περιορισμένης Επιλογής» και τα «Μαθήματα Ελεύθερης Επιλογής»;



PDF ΑΙ PDF

Στο πρόγραμμα σπουδών της Πληροφορικής, τα **Μαθήματα Περιορισμένης Επιλογής** και τα **Μαθήματα Ελεύθερης Επιλογής** έχουν τις εξής ορισμούς:

- 1. Μαθήματα Περιορισμένης Επιλογής:** Αυτά τα μαθήματα είναι μέρος του προπτυχιακού προγράμματος και προσφέρονται από το Τμήμα Πληροφορικής. Οι φοιτητές επιλέγουν μαθήματα από αυτή την κατηγορία για να εστιάσουν σε μια συγκεκριμένη ειδίκευση ή για να αποκτήσουν ευρύτερες γνώσεις στον τομέα της Πληροφορικής. Αυτά τα μαθήματα επιτρέπουν στους φοιτητές να προσαρμόσουν το πρόγραμμά τους ανάλογα με τα επαγγελματικά τους σχέδια και τα ενδιαφέροντά τους.
- 2. Μαθήματα Ελεύθερης Επιλογής:** Αυτά τα μαθήματα προσφέρονται από άλλα τμήματα του πανεπιστημίου και δίνουν τη δυνατότητα στους φοιτητές να εξερευνήσουν γνωστικά πεδία πέρα από την Πληροφορική. Οι φοιτητές συμβουλευόμενοι τον ακαδημαϊκό τους σύμβουλο, επιλέγουν αυτά τα μαθήματα με βάση τα προσωπικά τους ενδιαφέροντα και στόχους. Αυτή η ευελιξία βοηθά στην καλλιέργεια μιας πιο ολοκληρωμένης παιδείας και στην ανάπτυξη δεξιοτήτων που ωφελούν την επαγγελματική και προσωπική ανάπτυξη των φοιτητών.

Και οι δύο κατηγορίες μαθημάτων είναι σημαντικές για την παροχή μιας ισορροπημένης και πλήρους εκπαίδευσης στους φοιτητές, προσφέροντας τόσο εξειδίκευση σε συγκεκριμένα επιστημονικά πεδία όσο και ευκαιρίες για διαθεματικές σπουδές **【4*source】**.

13



You

Ποιες ειδικότητες έχουν στη διάθεσή τους οι φοιτητές από το ακαδημαϊκό έτος 2016/2017;



PDF ΑΙ PDF

Από το ακαδημαϊκό έτος 2016/2017, το Τμήμα Πληροφορικής προσφέρει στους φοιτητές του τις ακόλουθες εξειδικεύσεις:

- 1. Δίκτυα Υπολογιστών:** Εστιάζει σε θέματα σχεδίασης, ασφάλειας και διαχείρισης δικτυακών συστημάτων και τεχνολογιών.
- 2. Θεμελιώσεις της Πληροφορικής:** Καλύπτει βασικές έννοιες και θεωρητικά θέματα στην Πληροφορική, όπως αλγόριθμοι και λογισμός.
- 3. Μεγάλα Δεδομένα και Διαδικτυακός Υπολογισμός:** Εστιάζει στην ανάλυση, διαχείριση και επεξεργασία μεγάλων όγκων δεδομένων και τεχνολογίες στο διαδίκτυο.
- 4. Τεχνητή Νοημοσύνη:** Αναπτύσσει γνώσεις σε συστήματα που επιδεικνύουν έξυπνη συμπεριφορά και αυτοματισμούς.
- 5. Τεχνολογία Λογισμικού:** Καλύπτει την ανάπτυξη, σχεδίαση και διατήρηση λογισμικού.
- 6. Υπολογισμός Πραγματικού Κόσμου:** Εστιάζει στην εφαρμογή τεχνολογιών υπολογιστών σε πρακτικά προβλήματα της καθημερινότητας.
- 7. Ψηφιακά και Ενσωματωμένα Συστήματα:** Περιλαμβάνει την ανάπτυξη και τη διαχείριση ψηφιακών και ενσωματωμένων συστημάτων.

Κάθε εξειδίκευση προσφέρει μια σειρά από συγκεκριμένα μαθήματα που επιτρέπουν στους φοιτητές να βαθύνουν τις γνώσεις τους στον επιλεγμένο τομέα.

14



You

Πώς μπορεί ένας φοιτητής να επιτύχει μια εξειδίκευση σύμφωνα με τις οδηγίες του προγράμματος;



PDF Ai PDF

Για να επιτύχει ένας φοιτητής μια εξειδίκευση στο προπτυχιακό πρόγραμμα Πληροφορικής, πρέπει να ακολουθήσει τις εξής βήματα σύμφωνα με τις οδηγίες του τμήματος:

1. **Επιλογή Μαθημάτων Περιορισμένης Επιλογής:** Ο φοιτητής πρέπει να επιλέξει τουλάχιστον τρία μαθήματα από το σύνολο των περιορισμένων επιλογών που ανήκουν στην επιθυμητή εξειδίκευση.
2. **Ατομική Διπλωματική Εργασία:** Ο φοιτητής πρέπει να εκπονήσει μια ατομική διπλωματική εργασία στον τομέα της εξειδίκευσης, κάτω από την επίβλεψη ενός μέλους του ακαδημαϊκού προσωπικού του τμήματος.
3. **Υποβολή Αίτησης:** Αιτήματα για την αναγνώριση της εξειδίκευσης υποβάλλονται με ειδικό έντυπο στην Επιτροπή Προπτυχιακών Σπουδών μετά τις προσθαφαιρέσεις του τελευταίου εξαμήνου σπουδών των φοιτητών.

Με την ολοκλήρωση αυτών των βημάτων και την επιτυχή παρακολούθηση των σχετικών μαθημάτων, η εξειδίκευση θα αναγράφεται στην αναλυτική βαθμολογία του φοιτητή.

15



You

Ποιες προϋποθέσεις πρέπει να πληροί ένα μάθημα για να αναγνωριστεί ως εξειδίκευση;



PDF Ai PDF

Για να αναγνωριστεί ένα μάθημα ως μέρος μιας εξειδίκευσης στο προπτυχιακό πρόγραμμα Πληροφορικής, πρέπει να πληροί τις ακόλουθες προϋποθέσεις:

1. **Συνάφεια με την Εξειδίκευση:** Το μάθημα πρέπει να είναι ενταγμένο στην επιλογή μαθημάτων που προσφέρονται ως μέρος της συγκεκριμένης εξειδίκευσης, όπως ορίζεται από το τμήμα.
2. **Εγκρίσεις:** Θα πρέπει να έχει λάβει την απαραίτητη έγκριση από την Επιτροπή Προπτυχιακών Σπουδών και να συμφωνεί με τις ακαδημαϊκές προδιαγραφές του προγράμματος.
3. **Περιεχόμενο του Μαθήματος:** Το μάθημα πρέπει να καλύπτει βαθιά τα θεμελιώδη θέματα και τις προηγμένες γνώσεις που απαιτούνται για την εξειδίκευση.
4. **Αποτελεσματική Συμβολή:** Πρέπει να συνεισφέρει στην ακαδημαϊκή και επαγγελματική ανάπτυξη των φοιτητών σχετικά με την επιλεγμένη εξειδίκευση.

Αυτές οι προϋποθέσεις εξασφαλίζουν ότι τα μαθήματα της εξειδίκευσης είναι αυστηρά επιλεγμένα για να ανταποκρίνονται στους ακαδημαϊκούς και επαγγελματικούς στόχους των φοιτητών, προσδίδοντας έτσι μεγάλη αξία στις σπουδές τους στο πλαίσιο της εξειδίκευσης.

16



You

Περιγράψτε την εξάρτηση μεταξύ των μαθημάτων όπως περιγράφεται στο πρόγραμμα.



PDF Ai PDF

Στο πρόγραμμα σπουδών Πληροφορικής, η εξάρτηση μεταξύ των μαθημάτων δομείται με τρόπο που ενισχύει τη σταδιακή κατανόηση και εξέλιξη των γνώσεων των φοιτητών. Οι βασικοί τρόποι εξάρτησης περιλαμβάνουν:

1. **Προαπαιτούμενα Μαθήματα:** Ορισμένα μαθήματα απαιτούν την επιτυχή ολοκλήρωση άλλων μαθημάτων πριν από την εγγραφή σε αυτά. Αυτό εξασφαλίζει ότι οι φοιτητές έχουν τις απαραίτητες βασικές γνώσεις για να κατανοήσουν πιο προχωρημένα θέματα.
2. **Συναφή Μαθήματα:** Μαθήματα που καλύπτουν συγγενείς θεματικές περιοχές συχνά συνδέονται μεταξύ τους, επιτρέποντας στους φοιτητές να εξελίσουν την κατανόησή τους μέσα από διαφορετικές προοπτικές.
3. **Αλληλεπιδράσεις Μαθημάτων:** Μαθήματα σε διαφορετικές εξειδικεύσεις μπορεί να ενσωματώνουν γνώσεις ή να αναφέρονται σε θέματα που καλύπτονται σε άλλα μαθήματα, προσφέροντας έναν πιο ολοκληρωμένο τρόπο εκπαίδευσης.
4. **Ενσωμάτωση Θεωρίας και Πρακτικής:** Πολλά μαθήματα περιλαμβάνουν θεωρητική διδασκαλία και πρακτικές εφαρμογές (π.χ., εργαστήρια, προγράμματα, έργα), επιτρέποντας στους φοιτητές να εφαρμόζουν τη θεωρία σε πραγματικά προβλήματα.

Αυτές οι εξαρτήσεις είναι καίριες για τη δημιουργία ενός συνεκτικού προγράμματος σπουδών που προετοιμάζει τους φοιτητές για τις προκλήσεις του επαγγελματικού και ακαδημαϊκού κόσμου στον τομέα της Πληροφορικής. **[4*source]** .

**You**

Ποια είναι τα βασικά μαθήματα που πρέπει να παρακολουθήσουν όλοι οι φοιτητές της Πληροφορικής;

**PDF Ai PDF**

Τα βασικά μαθήματα που πρέπει να παρακολουθήσουν όλοι οι φοιτητές της Πληροφορικής στο προπτυχιακό πρόγραμμα είναι:

- 1. Εισαγωγή στην Πληροφορική:** Παρέχει μια γενική εισαγωγή στις βασικές έννοιες και τις αρχές της πληροφορικής.
- 2. Προγραμματισμός I & II:** Καλύπτει τις βασικές τεχνικές και μεθοδολογίες προγραμματισμού, με έμφαση στον αντικειμενοστρεφή προγραμματισμό.
- 3. Δομές Δεδομένων και Αλγόριθμοι:** Εστιάζει στη σχεδίαση, ανάλυση και υλοποίηση αποτελεσματικών δομών δεδομένων και αλγορίθμων.
- 4. Βάσεις Δεδομένων I:** Διδάσκει τα βασικά στοιχεία και τις τεχνικές για τη σχεδίαση και χρήση βάσεων δεδομένων.
- 5. Θεωρία Υπολογισμού:** Αναλύει τη θεωρία που καθορίζει τι μπορεί και τι δεν μπορεί να υπολογιστεί.
- 6. Οργάνωση και Αρχιτεκτονική Υπολογιστών:** Εξετάζει την αρχιτεκτονική των συστημάτων υπολογιστών, εστιάζοντας στην επεξεργασία και την αποθήκευση πληροφοριών.
- 7. Λειτουργικά Συστήματα:** Διδάσκει τα θεμελιώδη στοιχεία των λειτουργικών συστημάτων, συμπεριλαμβανομένων της διαχείρισης πόρων και του πολυπρογραμματισμού.

Αυτά τα μαθήματα αποτελούν την καρδιά του προπτυχιακού προγράμματος στην Πληροφορική και παρέχουν στους φοιτητές τις απαραίτητες γνώσεις και δεξιότητες για την περαιτέρω εξειδίκευσή τους σε πιο συγκεκριμένους τομείς του αντικείμενου **[4^{source}]** .

**You**

Ποια είναι τα κύρια χαρακτηριστικά της συνιστώσας της Μεμονωμένης Διατριβής;

**PDF Ai PDF**

Η συνιστώσα της Μεμονωμένης Διατριβής στο πρόγραμμα σπουδών της Πληροφορικής έχει τα εξής κύρια χαρακτηριστικά:

- 1. Ερευνητική Εργασία:** Η Μεμονωμένη Διατριβή είναι μια εκτενής ερευνητική εργασία που απαιτεί από τον φοιτητή να διεξάγει αυτόνομη έρευνα σε ένα θέμα σχετικό με την Πληροφορική.
- 2. Επίβλεψη:** Οι φοιτητές δουλεύουν κάτω από την επίβλεψη ενός ή περισσότερων μελών του διδακτικού προσωπικού, οι οποίοι παρέχουν καθοδήγηση και υποστήριξη καθ' όλη τη διάρκεια της διαδικασίας.
- 3. Ανάπτυξη Δεξιοτήτων:** Η διαδικασία της διατριβής ενθαρρύνει τους φοιτητές να αναπτύξουν κρίσιμες σκέψεις, αναλυτικές ικανότητες, και επιστημονικές δεξιότητες συγγραφής.
- 4. Τελική Παρουσίαση:** Στο τέλος της διαδικασίας, οι φοιτητές συνήθως παρουσιάζουν τα ευρήματά τους σε μορφή παρουσίασης ενώπιον ενός πάνελ ακαδημαϊκών, παρέχοντας μια ευκαιρία για κριτική αξιολόγηση και συζήτηση.
- 5. Εξειδίκευση:** Η διατριβή δίνει την ευκαιρία στους φοιτητές να εξειδικευτούν σε έναν συγκεκριμένο τομέα της Πληροφορικής, συχνά συμβάλλοντας στην επαγγελματική τους πορεία ή στις περαιτέρω ακαδημαϊκές τους σπουδές.

Η Μεμονωμένη Διατριβή αποτελεί ένα σημαντικό μέρος του προπτυχιακού προγράμματος καθώς προσφέρει στους φοιτητές την ευκαιρία να αναλάβουν μια προχωρημένη, αυτόνομη έρευνα σε ένα θέμα της επιλογής τους, ενισχύοντας ταυτόχρονα τις ερευνητικές και αναλυτικές τους ικανότητες



You

Πώς είναι δομημένα τα μαθήματα επιλογής σε διαφορετικά εξάμηνα;



PDF Ai PDF

Τα μαθήματα επιλογής στο πρόγραμμα σπουδών της Πληροφορικής είναι δομημένα σε διαφορετικά εξάμηνα με τρόπο που παρέχει στους φοιτητές την ευκαιρία να εξειδικευτούν σε συγκεκριμένους τομείς της πληροφορικής και να επιλέξουν μαθήματα ανάλογα με τα ενδιαφέροντά τους και την ακαδημαϊκή τους πρόοδο. Η δομή αυτή συνήθως περιλαμβάνει:

- 1. Βασικά Μαθήματα:** Τα πρώτα εξάμηνα συνήθως επικεντρώνονται στα βασικά μαθήματα που προετοιμάζουν τους φοιτητές με τις απαραίτητες γνώσεις και δεξιότητες για πιο προχωρημένες σπουδές.
- 2. Μαθήματα Περιορισμένης Επιλογής:** Μετά τα βασικά μαθήματα, στα επόμενα εξάμηνα προσφέρονται μαθήματα περιορισμένης επιλογής που αφορούν ειδικούς τομείς της πληροφορικής. Αυτά τα μαθήματα επιτρέπουν στους φοιτητές να αρχίσουν να εξειδικεύονται σε συγκεκριμένες περιοχές ενδιαφέροντος.
- 3. Μαθήματα Ελεύθερης Επιλογής:** Στα ανώτερα εξάμηνα, τα μαθήματα ελεύθερης επιλογής προσφέρουν τη μεγαλύτερη ευελιξία, καθώς οι φοιτητές μπορούν να επιλέξουν από μια ευρεία γκάμα μαθημάτων που προσφέρονται από το τμήμα πληροφορικής ή άλλα σχετικά τμήματα.
- 4. Μεμονωμένη Διατριβή:** Συνήθως στα τελευταία εξάμηνα, οι φοιτητές εργάζονται πάνω σε μια μεμονωμένη διατριβή, που είναι μια ευκαιρία να εφαρμόσουν τις γνώσεις και δεξιότητες που έχουν αποκτήσει σε ένα συγκεκριμένο ερευνητικό ή πρακτικό πρόβλημα.

Η δομή αυτή επιτρέπει την ομαλή μετάβαση από την γενική εκπαίδευση σε πιο εξειδικευμένες σπουδές, διασφαλίζοντας ότι κάθε φοιτητής αποκτά μια στέρεη βάση γνώσεων προτού μεταβεί στην εξειδίκευση και την ανεξάρτητη έρευνα **【4th source】** .

↑ ↓ ↶ ↷ ↻ ↺



20



You

Ποια μαθήματα μαθηματικών είναι υποχρεωτικά για τους φοιτητές της Πληροφορικής;



PDF Ai PDF

Για τους φοιτητές της Πληροφορικής, τα υποχρεωτικά μαθήματα μαθηματικών που είναι βασικά για την κατανόηση των θεωρητικών και πρακτικών ασκήσεων του τομέα περιλαμβάνουν:

- 1. Απειροστικός Λογισμός:** Εισαγωγή στις βασικές αρχές του λογισμού, με εφαρμογές στην ανάλυση συναρτήσεων, παραγώγους και ολοκληρώματα, χρήσιμο για την κατανόηση αλγορίθμων και συστημάτων που βασίζονται σε συνεχείς διαδικασίες.
- 2. Γραμμική Άλγεβρα:** Καλύπτει τη μελέτη διανυσματικών χώρων, γραμμικών μετασχηματισμών, συστημάτων γραμμικών εξισώσεων και μητρώων, στοιχεία που είναι ζωτικής σημασίας για τον προγραμματισμό και την ανάλυση αλγορίθμων.
- 3. Θεωρία Πιθανοτήτων και Στατιστική:** Παρέχει τις βάσεις για την κατανόηση των πιθανοτήτων, των στατιστικών μεθόδων και των στοχαστικών διαδικασιών, που είναι απαραίτητες στην ανάλυση δεδομένων και στην τεχνητή νοημοσύνη.
- 4. Διακριτά Μαθηματικά:** Εστιάζει στη μελέτη διακριτών δομών όπως γραφήματα, θεωρία συνόλων, λογική και κομβιοτόρκα, που είναι θεμελιώδη για την θεωρία αλγορίθμων και την οργάνωση δεδομένων.

Αυτά τα μαθήματα ενισχύουν τις αναλυτικές ικανότητες των φοιτητών και τους παρέχουν τα απαραίτητα μαθηματικά εργαλεία για την επίλυση προβλημάτων και την εφαρμογή θεωρητικών γνώσεων σε πρακτικά ζητήματα στην Πληροφορική **【4th source】** .

5.4 Συμπεράσματα από την σύγκριση των μοντέλων

Μετά από την αξιολόγηση του κάθε μοντέλου ξεχωριστά κατέληξα στα παρακάτω συμπεράσματα. Τα συμπεράσματα μου αφορούν την επίδοση, το κόστος και την ποιότητα των απαντήσεων που παρέχει το κάθε μοντέλο.

Επίδοση: Η επίδοση όλων των μοντέλων ήταν αρκετά καλή καθώς μπορούν να απαντήσουν σε ένα εύλογο χρονικό διάστημα.

Κόστος: Το κάθε μοντέλο προϋποθέτει ένα οικονομικό κόστος που πρέπει να καλυφθεί στην κάθε περίπτωση ξεχωριστά.

Κόστος Εφαρμογής με τη χρήση διανυσματικής βάσης ChromaDB:

Σε αυτή την περίπτωση χρησιμοποίησα το API της OpenAI και συγκεκριμένα το gpt-3.5-turbo api.

Model	Training	Input Usage	Output usage
gpt-3.5-turbo	US\$8.00 / 1M tokens	US\$3.00 / 1M tokens	US\$6.00 / 1M tokens

Χρεώσεις για την χρήση του Api

Στην πιο πάνω εικόνα βάλε αριθμό εικόνας) φαίνονται οι χρεώσεις που έχει ο χρήστης από την χρήση αυτού του μοντέλου. Αυτά που αφορούν το μοντέλο που χρησιμοποίησα είναι το Input Usage και το Output Usage.

Το Input Usage είναι το κόστος που επιβαρύνομαι κάθε φορά για να στείλω δεδομένα κειμένου στο μοντέλο για να πάρω απάντηση. Όπως αναγράφεται στην εικόνα η τιμή ανέρχεται \$3.00 δολάρια για 1M tokens.

Το Output usage είναι το κόστος που εξάγει το μοντέλο μετά την επεξεργασία των δεδομένων, το οποίο ανέρχεται στα \$6.00 δολάρια για κάθε 1M tokens.

Κόστος έτοιμης εφαρμογής GPTs/ καινούργιας εφαρμογής GPTs:

Η χρήση κάποιου GPT είτε η δημιουργία κάποιου προσωπικού GPT προσαρμοσμένο σε συγκεκριμένες ανάγκες το οποίο περιλαμβάνει προγραμματισμό ή εκπαίδευση μοντέλου με συγκεκριμένα δεδομένα (π.χ. Οδηγός σπουδών του Πανεπιστημίου Κύπρου) προϋποθέτει κάποιο κόστος.

Plus

For individuals looking to amplify their productivity

- ✓ Access to GPT-4, our most capable model
- ✓ Browse, create, and use GPTs
- ✓ Access to additional tools like DALL-E, Browsing, Advanced Data Analysis, and more

US\$20 per user, billed monthly

Εικόνα 22 Κόστος για την αγορά Chat GPT PLUS

Όπως φαίνεται στην εικόνα (βάε αριθμό εικόνας) το κόστος για την αγορά του Chat GPT PLUS το οποίο μπορεί να χρησιμοποιήσει τα πιο πάνω χαρακτηριστικά ανέρχεται στα \$20 δολάρια τον μήνα.

Σύγκριση Ποιότητας αποτελεσμάτων των μοντέλων:

Ένα γενικό συμπέρασμα στο οποίο μπορούμε να προβούμε μετά από την παράθεση των αποτελεσμάτων από τις ερωτήσεις των χρηστών είναι ότι το GPT που δημιούργησα με την χρήση του οδηγού σπουδών από την OpenAI επιστρέφει τα καλύτερα αποτελέσματα σε ερωτήσεις των χρηστών. Δεύτερο, στην ποιότητα των αποτελεσμάτων, έρχεται το έτοιμο μοντέλο GPT της OpenAI που του τροφοδοτήθηκε pdf με τον οδηγό σπουδών. Μπορεί να απαντήσει με ακρίβεια στις ερωτήσεις αλλά όχι τόσο καλά όσο το προηγούμενο μοντέλο. Τρίτο είναι το μοντέλο που δημιούργησα με τη χρήση της διανυσματικής βάσης ChromaDB το οποίο επιστέφει τα λιγότερο ποιοτικά αποτελέσματα σε σχέση με τα προηγούμενα δύο μοντέλα.

Βιβλιογραφία

// τοποθέτηση σωστά όλων των πηγών και λεζάντα για εικόνες.

[1] H. Naveed et al., ‘A Comprehensive Overview of Large Language Models’. arXiv, Apr. 09, 2024. Accessed: May 09, 2024. [Online]. Available: <http://arxiv.org/abs/2307.06435>.

[2] W. X. Zhao et al., 'A Survey of Large Language Models'. arXiv, Nov. 24, 2023. Accessed: May 09, 2024. [Online]. Available: <http://arxiv.org/abs/2303.18223>.

[3] 'A History of Chatbots'. [Online]. Available: https://www.chatbotpack.com/a-history-of-chatbots/?ref=quuu&utm_source=quuu.

[4] T. Zemčík, 'A Brief History of Chatbots', DEStech Transactions on Computer Science and Engineering, Oct. 2019, doi: 10.12783/dtcse/aicae2019/31439.

[5] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, 'Natural language processing: an introduction', J Am Med Inform Assoc, vol. 18, no. 5, pp. 544–551, Sep. 2011, doi: 10.1136/amiajnl-2011-000464.

[6] R. M. Schmidt, 'Recurrent Neural Networks (RNNs): A gentle Introduction and Overview'. arXiv, Nov. 23, 2019. Accessed: May 09, 2024. [Online]. Available: <http://arxiv.org/abs/1912.05911>

[7] 'BERT: A Comprehensive Guide to the Groundbreaking NLP Framework'. [Online]. Available: <https://www.analyticsvidhya.com/blog/2019/09/demystifying-bert-groundbreaking-nlp-framework/>.

[8] 'Recent Advances in Natural Language Processing via Large Pre-Trained Language Models: A Survey', ar5iv. Accessed: May 17, 2024. [Online]. Available: <https://ar5iv.labs.arxiv.org/html/2111.01243>

[9] 'What is Google Gemini? Everything you need to know about Google's next-gen AI'. [Online]. Available: <https://www.techradar.com/computing/artificial-intelligence/what-is-google-gemini>

Παράρτημα Α

Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό.

Παράρτημα Β

Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό.

Παράρτημα Γ

Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό. Εισάγεται το θέμα που προτίθεστε να διατυπώσετε στο παράρτημα αυτό.