



γλωσσAPI: προς ένα OSS LLM



Αλέξανδρος Μελίδης

Γενικός Διευθυντής **Οργανισμού Ανοιχτών Τεχνολογιών - ΕΕΛΛΑΚ**

Κώστας Παπαδήμας

Content and Community Manager **Οργανισμού Ανοιχτών Τεχνολογιών - ΕΕΛΛΑΚ**



Οργανισμός Ανοιχτών Τεχνολογιών

- Ο **Οργανισμός Ανοιχτών Τεχνολογιών (ΕΕΛΛΑΚ)** ιδρύθηκε το 2008.
- **Κύριος στόχος:** Να συμβάλλει στην ανοιχτότητα και ειδικότερα στην προώθηση και ανάπτυξη των **Ανοιχτών Προτύπων**, του **ΕΛΛΑΚ**, του **Ανοιχτού Περιεχομένου**, των **Ανοιχτών Δεδομένων** και των **Τεχνολογιών Ανοιχτής Αρχιτεκτονικής**.
- **Τομείς:** Εκπαίδευσης, Δημόσιου τομέα, Επιχειρήσεων και Κοινωνικής Οικονομίας στην Ελλάδα.
- **Φιλοδοξία:** Να αποτελέσει κέντρο γνώσης και πλατφόρμα διαλόγου για τις ανοιχτές τεχνολογίες.



- **Αποστολή:** Να καταστήσει εύκολα κι άμεσα κατανοητό και αναγνωρίσιμο το έργο ανάπτυξης GREEK OSS LLM στην ελληνική κοινότητα ΕΛΛΑΚ που θα θελήσει να εμπλακεί
- συνδυάζει τη γλώσσα με την τεχνολογία (γλώσσα + API) συνδυάζει τη γλώσσα με την τεχνολογία (γλώσσα + API)
- μεταδίδει καθαρά περί τίνος πρόκειται, ένα γλωσσικό API
- σαν λογότυπο μπορεί να σχεδιαστεί απλά και όμορφα με κατάλληλη χρήση τυπογραφίας



Πηγές δεδομένων “επίσημες”

Βουλή των Ελλήνων_Πρακτικά:

<https://www.hellenicparliament.gr/Praktika/Synedriaseis-Olomeleias>

Δίκαιο της ΕΕ:

<https://eur-lex.europa.eu/homepage.html?locale=el>

Ελληνική Βικιπαίδεια:

<https://el.wikipedia.org>



Πηγές δεδομένων “επίσημες”

Εθνικό Τυπογραφείο: <https://www.et.gr/SearchNomoi>

Διαρκής Κώδικας Νομοθεσίας: <https://raptarchis.gov.gr/>

Εθνικό Μητρώο Διαδικασιών: <https://mitos.gov.gr/>

Άλλα κυβερνητικά: <http://www.opengov.gr/>



Πηγές δεδομένων “επίσημες”

Ελληνική Φιλολογία:

<https://www.gutenberg.org/browse/languages/el>

Εθνικό Κέντρο Τεκμηρίωσης:

<https://www.didaktorika.gr/eadd/>

Ανοιχτές Ακαδημαϊκές Εκδόσεις:

<https://repository.kallipos.gr/>



Πηγές δεδομένων “επίσημες”

Σχολικά Βιβλία: <http://iep.edu.gr/el/>

Ενημέρωση-Εφημερίδες:

<https://www.amna.gr/>

Πολιτισμός-Ιστορία: <http://www.elia.org.gr/>



Άλλες πηγές δεδομένων

Social Media: twitter, facebook, forums, blogs etc

Websites:

<https://commoncrawl.org/>

Open Subtitles: <https://www.opensubtitles.org/el/>



Ενότητες

1. Επιλογή δεδομένων εκπαίδευσης
2. Επιλογή προεκπαιδευμένου μοντέλου
3. Ανάπτυξη πρωτοτύπου
4. Εκπαίδευση στην Ελληνική Γλώσσα
5. Αναπτυξη υποσυνόλων δεδομένων και έργων επαλήθευσης
6. Εκπαίδευση σε ειδικές γνωστικές περιοχές
7. Ανάπτυξη διαλογικού συστήματος
8. Βαθμονήμηση με ανθρώπινη αξιολόγηση και ανάδραση



Ειδικότητες

1. Ειδικοί AI/Γνωσιακής Επιστήμης
2. Ειδικοί Επεξεργασίας Φυσικού Λόγου/Υπολογιστικής Γλωσσολογίας
3. Ειδικοί Ανάλυσης Δεδομένων (R, Python)
4. Ειδικοί Πληροφοριακών Συστημάτων, Μεταδεδομένων, Οντολογιών
5. Διαχειριστές Συστημάτων με έμφαση σε Deployment CI/CD
6. Ειδικοί Κοινωνιογλωσσολογίας, Εθνογραφίας της Επικοινωνίας
7. Ειδικοί Ποιοτικής Έρευνας, Ανάλυσης Περιεχομένου Grounded Theory
8. Συντάκτες Περιεχομένου

Άδειες

Creative Commons, EUPL

<https://mathe.ellak.gr>



Συνεργασία

- Αποθετήριο στο Github
<https://github.com/eellak/glossAPI/>
- Κοινότητα στο Lemmy
https://lemmy.ml/c/greek_foss
- Mailing List
- Μειλοντική δημιουργία δωματίου στο στο Matrix/element

