



Χρηματοδοτείται από την Ευρωπαϊκή Ένωση. Οι απόψεις και οι γνώμες που εκφράζονται είναι αποκλειστικά του/των συγγραφέα/ων και δεν αντανακλούν κατ' ανάγκη τις απόψεις και τις γνώμες της Ευρωπαϊκής Ένωσης ή του Ευρωπαϊκού Εκτελεστικού Οργανισμού Εκπαίδευσης και Πολιτισμού (EACEA). Ούτε η Ευρωπαϊκή Ένωση ούτε ο EACEA μπορούν να θεωρηθούν υπεύθυνοι γι' αυτές.

STEAME ACADEMY

ΣΧΕΔΙΟ ΜΑΘΗΣΗΣ ΚΑΙ ΔΗΜΙΟΥΡΓΙΚΟΤΗΤΑΣ ΔΙΕΥΚΟΛΥΝΣΗΣ ΤΗΣ ΔΙΔΑΣΚΑΛΙΑΣ (ΣΧΕΔΙΟ ΜΑΘΗΣΗΣ ΚΑΙ ΔΗΜΙΟΥΡΓΙΚΟΤΗΤΑΣ) - ΕΠΙΠΕΔΟ 2 ΕΚΠΑΙΔΕΥΤΙΚΟΙ Εξόρυξη κειμένου: είναι αυτά τα έγγραφα τα ίδια;

S

T

Eng

A

M

Ent



1. Επισκόπηση

Τίτλος	Εξόρυξη κειμένου: είναι αυτά τα κείμενα ίδια;		
Ερώτηση ή θέμα αφόρμησης	Πώς μπορούν οι μηχανές αναζήτησης να βρίσκουν αποτελέσματα για μια αναζήτηση χρήστη βάσει λέξεων-κλειδιών; Πώς οι υπολογιστές εντοπίζουν έγγραφα κειμένου που επικεντρώνονται στα ίδια θέματα; Πώς οι αλγόριθμοι των υπολογιστών μοντελοποιούν μη δομημένα δεδομένα για ψηφιακή επεξεργασία;		
Ηλικίες, βαθμίδες, ...	16 έως 18 χρονών	Α' - Γ' Λυκείου	
Διάρκεια, χρονοδιάγραμμα, δραστηριότητες	18 ώρες	9 δίωρα μαθήματα	21 δραστηριότητες
Ευθυγράμμιση με το πρόγραμμα σπουδών	Εξόρυξη δεδομένων, μηχανική μάθηση, μοντελοποίηση μη δομημένων δεδομένων, προγραμματισμός υπολογιστών		
Συνεισφέροντες, Συνεργάτες			
Περίληψη - Σύνοψη	Οι μαθητές εισάγονται στην Εξόρυξη Δεδομένων (Data Mining) και στη Μηχανική Μάθηση (Machine Learning), με επίκεντρο τα βασικά θέματα της ψηφιακής επεξεργασίας κειμένου. Εξερευνάται η έννοια της ομοιότητας κειμένων, με την παρουσίαση των μαθηματικών της βάσεων, όπως η τομή συνόλων και το συνημίτονο της γωνίας μεταξύ δύο διανυσμάτων (cosine similarity). Οι μαθητές εργάζονται σε ομάδες για να υλοποιήσουν ένα απλό εργαλείο που μετρά την ομοιότητα μεταξύ δύο εγγράφων κειμένου. Στις τελευταίες συνεδρίες, οι μαθητές συμμετέχουν σε έναν διαγωνισμό για να εντοπίσουν την καλύτερη υλοποίηση. Καθ' όλη τη διάρκεια των συνεδριών, εισάγονται στις βασικές μεθόδους προ-επεξεργασίας κειμένου, όπως η		

Αναφορές, παραθέσεις	αφαίρεση κοινών λέξεων (stop-words) και το stemming. Η τελευταία συνεδρία ολοκληρώνεται με συζήτηση, όπου οι μαθητές εντοπίζουν τις ομοιότητες της υλοποίησής τους με μια μηχανή αναζήτησης και από εκεί σχεδιάζουν μια μηχανή αναζήτησης βασισμένη στην υλοποίηση ομοιότητας κειμένου που ανέπτυξαν.
----------------------	---

2. Πλαίσιο STEAME ACADEMY*

Συνεργασία των εκπαιδευτικών	Εκπαιδευτικός 1 (Φυσικές Επιστήμες) ● Μηχανική μάθηση και εξόρυξη δεδομένων: επισκόπηση του πεδίου, γενική αρχιτεκτονική, συνήθεις εφαρμογές στην καθημερινή ζωή, συχνά προβλήματα ● Μοντελοποίηση δεδομένων για ψηφιακή επεξεργασία: αναπαράσταση δομημένων δεδομένων για μηχανική μάθηση και εξόρυξη δεδομένων, μοντελοποίηση μη δομημένων δεδομένων για μηχανική μάθηση και εξόρυξη δεδομένων ● Εξόρυξη κειμένου (Text Mining): επισκόπηση του πεδίου, βασικές έννοιες, μοντελοποίηση (δυαδικό μοντέλο, μοντέλο TF, μοντέλο TFxIDF), ομοιότητα κειμενικών εγγράφων, προεπεξεργασία, βασικές εργασίες και εφαρμογές Εκπαιδευτικός 2 (Μηχανική) ● Προγραμματισμός με Python programming για εξόρυξη κειμένου ● Διανυσματική άλγεβρα στο Excel Εκπαιδευτικός 3 (Μαθηματικά) ● Συνάρτηση συνημίτονου, διανυσματική άλγεβρα ● Ορισμός τελεστών, τομή Ο εκπαιδευτικός 1 συνεργάζεται με τον εκπαιδευτικό 2 και τον εκπαιδευτικό 3 για να: - Αναγνωρίσουν βιβλιοθήκες στην Python που θα χρησιμοποιήσουν για εξόρυξη κειμένου - Προσδιορίσουν τις συναρτήσεις του Excel που θα χρησιμοποιηθούν για τη διανυσματική άλγεβρα - Δημιουργήσουν τις ασκήσεις και την πρόκληση Ο εκπαιδευτικός 1 συνεργάζεται με τον εκπαιδευτικό 2 για να:
------------------------------	--

- Συγκεντρώνουν τα απαραίτητα σώματα κειμένων για τις πρακτικές δραστηριότητες
- Επισημειώνουν τα απαραίτητα σώματα κειμένων για κάθε άσκηση

Ο εκπαιδευτικός 1 συνεργάζεται με τον εκπαιδευτικό 3 για να:

- Εισάγουν τελεστές συνόλων σε περιβάλλον επεξεργασίας κειμένου
- Εισάγουν διανυσματική άλγεβρα σε περιβάλλον επεξεργασίας κειμένου

Οργάνωση STEAME in Life (SiL)

Η τελευταία συνεδρία ολοκληρώνεται με την παρότρυνση των μαθητών να συζητήσουν και να εντοπίσουν τις ομοιότητες ανάμεσα στη δική τους υλοποίηση και σε μια μηχανή αναζήτησης, όπως η Google, και από εκεί να σχεδιάσουν μια μηχανή αναζήτησης χρησιμοποιώντας την προηγούμενη υλοποίησή τους για την ομοιότητα κειμένων.

Διαμόρφωση Σχεδίου Δράσης

Προ-παρασκευαστική φάση

1. Έρευνα σχετικά με τις παραδοσιακές και σύγχρονες εφαρμογές της εξόρυξης δεδομένων (data mining) και της μηχανικής μάθησης (machine learning)· σύνδεση με μηχανές αναζήτησης και γενετική τεχνητή νοημοσύνη (generative AI)· αναφορά σε περιπτώσεις δομημένων και μη δομημένων δεδομένων· ανασκόπηση βασικών προκλήσεων της εξόρυξης δεδομένων και της μηχανικής μάθησης (μοντελοποίηση δεδομένων, πολυδιάστατοτητα, υπερεκπαίδευση, ελλιπή δεδομένα, όγκος δεδομένων, μεγάλα δεδομένα, ερμηνεία έναντι πρόβλεψης, κ.ά.)
2. Συγκέντρωση και σχολιασμός σωμάτων κειμένων (corpora) για ασκήσεις
3. Ρύθμιση του προγραμματιστικού περιβάλλοντος Python (docker, αποθετήριο στο Github για κλωνοποίηση, άλλα εργαλεία)

Δομή του σεμιναρίου

1. Εισαγωγή
 - 1.1. Επισκόπηση της εξόρυξης δεδομένων και της μηχανικής μάθησης: ιστορική αναδρομή, βασικά προβλήματα/εργασίες, εφαρμογές, καθοδήγηση της συζήτησης προς τη μοντελοποίηση δεδομένων (ξεκινώντας από πίνακες δεδομένων – tabular datasets – και παρουσιάζοντας παραδείγματα, στη συνέχεια τίθεται το ερώτημα για το πως αντιμετωπίζονται τα μη δομημένα δεδομένα όπως εικόνες και κείμενο). Συζήτηση αυτών των λύσεων με τους μαθητές.
2. Μοντελοποίηση κειμένου
 - 2.1. Εξήγηση του σώματος κειμένων (corpora) που θα χρησιμοποιηθεί (πρέπει να περιλαμβάνει μικρά κείμενα με περιορισμένο λεξιλόγιο, να περιέχει κείμενα όπου η Συχνότητα Όρων – TF – κάνει τη διαφορά σε σύγκριση με το Δυναμικό Μοντέλο – Boolean model).
 - 2.2. Επανάληψη της μοντελοποίησης κειμένου και εισαγωγή του Δυναμικού Μοντέλου, παρουσίαση παραδειγμάτων, ερώτηση: πώς βρίσκουμε ομοιότητες μεταξύ κειμένων;
 - 2.3. Εισαγωγή τελεστών συνόλων (ένωση, τομή κ.λπ.).
 - 2.4. Οι μαθητές, σε ομάδες των 3 ή 4, καλούνται να υλοποιήσουν ένα

εργαλείο στο Excel για να εφαρμόσουν τις λύσεις τους και να τις δοκιμάσουν· κάθε ομάδα εξηγεί ποιοι τελεστές συνόλων χρησιμοποιούνται και γιατί.

- 2.5. Εισαγωγή της διανυσματικής άλγεβρας: εσωτερικό γινόμενο και συνημίτονο (cosine). Οι μαθητές συζητούν και προσπαθούν να εντοπίσουν σχέσεις μεταξύ των τελεστών συνόλων και της διανυσματικής άλγεβρας (τομή και εσωτερικό γινόμενο, ένωση και άθροισμα). Επανασχεδιάζουν την υλοποίηση στο Excel χρησιμοποιώντας συνδυασμό τελεστών συνόλων και διανυσμάτων.
- 2.6. Εμφάνιση της σημασίας/συνεισφοράς της Συχνότητας Όρων (TF) σε σχέση με το Δυναμικό Μοντέλο· οι μαθητές τροποποιούν την υλοποίηση για να εφαρμόσουν το μοντέλο TF και να υπολογίσουν ομοιότητες χρησιμοποιώντας το ίδιο σώμα κειμένων. Πλέον θα είναι σαφές ότι το συνημίτονο λειτουργεί καλά.
- 2.7. Συζήτηση των αλλαγών και των προβλημάτων που ενδέχεται να προκύψουν όταν η ομοιότητα βασίζεται μόνο στο TF (οι όροι που υπάρχουν σε όλα τα κείμενα δεν έχουν διακριτική δύναμη)· οι μαθητές καλούνται να προτείνουν λύσεις.
- 2.8. Εισαγωγή του IDF και του TFxIDF· συζήτηση, σχεδίαση και υλοποίηση στο Excel ενός μέτρου ομοιότητας βασισμένου στο μοντέλο TFxIDF για χρήση σε απλοποιημένα κείμενα με περιορισμένο λεξιλόγιο.

3. Εφαρμογή

- 3.1. Εισαγωγή στις R, Python ή άλλες βιβλιοθήκες για εξόρυξη κειμένου.
- 3.2. Υλοποίηση μιας συνάρτησης υπολογισμού της ομοιότητας μεταξύ δύο εγγράφων χρησιμοποιώντας R, Python ή άλλη γλώσσα προγραμματισμού. Όλες οι ομάδες μαθητών θα πρέπει να έχουν μια λειτουργική συνάρτηση στο τέλος αυτής της φάσης· σε διαφορετική περίπτωση, να παρέχεται μια βασική υλοποίηση για όλους.

4. Μαθηματική εξερεύνηση

- 4.1. Εμπλέξτε τους μαθητές σε πρακτικές δραστηριότητες που εξερευνούν τους τελεστές συνόλων και την διανυσματική άλγεβρα για να υπολογίσουν την ομοιότητα μεταξύ εγγράφων κειμένου σε ένα σώμα κειμένων.
- 4.2. Διευκόλυνση της συζήτησης για τις μαθηματικές αρχές πίσω από την επεξεργασία κειμένων.

5. Ολοκληρωμένα έργα

- 5.1. Οι ομάδες των μαθητών καλούνται να αναζητήσουν, να αναδιαμορφώσουν και να υλοποιήσουν τον καλύτερο αλγόριθμο για μια συνάρτηση ομοιότητας.
- 5.2. Δίνεται στους μαθητές ένα σώμα κειμένων (corpus) που έχει ήδη επισημειωθεί (annotated) ως προς την ομοιότητα μεταξύ στατικών ερωτημάτων/κειμένων (κάθε έγγραφο στο σώμα συνοδεύεται από μια κατάταξη ομοιότητας για κάθε ένα από τα test cases). Επίσης διατίθεται ένα σύνολο επικύρωσης (validation set), επίσης σχολιασμένο – αυτό μπορεί να αποτελείται από δύο ή τρία κείμενα, το καθένα με λίγους όρους, σαν να ήταν ερωτήματα προς μια μηχανή

αναζήτησης. Για κάθε ένα από αυτά τα “ερωτήματα” παρέχεται η κατάταξη των εγγράφων του corpus των μαθητών, η οποία θα χρησιμοποιηθεί για τον υπολογισμό του δείκτη F1 στο τέλος και για την ανάδειξη της νικήτριας ομάδας.

- 5.3. Παρέχεται χρόνος στους μαθητές ώστε να υλοποιήσουν και να βελτιστοποιήσουν τη συνάρτηση ομοιότητας με στόχο τα καλύτερα αποτελέσματα στο δικό τους corpus. Κατά τη διάρκεια αυτής της φάσης, δίνεται ανατροφοδότηση και καθοδήγηση από τους εκπαιδευτές, οι οποίοι εισάγουν τεχνικές προ-επεξεργασίας κειμένου, όπως η αφαίρεση stop words (stop-word removal), η αποκοπή καταλήξεων (stemming), κ.ά.

6. Σύνδεσμοι σε μηχανές αναζήτησης

- 6.1. Ένα «έγγραφο» μπορεί να λειτουργήσει ως ερώτημα, όπως αυτά που χρησιμοποιούμε σε μηχανές αναζήτησης όπως η Google, π.χ. «καμέλιες Πόρτο» ή «υπολογιστική όραση». Δείτε το αυτό στην πράξη χρησιμοποιώντας τη Google. Δώστε στους μαθητές το σύνολο επικύρωσης (validation set), δηλαδή τρία έγγραφα, το καθένα με λίγες λέξεις-κλειδιά, σαν να ήταν ένα ερώτημα προς μηχανή αναζήτησης. Εκτελέστε τις συναρτήσεις ομοιότητας για να παραχθεί μια κατάταξη των εγγράφων στο corpus και δώστε στους μαθητές την καλύτερη δυνατή κατάταξη. Υπολογίστε τον νικητή χρησιμοποιώντας το μέτρο F1. Εξηγήστε στους μαθητές τι είναι το F1, η Ανάκληση (Recall) και η Ακρίβεια (Precision).

Αξιολόγηση και Αναστοχασμός

1. Αξιολογήστε την κατανόηση και την εφαρμογή των μαθηματικών εννοιών και τελεστών της θεωρίας συνόλων και της διανυσματικής άλγεβρας από τους μαθητές, μέσω αξιολογήσεων βασισμένων σε έργα, παρουσιάσεων και γραπτών αναστοχασμών.
2. Ενθαρρύνετε τους μαθητές να αναστοχαστούν πάνω στις μαθησιακές τους εμπειρίες, τονίζοντας τη σύνδεση μεταξύ των μαθηματικών και της εξόρυξης κειμένου (text mining).

Ζητήστε από τους μαθητές να σχεδιάσουν και να παρουσιάσουν τη μεθοδολογία μιας μηχανής αναζήτησης και ένα μη λειτουργικό πρωτότυπο χρησιμοποιώντας όσα έχουν μάθει.

* υπό ανάπτυξη τα τελικά στοιχεία του πλαισίου

3. Στόχοι και μεθοδολογίες

Μαθησιακοί σκοποί και στόχοι

1. Κατανόηση των γενικών εννοιών, μοντέλων και τεχνικών επεξεργασίας που χρησιμοποιούνται στην εξόρυξη κειμένου (text mining).
2. Διερεύνηση των διεπιστημονικών συνδέσεων μεταξύ εξόρυξης κειμένου, μηχανών αναζήτησης και διανυσματικής άλγεβρας.
3. Απεικόνιση της ομοιότητας μεταξύ εγγράφων κειμένου και άλλων μη δομημένων συνόλων δεδομένων ως εφαρμογές της διανυσματικής άλγεβρας.

Μαθησιακά αποτελέσματα και αναμενόμενα αποτελέσματα	Μαθησιακοί στόχοι - Συζήτηση θεμάτων υψηλού επιπέδου που σχετίζονται με τους τομείς της εξόρυξης κειμένου (text mining) και των μηχανών αναζήτησης. - Περιγραφή της σχέσης μεταξύ διανυσματικής άλγεβρας, θεωρίας συνόλων και εξόρυξης κειμένου. - Εφαρμογή βασικών τεχνικών εξόρυξης κειμένου για την αντιμετώπιση απλών περιπτώσεων χρήσης. Αναμενόμενα αποτελέσματα - Η συνάρτηση ομοιότητας κειμένου στην R, την Python ή άλλη γλώσσα προγραμματισμού
Προηγούμενες γνώσεις και Προαπαιτούμενα	- Βασικές γνώσεις στη διανυσματική άλγεβρα - Εξοικείωση με το Excel - Βασικές δεξιότητες προγραμματισμού - Άριστη γνώση χρήσης IT εργαλείων
Κίνητρα, Μεθοδολογία, Στρατηγικές, Υποστήριξη	- Δημιουργήστε μικρές ομάδες μαθητών (3 ή 4 μαθητών). - Σχεδιάστε μία λύση, εφαρμογή, δοκιμή και βελτιστοποίηση με έναν επαναλαμβανόμενο τρόπο. Χρησιμοποιήστε μια επαναληπτική μεθοδολογία ανάπτυξης. - Τονίστε τις συνδέσεις μεταξύ διανυσματικής άλγεβρας, μοντέλων κειμένων, συνημίτονου και ομοιότητας. - Εξερευνήστε μηχανές αναζητήσεων για να παρουσιάσετε τη σχέση μεταξύ ομοιότητας κειμένου και αποτελεσμάτων της μηχανής αναζήτησης. - Καθοδήγηση των μαθητών μέσα από μία εξελικτική πορεία από τα πιο απλά μοντέλα (Δυαδικό - Boolean) προς πιο σύνθετες προσεγγίσεις (TFxIDF), με σταδιακή εισαγωγή προκλήσεων, δοκιμάζοντας στο Excel βασικές υλοποιήσεις για μικρά κείμενα με μία ή δύο σύντομες προτάσεις και λίγους διαφορετικούς όρους από ένα λεξιλόγιο 10 ή 20 λέξεων.

4. Προετοιμασία και μέσα

Προετοιμασία, Διευθέτηση χώρου, Συμβουλές αντιμετώπισης προβλημάτων	Το εργαστήριο θα πραγματοποιηθεί σε μία αίθουσα για περίπου 20 μαθητές, οργανωμένους σε ομάδες των 3 έως 4 ατόμων. Ιδανικά, η διάταξη της αίθουσας θα περιλαμβάνει 5 έως 7 ομάδες τραπεζιών, ώστε τα μέλη κάθε ομάδας να κάθονται αντικριστά. Η αίθουσα χρειάζεται προβολέα (beamer) και έναν τοίχο για παρουσιάσεις προς όλους, καθώς και έναν πίνακα με μαρκαδόρους για τη συζήτηση ιδεών.
Πόροι, Εργαλεία, Υλικά, Επισυναπτόμενα,	Ένα αποθετήριο σε GDrive, Teams, Github ή άλλη πλατφόρμα θα πρέπει να έχει προετοιμαστεί εκ των προτέρων με όλο το απαιτούμενο προγραμματιστικό

Εξοπλισμός	περιβάλλον (R, Python, κ.λπ.) και τα σώματα κειμένων (corpora) που θα χρησιμοποιηθούν στις πρακτικές συνεδρίες. Επιπλέον, θα πρέπει να δοθεί στους μαθητές ένα έγγραφο καθοδήγησης για όλη τη διάρκεια του μαθήματος/εργαστηρίου, το οποίο να εξηγεί λεπτομέρειες, αναμενόμενα αποτελέσματα, αξιολόγηση και μαθησιακά αποτελέσματα ανά συνεδρία.
Υγιεινή και Ασφάλεια	

5. Εφαρμογή

Μαθησιακές δραστηριότητες, Διαδικασίες, Αναστοχασμοί	Δομή του εργαστηρίου - Εισαγωγή [Μάθημα 1: 2 ώρες, 3 δραστηριότητες] - Επισκόπηση της Εξόρυξης Δεδομένων (Data Mining) και της Μηχανικής Μάθησης (Machine Learning): ιστορική αναδρομή, είδη προβλημάτων και εργασιών, μεθοδολογίες, τεχνικές και εργαλεία (Weka, R, κ.ά.), βασικές εργασίες (ταξινόμηση, ομαδοποίηση, κ.λπ.), εφαρμογές· καθοδήγηση της συζήτησης προς τη μοντελοποίηση δεδομένων (ξεκινώντας από πίνακες δεδομένων και παρουσιάζοντας παραδείγματα). [40 λεπτά debate] - Συζήτηση με τους μαθητές σχετικά με τη μοντελοποίηση μη δομημένων δεδομένων (όπως εικόνες και κείμενα) για αυτόματη επεξεργασία. Ανάλυση εναλλακτικών λύσεων μαζί με τους μαθητές, με εισαγωγή σχετικών εννοιών όπως: σακίδιο λέξεων (bag of words), συνώνυμα, τοποθεσία του κειμένου (τίτλος, περίληψη κ.λπ.). [40 λεπτά, debate] - Εισαγωγή στην ομοιότητα εγγράφων και η σχετικότητα τους στην εξόρυξη κειμένου. [40 λεπτά, debate] - Μοντελοποίηση Κειμένου 1 [Μάθημα 2: 2 ώρες, 4 δραστηριότητες] - Αναφορά στη μοντελοποίηση κειμένου και έναρξη επίδειξης μιας συνάρτησης ομοιότητας κειμένων με χρήση του Δυαδικού Μοντέλου (Boolean model), παρουσίαση παραδειγμάτων και ερώτηση προς τους μαθητές: πώς μπορούμε να εντοπίσουμε ομοιότητες μεταξύ εγγράφων; [40 λεπτά παρουσίαση, πρακτική] - Εξήγηση του σώματος κειμένων (corpora) που θα χρησιμοποιηθεί (θα πρέπει να περιλαμβάνει μικρά κείμενα με πολύ περιορισμένο λεξιλόγιο· να περιέχει κείμενα στα οποία η Συχνότητα Όρων – TF – μπορεί να κάνει διαφορά σε σύγκριση με το Δυαδικό Μοντέλο – Boolean model). [20 λεπτά, συνεδρία παρουσίασης] - Εισαγωγή τελεστών συνόλου (ένωση, τομή, κ.λπ.). [20 λεπτά, παρουσίαση, επίδειξη] - Οι μαθητές, σε ομάδες των 3 ή 4, καλούνται να υλοποιήσουν μια συνάρτηση υπολογισμού ομοιότητας κειμένων στο Excel και να τη δοκιμάσουν· κάθε ομάδα εξηγεί ποιους τελεστές συνόλων χρησιμοποιεί και γιατί. [40 λεπτά, πρακτική] - Μοντελοποίηση κειμένου 2 [Μάθημα 3: 2 ώρες, 4 δραστηριότητες] - Εισαγωγή της διανυσματικής άλγεβρας: εσωτερικό γινόμενο και
--	---

- συνημίτονο. Διευκόλυνση συζήτησης για τις μαθηματικές αρχές πίσω από την επεξεργασία κειμένου, [20 λεπτά, παρουσίαση, επίδειξη]
- 3.2. Πρακτικές δραστηριότητες όπου οι μαθητές εξερευνούν τελεστές συνόλων και διανυσματική άλγεβρα για να υπολογίσουν την ομοιότητα μεταξύ κειμένων σε ένα σώμα κειμένων. Οι ομάδες συζητούν τη σχέση των τελεστών συνόλων με τη διανυσματική άλγεβρα (τομή με εσωτερικό γινόμενο, ένωση με άθροισμα). Ανασχεδιάζουν την Excel υλοποίηση χρησιμοποιώντας συνδυασμό τελεστών. [40 λεπτά, παρουσίαση]
- 3.3. Παρουσίαση της συνεισφοράς της συχνότητας όρων (TF) σε σχέση με το Boolean μοντέλο για την ομοιότητα κειμένων. [20 λεπτά, παρουσίαση]
- 3.4. Οι μαθητές ανασχεδιάζουν την υλοποίησή τους εφαρμόζοντας το μοντέλο TF και υπολογίζουν ομοιότητες στα ίδια corpora· το συνημίτονο αποδεικνύεται λειτουργικό. [40 λεπτά, παρουσίαση]
4. Μοντελοποίηση κειμένου 3 [Μάθημα 4: 2 ώρες, 3 δραστηριότητες]
- 4.1. Συζήτηση αλλαγών και προβλημάτων που προκύπτουν όταν η ομοιότητα βασίζεται μόνο στο TF (π.χ. κοινές λέξεις σε όλα τα κείμενα δεν έχουν διακριτική ισχύ). [20 λεπτά, παρουσίαση]
- 4.2. Οι μαθητές προτείνουν λύσεις. [60 λεπτά, παρουσίαση]
- 4.3. Εισαγωγή IDF και TFxIDF. Συζήτηση, σχεδιασμός και υλοποίηση ενός μέτρου ομοιότητας με βάση το μοντέλο TFxIDF για μικρά κείμενα με περιορισμένο λεξιλόγιο. [40 λεπτά, επίδειξη]
5. Εφαρμογή [Μάθημα 5: 2 ώρες, 3 δραστηριότητες]
- 5.1. Εισαγωγή στις βιβλιοθήκες R, Python ή άλλες για εξόρυξη κειμένου. [30 λεπτά, παρουσίαση]
- 5.2. Ρύθμιση περιβάλλοντος ανάπτυξης για εξόρυξη κειμένου. Ο εκπαιδευτικός έχει προετοιμάσει οδηγό και τον παρέχει στους μαθητές. [30 λεπτά, ομαδική εργασία]
- 5.3. Οι ομάδες υλοποιούν συνάρτηση ομοιότητας δύο εγγράφων σε R, Python ή άλλη γλώσσα. Στο τέλος της φάσης όλες οι ομάδες πρέπει να έχουν λειτουργική συνάρτηση (ή να δοθεί έτοιμη αν χρειαστεί). [60 λεπτά, πρακτική]
6. Τελικό έργο 1 [Μάθημα 6: 2 ώρες, 1 δραστηριότητα]
- 6.1. Οι ομάδες καλούνται να εντοπίσουν, ανασχεδιάσουν και υλοποιήσουν τον καλύτερο αλγόριθμο για συνάρτηση ομοιότητας. [120 λεπτά, έρευνα, ομαδική εργασία]
7. Τελικό έργο 2 [Μαθήματα 7 και 8: 4 ώρες, 1 δραστηριότητα]
- 7.1. Παροχή ενός επισημειωμένου σώματος εγγράφων για δοκιμαστικές ερωτήσεις/εγγραφές. Παροχή και του συνόλου επικύρωσης (π.χ. 2-3 έγγραφα με λίγους όρους ως ερωτήματα αναζήτησης). Οι μαθητές χρησιμοποιούν τις συναρτήσεις ομοιότητας τους για να κατατάξουν τα έγγραφα· υπολογίζεται το F1 για να ανακοινωθεί νικητής. Παράλληλα, διδάσκονται τεχνικές προεπεξεργασίας όπως stop-words, stemming. [240 λεπτά, ομαδική εργασία, πρακτική]
8. Σύνδεση με μηχανές αναζήτησης [Μάθημα 9: 2 ώρες, 2 δραστηριότητες]

	8.1. Παρουσίαση παραδείγματος όπου ένα “έγγραφο” είναι ένα ερώτημα όπως “camélias porto” ή “computer vision”. Επίδειξη μέσω Google. Εκτέλεση των συναρτήσεων ομοιότητας με τα ίδια queries. Υπολογισμός F1, εξήγηση του F1, Recall και Precision. [60 λεπτά, παρουσίαση]. [60 λεπτά παρουσίαση] 8.2. Συζήτηση, αναστοχασμός και συμπεράσματα. [60 λεπτά, παρουσίαση]
Αξιολόγηση	Αξιολόγηση και αναστοχασμός 1. Αξιολόγηση της κατανόησης και εφαρμογής των τελεστών άλγεβρας συνόλων και διανυσματικής άλγεβρας μέσω project-based εργασιών, παρουσιάσεων και γραπτών αναστοχασμών 2. Ενθάρρυνση των μαθητών να αναστοχαστούν τη μαθησιακή τους εμπειρία, αναδεικνύοντας τη σχέση μεταξύ μαθηματικών και εξόρυξης κειμένου. 3. Ζητήστε από τους μαθητές να σχεδιάσουν και να παρουσιάσουν μια μεθοδολογία και ένα μη λειτουργικό πρωτότυπο μηχανής αναζήτησης βασισμένο σε όσα έμαθαν.
Παρουσίαση - Αναφορά - Διαμοιρασμός	1. Μια συνάρτηση που υπολογίζει την ομοιότητα μεταξύ δύο εγγράφων κειμένου. 2. Μια παρουσίαση σε PowerPoint που περιγράφει μια μεθοδολογία και ένα μη λειτουργικό πρωτότυπο μιας νέας μηχανής αναζήτησης που προτείνεται από τους μαθητές, χρησιμοποιώντας όσα αναπτύχθηκαν στο εργαστήριο (τη συνάρτηση ομοιότητας εγγράφων).
Επεκτάσεις - Άλλες πληροφορίες	

STEAME ACADEMY Πρωτότυπο/Οδηγός για τη μάθηση και τη δημιουργικότητα Προσέγγιση
Διαμόρφωση σχεδίου δράσης

Σημαντικά βήματα της μαθησιακής προσέγγισης STEAME:

ΣΤΑΔΙΟ Ι: Προετοιμασία από έναν ή περισσότερους εκπαιδευτικούς

1. Διατύπωση αρχικών σκέψεων σχετικά με τους θεματικούς τομείς/περιοχές που πρέπει να καλυφθούν
2. Σύνδεση με την πραγματικότητα του ευρύτερου περιβάλλοντος / εργασία / επιχειρήσεις / γονείς / κοινωνία / περιβάλλον / ηθική
3. Ηλικιακή ομάδα-στόχος των μαθητών - Σύνδεση με το επίσημο πρόγραμμα σπουδών - Καθορισμός στόχων και σκοπών
4. Οργάνωση των εργασιών των εμπλεκόμενων μερών - Ορισμός Συντονιστή - Χώροι εργασίας κ.λπ.

ΣΤΑΔΙΟ ΙΙ: Διαμόρφωση σχεδίου δράσης (Βήματα 1-18)

Προετοιμασία (από τους εκπαιδευτικούς)

1. Σύνδεση με τον πραγματικό κόσμο - Αναστοχασμός
2. Κινητοποίηση - Κίνητρα
3. Διατύπωση ενός προβλήματος (ενδεχομένως σε στάδια ή φάσεις) που προκύπτει από τα παραπάνω

Ανάπτυξη (από τους μαθητές) - Καθοδήγηση και αξιολόγηση (στο 9-11, από τους εκπαιδευτικούς)

4. Ανάπτυξη υποβάθρου- Αναζήτηση / Συγκέντρωση πληροφοριών
5. Απλοποίηση του ζητήματος - Διαμόρφωση του προβλήματος σε έναν περιορισμένο αριθμό απαιτήσεων
6. Διατύπωση περίπτωσης - Σχεδιασμός - προσδιορισμός υλικών για οικοδόμηση / ανάπτυξη / δημιουργία
7. Κατασκευή - Ροή εργασιών – εφαρμογή των project
8. Παρατήρηση -Πειραματισμός - Αρχικά συμπεράσματα
9. Τεκμηρίωση - Αναζήτηση θεματικών περιοχών (πεδία ΤΝ) που σχετίζονται με το υπό μελέτη θέμα - Επεξήγηση με βάση τις υπάρχουσες θεωρίες ή/και τα εμπειρικά αποτελέσματα
10. Συγκέντρωση αποτελεσμάτων/πληροφοριών με βάση τα σημεία 7, 8, 9
11. Πρώτη ομαδική παρουσίαση από τους μαθητές

Διαμόρφωση και αποτελέσματα (από τους μαθητές) - Καθοδήγηση και αξιολόγηση (από τους εκπαιδευτικούς)

12. Διαμόρφωση μοντέλων STEAME για την περιγραφή / αναπαράσταση / απεικόνιση των αποτελεσμάτων
13. Μελέτη των αποτελεσμάτων στο σημείο 9 και εξαγωγή συμπερασμάτων, χρησιμοποιώντας το σημείο 12

14. Εφαρμογές στην καθημερινή ζωή - Προτάσεις για την ανάπτυξη του σημείου 9
(Επιχειρηματικότητα - SIL ημέρες)

Ανασκόπηση (από εκπαιδευτικούς)

15. Επανεξέταση του προβλήματος και επανεξέταση και υπό περισσότερο απαιτητικές συνθήκες

Ολοκλήρωση έργου (από τους μαθητές) - Καθοδήγηση και αξιολόγηση (από τους εκπαιδευτικούς)

16. Επανάληψη των βημάτων 5 έως 11 με πρόσθετες ή νέες απαιτήσεις όπως διατυπώνονται στο σημείο 15
17. Διερεύνηση - Μελέτες περιπτώσεων - Επέκταση - Νέες θεωρίες - Έλεγχος νέων συμπερασμάτων
18. Παρουσίαση συμπερασμάτων - Τακτικές επικοινωνίας.

ΣΤΑΔΙΟ III: STEAME ACADEMY Δράσεις και συνεργασία σε δημιουργικά έργα για μαθητές σχολείων

Τίτλος του έργου: _____

Σύντομη περιγραφή/προγραμματισμός των οργανωτικών διευθετήσεων/αρμοδιοτήτων για δράση

ΣΤΑΔΙΟ	Δραστηριότητες/Βήματα Εκπαιδευτικός 1(T1) Συνεργασία με τον T2 και καθοδήγηση των μαθητών	Δραστηριότητες /Βήματα Από Μαθητές Ηλικιακή ομάδα: _____	Δραστηριότητες /Βήματα Εκπαιδευτικός 2 (T2) Συνεργασία με T1 και καθοδήγηση των μαθητών
A	Προετοιμασία των σταδίων 1,2,3		Συνεργασία στο βήμα 3
B	Καθοδήγηση στο βήμα 9	4,5,6,7,8,9,10	Καθοδήγηση και υποστήριξη στο βήμα 9
C	Δημιουργική αξιολόγηση	11	Δημιουργική αξιολόγηση
D	Καθοδήγηση	12	Καθοδήγηση
E	Καθοδήγηση	13 (9+12)	Καθοδήγηση
F	Οργάνωση (SIL) STEAME στην καθημερινή ζωή	14 Συνάντηση με εκπροσώπους φορέων	Οργάνωση (SIL) STEAME στην καθημερινή ζωή
G	Προετοιμασία του βήματος 15		Συνεργασία στο βήμα 15
H	Καθοδήγηση	16 (επανάληψη 5-11)	Καθοδήγηση και υποστήριξη
I	Καθοδήγηση	17	Καθοδήγηση και υποστήριξη
K	Δημιουργική αξιολόγηση	18	Δημιουργική αξιολόγηση

--	--	--	--

