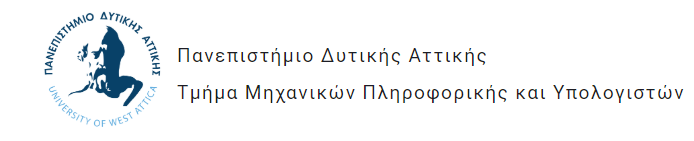
****

**Ανάκτηση**

**Πληροφορίας**

**Εργασία Εξαμήνου**

**2025**

**Ονοματεπώνυμο - 1** : Βαρσάκης Νίκων - Αλκίνοος

**ΑΜ - 1** : 18390040

**Ονοματεπώνυμο - 2** : Ψυχόπαιδας Αγησίλαος

**ΑΜ - 2** : 18390107

**Ημερομηνία** : 19/01/2025

<https://github.com/NikwnasV/Varsakis18390040-Psychopaidas18390107>

Περιεχόμενα

[Εισαγωγή 3](#_Toc187682162)

[Στόχοι της εργασίας 3](#_Toc187682163)

[Περιγραφή εργασίας 3](#_Toc187682164)

# Εισαγωγή

Η εργασία του μαθήματος «Ανάκτηση Πληροφορίας» αφορά στην πρακτική εξάσκηση και εμπειρία στη δημιουργία μιας απλής μηχανής αναζήτησης, με σκοπό την κατανόηση των θεμελιωδών εννοιών της ανάκτησης πληροφορίας, την ευρετηρίασης, της κατάταξης και της αναζήτησης πληροφορίας, καθώς και πρακτικές δεξιότητες στην επεξεργασία φυσικής γλώσσας και την εφαρμογή αλγορίθμων αναζήτησης.

## Στόχοι της εργασίας

• Ο σχεδιασμός και υλοποίηση ενός συστήματος ανάκτησης εγγράφων που μπορεί να

ευρετηριάσει και να αναζητήσει αποτελεσματικά σε μια συλλογή εγγράφων κειμένου.

• Η ανάπτυξη αλγορίθμων ανάκτησης και η αξιολόγηση της απόδοσή τους χρησιμοποιώντας

κοινές μετρήσεις αξιολόγησης, όπως ακρίβεια, ανάκληση και βαθμολογία F1.

• Η παροχή μιας φιλικής προς το χρήστη διεπαφής για τους χρήστες να εισάγουν ερωτήματα

και να ανακτούν σχετικά έγγραφα.

• Η απόκτηση πρακτικής εμπειρίας σε διάφορες τεχνικές ανάκτησης πληροφοριών, όπως

Boolean retrieval, Vector Space Model και Probabilistic retrieval models.

## Περιγραφή εργασίας

Σε αυτή την εργασία θα υλοποιήσουμε μια μηχανή αναζήτησης που ανακτά έγγραφα από το Wikipedia με βάση τα ερωτήματα των χρηστών.

# Υλοποίηση της Εργασίας

Σε αυτό το κομμάτι της εργασίας θα παρουσιάσουμε και θα επεξηγήσουμε τον τρόπο με τον οποίο υλοποιήθηκε, ακολουθώντας τα βήματα, όπως δίνονται στην εκφώνηση της εργασίας.

## Βήμα 1. Συλλογή δεδομένων

α. Αρχικά, θα πρέπει να γίνει η συλλογή των εγγράφων από το Wikipedia.

Επιλέξαμε για τις ανάγκες της εργασίας να συλλέξουμε 21 έγγραφα τα οποία αναφέρονται σε αυτοκινητοβιομηχανίες.

Τα έγγραφα που θα χρησιμοποιηθούν είναι τα εξής :

<https://en.wikipedia.org/wiki/BMW>

<https://en.wikipedia.org/wiki/Volkswagen>

<https://en.wikipedia.org/wiki/Audi>

<https://en.wikipedia.org/wiki/Mercedes_Benz>

<https://en.wikipedia.org/wiki/Opel>

<https://en.wikipedia.org/wiki/Kia>

<https://en.wikipedia.org/wiki/LADA>

<https://en.wikipedia.org/wiki/Datsun>

<https://en.wikipedia.org/wiki/Toyota>

<https://en.wikipedia.org/wiki/Hyundai_Motor_Company> <https://en.wikipedia.org/wiki/Ferrari>

<https://en.wikipedia.org/wiki/McLaren_Automotive>

<https://en.wikipedia.org/wiki/Ford_Motor_Company>

<https://en.wikipedia.org/wiki/Lancia>

<https://en.wikipedia.org/wiki/Lamborghini>

<https://en.wikipedia.org/wiki/Porsche>

<https://en.wikipedia.org/wiki/Škoda_Auto>

<https://en.wikipedia.org/wiki/Fiat>

<https://en.wikipedia.org/wiki/Volvo>

<https://en.wikipedia.org/wiki/Honda>

<https://en.wikipedia.org/wiki/SEAT>