# Περιγραφή Scripts Πτυχιακής

## 1. download\_coco\_val\_images.py

Κατεβάζει το COCO val2017 image dataset (~6GB) και το αποσυμπιέζει στον φάκελο data/images/val2017/.

## 2. download\_coco\_val\_captions.py

Κατεβάζει το annotations\_trainval2017.zip (~250MB), το αποσυμπιέζει και αποθηκεύει το captions\_val2017.jsonc μέ 5 λεζάντες ανά εικόνα.

## 3. extract\_image\_embeddings.py

Χρησιμοποιεί το CLIP μοντέλο ViT-B/32 για να υπολογίσει image embeddings για κάθε εικόνα από το val2017.  
Αποθηκεύει τα embeddings στο αρχείο data/embeddings/coco\_val\_image\_embeddings.pt

## 4. extract\_caption\_embeddings.py

Διαβάζει το captions\_val2017.json, δημιουργεί text embeddings για κάθε caption,  
κάνει μέσο όρο ανά εικόνα και αποθηκεύει τα αποτελέσματα στο data/embeddings/coco\_val\_text\_embeddings.pt

## 📌 Τελική Κατάσταση

Έχεις δύο αρχεία με embeddings:  
- coco\_val\_image\_embeddings.pt (εικόνες)  
- coco\_val\_text\_embeddings.pt (captions)  
Έτοιμα για αναζήτηση εικόνας με βάση κείμενο ή παρόμοια εικόνα.

## 5. search\_by\_text.py

Επιτρέπει αναζήτηση εικόνων με βάση φράση (π.χ. "a man riding a bike").  
1. Φορτώνει τα image embeddings.  
2. Ζητά από τον χρήστη κείμενο αναζήτησης.  
3. Υπολογίζει το text embedding του query.  
4. Υπολογίζει cosine similarity με όλες τις εικόνες.  
5. Επιστρέφει τις πιο σχετικές εικόνες (top 5).