

Περιγραφή Προγραμματιστικής Άσκησης

Retrieval Augmented Generation (RAG) με το “Μελτέμι”

Ο σκοπός της προγραμματιστικής άσκησης είναι να κατανοήσετε την τεχνική του Retrieval Augmented Generation (RAG) και να εφαρμόσουμε τις γνώσεις που αποκομίσατε από την σημερινή ημερίδα στην πράξη. Η εργασία σας είναι να υλοποιήσετε μία από τις παρακάτω θεματικές ενότητες, χρησιμοποιώντας τις δυνατότητες του “Μελτέμι” για την παραγωγή και τη σύνθεση απαντήσεων.

Οι ομάδες θα αποτελούνται από 2-3 άτομα, και τα αποτελέσματά σας θα πρέπει να παραδοθούν στο GitHub του οργανισμού (το σχετικό link έχει σταλεί). Οι τρεις καλύτερες λύσεις, που θα αξιολογηθούν από την ομάδα του “Μελτέμι”, θα δημοσιευτούν στη σελίδα του οργανισμού.

Γενικές Οδηγίες

- Διάρκεια Υλοποίησης:** Η εργασία εντός της ημερίδας μπορεί να μην ολοκληρωθεί, οπότε θα έχετε πρόσβαση στους servers για επιπλέον 1-2 μέρες ώστε να την ολοκληρώσετε ασύγχρονα.
- Παρουσίαση Λύσης:** Μπορείτε να φτιάξετε μια απλή διεπαφή (frontend) για καλύτερη παρουσίαση. Απαιτείται επίσης ένα αρχείο **README.md** στο repository σας, ώστε οι αξιολογητές να μπορούν να εκτελέσουν και να κατανοήσουν τη λύση σας.
- Επιλογή Θεματικής:** Κατά την επιλογή θέματος μπορείτε να ρωτήσετε τους βοηθούς που θα βρίσκονται στον χώρο ώστε να σας καθοδηγήσουν με βάση την εμπειρία και το επίπεδο σας.

Θεματικές Ενότητες

Θεματική 1 - RAG για τη Wikipedia σελίδα της επιλογής σας

- Περιγραφή:** Επιλέξτε μια σελίδα της Wikipedia, όπως για παράδειγμα την [Ελληνική Επανάσταση του 1821](#). Ανακτήστε το περιεχόμενο της σελίδας και ενσωματώστε το στο RAG σύστημά σας. Το “Μελτέμι” θα χρησιμοποιηθεί για να παράγει απαντήσεις σε ερωτήματα σχετικά με το περιεχόμενο της σελίδας.

- **Στόχοι:**
 - Ανάκτηση δεδομένων από τη Wikipedia.
 - Χρήση του “Μελτέμι” για τη δημιουργία φυσικών και περιεκτικών απαντήσεων.
 - Εφαρμογή RAG για ερωτήσεις που συνδυάζουν ανακτώμενη και παραγόμενη πληροφορία.
-

Θεματική 2 - RAG για την Ελληνική Μυθολογία

- **Περιγραφή:** Εξερευνήστε δεδομένα της Ελληνικής Μυθολογίας από αξιόπιστες πηγές (π.χ. ιστοσελίδες ή βιβλία σε ψηφιακή μορφή, όπως [GreekMythology.gr](https://greekmythology.gr), [HellenicMythology.gr](https://hellenicmythology.gr)). Το “Μελτέμι” θα απαντά σε ερωτήματα σχετικά με θεούς, μύθους και ήρωες.
 - **Στόχοι:**
 - Ενσωμάτωση μύθων και ιστοριών της Ελληνικής Μυθολογίας στο μοντέλο.
 - Δημιουργία απαντήσεων με αφηγηματικό χαρακτήρα, αξιοποιώντας τη δύναμη παραγωγής του “Μελτέμι”.
 - Δοκιμή RAG με ανοιχτές ερωτήσεις που απαιτούν συνδυαστική σκέψη.
-

Θεματική 3 - RAG για το Paper του “Μελτέμι”

- **Περιγραφή:** Αναλύστε το [ερευνητικό άρθρο](#) του “Μελτέμι” και ανακτήστε πληροφορίες για την αρχιτεκτονική, τις μεθόδους και τα αποτελέσματά του. Το RAG θα απαντά σε ερωτήματα, ενώ το “Μελτέμι” θα συνθέτει λεπτομερείς, τεκμηριωμένες απαντήσεις.
 - **Στόχοι:**
 - Ανάκτηση δεδομένων από PDF αρχεία με χρήση **pdfplumber**.
 - Ενσωμάτωση των δεδομένων στο RAG.
 - Παροχή ακριβών απαντήσεων για το ερευνητικό έργο του “Μελτέμι”.
-

Θεματική 4 - Ελεύθερη Επιλογή

- **Περιγραφή:** Αυτή η θεματική σας επιτρέπει να πειραματιστείτε με οποιοδήποτε σύνολο δεδομένων και εφαρμογή του RAG. Για παράδειγμα, μπορείτε να επιλέξετε εφαρμογές στην υγεία, την εκπαίδευση ή την ψυχαγωγία, αξιοποιώντας τη δύναμη παραγωγής του “Μελτέμι”.
- **Στόχοι:**

- Δημιουργία ενός καινοτόμου συστήματος RAG.
 - Εξερεύνηση της ικανότητας παραγωγής κειμένου του “Μελτέμι” σε νέα πεδία.
-

Χρήσιμα Πακέτα

- **Γλώσσα Προγραμματισμού:** Python
- **Ανάλυση Δεδομένων:** Numpy, Pandas
- **Οπτικοποίηση Δεδομένων:** Matplotlib, Seaborn
- **Εξαγωγή Πληροφορίας από Ιστοσελίδες:** BeautifulSoup, Selenium
- **Ανάλυση PDF Αρχείων:** pdfplumber
- **Ενσωμάτωση RAG:** LangChain, LlamaIndex