# ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΠΛΗΡΟΦΟΡΙΑΚΑ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΑ ΣΥΣΤΗΜΑΤΑ


**Ανάκτηση Απαντήσεων σε Μαθηματικά Προβλήματα**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της


Αναστασία Λεοντή


**Επιβλέπων :** Σταματάτος Ευστάθιος

**Μέλη εξεταστικής επιτροπής:**

Σάμος, [06 / 2023]

# Πίνακας περιεχομένων


**1 Εισαγωγή 7**
1.1 Ανάκτηση πληροφοριών 7
1.2 Μελέτη του διαγωνισμού ArqMath και στοχευμένη ανάλυση στην τεχνική “Approach 0” 8
1.3 Δομή της διπλωματικής 10
**2 Ανάκτηση πληροφορίας και δεδομένα 11**
2.1 Τι σημαίνει ανάκτηση πληροφορίας 11
2.2 Αδόμητα και δομημένα δεδομένα 12
2.3 Ταξινόμηση (Classification) 14
2.3.1 Ταξινόμηση κειμένου 14
2.3.4 Ταξινόμηση εγγράφων 14
2.3.5 Ιεραρχική ταξινόμηση 15
2.3.6 Ταξινόμηση ιστοσελίδων 15
Πλεονεκτήματα 15
Μειονεκτήματα 15
2.4 Κλίμακα 16
2.4.1 Ευρετηρίαση μεγάλης κλίμακας (Large-Scale Indexing): 16
2.4.2 Ανάκτηση κατανεμημένων πληροφοριών (Distributed Information Retrieval) 16
2.4.3 Αποδοτικότητα και καθυστέρηση (Efficiency and Latency) 16
2.4.4 Ποιότητα σε κλίμακα (Quality at scale) 17
2.4.5 Διαχείριση πόρων (Resource Management) 17
2.4.6 Η Κλίμακα αποτελεί βασικό στοιχείο σε διάφορα περιβάλλοντα. 17
2.4.6.1 Αναζήτηση διαδικτύου ‘web search’ 17
2.4.6.2 Ανάκτηση προσωπικών πληροφοριών ‘personal information retrieval’ 18
2.4.6.3 Ανάκτηση πληροφοριών σε εταιρίες, ακαδημαϊκά και συγκεκριμένου τομέα 19
**3 Αναζήτηση πληροφοριών στα μαθηματικά 20**
3.1 Στόχος αναζήτησης πληροφοριών Math IR 20
3.1.1 Πρόσβαση στη Μαθηματική Βιβλιογραφία 21
3.1.2 Σημασιολογική αναζήτηση 21
3.1.3 Συμβολική Κατανόηση 21
3.1.4 Διασταυρούμενη αναφορά 21
3.1.5 Ενίσχυση της συνεργασίας και της ανακάλυψης γνώσης 22
3.1.6 Εκπαιδευτική Υποστήριξη 22
3.2 Περιπτώσεις χρήσης 22
3.2.1 Κατανόηση Μαθηματικών Ερωτημάτων 22
3.2.3 Αναπαράσταση και Ευρετηρίαση Μαθηματικού Περιεχομένου 23
3.2.4 Συνάφεια κρίση και κατάταξη (Relevance Judgement and Ranking) 25
3.2.5 Αλληλεπίδραση και αξιολόγηση χρήστη 26
**4 Μέτρηση της σχετικότητας των αποτελεσμάτων 27**
4.1 Μέτρηση της σχετικότητας 27
4.2 Precision and Recall 27
4.3 F1 Score 28

4.4 Human evaluation 29
4.5 Embedding-based metrics 31
4.6 Mean Average Precision (MAP) 32
4.7 Precision at K (P@K) 33
4.8 NDCG score 34
**5 Οι διαγωνισμοί ArqMath 36**
5.1 Οι εργασίες του ARQMath 36
5.2 Μετρα σχετικότητας που χρησιμοποίησαν οι ομάδες 39
5.3 Θετικά και αρνητικά των μεθόδων μέτρησης σχετικότητας 41
5.3.1 Πως αξιολόγησαν τα αποτελέσματα οι αξιολογητές 42
5.4 Υπολογισμός μέτρων αποτελεσματικότητας στο ArqMath 43
Εν κατακλείδι 44
5.5 Task 1 45
5.6 Αποτελέσματα συστήματος των ομάδων 45
5.6.1 Πως προσέγγισαν το πρόβλημα οι ομάδες με τα καλύτερα αποτελέσματα 47
5.6.1.1 Η ομάδα TU_DBS 47
5.6.1.2 Η ομάδα MathDowsers 48
5.6.1.3 Η ομάδα MSM 51
5.7 Το σύστημα “Approach 0” 52
5.7.1 Πώς λειτουργεί η Zero-Pass Retrieval: 53
5.7.2 Πώς λειτουργεί το ColBERT: 54
5.7.3 Ένωση (Fusion) 55
**6 Μελέτη συστήματος “Approach 0” 57**
6.1 Πως λειτουργεί η σύγκριση ερωτήματος-δεδομένων 58
6.1.1 Η συνάρτηση msearch 58
6.1.2 Υπηρεσία αναζήτησης που χρησιμοποιεί το msearch: 59
6.1.3 Πως λειτουργεί η βιβλιοθήκη pya0 59
6.2 Τροποποίηση της λογικής ή του μοντέλου της μηχανής αναζήτησης 60
6.2.1 Τροποποίηση της συνάρτησης msearch 60
6.2.1.1 Αλλαγή του αλγόριθμου αναζήτησης: 60
6.2.1.2 Τροποποίηση παραμέτρων αναζήτησης: 61
6.2.1.3 Αποτελέσματα μετά τη διαδικασία: 62
6.2.2 Τροποποίηση της συνάρτησης task1_getresults 62
6.2.3 Τροποποίηση της συνάρτησης arqmath_to_prim_task1 63
6.2.3.1 Αλλαγή στον αλγόριθμο ωστε να μην αποκλείει τα topic_ids αν υπάρχουν 64
6.2.3.2Εισαγωγή νέας παραμέτρου. της -pic 68
6.2.4 Τροποποίηση της συνάρτησης indexer 72
**7 Συμπεράσματα 75**
7.1 Προτεινόμενο εύρος καινοτόμων τεχνικών στην ανάκτηση μαθηματικών πληροφοριών 75
7.2 Περιοχές που θα βοηθήσουν την εξέλιξη στην ανάκτηση πληροφοριών 76
**Βιβλιογραφία 79**

# Ακρωνύμια

| | |
|---|---|
| IR | Information Retrieval / Ανάκτηση Πληροφοριών |
| MIR | Μαθηματική Ανάκτηση Πληροφοριών |
| CLEF | Συνέδριο και τα Εργαστήρια του Φόρουμ Αξιολόγησης |
| MAP | Mean Average Precision / Ακρίβεια μέσου όρου |
| F1 score | Βαθμολογία F1 |
| nDCG | Κκανονικοποιημένο εκπτωτικό άθροισμα κέρδους |
| SQL | Δομημένη γλώσσα ερωτημάτων |
| NLP | Επεξεργασία Φυσικής Γλώσσας |
| CNN | Συνελικτικά Νευρωνικά Δίκτυα |
| RNN | Επαναλαμβανόμενα Νευρωνικά Δίκτυα |
| MOOCs | Massive Open Online Courses |
| SLT | Δέντρα διάταξης συμβόλων |
| OPT | Δέντρα χειριστή |

# Περίληψη

Η παρούσα διατριβή εστιάζει σε τεχνικές ανάκτησης πληροφοριών που είναι προσαρμοσμένες ειδικά για μαθηματικά προβλήματα. Το μαθηματικό περιεχόμενο παρουσιάζει μοναδικές προκλήσεις στην ανάκτηση πληροφοριών λόγω της πολύπλοκης δομής, των συμβολικών αναπαραστάσεων και του εξειδικευμένου λεξιλογίου του. Ο στόχος αυτής της έρευνας είναι να μελετήσει υπάρχουσες μεθόδους με στόχο την ενίσχυση της ανάκτησης σχετικών μαθηματικών πληροφοριών, βοηθώντας τους χρήστες να βρουν ακριβείς λύσεις, εξηγήσεις και σχετικούς πόρους.
Συγκεκριμένα, αυτή η διατριβή εστιάζει στη βελτίωση της αποτελεσματικότητας των συστημάτων ανάκτησης πληροφοριών για μαθηματικά προβλήματα αξιοποιώντας το σύνολο δεδομένων ARQMath. Το σύνολο δεδομένων ARQMath παρέχει έναν πολύτιμο πόρο που καταγράφει την ποικιλόμορφη φύση των μαθηματικών ερωτήσεων, των απαντήσεων και των σχετικών μεταδεδομένων. Αυτή η διατριβή διερευνά νέες προσεγγίσεις και τεχνικές για τη βελτίωση της απόδοσης ανάκτησης στον τομέα της επίλυσης μαθηματικών προβλημάτων.

Το πρώτο μέρος της διατριβής περιλαμβάνει μια εις βάθος ανάλυση της ανάκτησης πληροφοριών και στη συνέχεια αναλύει τη μέτρηση της σχετικότητας σε ένα γενικό επίπεδο αλλά και πιο συγκεκριμένο με στόχο να καταλήξει στη λειτουργία του διαγωνισμού ARQMath CLEF. Περιλαμβάνεται μια ανάλυση του διαγωνισμού ARQMath αλλά και των ομάδων με πιο συγκεκριμένη προσοχή στην προσέγγιση της ομάδας “Approach 0”. Στη συνέχεια υπάρχουν διαφορετικές προσεγγίσεις και αλλαγές στην προσέγγιση της ομάδας και παρατήρηση των αποτελεσμάτων με ανάλυση στις διαφορές των αποδόσεων nDCG’, mAP’, p@10’ και BPref. Αυτή η ανάλυση θέτει τα θεμέλια για την κατανόηση της πολυπλοκότητας της μαθηματικής ανάκτησης πληροφοριών και τον εντοπισμό βασικών τομέων για βελτίωση.

Με βάση την ανάλυση δεδομένων, η διατριβή προτείνει μια σειρά από υπάρχουσες τεχνικές για την αντιμετώπιση των προκλήσεων της μαθηματικής ανάκτησης πληροφοριών.

# 1 *Εισαγωγή*

## *1.1 Ανάκτηση πληροφοριών*

Η ψηφιακή εποχή έχει εγκαινιάσει μια άνευ προηγουμένου έκρηξη δεδομένων, που εκτείνεται σε κάθε πιθανό τομέα, από την επιστήμη έως τις ανθρωπιστικές επιστήμες, από τις παγκόσμιες επιχειρήσεις έως την ατομική δραστηριότητα των μέσων κοινωνικής δικτύωσης. Μέσα σε αυτή την απέραντη θάλασσα πληροφοριών, η ανάγκη εντοπισμού σχετικών και ακριβών δεδομένων έχει αναδειχθεί ως θεμελιώδης και ολοένα και πιο περίπλοκη πρόκληση. Αυτή η εργασία είναι μια εξερεύνηση της Ανάκτησης Πληροφοριών, μιας βασικής επιστήμης στο επίκεντρο αυτής της πρόκλησης.

Η ανάκτηση πληροφοριών, στον πυρήνα της, ασχολείται με την οργάνωση και την εξαγωγή σχετικών πληροφοριών από ένα μεγαλύτερο σώμα. Αυτό περιλαμβάνει, αλλά δεν περιορίζεται σε, την αναζήτηση σχετικών εγγράφων μέσα σε μια βάση δεδομένων, τον εντοπισμό πληροφοριών μέσα σε έγγραφα, ακόμη και την αναγνώριση μεταδεδομένων που παρέχουν το πλαίσιο για άλλα δεδομένα [6]. Η ανάκτηση πληροφοριών (Information Retrieval η αλλιώς IR) είναι θεμελιώδες για πολλές από τις καθημερινές μας δραστηριότητες, τροφοδοτώντας τις μηχανές αναζήτησης ιστού, τις ψηφιακές βιβλιοθήκες, τα συστήματα συστάσεων ηλεκτρονικού εμπορίου και πολλά άλλα.

Ενώ η εφαρμογή του πεδίου είναι ευρέως αναγνωρίσιμη, οι αλγόριθμοι και οι αρχές σχεδιασμού είναι πολύπλοκες και πολύπλευρες. Αυτή η εργασία στοχεύει να απομυθοποιήσει αυτά τα στοιχεία, παρέχοντας μια σαφή, περιεκτική και προσβάσιμη επισκόπηση της Ανάκτησης Πληροφοριών.

Ξεκινάμε διερευνώντας τα βασικά του IR, συμπεριλαμβανομένης της έννοιας της «συνάφειας» και των παραδοσιακών μοντέλων ανάκτησης πληροφοριών. Στη συνέχεια ταξιδεύουμε στον κόσμο της ευρετηρίασης, εξερευνώντας πώς τα συστήματα IR διαχειρίζονται και δομούν

τεράστιες αποθήκες δεδομένων για αποτελεσματική αναζήτηση. Μια βαθιά κατάδυση στους κεντρικούς αλγόριθμους και τις τεχνικές που χρησιμοποιούνται στην ανάκτηση πληροφοριών θα παρέχει μια κατανόηση του τρόπου με τον οποίο αυτά τα συστήματα ταξινομούν και ανακτούν δεδομένα.

Η εργασία θα επισημάνει επίσης τις εγγενείς προκλήσεις που αντιμετωπίζει η Ανάκτηση Πληροφοριών, όπως ο χειρισμός της ασάφειας, η αντιμετώπιση της δυναμικής φύσης των δεδομένων και η αντιμετώπιση της υποκειμενικότητας της συνάφειας. Θα διερευνήσουμε επίσης πώς η πρόοδος στην τεχνολογία και την έρευνα καθοδηγεί το μέλλον του IR, με ιδιαίτερη έμφαση στη μηχανική μάθηση και τον αυξανόμενο ρόλο της τεχνητής νοημοσύνης στη βελτίωση της ακρίβειας ανάκτησης και στην εξατομίκευση των εμπειριών αναζήτησης.

Μέσω αυτής της μελέτης, στόχος μας είναι να παρέχουμε μια ευρεία, αλλά και λεπτομερή, κατανόηση της Ανάκτησης Πληροφοριών, εξοπλίζοντας έτσι τους αναγνώστες με την εκτίμηση των εξελιγμένων διαδικασιών που οδηγούν τις καθημερινές μας αλληλεπιδράσεις με τα συστήματα πληροφοριών.[30]

.

## *1.2 Μελέτη του διαγωνισμού ArqMath και στοχευμένη ανάλυση στην τεχνική "Approach 0"*

Η διαρκώς αυξανόμενη συλλογήμαθηματικών γνώσεων, που διατηρείται σε άρθρα, σχολικά βιβλία, ιστοσελίδες και πολλά άλλα, παρουσιάζει μια σημαντική πρόκληση: πώς να ανακτήσετε αποτελεσματικά και αποδοτικά τις μαθηματικές πληροφορίες που χρειάζεται ένα άτομο ή ένα σύστημα σε μια δεδομένη στιγμή. Αυτή η εργασία στοχεύει να φωτίσει τις πολυπλοκότητες και τις δυνατότητες της Ανάκτησης Πληροφοριών στη σφαίρα των μαθηματικών, μια εργασία που απαιτεί συγκεκριμένες τροποποιήσεις σε σύγκριση με τις πιο συμβατικές IR βασισμένες σε κείμενο.

Η Μαθηματική Ανάκτηση Πληροφοριών έχει το ουσιαστικό έργο της κατανόησης, αναπαράστασης και ανάκτησης μαθηματικά πλούσιου περιεχομένου, συμπεριλαμβανομένων

εξισώσεων, αποδείξεων, θεωρημάτων και μαθηματικών συμβόλων, που συχνά ενσωματώνονται σε κειμενικά πλαίσια [27]. Είτε είναι ένας ερευνητής που αναζητά μια συγκεκριμένη μαθηματική απόδειξη, ένας μαθητής που αναζητά εξηγήσεις για μια μαθηματική έννοια ή ένα σύστημα τεχνητής νοημοσύνης που στοχεύει να απαντήσει σε ερωτήματα που σχετίζονται με τα μαθηματικά, η ζήτηση για αποτελεσματικά συστήματα MIR είναι αναμφισβήτητα κρίσιμη και αυξάνεται.

Το Συνέδριο και τα Εργαστήρια του Φόρουμ Αξιολόγησης (CLEF) φιλοξενούν τον διαγωνισμό ARQMath, μια πρόκληση βασισμένη σε εργασίες που αποσκοπεί στην τόνωση της προόδου στο MIR. Παρέχοντας μια κοινή πλατφόρμα και σύνολο δεδομένων, το ARQMath προτρέπει τους ερευνητές και τους προγραμματιστές να σχεδιάσουν και να αξιολογήσουν μεθόδους MIR, ωθώντας τα όρια του τι είναι δυνατό σε αυτόν τον εξειδικευμένο τομέα [29].

Σε αυτη τη διπλωματική εργασία, εμβαθύνουμε στις μοναδικές πτυχές του μαθηματικού IR, διερευνώντας τις προκλήσεις και τις τρέχουσες μεθόδους που χρησιμοποιούνται για την αναπαράσταση, την ευρετηρίαση και την ανάκτηση μαθηματικού περιεχομένου. Στη συνέχεια εμβαθύνουμε στις ιδιαιτερότητες του διαγωνισμού ARQMath, στη δομή εργασιών του και στις αξιοσημείωτες προσεγγίσεις που χρησιμοποιούνται από τις συμμετέχουσες ομάδες. Εξετάζοντας τόσο επιτυχημένες όσο και λιγότερο αποτελεσματικές στρατηγικές, μπορούμε να αντλήσουμε σημαντικά μαθήματα σχετικά με την αποτελεσματικότητα στα μαθηματικά IR.

Στη συνέχεια στοχεύουμε στην εργασία της ομάδας “Approach 0” διερευνώντας τα αποτελέσματα και μελετώντας μελετώντας την επίδραση συγκεκριμένων παραμέτρων στην επίδοσή της [33]. Στη συνέχεια αναλύουμε κομμάτι προς κομμάτι τα ενδιαφέροντα σημεία της εργασίας και παρατηρούμε τις αλλαγές στα αποτελέσματα.

Επιπλέον, στοχεύουμε να συμβάλουμε στην εξέλιξη του διαγωνισμού ARQMath και του ευρύτερου πεδίου του MIR. Προτείνοντας βελτιώσεις στις τρέχουσες τεχνικές, προτείνοντας νέες προσεγγίσεις και επισημαίνοντας ανεξερεύνητες περιοχές, αυτό το έγγραφο προσπαθεί να παρέχει έναν οδικό χάρτη για μελλοντική έρευνα και ανάπτυξη στην ανάκτηση μαθηματικών πληροφοριών [32].

## *1.3 Δομή της διπλωματικής*

Στο Κεφάλαιο 2 γίνεται μια ανάλυση για την ανάκτηση πληροφορίας και τα δεδομένα και παρουσιάζονται οι κατηγορίες τους.

Το Κεφάλαιο 3 στοχεύει στην ανάκτηση πληροφορίας στα Μαθηματικά και τις περιπτώσεις χρήσης του.

Στο Κεφάλαιο 4 αναπτύσσουμε τη μέτρηση της σχετικότητας των αποτελεσμάτων και με μία ανάλυση για κάποια μέτρα που είναι σχετικά με το στόχο της εργασίας.

Στο Κεφάλαιο 5 αναλύουμε τον διαγωνισμό ArqMath@CLEF , τις απαιτήσεις του διαγωνισμού και την προσέγγιση της ενδιαφερόμενης ομάδας.

Στο Κεφάλαιο 6 ξεκινάει και ολοκληρώνεται το πρακτικό κομμάτι με ουσιαστικές αλλαγές στην εργασία της ομάδας “Approach 0” αλλά και παρατηρήσεις για τη διαφορετικότητα των αποτελεσμάτων.

Στο Κεφάλαιο 7 ολοκληρώνουμε με προτάσεις εξέλιξης του τομέα.

# 2 *Ανάκτηση πληροφορίας και δεδομένα*

## *2.1 Τι σημαίνει ανάκτηση πληροφορίας*

Η ανάκτηση πληροφοριών είναι ένα θεμελιώδες πεδίο που ασχολείται με τη διαδικασία ανάκτησης σχετικών πληροφοριών από μεγάλες συλλογές δεδομένων, όπως έγγραφα, ιστοσελίδες ή περιεχόμενο πολυμέσων. Με τον συνεχώς αυξανόμενο όγκο των διαθέσιμων ψηφιακών πληροφοριών, η αποτελεσματική ανάκτηση πληροφοριών έχει καταστεί ζωτικής σημασίας σε διάφορους τομείς, συμπεριλαμβανομένων των μηχανών αναζήτησης, των ψηφιακών βιβλιοθηκών, του ηλεκτρονικού εμπορίου και της ανάλυσης δεδομένων.

Η ανάκτηση πληροφοριών περιλαμβάνει την εύρεση και παρουσίαση σχετικών πληροφοριών στους χρήστες με βάση τις ανάγκες τους σε πληροφορίες. Ο πρωταρχικός στόχος είναι η αντιστοίχιση των ερωτημάτων των χρηστών με σχετικά έγγραφα ή πόρους, λαμβάνοντας υπόψη παράγοντες όπως η συνάφεια, η κατάταξη και η ικανοποίηση των χρηστών. Η διαδικασία περιλαμβάνει τυπικά μηχανισμούς ευρετηρίασης, αναζήτησης και κατάταξης για την αποτελεσματική ανάκτηση και παρουσίαση των επιθυμητών πληροφοριών.

Έχουν αναπτυχθεί διάφορα μοντέλα ανάκτησης για την τυποποίηση της διαδικασίας ανάκτησης πληροφοριών. Το Boolean μοντέλο, το διανυσματικό μοντέλο χώρου, τα πιθανοτικά μοντέλα (π.χ., BM25) και τα μοντέλα γλώσσας (π.χ., Okapi BM25, Dirichlet Language Model) είναι ευρέως χρησιμοποιούμενα μοντέλα ανάκτησης που εκχωρούν βαθμολογίες συνάφειας ή πιθανότητες σε έγγραφα με βάση τις αντιστοιχίσεις ερωτήματος-εγγράφου(query-document) και τα σχετικά κριτήρια [6].

Για να διευκολυνθεί η αποτελεσματική ανάκτηση, χρησιμοποιούνται τεχνικές ευρετηρίασης για την οργάνωση και την αναπαράσταση του περιεχομένου σε μορφή αναζήτησης. Η ανεστραμμένη ευρετηρίαση είναι μια κοινή μέθοδος όπου οι όροι ή τα χαρακτηριστικά ευρετηριάζονται μαζί με τα σχετικά έγγραφα ή εγγραφές. Η αναζήτηση περιλαμβάνει την αντιστοίχιση ερωτημάτων χρήστη με το περιεχόμενο ευρετηρίου χρησιμοποιώντας κατάλληλους αλγόριθμους, όπως αντιστοίχιση με βάση όρους, αντιστοίχιση φράσεων ή σχόλια σχετικότητας [6].

Τα μέτρα αξιολόγησης χρησιμοποιούνται για την αξιολόγηση της αποτελεσματικότητας των συστημάτων ανάκτησης πληροφοριών. Τα κοινά μέτρα περιλαμβάνουν την ακρίβεια, την ανάκληση, το μέτρο F1, τη MAP, το κανονικοποιημένο εκπτωτικό άθροισμα κέρδους (nDCG) και τις καμπύλες ακριβείας-ανάκλησης. Αυτά τα μέτρα βοηθούν στην αξιολόγηση της συνάφειας και της κατάταξης των ανακτημένων αποτελεσμάτων σε σύγκριση με μια βασική αλήθεια ή ανθρώπινες κρίσεις.

Η ανάκτηση πληροφοριών αντιμετωπίζει πολλές προκλήσεις, συμπεριλαμβανομένου του προβλήματος της ασάφειας των ερωτημάτων, του χειρισμού συλλογών μεγάλης κλίμακας, της ενσωμάτωσης των σχολίων από τους χρήστες και της αντιμετώπισης του σημασιολογικού χάσματος μεταξύ των ερωτημάτων των χρηστών και του περιεχομένου που έχει καταχωρηθεί στο ευρετήριο. Η πρόοδος στην επεξεργασία φυσικής γλώσσας και της βαθιάς μάθησης έχουν ανοίξει το δρόμο για πιο εξελιγμένα μοντέλα και τεχνικές ανάκτησης που μπορούν να χειριστούν περίπλοκα ερωτήματα, σημασιολογικές σχέσεις και πολυτροπικά δεδομένα.[6][17]

## *2.2 Αδόμητα και δομημένα δεδομένα*

Δομημένα Δεδομένα (Structured Data)

Τα δομημένα δεδομένα είναι καλά οργανωμένα και μορφοποιημένα με τρόπο που να είναι εύκολα αναζητήσιμα σε σχεσιακές βάσεις δεδομένων. Αυτή η κατηγορία δεδομένων μπορεί να αποθηκευτεί, να ερωτηθεί και να αναλυθεί αβίαστα. Οι παραδοσιακές βάσεις δεδομένων που περιέχουν πίνακες με γραμμές και στήλες, αρχεία CSV και υπολογιστικά φύλλα είναι κοινά παραδείγματα δομημένων δεδομένων. Κάθε πίνακας περιέχει μία ή περισσότερες κατηγορίες δεδομένων σε στήλες και κάθε σειρά περιέχει μια συγκεκριμένη παρουσία δεδομένων για τις κατηγορίες που ορίζονται από τις στήλες [19].

Πλεονεκτήματα

Είναι πιο εύκολο να αναζητηθούν και να αναλυθούν λόγω της οργανωμένης μορφής τους.

Η δομημένη γλώσσα ερωτημάτων (SQL) επιτρέπει σε πολύπλοκα ερωτήματα να εξάγουν συγκεκριμένες πληροφορίες.

Οι αλλαγές στη δομή μπορούν να αντιμετωπιστούν χωρίς να επηρεαστούν τα υπάρχοντα ερωτήματα ή οι αναφορές[19].

Μειονεκτήματα

Απαιτεί σημαντικό όγκο εργασιών εισαγωγής και εγκατάστασης για τη διατήρηση της δομής και γι'αυτο το λόγο μπορεί να μην καταγράφει όλα τα δεδομένα λόγω άκαμπτης μορφοποίησης.

Η αναντιστοιχία μεταξύ οντοτήτων του πραγματικού κόσμου και του μοντέλου δεδομένων μπορεί να προκαλέσει προβλήματα[19].

Μη δομημένα δεδομένα(unstructured data)

Από την άλλη πλευρά, τα μη δομημένα δεδομένα δεν ακολουθούν μια συγκεκριμένη μορφή ή οργάνωση. Είναι συνήθως πλούσια σε κείμενο, αλλά μπορεί επίσης να περιέχει δεδομένα όπως ημερομηνίες, αριθμούς και γεγονότα. Παραδείγματα περιλαμβάνουν αρχεία κειμένου, email, αναρτήσεις μέσων κοινωνικής δικτύωσης, αρχεία ήχου, αρχεία βίντεο και ιστοσελίδες.

Πλεονεκτήματα

Είναι ευέλικτα και δεν απαιτεί προκαθορισμένο σχήμα για αποθήκευση.

Μπορεί να συλλάβει αποχρώσεις της ανθρώπινης γλώσσας και επικοινωνίας και αποτελεί σημαντικό ποσοστό των παγκόσμιων δεδομένων (κάποιες εκτιμήσεις υποδηλώνουν έως και 80-90%)[20].

Μειονεκτήματα

Μπορεί να είναι δύσκολη η εξαγωγή, η κατανόηση και η ανάλυση λόγω της μη οργανωμένης φύσης του.

Απαιτούνται πιο εξελιγμένα εργαλεία και τεχνικές, όπως η Επεξεργασία Φυσικής Γλώσσας (NLP), για αποτελεσματική επεξεργασία και ανάλυση.

Η αποθήκευση και η ανάκτηση μη δομημένων δεδομένων μπορεί να απαιτεί ένταση πόρων[20].

Συμπέρασμα

Παρά τις προκλήσεις, έχουν γίνει πολλές πρόοδοι για τη βελτίωση της ανάκτησης πληροφοριών τόσο από δομημένα όσο και από μη δομημένα δεδομένα. Τεχνικές όπως η μηχανική εκμάθηση, το NLP και οι βελτιωμένες τεχνολογίες βάσεων δεδομένων έχουν αυξήσει σημαντικά την ικανότητα εξαγωγής χρήσιμων πληροφοριών και από τους δύο τύπους δεδομένων[6].

## *2.3 Ταξινόμηση (Classification)*

Η ταξινόμηση στην ανάκτηση πληροφοριών είναι η διαδικασία κατηγοριοποίησης ή επισήμανσης εγγράφων ή άλλων ενοτήτων πληροφοριών. Μπορεί να παίξει κρίσιμο ρόλο σε διάφορες εργασίες όπως το φιλτράρισμα ανεπιθύμητων μηνυμάτων, η δρομολόγηση εγγράφων σε κατάλληλους προορισμούς και η οργάνωση περιεχομένου για περιήγηση. Οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται συχνά για εργασίες αυτόματης ταξινόμησης.

**Ανάλυση Ταξινόμησης στην Ανάκτηση Πληροφοριών**

### *2.3.1 Ταξινόμηση κειμένου*

Ταξινόμηση κειμένου είναι η διαδικασία απόδοσης προκαθορισμένων κατηγοριών (ετικέτες) σε έγγραφα κειμένου. Είναι ένα βασικό συστατικό των συστημάτων ανάκτησης πληροφοριών, που χρησιμοποιείται για φιλτράρισμα ανεπιθύμητων μηνυμάτων, ανάλυση συναισθημάτων και επισήμανση θεμάτων. Οι τεχνικές που χρησιμοποιούνται για την ταξινόμηση κειμένου κυμαίνονται από παραδοσιακές μεθόδους όπως οι Naive Bayes έως πιο προηγμένες τεχνικές βαθιάς μάθησης όπως τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) και τα Νευρωνικά Δίκτυα με ανατροφοδότηση (RNN)[21].

### *2.3.4 Ταξινόμηση εγγράφων*

Είναι παρόμοια με την ταξινόμηση κειμένου, αλλά εστιάζει συγκεκριμένα στην κατηγοριοποίηση των εγγράφων (τα οποία μπορεί να περιέχουν μη κειμενικά στοιχεία όπως εικόνες, γραφήματα κ.λπ.) με βάση το περιεχόμενό τους. Η ταξινόμηση εγγράφων έχει ευρείες εφαρμογές, συμπεριλαμβανομένης της ηλεκτρονικής ανακάλυψης για νομικές διαδικασίες, της αυτόματης κατηγοριοποίησης ειδήσεων και της ταξινόμησης ιατρικών αρχείων[22].

### *2.3.5 Ιεραρχική ταξινόμηση*

Σε αυτήν την ταξινόμηση, οι κατηγορίες οργανώνονται σε μια ιεραρχική δομή, συνήθως ένα δέντρο ή ένα κατευθυνόμενο άκυκλο γράφημα. Αυτό είναι ιδιαίτερα χρήσιμο όταν υπάρχει ασχολία με μεγάλο αριθμό τάξεων, μερικές από τις οποίες μπορεί να μοιάζουν πολύ μεταξύ τους. Η ιεραρχική ταξινόμηση μπορεί επίσης να είναι πιο αποτελεσματική όσον αφορά τον υπολογισμό και να παρέχει πιο ερμηνεύσιμα αποτελέσματα[23].

### *2.3.6 Ταξινόμηση ιστοσελίδων*

Αυτός ο τύπος ταξινόμησης εστιάζει στην ταξινόμηση ιστοσελίδων, οι οποίες συχνά περιέχουν ένα συνδυασμό περιεχομένου κειμένου, οπτικού περιεχομένου και μεταδεδομένων. Οι τεχνικές συχνά λαμβάνουν υπόψη τους υπερσυνδέσμους μεταξύ ιστοσελίδων και τη δομή του ίδιου του Ιστού[24].

### *Πλεονεκτήματα*

Η ταξινόμηση βοηθά στην οργάνωση μεγάλων συνόλων μη δομημένων δεδομένων, διευκολύνοντας τους χρήστες να βρίσκουν σχετικές πληροφορίες.

Μπορεί να διευκολύνει την ανάκτηση πληροφοριών βελτιώνοντας την ακρίβεια αναζήτησης.

Η αυτοματοποιημένη ταξινόμηση μπορεί να εξοικονομήσει σημαντικά χρόνο και πόρους σε σχέση με τη μη αυτόματη κατηγοριοποίηση.

### *Μειονεκτήματα*

Η ακρίβεια ταξινόμησης μπορεί να ποικίλλει ανάλογα με την ποιότητα των δεδομένων και το επιλεγμένο μοντέλο.

Η εκπαίδευση ενός ταξινομητή σε ένα μεγάλο σύνολο δεδομένων μπορεί να είναι υπολογιστικά εντατική.

Μπορεί να προκύψει εσφαλμένη ταξινόμηση, ιδιαίτερα σε περιπτώσεις όπου οι κατηγορίες δεν είναι καλά καθορισμένες ή αλληλοεπικαλύπτονται[21].

## *2.4 Κλίμακα*

Η κλίμακα στην ανάκτηση πληροφοριών αναφέρεται στο μέγεθος των δεδομένων που πρόκειται να υποβληθούν σε επεξεργασία για την ανάκτηση σχετικών πληροφοριών. Καθώς ο όγκος των ψηφιακών δεδομένων συνεχίζει να αυξάνεται εκθετικά, η επεκτασιμότητα έχει γίνει ένα κρίσιμο μέλημα για το σχεδιασμό αποτελεσματικών και αποδοτικών συστημάτων IR.

**Ανάλυση της Κλίμακας στην Ανάκτηση Πληροφοριών.**

### *2.4.1 Ευρετηρίαση μεγάλης κλίμακας (Large-Scale Indexing):*

Η καρδιά ενός συστήματος είναι το ευρετήριό του, μια δομή δεδομένων που βελτιώνει την ταχύτητα των λειτουργιών ανάκτησης δεδομένων. Για μεγάλα σύνολα δεδομένων, η δημιουργία και η διατήρηση αυτών των ευρετηρίων παρουσιάζει σημαντικές προκλήσεις. Τα ανεστραμμένα ευρετήρια, τα κατανεμημένα ευρετήρια και τα κλιμακωτά ευρετήρια είναι μερικές μέθοδοι που χρησιμοποιούνται για τη διαχείριση ευρετηρίασης μεγάλης κλίμακας[25].

### *2.4.2 Ανάκτηση κατανεμημένων πληροφοριών (Distributed Information Retrieval)*

Καθώς τα σύνολα δεδομένων αυξάνονται πέρα από τη χωρητικότητα μεμονωμένων μηχανών, τα κατανεμημένα συστήματα ανάκτησης πληροφοριών καθίστανται απαραίτητα. Αυτά τα συστήματα περιλαμβάνουν τη διανομή των δεδομένων και την επεξεργασία σε πολλούς κόμβους, συχνά σε περιβάλλον συμπλέγματος ή cloud. Για το σκοπό αυτό χρησιμοποιούνται συνήθως τεχνικές όπως το MapReduce και πλατφόρμες όπως το Hadoop και το Spark[17].

### *2.4.3 Αποδοτικότητα και καθυστέρηση (Efficiency and Latency)*

Καθώς η κλίμακα αυξάνεται, καθίσταται ζωτικής σημασίας να διατηρηθεί η αποτελεσματικότητα της ανάκτησης και να ελαχιστοποιηθεί η καθυστέρηση. Τεχνικές όπως η βελτιστοποίηση ερωτημάτων, η προσωρινή αποθήκευση και ο προ-υπολογισμός χρησιμοποιούνται για την αντιμετώπιση αυτών των ζητημάτων[6].

### *2.4.4 Ποιότητα σε κλίμακα (Quality at scale)*

Η διατήρηση της ποιότητας των αποτελεσμάτων ανάκτησης σε μεγάλη κλίμακα είναι μια πρόκληση. Καθώς το μέγεθος του συνόλου δεδομένων αυξάνεται, υπάρχει κίνδυνος αυξημένου θορύβου και μειωμένης ακρίβειας στα αποτελέσματα αναζήτησης. Οι τεχνικές μηχανικής μάθησης και τεχνητής νοημοσύνης χρησιμοποιούνται όλο και περισσότερο για τη βελτίωση της ποιότητας των αποτελεσμάτων[26].

### *2.4.5 Διαχείριση πόρων (Resource Management)*

Τα μεγαλύτερα σύνολα δεδομένων απαιτούν περισσότερους υπολογιστικούς πόρους και χωρητικότητα αποθήκευσης, με αποτέλεσμα αυξημένο κόστος. Οι αποτελεσματικές στρατηγικές διαχείρισης πόρων είναι απαραίτητες για την επεκτασιμότητα[26].

### *2.4.6 Η Κλίμακα αποτελεί βασικό στοιχείο σε διάφορα περιβάλλοντα.*

#### *2.4.6.1 Αναζήτηση διαδικτύου 'web search'*

Ένα σύστημα IR σε μια αναζήτηση διαδικτύου παρουσιάζει αποτελέσματα αναζήτησης από δισεκατομμύρια αρχεία σε όλο τον κόσμο. Ένα από τα βασικότερα ζητήματα στην αναζήτηση στο διαδίκτυο είναι η σωστή αρίθμηση των αποτελεσμάτων ανάλογα με την αναζήτηση του χρήστη. Ένα ιδανικό σύστημα θα πρέπει να είναι σε θέση να επιστρέφει αποτελέσματα που ικανοποιούν στο μέγιστο την αναζήτηση του χρήστη σε ένα τόσο μεγάλο σκέλος όπως τα δισεκατομμύρια αποτελέσματα που έχει να χειριστεί. Θα πρέπει να μπορεί να χειριστεί κάποια ιδιαίτερα σημεία όπως την εκμετάλλευση του υπερκειμένου 'hypertext' αλλά και να μπορεί να χειρίζεται την εξαπάτηση περιεχομένου ιστοσελίδων σωστά, το οποίο πολλοί προσπαθούν να εντάξουν μιας και εξαπατούν την κατάταξη τους στον παγκόσμιο ιστό μέσω των κλικ, προστατεύοντας τον χρήστη από λανθασμένη ανάκτηση δεδομένων και είσοδο σε ιστοσελίδες που λανθασμένα βρίσκονται στο αποτέλεσμα της αναζήτησης τους. [25]

Βασικά χαρακτηριστικά της μεθόδου

Ευρετηρίαση μεγάλης κλίμακας: Οι αποτελεσματικές στρατηγικές ευρετηρίασης, όπως η κατανεμημένη ευρετηρίαση και η κλιμακωτή ευρετηρίαση, είναι ζωτικής σημασίας για τη διαχείριση τεράστιων ποσοτήτων δεδομένων ιστού[25].

Αποδοτικότητα ερωτημάτων: Ο χειρισμός εκατομμυρίων ταυτόχρονων ερωτημάτων με χαμηλό λανθάνοντα χρόνο είναι μια σημαντική πρόκληση. Για την αντιμετώπιση αυτού του προβλήματος χρησιμοποιούνται τεχνικές όπως η προσωρινή αποθήκευση, η βελτιστοποίηση ερωτημάτων και ο προ-υπολογισμός των αποτελεσμάτων των ερωτημάτων[25].

Ποιοτικός έλεγχος: Με ένα τόσο τεράστιο σώμα, η διατήρηση της ποιότητας και της συνάφειας των αποτελεσμάτων αναζήτησης είναι δύσκολη. Για τη βελτίωση της ποιότητας των αποτελεσμάτων χρησιμοποιούνται προηγμένοι αλγόριθμοι κατάταξης, που ενσωματώνουν τη συμπεριφορά των χρηστών και την εκμάθηση τεχνικών κατάταξης[25].

*2.4.6.2 Ανάκτηση προσωπικών πληροφοριών 'personal information retrieval'*

Τα τελευταία χρόνια τα εμπορικά συστήματα υπολογιστών όπως Windows, Mac OS έχουν υιοθετήσει το δικό τους τρόπο στην ανάκτηση πληροφοριών. Ένα παράδειγμα αυτού είναι τα e-mail τα οποία κατηγοριοποιούν με βάση το κείμενο αλλά και προσφέρουν κατηγοριοποίηση με βάση την ανάγκη του χρήστη για το συγκεκριμένο γράμμα τοποθετώντας τα στα αχρείαστα 'junk' αλλά και άλλους φακέλους με αυτοματοποιημένο τρόπο χωρίς κάποια παραπάνω ενέργεια από το χρήστη. Κάποια βασικά θέματα στη διαδικασία αυτή, είναι ο αλγόριθμος να είναι ελαφρύς και να χειρίζεται διακριτικά τις πληροφορίες χωρίς να απασχολεί τη μνήμη του υπολογιστή αλλά και το χρήστη. [25]

Βασικά χαρακτηριστικά της μεθόδου:

Σε μία τέτοια κατηγορία μπορούν να υπάρχουν πολλοί διαφορετικοί τύποι δεδομένων. Τα προσωπικά δεδομένα μπορεί να διαφέρουν πολύ, από μηνύματα ηλεκτρονικού ταχυδρομείου και έγγραφα έως φωτογραφίες και αναρτήσεις στα μέσα κοινωνικής δικτύωσης. Αυτή η ποικιλομορφία απαιτεί ευέλικτες και ισχυρές τεχνικές ευρετηρίασης και ανάκτησης[25].

Επίσης, είναι σημαντικό να διασφαλίζεται το απόρρητο. Καθώς τα προσωπικά δεδομένα είναι ευαίσθητα, η εξισορρόπηση της αποτελεσματικής ανάκτησης με την προστασία του απορρήτου είναι ζωτικής σημασίας[25].

Ακόμη, είναι απαραίτητο να υπάρχει μία σχεδίαση διεπαφής με το χρήστη. Ο λόγος είναι τα προσωπικά δεδομένα που συχνά διαδίδονται σε πολλές συσκευές και πλατφόρμες, μια φιλική προς τον χρήστη διεπαφή για αναζήτηση και ανάκτηση είναι σημαντική για την επιτυχία[25].

*2.4.6.3 Ανάκτηση πληροφοριών σε εταιρίες, ακαδημαϊκά και συγκεκριμένου τομέα*

Η ανάκτηση πληροφορίας σε αυτές τις κατηγορίες πραγματοποιείται σε συλλογές αρχείων συνήθως στου οργανισμού τα εσωτερικά αρχεία, κάποια εσωτερική βάση δεδομένων αλλά και άρθρα τα οποία βρίσκονται σε κάποιο χώρο σε κάποια εσωτερική ηλεκτρονική βιβλιοθήκη. Σε αυτές τις περιπτώσεις τα αρχεία βρίσκονται συγκεντρωμένα σε κάποιο σύστημα και η ανάκτηση από αυτά θα γίνεται από ένα η και από πολλά μηχανήματα αναζήτησης πληροφορίας.

Στον ακαδημαϊκό χώρο και τις εταιρείες, η ανάκτηση πληροφοριών συχνά περιλαμβάνει σύνολα δεδομένων για συγκεκριμένους τομείς διαφορετικών μεγεθών. Η κλίμακα μπορεί να κυμαίνεται από σχετικά μικρές συλλογές ερευνητικών εργασιών ή εταιρικών εγγράφων έως μεγαλύτερες βάσεις δεδομένων, όπως δίκτυα ακαδημαϊκών αναφορών ή ενδοδίκτυα σε όλη την εταιρεία.

Βασικά χαρακτηριστικά της μεθόδου:

Ιδιαιτερότητα: Τα ακαδημαϊκά και εταιρικά σύνολα δεδομένων συχνά περιέχουν ορολογία και δομές για συγκεκριμένο τομέα, που απαιτούν εξειδικευμένες τεχνικές IR[6].

Έλεγχος πρόσβασης: Σε αυτές τις ρυθμίσεις, η πρόσβαση στα δεδομένα χρειάζεται συχνά να ελέγχεται με βάση τους ρόλους των χρηστών ή τα επίπεδα ασφάλειας[6].

Ενοποίηση: Οι επιχειρήσεις και οι ακαδημαϊκοί IR απαιτούν συχνά την ενοποίηση διαφορετικών πηγών δεδομένων, όπως βάσεις δεδομένων, αποθετήρια εγγράφων, ακόμη και δεδομένα ιστού[6].

# 3 *Αναζήτηση πληροφοριών στα μαθηματικά*

## *3.1 Στόχος αναζήτησης πληροφοριών Math IR*

Η αναζήτηση πληροφοριών στα μαθηματικά στοχεύει στη διευκόλυνση της πρόσβασης, της ανάκτησης αλλα και την ανακάλυψη των πηγών των μαθηματικών προβλημάτων.[2] Ένα παράδειγμα είναι τα αρκετά παραδοσιακά μαθήματα και Massive Open Online Courses (MOOCs) τα οποία προσφέρουν τις πηγές τους, όπως σημειώσεις, διαλέξεις και βιβλία, σαν αρχεία τα οποία μπορούν να επεξεργαστούν απο υπολογιστή με τη μορφή HTML η και XML.

Ένας στόχος της αναζήτησης σε μαθηματικά προβλήματα είναι οι μαθηματικού τύποι.

Σε αντίθεση με απλές λέξεις αναζήτησης, οι μαθηματικοί τύποι έχουν συνήθως διάφορες ιδιότητες και σχέσεις οι οποίες θα οδηγήσουν σε κάποιο αποτέλεσμα. Υπάρχουν πολλοί μαθηματικοί τύποι που είναι δομημένοι αρκετά διαφορετικά. Στην ερώτηση που αφορά την έρευνα του χρήστη και τι απάντηση ψάχνει όταν αναζητά το $x^2+y^2 = 1$ η απάντηση είναι το χ και y που ικανοποιούν την εξίσωση, όλες τις απαντήσεις που ικανοποιούν επίσης το $a^2 + b^2 = 1$ αλλά ακόμα και το διάγραμμα που αντικατοπτρίζει την εξίσωση.

Παρόλα αυτά, μιας και υπάρχουν πολύ συγκεκριμένα χαρακτηριστικά σε κάποιους μαθηματικούς τύπους, οι κλασικές μηχανές αναζήτησης πληροφορίας δεν μπορούν να δουλέψουν με σωστή ακρίβεια και δεν μπορούν να αναζητήσουν σωστά την ζητούμενη πληροφορία που αναζητά ο χρήστης.

Ο στόχος της ανάκτησης πληροφοριών στα μαθηματικά μπορεί να γίνει κατανοητός από διάφορες προοπτικές:

### *3.1.1 Πρόσβαση στη Μαθηματική Βιβλιογραφία*

Δεδομένου του μεγάλου και αυξανόμενου σώματος της μαθηματικής βιβλιογραφίας, ο βασικός στόχος του IR στα μαθηματικά είναι να παρέχει εύκολη πρόσβαση σε σχετικά μαθηματικά έγγραφα - όπως ερευνητικές εργασίες, βιβλία ή σημειώσεις διαλέξεων - ως απάντηση σε ερωτήματα των χρηστών. Πρόκειται για την παροχή των σωστών πληροφοριών στον χρήστη που αναζητά την πληροφορία την στιγμή που την αναζητά.

### *3.1.2 Σημασιολογική αναζήτηση*

Σε αντίθεση με τα παραδοσιακά έγγραφα κειμένου, τα μαθηματικά έγγραφα περιέχουν συχνά μεγάλο αριθμό μαθηματικών εκφράσεων και συμβόλων. Κατά συνέπεια, ένας από τους πρωταρχικούς στόχους του μαθηματικού IR είναι να υποστηρίξει τη σημασιολογική αναζήτηση, δηλαδή να δώσει τη δυνατότητα στον χρήστη να αναζητήσει όχι μόνο με λέξεις-κλειδιά αλλά και με μαθηματικούς τύπους και δομές[27].

### *3.1.3 Συμβολική Κατανόηση*

Το IR στα μαθηματικά επιδιώκει να διευκολύνει τη βαθιά κατανόηση των μαθηματικών συμβόλων και τύπων πέρα από την κυριολεκτική αναπαράστασή τους, περιλαμβάνοντας την αποσαφήνιση των συμβόλων και τον προσδιορισμό της έννοιας των μαθηματικών εκφράσεων που εξαρτάται από τα συμφραζόμενα [27].

### *3.1.4 Διασταυρούμενη αναφορά*

Το IR στα μαθηματικά στοχεύει στη σύνδεση μαθηματικών εννοιών σε διαφορετικά έγγραφα και πόρους. Αυτό περιλαμβάνει τον προσδιορισμό του σημείου που εισήχθη για πρώτη φορά ένα θεώρημα ή τύπος, τον εντοπισμό των παραγώγων και των χρήσεών του σε διαφορετικά περιβάλλοντα και την κατανόηση των διαφόρων αποδείξεών του[27].

### *3.1.5 Ενίσχυση της συνεργασίας και της ανακάλυψης γνώσης*

Παρέχοντας ένα ισχυρό σύστημα για την ανάκτηση μαθηματικών πληροφοριών, το IR στα μαθηματικά επιδιώκει να ενισχύσει τις συνεργασίες μεταξύ μαθηματικών και να επιτρέψει την ανακάλυψη νέας μαθηματικής γνώσης.

### *3.1.6 Εκπαιδευτική Υποστήριξη*

Τα αποτελεσματικά συστήματα IR στα μαθηματικά μπορούν να υποστηρίξουν σε μεγάλο βαθμό την εκπαίδευση παρέχοντας στους μαθητές και τους δασκάλους εύκολη πρόσβαση σε μαθηματικούς πόρους, λύσεις σε προβλήματα και επεξηγήσεις περίπλοκων μαθηματικών εννοιών[27].

## *3.2 Περιπτώσεις χρήσης*

Σε περιπτώσεις εκπαίδευσης, τα συστήματα αναζήτησης πληροφορίας Math-aware, μπορούν να φανούν χρήσιμα αν οι σπουδαστές αναζητούν αναφορές που θα βοηθήσουν να λύσουν προβλήματα και να αναπτύξουν τις γνώσεις τους.

Οι καθηγητές μπορούν να επωφεληθούν εξίσου από τέτοιου είδους συστήματα ωθώντας τους μαθητές να δημιουργήσουν γνωσιακές ομάδες και να αναπτύξουν την ομαδικότητα μαζί με τις γνώσεις τους.

Άλλη μια κατηγορία είναι οι ερευνητές οι οποίοι επωφελούνται απο τις μηχανές αναζήτησης Math-aware μιας και μπορούν να εντοπίσουν κάποια πιθανά χρήσιμα συστήματα, χώρους έρευνας αλλά και νέους συνεργάτες. Ένα καλό παράδειγμα είναι το όφελος που υπάρχει στην αναζήτηση προβλημάτων φυσικής τα οποία περιλαμβάνουν τα AdS/CFT και τις θεωρίες holographic duality.

### *3.2.1 Κατανόηση Μαθηματικών Ερωτημάτων*

Ένας από τους θεμελιώδεις στόχους του μαθηματικού IR είναι η σωστή κατανόηση των μαθηματικών ερωτημάτων. Η ερμηνεία των μαθηματικών ερωτημάτων μπορεί να είναι

δύσκολη λόγω της εγγενούς ασάφειας και πολυπλοκότητας της μαθηματικής σημειογραφίας [29].

Έτσι, ένας σημαντικός στόχος του μαθηματικού IR είναι η ανάπτυξη τεχνικών για την ανάλυση και την ερμηνεία των μαθηματικών ερωτημάτων με τρόπο που να συλλαμβάνει το επιδιωκόμενο νόημα του χρήστη. Αυτό περιλαμβάνει μεθόδους ανάλυσης μαθηματικών σημειώσεων, καθώς και τεχνικές για την αποσαφήνιση συμβόλων και όρων. Επιπλέον, δεδομένου ότι τα μαθηματικά ερωτήματα συχνά περιέχουν κείμενο και μαθηματικές εκφράσεις (όπως για παράδειγμα ένα ερώτημα μαθηματικού προβλήματος το οποίο περιέχει το κείμενο του ερωτήματος αλλά και κάποιο μαθηματικό τύπο), τα μαθηματικά συστήματα IR πρέπει επίσης να χειρίζονται αποτελεσματικά την αλληλεπίδραση μεταξύ κειμενικού και μαθηματικού περιεχομένου [30].

### *3.2.3 Αναπαράσταση και Ευρετηρίαση Μαθηματικού Περιεχομένου*

Ένας άλλος βασικός στόχος του μαθηματικού IR είναι η αποτελεσματική αναπαράσταση και ευρετηρίαση του μαθηματικού περιεχομένου. Τα μαθηματικά έγγραφα μπορούν να περιέχουν ένα ευρύ φάσμα μαθηματικού περιεχομένου, συμπεριλαμβανομένων εξισώσεων, θεωρημάτων, αποδείξεων και διαγραμμάτων, και αυτά μπορούν να αναπαρασταθούν με διάφορους τρόπους, όπως LaTeX, MathML ή εικόνες [30]. Ένα παράδειγμα αναπαράστασης μιας μαθηματικής εξίσωσης ax^2+bx+c=0 , όπου ο τύπος αναζητά το x που θα ικανοποιήσει την εξίσωση είναι το εξής:

**LaTeX:**

```
n\documentclass{article}
\usepackage{amsmath}

\begin{document}

The quadratic formula is given by:
\[
x = \frac{{-b \pm \sqrt{{b^2 - 4ac}}}}{{2a}}
```

\]

\end{document}

**MathML:**

```
<math xmlns="http://www.w3.org/1998/Math/MathML">
  <mtext>The quadratic formula is given by:</mtext>
  <mrow>
    <mi>x</mi>
    <mo>=</mo>
    <mfrac linethickness="0">
      <mrow>
        <mo>-</mo>
        <mi>b</mi>
        <mo>&#177;</mo>
        <msqrt>
          <mrow>
            <msup>
              <mi>b</mi>
              <mn>2</mn>
            </msup>
            <mo>-</mo>
            <mn>4</mn>
            <mi>a</mi>
            <mi>c</mi>
          </mrow>
        </msqrt>
      </mrow>
      <mrow>
        <mn>2</mn>
```

```
      <mi>a</mi>
    </mrow>
  </mfrac>
 </mrow>
</math>
```

Το παράδειγμα LaTeX είναι ένα απόσπασμα κώδικα που μπορεί να ‘τρέξει’ χρησιμοποιώντας λογισμικό LaTeX, ενώ το παράδειγμα MathML βασίζεται σε XML ειδικά σχεδιασμένη για την αναπαράσταση μαθηματικών εκφράσεων στον Ιστό.

Μια σημαντική πρόκληση για το μαθηματικό IR είναι η ανάπτυξη μεθόδων αναπαράστασης και ευρετηρίασης που αποτυπώνουν τον σημασιολογικό πλούτο αυτού του μαθηματικού περιεχομένου και επιτρέπουν την αποτελεσματική ανάκτηση. Αυτό περιλαμβάνει μεθόδους εξαγωγής μαθηματικών παραστάσεων από έγγραφα, αναπαράσταση μαθηματικών παραστάσεων σε μορφή κατάλληλη για ευρετηρίαση και ανάκτηση και σχεδιασμό δομών ευρετηρίασης που υποστηρίζουν αποτελεσματική αναζήτηση και ανάκτηση μαθηματικού περιεχομένου [30].

### *3.2.4 Συνάφεια κρίση και κατάταξη (Relevance Judgement and Ranking)*

Η κρίση συνάφειας και η κατάταξη είναι ένας κρίσιμος στόχος στη μαθηματική IR. Μόλις γίνει κατανοητό το μαθηματικό ερώτημα και αντιστοιχιστεί με τα ευρετηριασμένα έγγραφα, ο προσδιορισμός της συνάφειας των ανακτημένων εγγράφων και η κατάταξή τους είναι μια πρόκληση. Οι παραδοσιακές μετρήσεις, όπως η ακρίβεια, η ανάκληση και η βαθμολογία F1, ενδέχεται να μην αποτυπώνουν πλήρως τις αποχρώσεις της συνάφειας στο πλαίσιο του μαθηματικού περιεχομένου [31].

Επομένως, ένας στόχος του μαθηματικού IR είναι να αναπτύξει νέα μοντέλα συνάφειας και αλγόριθμους κατάταξης που μπορούν να αξιολογήσουν με ακρίβεια τη συνάφεια του μαθηματικού περιεχομένου και να παρέχουν στον χρήστη τα πιο σχετικά αποτελέσματα.

Αυτό μπορεί να περιλαμβάνει την εξέταση παραγόντων όπως η δομική ομοιότητα των μαθηματικών εκφράσεων, η σημασιολογική συνάφεια του μαθηματικού περιεχομένου και το πλαίσιο στο οποίο εμφανίζεται το μαθηματικό περιεχόμενο [31].

### *3.2.5 Αλληλεπίδραση και αξιολόγηση χρήστη*

Τέλος, η αλληλεπίδραση και πως αξιολογείται ένας χρήστης είναι ένας κρίσιμος στόχος του μαθηματικού IR. Ο σχεδιασμός διαισθητικών και αποτελεσματικών διεπαφών χρήστη για μαθηματική ανάκτηση πληροφοριών και η διεξαγωγή αυστηρών μελετών και αξιολογήσεων χρηστών για την αξιολόγηση της αποτελεσματικότητας αυτών των συστημάτων είναι κρίσιμες πτυχές του μαθηματικού IR [31].

Στόχος δεν είναι μόνο η ανάπτυξη εξελιγμένων αλγορίθμων και μοντέλων, αλλά και η διασφάλιση ότι αυτά τα μοντέλα ανταποκρίνονται στις ανάγκες των χρηστών και βελτιώνουν την εμπειρία τους στην εύρεση μαθηματικών πληροφοριών. Αυτό περιλαμβάνει μια βαθιά κατανόηση της συμπεριφοράς και των προτιμήσεων αναζήτησης πληροφοριών των χρηστών, καθώς και αυστηρών μεθοδολογιών αξιολόγησης για την αξιολόγηση της αποτελεσματικότητας των μαθηματικών συστημάτων IR [31].

# 4 *Μέτρηση της σχετικότητας των αποτελεσμάτων*

## *4.1 Μέτρηση της σχετικότητας*

Η αξιολόγηση συνάφειας (relevance evaluation) είναι ένα κρίσιμο στοιχείο των συστημάτων Ανάκτησης Πληροφοριών, που διαδραματίζει ουσιαστικό ρόλο στον προσδιορισμό της ποιότητας και της χρησιμότητας αυτών των συστημάτων. Περιλαμβάνει τη διαδικασία αξιολόγησης και κατάταξης των ανακτημένων πληροφοριών με βάση τη συνάφειά τους με ένα δεδομένο ερώτημα χρήστη. Η πολυπλοκότητα της αξιολόγησης συνάφειας πηγάζει από την υποκειμενική φύση της «συνάφειας», η οποία μπορεί να εξαρτάται από διάφορους παράγοντες όπως η πρόθεση του χρήστη, το πλαίσιο της αναζήτησης και η συγκεκριμένη ανάγκη πληροφοριών. Επιπλέον, καθώς αυξάνεται το μέγεθος των διαθέσιμων δεδομένων, η πρόκληση της ακριβούς αξιολόγησης της συνάφειας γίνεται ολοένα και πιο σημαντική. Παρά αυτές τις πολυπλοκότητες, η αποτελεσματική αξιολόγηση συνάφειας παραμένει θεμελιώδης για την επιτυχία των συστημάτων IR, στηρίζοντας την εμπειρία αναζήτησης του χρήστη και οδηγώντας τελικά στην ικανοποίηση των χρηστών.

Η μέτρηση της σχετικότητας των αποτελεσμάτων για τη σχετικότητα όσον αφορά την επίλυση μαθηματικών προβλημάτων και τον συλλογισμό, μπορεί να είναι ένα δύσκολο έργο και μπορεί να απαιτεί συνδυασμό διαφορετικών μετρήσεων. Μερικές κοινές μέθοδοι για τη μέτρηση της σχετικότητας περιλαμβάνουν τις παρακάτω:

## *4.2 Precision and Recall*

Η ακρίβεια και η ανάκληση είναι αναπόσπαστες μετρήσεις για την αξιολόγηση της αποτελεσματικότητας των συστημάτων ανάκτησης πληροφοριών. Αποτελούν τη βάση για την κατανόηση της αντιστάθμισης μεταξύ της συνάφειας και της πληρότητας των ανακτημένων αποτελεσμάτων[6].

Η ακρίβεια (precision) αναφέρεται στην αναλογία των ανακτημένων εγγράφων που σχετίζονται με το ερώτημα του χρήστη. Ορίζεται ως ο αριθμός των σχετικών εγγράφων που ανακτήθηκαν διαιρεμένος με τον συνολικό αριθμό των εγγράφων που ανακτήθηκαν. Υψηλή ακρίβεια σημαίνει ότι τα περισσότερα από τα ανακτημένα έγγραφα είναι σχετικά, μειώνοντας την προσπάθεια του χρήστη να διερευνήσει άσχετα αποτελέσματα. Αυτό είναι ιδιαίτερα σημαντικό σε σενάρια όπου οι χρήστες απαιτούν ένα περιορισμένο σύνολο αποτελεσμάτων υψηλής ακρίβειας, όπως η λήψη κλινικών αποφάσεων ή η ανάκτηση νομικών εγγράφων. Ωστόσο, η εστίαση μόνο στην ακρίβεια μπορεί να έχει ως αποτέλεσμα να μην ανακτώνται πολλά σχετικά έγγραφα, ειδικά σε μεγάλες βάσεις δεδομένων[6].

Η ανάκληση (recall), από την άλλη πλευρά, αναφέρεται στην αναλογία όλων των σχετικών εγγράφων στη βάση δεδομένων που ανακτώνται με επιτυχία. Υπολογίζεται ως ο αριθμός των σχετικών εγγράφων που ανακτήθηκαν διαιρεμένος με τον συνολικό αριθμό των σχετικών εγγράφων στη βάση δεδομένων. Η υψηλή ανάκληση είναι σημαντική σε σενάρια όπου η έλλειψη σχετικού εγγράφου μπορεί να έχει σοβαρές συνέπειες, όπως συστηματικές ανασκοπήσεις βιβλιογραφίας ή αναζήτηση διπλωμάτων ευρεσιτεχνίας. Ωστόσο, η προσπάθεια για υψηλή ανάκληση μπορεί να οδηγήσει στην ανάκτηση πολλών άσχετων εγγράφων, συντρίβοντας έτσι τον χρήστη με δυνητικά χαμηλής ποιότητας αποτελέσματα[6].

Η αντιστάθμιση μεταξύ ακρίβειας και ανάκλησης αποτελεί κεντρικό μέλημα στο σχεδιασμό συστημάτων IR. Ανάλογα με την εφαρμογή, οι σχεδιαστές συστημάτων μπορεί να δώσουν προτεραιότητα το ένα έναντι του άλλου. Η βαθμολογία F1, ο αρμονικός μέσος όρος ακρίβειας και ανάκλησης, χρησιμοποιείται συχνά ως ενιαία μέτρηση που εξισορροπεί και τις δύο πτυχές[6].

## *4.3 F1 Score*

Το σκορ F1 είναι μια μέτρηση που χρησιμοποιείται ευρέως στην Ανάκτηση Πληροφοριών και σε άλλα πεδία όπως η Μηχανική Μάθηση για την αξιολόγηση της αντιστάθμισης μεταξύ ακρίβειας και ανάκλησης. Γνωστό και ως balanced F-score ή F-measure, το σκορ F1 είναι ο αρμονικός μέσος όρος ακρίβειας και ανάκλησης, και ως εκ τούτου, δίνει ίση βαρύτητα σε αυτά τα δύο μέτρα. Σε αντίθεση με τον αριθμητικό μέσο όρο, ο αρμονικός μέσος όρος τείνει προς τη μικρότερη από τις δύο τιμές, που σημαίνει ότι τα υψηλά σκορ F1 μπορούν να επιτευχθούν μόνο εάν τόσο η ακρίβεια όσο και η ανάκληση είναι υψηλές[6].

Υψηλή ακρίβεια αλλά χαμηλή ανάκληση σημαίνει ότι χάνονται πολλά σχετικά έγγραφα (high false negatives), ενώ υψηλή ανάκληση αλλά χαμηλή ακρίβεια σημαίνει ότι ανακτώνται πολλά άσχετα έγγραφα (high false positives). Συνδυάζοντας την ακρίβεια και την ανάκληση σε μια ενιαία μέτρηση, το σκορ F1

παρέχει μια πιο ολιστική εικόνα της απόδοσης ενός συστήματος IP, ειδικά σε καταστάσεις όπου είναι ζωτικής σημασίας να εξισορροπηθούν τα ψευδώς θετικά και τα ψευδώς αρνητικά. Υπό αυτή την έννοια, το σκορ F1 χρησιμεύει ως μέτρο της ακρίβειας και της πληρότητας του συστήματος IR[28].

Ωστόσο,το σκορ F1 δεν είναι χωρίς περιορισμούς. Μια βασική υπόθεση πίσω από το σκορ της F1 είναι ότι η ακρίβεια και η ανάκληση είναι εξίσου σημαντικές, αλλά αυτό μπορεί να μην ισχύει σε όλα τα πλαίσια. Για παράδειγμα, σε ορισμένα σενάρια υψηλού κινδύνου, όπως η ανάκτηση ιατρικών ή νομικών πληροφοριών, μπορεί να δοθεί προτεραιότητα στην ανάκληση έναντι της ακρίβειας για να διασφαλιστεί ότι δεν παραβλέπονται κρίσιμες πληροφορίες. Αντίθετα, σε ένα σενάριο περιστασιακής αναζήτησης ιστού, η ακρίβεια μπορεί να αποτιμηθεί περισσότερο για να αποφευχθεί η η δημιουργία πολύ εκτεταμένων λιστών αποτελεσμάτων [28].

Επιπλέον, το σκορ F1 παρέχει ένα συνολικό μέτρο απόδοσης και ως εκ τούτου μπορεί να κρύψει τις συγκεκριμένες πτυχές της συμπεριφοράς ενός συστήματος IR. Για παράδειγμα, ένα σύστημα θα μπορούσε να έχει υψηλό σκορ F1 με μέτρια καλή απόδοση τόσο στην ακρίβεια όσο και στην ανάκληση, ή θα μπορούσε να έχει υψηλό σκορ F1 αποδίδοντας εξαιρετικά καλά στο ένα μέτρο και απλώς επαρκώς στο άλλο. Ως εκ τούτου, ενώ το σκορ F1 παρέχει ένα πολύτιμο συνοπτικό μέτρο, είναι συχνά ωφέλιμο να εξετάζουμε επίσης την ακρίβεια και την ανάκληση ξεχωριστά για να αποκτήσουμε μια πιο λεπτή κατανόηση της απόδοσης του συστήματος[28].

## *4.4 Human evaluation*

Η ανθρώπινη αξιολόγηση στην ανάκτηση πληροφοριών αναφέρεται στη διαδικασία αξιολόγησης και μέτρησης της αποτελεσματικότητας και της ποιότητας των συστημάτων ή αλγορίθμων ανάκτησης πληροφοριών με τη συμμετοχή ανθρώπινων κρίσεων. Περιλαμβάνει τη συλλογή ανατροφοδότησης και αξιολογήσεων από ανθρώπους αξιολογητές για την αξιολόγηση διαφόρων πτυχών του συστήματος ανάκτησης, όπως η σχετικότητα, η χρησιμότητα και η ικανοποίηση των χρηστών. Η ανθρώπινη αξιολόγηση είναι ζωτικής σημασίας για την κατανόηση της απόδοσης των συστημάτων ανάκτησης και τη βελτίωση της αποτελεσματικότητάς τους.

Η ανθρώπινη αξιολόγηση στην ανάκτηση πληροφοριών περιλαμβάνει κάποια βήματα, ένα από αυτά είναι συνήθως ο σχεδιασμός εργασιών αξιολόγησης (designing evaluation tasks).

Ο σχεδιασμός εργασιών αξιολόγησης αντικατοπτρίζει τις ανάγκες πληροφόρησης του πραγματικού κόσμου. Αυτές οι εργασίες μπορεί να περιλαμβάνουν εργασίες όπως ανάκτηση εγγράφων, απάντηση ερωτήσεων ή κατάταξη σχετικότητας (relevance ranking).

Έπειτα βρίσκεται η επιλογή αξιολογητών. Οι ανθρώπινοι αξιολογητές επιλέγονται με βάση την τεχνογνωσία τους, τις γνώσεις τους στον τομέα και την εξοικείωση τους με τις εργασίες αξιολόγησης. Μπορούν να είναι ειδικοί ανάκτησης πληροφοριών, ειδικοί τομέα ή τελικοί χρήστες.

Αυτό ακολουθείται από την ανάπτυξη κατευθυντήριων γραμμών αξιολόγησης. Οι κατευθυντήριες γραμμές δημιουργούνται για να παρέχουν οδηγίες και κριτήρια στους αξιολογητές για να αξιολογήσουν τη σχετικότητα και την ποιότητα των ανακτημένων πληροφοριών. Αυτές οι κατευθυντήριες γραμμές διασφαλίζουν συνέπεια και τυποποίηση στη διαδικασία αξιολόγησης.

Στη συνέχεια γίνεται διεξαγωγή της αξιολόγησης. Παρέχονται στους αξιολογητές οι κατευθυντήριες γραμμές αξιολόγησης και η πρόσβαση στο σύστημα ανάκτησης που αξιολογείται. Εκτελούν τις εργασίες που έχουν ανατεθεί, εξετάζουν τις πληροφορίες που έχουν αποκτηθεί και παρέχουν κρίσεις ή αξιολογήσεις με βάση τα προκαθορισμένα κριτήρια.

Στην ανάλυση των αποτελεσμάτων, οι συλλεγόμενες κρίσεις αναλύονται για τη μέτρηση της επίδοσης του συστήματος. Οι συνήθεις μετρήσεις αξιολόγησης περιλαμβάνουν την ακρίβεια(precision), την ανάκληση(recall), τη βαθμολογία F1, τη mean average precision και το normalized discounted cumulative gain. Οι τεχνικές στατιστικής ανάλυσης χρησιμοποιούνται συχνά για την εξαγωγή γνώσεων και την εξαγωγή συμπερασμάτων.

Για τα βέλτιστα αποτελέσματα στη συνέχεια πραγματοποιείται επαναληπτική βελτίωση (Iterative refinement). Με βάση την ανάλυση των αποτελεσμάτων αξιολόγησης, οι σχεδιαστές συστημάτων μπορούν να κάνουν τροποποιήσεις στους αλγόριθμους ανάκτησης ή στις διεπαφές χρήστη για να βελτιώσουν την απόδοση του συστήματος. Η διαδικασία αξιολόγησης μπορεί να επαναληφθεί επαναλαμβανόμενα για να αξιολογηθεί η αποτελεσματικότητα αυτών των τροποποιήσεων.

Η ανθρώπινη αξιολόγηση στην ανάκτηση πληροφοριών είναι ένα ουσιαστικό συστατικό για την ανάπτυξη και τη βελτίωση των συστημάτων ανάκτησης. Βοηθά τους ερευνητές και τους επαγγελματίες να κατανοήσουν πόσο καλά τα συστήματά τους ανταποκρίνονται στις ανάγκες των χρηστών, να εντοπίζουν δυνατά και αδύνατα σημεία, να συγκρίνουν διαφορετικούς αλγόριθμους ανάκτησης και να καθοδηγούν τις βελτιώσεις του συστήματος.[6][7][8]

## *4.5 Embedding-based metrics*

Οι μετρήσεις που βασίζονται στην ενσωμάτωση στην ανάκτηση πληροφοριών αναφέρονται στις τεχνικές αξιολόγησης που χρησιμοποιούν ενσωματώσεις λέξεων ή εγγράφων για τη μέτρηση της ομοιότητας ή της σχετικότητας μεταξύ των ερωτημάτων και των ανακτημένων εγγράφων. Αυτές οι μετρήσεις αξιοποιούν τις σημασιολογικές σχέσεις που καταγράφονται από τις ενσωματώσεις για την αξιολόγηση της ποιότητας των συστημάτων ανάκτησης.

Ένα βασικό μέτρο ανάλυσης είναι οι ενσωματώσεις λέξεων. Οι ενσωματώσεις λέξεων είναι πυκνές διανυσματικές αναπαραστάσεις λέξεων σε ένα χώρο υψηλών διαστάσεων, όπου παρόμοιες λέξεις είναι πιο κοντά η μία στην άλλη. Τα δημοφιλή μοντέλα ενσωμάτωσης λέξεων περιλαμβάνουν τα Word2Vec, GloVe και FastText [9]. Αυτές οι ενσωματώσεις καταγράφουν σημασιολογικές και συντακτικές σχέσεις μεταξύ των λέξεων, επιτρέποντας περισσότερες συγκρίσεις.

Έπειτα βρίσκονται οι ενσωματώσεις εγγράφων. Εκτός από τις ενσωματώσεις λέξεων, οι ενσωματώσεις εγγράφων αντιπροσωπεύουν ολόκληρα έγγραφα ως διανύσματα. Οι ενσωματώσεις εγγράφων μπορούν να δημιουργηθούν χρησιμοποιώντας τεχνικές όπως το Διάνυσμα παραγράφου (Doc2Vec), ο Κωδικοποιητής καθολικής πρότασης ή τα μοντέλα μετασχηματιστή (π.χ. BERT) [9]. Οι ενσωματώσεις εγγράφων αποτυπώνουν το σημασιολογικό περιεχόμενο και το πλαίσιο ολόκληρου του εγγράφου.

Ένα επόμενο μέτρο ανάλυσης είναι οι μετρήσεις ομοιότητας. Οι μετρήσεις που βασίζονται στην ενσωμάτωση υπολογίζουν την ομοιότητα μεταξύ των ερωτημάτων (queries) και εγγράφων για να καθορίσουν τη σχετικότητα των ανακτημένων εγγράφων. Η ομοιότητα συνημιτόνου είναι μια ευρέως χρησιμοποιούμενη μέτρηση που μετρά τη γωνία μεταξύ δύο διανυσμάτων στο χώρο ενσωμάτωσης. Μια υψηλότερη βαθμολογία ομοιότητας συνημιτόνου υποδηλώνει μεγαλύτερη ομοιότητα ή σχετικότητα.

Έπειτα βρίσκεται η αξιολόγηση με χρήση μετρήσεων που βασίζονται στην ενσωμάτωση.Για την αξιολόγηση συστημάτων ανάκτησης χρησιμοποιώντας μετρήσεις που βασίζονται στην ενσωμάτωση, ορίζεται ένα σύνολο ερωτημάτων και το σύστημα ανάκτησης επιστρέφει μια ταξινομημένη λίστα εγγράφων για κάθε ερώτημα. Οι ενσωματώσεις ερωτήματος και εγγράφων υπολογίζονται και οι βαθμολογίες ομοιότητας υπολογίζονται χρησιμοποιώντας την επιλεγμένη μέτρηση. Μέτρα αξιολόγησης όπως η MAP, το nDCG ή η ακρίβεια στο K μπορούν να χρησιμοποιηθούν για την αξιολόγηση της απόδοσης ανάκτησης.[10]

Υπάρχουν αρκετά πλεονεκτήματα των μετρήσεων που βασίζονται στην ενσωμάτωση. Οι μετρήσεις που βασίζονται σε ενσωμάτωση προσφέρουν πολλά πλεονεκτήματα. Μπορούν να συλλάβουν σημασιολογικές σχέσεις πέρα από την αντιστοίχιση λέξεων-κλειδιών, επιτρέποντας στα συστήματα ανάκτησης να διαχειρίζονται καλύτερα τη συνωνυμία και τις σχετικές έννοιες. Αυτές οι μετρήσεις μπορούν επίσης να χειριστούν λέξεις εκτός λεξιλογίου αξιοποιώντας τις πληροφορίες με βάση τα συμφραζόμενα των ενσωματώσεων. Επιπλέον, οι μετρήσεις που βασίζονται στην ενσωμάτωση μπορούν να εφαρμοστούν σε διάφορες γλώσσες.

Ενώ οι μετρήσεις που βασίζονται στην ενσωμάτωση παρέχουν πολύτιμες πληροφορίες, αντιμετωπίζουν επίσης προκλήσεις. Η ποιότητα των ενσωματώσεων εξαρτάται σε μεγάλο βαθμό από τα δεδομένα εκπαίδευσης και το συγκεκριμένο μοντέλο ενσωμάτωσης που χρησιμοποιείται. Η επιλογή της μετρικής ομοιότητας και του ορίου μπορεί να επηρεάσει τα αποτελέσματα της αξιολόγησης. Επιπλέον, οι ενσωματώσεις ενδέχεται να μην καταγράφουν αποχρώσεις του τομέα ή να παρουσιάζουν προκαταλήψεις που υπάρχουν στα δεδομένα εκπαίδευσης, οι οποίες μπορούν να επηρεάσουν την απόδοση ανάκτησης.[11][12]

## *4.6 Mean Average Precision (MAP)*

Η MAP είναι ένα δημοφιλές μέτρο αξιολόγησης στην Ανάκτηση Πληροφοριών που ενσωματώνει τόσο την ακρίβεια όσο και την ανάκληση, παρέχοντας έτσι ένα μονοψήφιο μέτρο ποιότητας σε όλα τα επίπεδα ανάκλησης. Συγκεκριμένα, παρέχει ένα μέτρο της ακρίβειας σε κάθε κατάταξη στο σύνολο αποτελεσμάτων όπου βρίσκεται ένα σχετικό έγγραφο, με μέσο όρο του συνολικού αριθμού των σχετικών εγγράφων[6].

Η Μέση Ακρίβεια (Average Precision - AP) για ένα μόνο ερώτημα υπολογίζεται λαμβάνοντας τον μέσο όρο των βαθμολογιών ακριβείας μετά την ανάκτηση κάθε σχετικού εγγράφου. Με άλλα λόγια, κάθε φορά που συναντάται ένα σχετικό έγγραφο στη λίστα κατάταξης ανάκτησης, υπολογίζεται η ακρίβεια σε αυτό το σημείο της λίστας και στη συνέχεια υπολογίζεται ο μέσος όρος αυτών των τιμών για να ληφθεί το AP. Το MAP είναι τότε ο μέσος όρος των βαθμολογιών AP για όλα τα ερωτήματα στη συλλογή δοκιμής. Αξίζει να σημειωθεί ότι εάν ένα σχετικό έγγραφο δεν ανακτηθεί, συνεισφέρει μηδενική ακρίβεια στο μέσο όρο, καθιστώντας το MAP ένα μέτρο ευαίσθητο στην ανάκληση[17].

Ένα από τα πλεονεκτήματα του MAP ως μέτρου αξιολόγησης είναι ότι λαμβάνει υπόψη ολόκληρη τη λίστα κατάταξης των ανακτημένων εγγράφων, σε αντίθεση με την ακρίβεια, την ανάκληση ή τη βαθμολογία F1, τα οποία συνήθως αξιολογούνται σε μια συγκεκριμένη κατάταξη ή σημείο αποκοπής. Ως

εκ τούτου, το MAP είναι ιδιαίτερα χρήσιμο σε σενάρια όπου η σειρά των ανακτημένων εγγράφων είναι σημαντική και ο χρήστης ενδιαφέρεται για κάτι περισσότερο από το έγγραφο που βρίσκεται στην πρώτη θέση[17].

Ωστόσο, ένας περιορισμός του MAP είναι ότι υποθέτει ότι όλα τα σχετικά έγγραφα είναι εξίσου συναφή, ενώ σε πολλά πρακτικά σενάρια, ορισμένα σχετικά έγγραφα μπορεί να είναι πιο σχετικά από άλλα. Επιπλέον, το MAP αντιμετωπίζει όλα τα ερωτήματα ως εξίσου σημαντικά, κάτι που μπορεί να μην ισχύει πάντα σε πραγματικές ρυθμίσεις[17].

## *4.7 Precision at K (P@K)*

Η ακρίβεια στο K (Precision at K - P@K) είναι ένα μέτρο απόδοσης στην Ανάκτηση Πληροφοριών που υπολογίζει την ακρίβεια των επιστρεφόμενων αποτελεσμάτων σε μια καθορισμένη θέση κατάταξης K στη λίστα αποτελεσμάτων. Για παράδειγμα, το Precision at 10 (P@10) είναι μια μέτρηση που χρησιμοποιείται συνήθως και ποσοτικοποιεί την αναλογία των σχετικών εγγράφων μεταξύ των 10 κορυφαίων ανακτημένων αποτελεσμάτων[6]. Αυτό το μέτρο είναι ιδιαίτερα σημαντικό σε σενάρια όπου οι χρήστες είναι πιθανό να λαμβάνουν υπόψη μόνο τα κορυφαία αποτελέσματα, όπως η αναζήτηση στον ιστό ή τα συστήματα συστάσεων.

Βασικό πλεονέκτημα του P@K ως μέτρου αξιολόγησης είναι η απλότητα και η ερμηνευτικότητά του. Παρέχει έναν απλό τρόπο αξιολόγησης της απόδοσης ενός συστήματος IR από την οπτική γωνία του χρήστη, καθώς εστιάζει στα κορυφαία K αποτελέσματα που είναι πιο πιθανό να λάβει υπόψη ο χρήστης. Για παράδειγμα, σε σενάρια αναζήτησης ιστού, οι χρήστες σπάνια υπερβαίνουν την πρώτη σελίδα των αποτελεσμάτων αναζήτησης, η οποία συνήθως περιέχει περίπου 10 αποτελέσματα. Ως εκ τούτου, το P@10 παρέχει ένα πρακτικό μέτρο της απόδοσης του συστήματος[6].

Ωστόσο και το P@K έχει τους περιορισμούς του. Πρώτον, δεν λαμβάνει υπόψη τη σειρά των εγγράφων στα κορυφαία K αποτελέσματα. Για παράδειγμα, το P@10 δεν κάνει διαφοροποίηση μεταξύ ενός σχετικού εγγράφου που κατατάσσεται στη θέση 1 ή στη θέση 10. Δεύτερον, το P@K μπορεί να ποικίλλει σημαντικά ανάλογα με την επιλογή του K. Σε ορισμένες περιπτώσεις, ένα μικρό K μπορεί να είναι πολύ αυστηρό, προκαλώντας πολλά σχετικά έγγραφα που πρέπει να αγνοηθούν, ενώ ένα μεγάλο K μπορεί να περιλαμβάνει πάρα πολλά άσχετα έγγραφα[17].

Τέλος, όπως και άλλα μέτρα ακριβείας, το P@K δεν είναι προσανατολισμένο στην ανάκληση. Δεν λαμβάνει υπόψη τον συνολικό αριθμό των σχετικών εγγράφων στη βάση δεδομένων, και ως εκ τούτου δεν παρέχει καμία ένδειξη για το πόσα σχετικά έγγραφα χάθηκαν. Κατά συνέπεια, το P@K θα πρέπει ιδανικά να χρησιμοποιείται σε συνδυασμό με μέτρα προσανατολισμένα στην ανάκληση για μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης του συστήματος IR[17].

## *4.8 NDCG score*

Το κανονικοποιημένο εκπτωτικό άθροισμα κέρδους (NDCG) υπολογίζεται διαιρώντας το εκπτωτικό άθροισμα κέρδους (DCG) της λίστας κατάταξης με το DCG της ιδανικής λίστας κατάταξης, που είναι η λίστα με τα σχετικά στοιχεία που κατατάσσονται με την βέλτιστη σειρά. Το NDCG κυμαίνεται από 0 έως 1, με υψηλότερες τιμές να υποδηλώνουν καλύτερη απόδοση.

Το NDCG είναι μια κανονικοποίηση του εκπτωτικού άθροισματος κέρδους (DCG). Το DCG είναι ένα σταθμισμένο άθροισμα του βαθμού σχετικότητας των στοιχείων που κατατάσσονται. Το βάρος είναι μια φθίνουσα συνάρτηση της κατάταξης (θέσης) του αντικειμένου, και επομένως ονομάζεται discount. Ο αρχικός λόγος για την εισαγωγή της έκπτωσης είναι ότι η πιθανότητα ο χρήστης να δει ένα έγγραφο μειώνεται σε σχέση με την κατάταξή του. Το NDCG κανονικοποιεί το DCG με το ιδανικό DCG(IDCG), το οποίο είναι απλώς το μέτρο DCG του καλύτερου αποτελέσματος κατάταξης. Επομένως, το μέτρο NDCG είναι πάντα ένας αριθμός στο [0,1].

Αυστηρά μιλώντας, το NDCG είναι μια οικογένεια μέτρων κατάταξης, αφού υπάρχει ευελιξία στην επιλογή της συνάρτησης discount. Η logarithmic discount 1/log(1+r), όπου r είναι η κατάταξη, κυριάρχησε στη βιβλιογραφία και τις εφαρμογές. [5]

Μια άλλη συνάρτηση discount που εμφανίζεται στη βιβλιογραφία είναι η r^ −1, η οποία ονομάζεται Zipfian στην Ανάκτηση Πληροφοριών. Τα συστήματα μηχανών αναζήτησης χρησιμοποιούν επίσης μια cut-off top-k έκδοση του NDCG. Δηλαδή, η discount έχει οριστεί να είναι μηδέν για τάξεις μεγαλύτερες από k. Αυτό το μέτρο NDCG αναφέρεται συνήθως ως NDCG@k. Δεδομένης της σημασίας και της δημοτικότητας του NDCG, έχουν γίνει εκτενείς μελέτες σχετικά με αυτό το μέτρο, κυρίως στον τομέα της Ανάκτησης Πληροφοριών. Όλες αυτές οι έρευνες διεξάγονται από εμπειρική προοπτική κάνοντας πειράματα σε σύνολα δεδομένων αναφοράς. Υπάρχουν επίσης εργασίες που εξέτασαν τις συναρτήσεις έκπτωσης εκμάθησης για το NDCG. Αν και αυτές οι εργασίες απέκτησαν πληροφορίες σχετικά με το NDCG, εξακολουθούν να υπάρχουν σημαντικά ζητήματα που δεν έχουν αντιμετωπιστεί.

Τεχνική αποτύπωση:

Για κάθε ερώτημα, κατατάσσει πρώτα όλα τα ανακτημένα έγγραφα με βάση τις βαθμολογίες συνάφειάς τους.

Για τα κορυφαία έγγραφα K (για παράδειγμα K=1000), υπολογίζει το DCG (Discounted Cumulative Gain) ως εξής:

DCG@K = rel1 + sum((rel_i) / log2(i+1) για i σε 2 έως K)

όπου rel_i είναι η συνάφεια του αποτελέσματος στη θέση i.

Για να ομαλοποιήσει αυτήν την τιμή (καθιστώντας την ανεξάρτητη από το ερώτημα), υπολογίζει το Ideal DCG (IDCG), που είναι η τιμή DCG@K μιας τέλειας κατάταξης (όπου τα έγγραφα ταξινομούνται με βάση τις κρίσεις συνάφειάς τους σε φθίνουσα σειρά).

Στη συνέχεια, το nDCG'@K υπολογίζεται ως nDCG'@K = DCG@K / IDCG@K.

Το σύμβολο (') υποδηλώνει ότι τα μη κριθέντα έγγραφα δεν λαμβάνονται υπόψη στον υπολογισμό. Απλώς αφαιρούνται από την κατάταξη πριν από τον υπολογισμό του nDCG'.

# 5 Οι διαγωνισμοί ArqMath

## 5.1 Οι εργασίες του ARQMath

Το ARQMath CLEF είναι μια εκστρατεία αξιολόγησης που διεξάγεται ως μέρος του Συνεδρίου και των Εργαστηρίων του Φόρουμ Αξιολόγησης (CLEF). Ο στόχος του ARQMath CLEF είναι να τονώσει την έρευνα στον τομέα της ανάκτησης μαθηματικών πληροφοριών. Επικεντρώνεται στο σύνολο δεδομένων ARQMath, το οποίο αποτελείται από μαθηματικές ερωτήσεις και απαντήσεις από το Mathematics StackExchange, ένα διαδικτυακό φόρουμ ερωτήσεων και απαντήσεων.

Σε γενικές γραμμές, το ARQMath CLEF περιλαμβάνει δύο κύριες εργασίες:

Task 1: Ανάκτηση Απαντήσεων για Ερωτήσεις - Ζητείται από τους συμμετέχοντες να αναπτύξουν συστήματα που, λαμβάνοντας υπόψη ένα ερώτημα που αποτελείται από έναν τίτλο και μια περιγραφή του προβλήματος, θα ανακτήσουν ένα σύνολο σχετικών απαντήσεων μέσα από μία έτοιμη συλλογή απαντήσεων.

Παράδειγμα ερωτήματος:

# How can I evaluate $\sum_{n=0}^{\infty}(n+1)x^n$?

Asked 8 years, 5 months ago  Active 4 months ago  Viewed 34k times

384

146

How can I evaluate

$$\sum_{n=1}^{\infty}\frac{2n}{3^{n+1}}$$

I know the answer thanks to Wolfram Alpha, but I'm more concerned with how I can derive that answer. It cites tests to prove that it is convergent, but my class has never learned these before so I feel that there must be a simpler method.

In general, how can I evaluate

$$\sum_{n=0}^{\infty}(n+1)x^n?$$

sequences-and-series  convergence  power-series  faq

edited Sep 24 '17 at 12:09
Parcly Taxel
51.7k ● 13 ■ 80 ▲ 120

asked Apr 3 '11 at 21:41
Backus
2,072 ● 3 ■ 12 ▲ 8

Παράδειγμα αναζήτησης και αποτελεσμάτων:

**1**

No need to use Taylor series, this can be derived in a similar way to the formula for geometric series. Let's find a general formula for the following sum:

$$S_m = \sum_{n=1}^{m} nr^n.$$

...

**2**

It is equivalent to $x(x+1)(x+5)(x+6)+96=0$

Now

...

$$(x^2+6x)(x^2+6x+5)+96=0$$

**3**

If you want a solution that doesn't require derivatives or integrals, notice that

$$\begin{aligned}1+2x+3x^2+4x^3+\cdots = 1+x+x^2+x^3+\ldots\\ +x+x^2+x^3+\ldots\\ +x^2+x^3+\ldots\end{aligned}$$

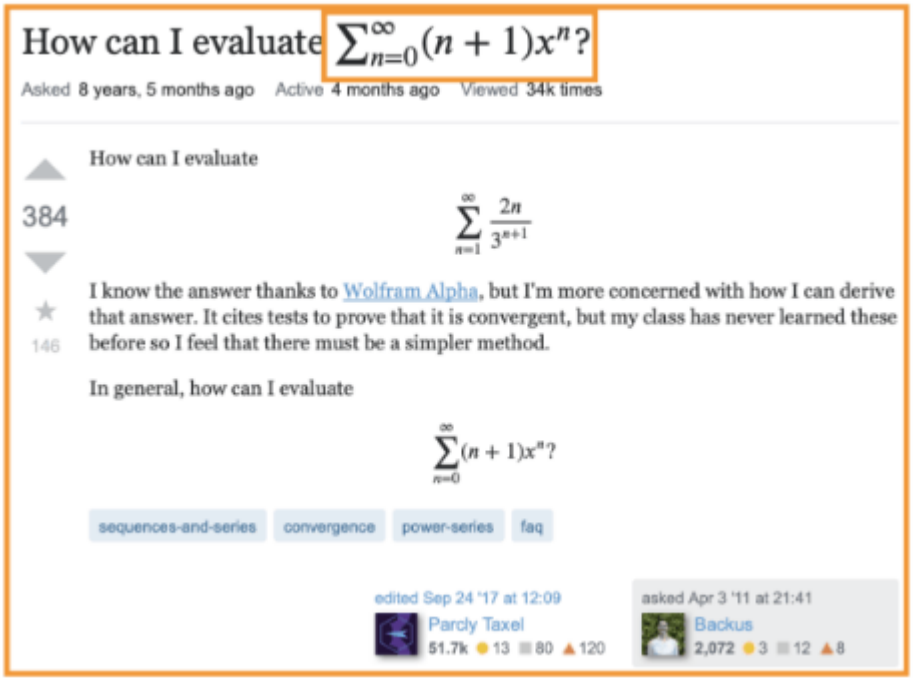
...

- 
- 
- 

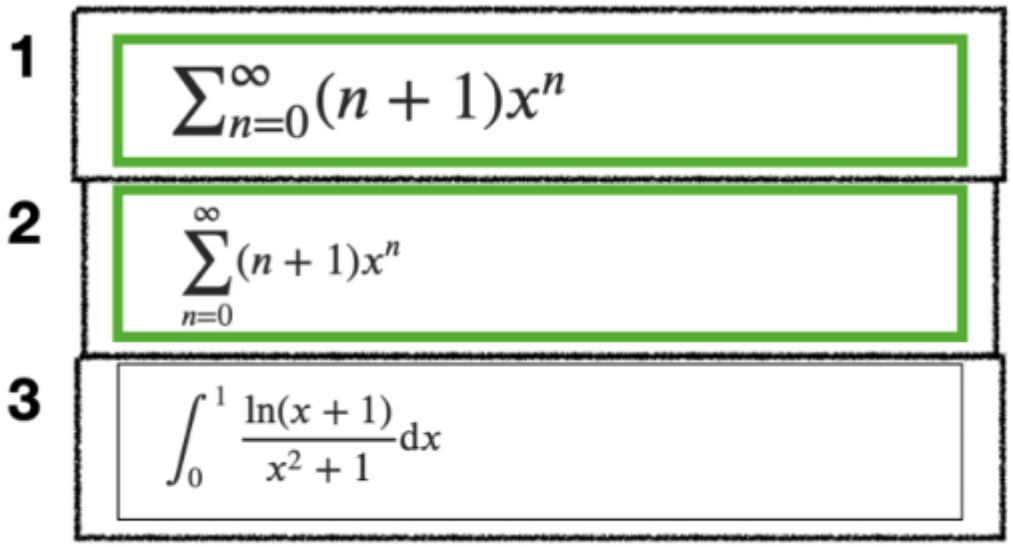
Activate Winc

Task 2: Αναζήτηση τύπου - Ζητείται από τους συμμετέχοντες να αναπτύξουν συστήματα που μπορούν να ανακτήσουν σχετικούς τύπους που δίνονται σε ένα ερώτημα τύπου μέσα από μία έτοιμη συλλογή απαντήσεων.

Παράδειγμα ερωτήματος:




Παράδειγμα αναζήτησης και αποτελεσμάτων:



Κάθε χρόνο από τον πρώτο διαγωνισμό που ξεκίνησε το 2019 μέχρι και το 2021, που έτρεξε ο τρίτος διαγωνισμός, οι ομάδες που συμμετέχουν στο ARQMath CLEF λαμβάνουν ένα σύνολο θεμάτων (δηλαδή ερωτήματα) και καλούνται να υποβάλουν ταξινομημένες λίστες εγγράφων (δηλαδή απαντήσεις ή τύπους) που σχετίζονται με κάθε θέμα. Στη συνέχεια, οι υποβολές των ομάδων αξιολογούνται με βάση πόσα από τα ανακτημένα έγγραφα είναι πραγματικά σχετικά.

Οι ομάδες χρησιμοποιούν μια ποικιλία τεχνικών για να επιτύχουν τα αποτελέσματά τους και οι ιδιαιτερότητες μπορεί να διαφέρουν πολύ μεταξύ των ομάδων. Ωστόσο, ορισμένες κοινές προσεγγίσεις περιλαμβάνουν:

Κειμενική και σημασιολογική ανάλυση: Το κείμενο που περιβάλλει τις μαθηματικές εκφράσεις μπορεί να αναλυθεί για να κατανοήσει το πλαίσιο και τη σημασιολογία.

Ανάλυση μαθηματικών εκφράσεων: Οι μαθηματικές εκφράσεις μπορούν να αναλυθούν και να συγκριθούν δομικά ή να μετατραπούν σε κανονική μορφή για ευκολότερη σύγκριση.

Τεχνικές μηχανικής μάθησης: Τα μοντέλα μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, μπορούν να εκπαιδευτούν στο έργο της ανάκτησης σχετικών εγγράφων που δίνονται σε ένα ερώτημα.

Ο τρίτος διαγωνισμός του ARQMath έχει τρεις κύριες εργασίες πέρα απο τις δύο πρώτες που αναφέρθηκαν για τα προηγούμενα χρόνια.
Εισάγεται μια νέα πιλοτική εργασία, την Απάντηση ερωτήσεων ανοιχτού τομέα (Task 3).

Για το Task 3, με τις ίδιες ερωτήσεις με το Task 1, τα συμμετέχοντα συστήματα παρέχουν επίσης μια απάντηση, αλλά δεν περιορίζονται στην αναζήτηση της συλλογής ARQMath για την εύρεση αυτής της απάντησης. Η συνάφεια (relevance) καθορίζεται από την αναμενόμενη χρησιμότητα κάθε επιστρεφόμενου αντικειμένου. Αυτές οι εργασίες επιτρέπουν στις συμμετέχουσες ομάδες να εξερευνήσουν αξιοποιώντας μαθηματικούς συμβολισμούς μαζί με κείμενο για να βελτιώσουν την ποιότητα των αποτελεσμάτων ανάκτησης.

## *5.2 Μετρα σχετικότητας που χρησιμοποίησαν οι ομάδες*

Με βάση προηγούμενους διαγωνισμούς ARQMath, οι συμμετέχουσες ομάδες χρησιμοποιούν γενικά μια ποικιλία μεθόδων αξιολόγησης, τα οποία είναι ορισμένα από τους διοργανωτές, για να τελειοποιήσουν τα

μοντέλα τους πριν από την υποβολή. Αυτές οι μέθοδοι αξιολόγησης βασίζονται συχνά σε παραδοσιακές μετρήσεις Ανάκτησης Πληροφοριών , όπως η MAP, το NDCG και η Ακρίβεια στο K (P@K)[29].

Όσον αφορά τα μοντέλα και τις μεθόδους που χρησιμοποιούν οι ομάδες, συχνά πειραματίζονται με διαφορετικές προσεγγίσεις χειρισμού μαθηματικών παραστάσεων, οι οποίες αποτελούν βασική πτυχή του ARQMath. Αυτό περιλαμβάνει μεθόδους ανάλυσης και αναπαράστασης μαθηματικών παραστάσεων, καθώς και τεχνικές για την ενσωμάτωση των μαθηματικών παραστάσεων στο μοντέλο ανάκτησης. Για παράδειγμα, ορισμένες ομάδες μπορεί να χρησιμοποιήσουν το LaTeX για να αναλύσουν μαθηματικές εκφράσεις και στη συνέχεια να τις ενσωματώσουν σε ένα μοντέλο ανάκτησης που βασίζεται σε κείμενο, ενώ άλλες μπορεί να χρησιμοποιήσουν πιο εξειδικευμένα μοντέλα ανάκτησης μαθηματικών πληροφοριών[29].

Εκτός από αυτές τις ειδικές για τα μαθηματικά μεθόδους, οι ομάδες χρησιμοποιούν συχνά προηγμένες τεχνικές IR, όπως η επέκταση ερωτήματος (query expansion), η ανάδραση σχετικότητας (relevance feedback) και η κατάταξη βάσει μηχανικής μάθησης (machine learning-based learning). Επιπλέον, οι ομάδες μπορούν να χρησιμοποιήσουν διαφορετικά βήματα προεπεξεργασίας, όπως κανονικοποίηση κειμένου (text normalization), διακοπή αφαίρεσης λέξεων (stop word removal) και απορρίψεις (stemming), τα οποία μπορούν επίσης να επηρεάσουν σημαντικά την απόδοση του συστήματος ανάκτησης[6].

Είναι σημαντικό να σημειωθεί ότι οι συγκεκριμένες μέθοδοι που χρησιμοποιούνται από τις κορυφαίες ομάδες μπορεί να ποικίλλουν ευρέως και θα εξαρτώνται από τις ιδιαιτερότητες του αγώνα και τις επιμέρους στρατηγικές του τις ομάδες.

## *5.3 Θετικά και αρνητικά των μεθόδων μέτρησης σχετικότητας*

Στο πλαίσιο του διαγωνισμού ARQMath CLEF και της ανάκτησης πληροφοριών γενικά, οι μέθοδοι που χρησιμοποιούνται συνήθως από τις συμμετέχουσες ομάδες (οι οποίες έχουν οριστεί από τους διοργανωτές) — όπως η χρήση παραδοσιακών μετρήσεων IR (MAP, NDCG, P@K), ο χειρισμός μαθηματικών εκφράσεων, η χρήση προηγμένων τεχνικών IR ( η επέκταση ερωτήματος, η ανάδραση συνάφειας, η κατάταξη βάσει μηχανικής μάθησης) και η χρήση διαφορετικών βημάτων προεπεξεργασίας—έχουν πλεονεκτήματα και μειονεκτήματα.

Ξεκινώντας με τις παραδοσιακές μετρήσεις ανάκτησης πληροφοριών, αυτές οι μέθοδοι είναι ωφέλιμες επειδή παρέχουν ένα συνεπές, τυποποιημένο μέσο αξιολόγησης της απόδοσης διαφορετικών συστημάτων ανάκτησης, επιτρέποντας συγκρίσεις μεταξύ διαφορετικών μεθόδων και προσεγγίσεων. Ωστόσο, έχουν ορισμένους περιορισμούς. Για παράδειγμα, μετρήσεις όπως το P@K και το MAP λαμβάνουν υπόψη μόνο τη σειρά κατάταξης των σχετικών εγγράφων, αγνοώντας τις πιθανές διαφορές στη συνάφεια μεταξύ τους. Επιπλέον, αυτά τα μέτρα ενδέχεται να μην αντικατοπτρίζουν πλήρως την ικανοποίηση των χρηστών, καθώς δεν λαμβάνουν υπόψη παράγοντες όπως η καινοτομία ή η ποικιλομορφία των επιστρεφόμενων αποτελεσμάτων[6].

Οι μέθοδοι που χρησιμοποιούνται για το χειρισμό μαθηματικών εκφράσεων, οι οποίες αποτελούν βασική πτυχή των εργασιών ARQMath, έχουν τα δικά τους δυνατά και αδύνατα σημεία. Τεχνικές που αναλύουν και αναπαριστούν μαθηματικές εκφράσεις μπορούν να συλλάβουν τον σημασιολογικό πλούτο αυτών των εκφράσεων, βοηθώντας στην ακριβέστερη ανάκτηση. Ωστόσο, οι μαθηματικές εκφράσεις μπορούν να αναπαρασταθούν με πολλούς διαφορετικούς τρόπους, θέτοντας μια πρόκληση για αυτές τις μεθόδους. Επιπλέον, η ανάλυση και η αναπαράσταση μαθηματικών παραστάσεων μπορεί να είναι υπολογιστικά ακριβή, επηρεάζοντας την επεκτασιμότητα αυτών των μεθόδων[6].

Οι προηγμένες τεχνικές IR, όπως η επέκταση ερωτήματος και η ανάδραση συνάφειας, μπορούν να βελτιώσουν την απόδοση ανάκτησης καθιστώντας το ερώτημα πιο αντιπροσωπευτικό της ανάγκης πληροφοριών του χρήστη ή ενσωματώνοντας τα σχόλια των χρηστών στη διαδικασία ανάκτησης. Ωστόσο, αυτές οι τεχνικές βασίζονται στην ποιότητα των διευρυμένων όρων ή στην ακρίβεια των σχολίων, κάτι που μπορεί να μην είναι πάντα εγγυημένο. Οι τεχνικές κατάταξης που βασίζονται στη μηχανική μάθηση, από

την άλλη πλευρά, μπορούν δυνητικά να αποφέρουν καλύτερα αποτελέσματα μαθαίνοντας από προηγούμενες σχετικές κρίσεις, αλλά απαιτούν μεγάλες ποσότητες δεδομένων εκπαίδευσης και μπορεί να είναι πολύπλοκες στην εφαρμογή τους[30].

Τέλος, τα βήματα προεπεξεργασίας, όπως η κανονικοποίηση κειμένου, η αφαίρεση συχνών όρων (stop word removal) και η δημιουργία προέλευσης μπορούν να κάνουν τη διαδικασία ανάκτησης πιο αποτελεσματική και ακριβή, μειώνοντας τη διάσταση των δεδομένων και εστιάζοντας στα πιο σημαντικά μέρη του κειμένου. Ωστόσο, αυτά τα βήματα μπορούν επίσης να οδηγήσουν σε απώλεια πληροφοριών. Για παράδειγμα, το stemming μπορεί να συγχέει λέξεις που έχουν διαφορετική σημασία αλλά μοιράζονταιτην ίδια ρίζα, οδηγώντας δυνητικά σε απώλεια της ακρίβειας ανάκτησης[30].

Συμπερασματικά, ενώ αυτές οι μέθοδοι μπορούν να βοηθήσουν σημαντικά στη διαδικασία ανάκτησης πληροφοριών, έχουν επίσης εγγενείς περιορισμούς που πρέπει να αντιμετωπιστούν για να βελτιωθεί περαιτέρω η αποτελεσματικότητα και η αποδοτικότητα των συστημάτων ανάκτησης.

### *5.3.1 Πως αξιολόγησαν τα αποτελέσματα οι αξιολογητές*

Οι αξιολογητές εκπαιδεύτηκαν στο Zoom σε τέσσερις συνεδρίες. Στην πρώτη συνεδρία, εξηγήθηκε η εργασία και τα κριτήρια συνάφειας. Τους έδειξαν μερικά παραδείγματα, ζήτησαν τη γνώμη τους σχετικά με τη συνάφεια και στη συνέχεια είχαν συζητήσεις με έναν ειδικό αξιολογητή, έναν καθηγητή μαθηματικών.

Στη συνέχεια, πραγματοποίησαν τρεις κύκλους εκπαίδευσης όπου οι αξιολογητές αξιολόγησαν μικρές ομάδες κρίσης για τέσσερα δείγματα θεμάτων από το ARQMath-2. Μετά από κάθε γύρο, είχαν μια συνεδρία Zoom για να συζητήσουν τις κρίσεις σχετικότητας και να διευκρινίσουν την κατανόησή τους για τα κριτήρια συνάφειας. Ο στόχος της εκπαίδευσης ήταν να βοηθήσει τους αξιολογητές να κάνουν αυτό συνεπείς σχολιασμούς, καθώς οι ερμηνείες των θεμάτων μπορεί να διαφέρουν μεταξύ των ατόμων.

Οι αξιολογητές βρήκαν τα νήματα ερωτήσεων στη διεπαφή Turkle χρήσιμα, ειδικά για θέματα που αφορούν ζητήματα που δεν καλύπτονται συνήθως σε προπτυχιακά μαθήματα. Υπολογίστηκε η αναφορά κάππα του Cohen για συμφωνία μεταξύ κάθε αξιολογητή και όλων των άλλων κατά τη διάρκεια της εκπαίδευσης. Η συμφωνία ήταν καλύτερη όταν μόνο η υψηλή και μεσαία συνάφεια θεωρούνταν σχετική.

Μετά την εκπαίδευση αξιολόγησαν 80 θέματα, με δύο από αυτά να αφαιρούνται από τη συλλογή καθώς είχαν μόνο μία υψηλή ή μεσαία σχετική απάντηση. Για τα υπόλοιπα 78 θέματα, αξιολογήθηκαν κατά μέσο όρο 446,8 απαντήσεις, με μέσο χρόνο αξιολόγησης 44,1 δευτερόλεπτα ανά ανάρτηση απάντησης. Ο μέσος αριθμός απαντήσεων με οποιοδήποτε βαθμό συνάφειας ήταν 100,8 ανά ερώτηση, που είναι διπλάσιος από αυτόν που παρατηρήθηκε στο ARQMath-2.

Τέλος, αφού έγιναν οι αξιολογήσεις, καθένας από τους αξιολογητές αξιολόγησε ένα θέμα που αξιολογήθηκε από άλλον αξιολογητή. Ο συντελεστής κάππα του Cohen, ένα μέτρο συμφωνίας μεταξύ των αξιολογητών, που επιτεύχθηκε ήταν 0,24 στην εργασία αξιολόγησης τεσσάρων κατευθύνσεων και με υψηλή και μεσαία δυαδοποίηση, η μέση τιμή κάπα ήταν 0,25.

## *5.4 Υπολογισμός μέτρων αποτελεσματικότητας στο ArqMath*

Ο διαγωνισμός ARQMath, ένας διαγωνισμός του Conference and Labs of Evaluation Forum, επικεντρώνεται στη βελτίωση της αυτοματοποιημένης απάντησης ερωτήσεων στα μαθηματικά. Όσον αφορά την αξιολόγηση της συνάφειας των αποτελεσμάτων, το ARQMath χρησιμοποιεί κυρίως τις παραδοσιακές μετρήσεις Ανάκτησης Πληροφοριών που μετρούν την ποιότητα των καταταγμένων λιστών εγγράφων, όπως η MAP, η Ακρίβεια στο K (P@K) και η Κανονική Προεξοφλημένο αθροιστικό κέρδος (NDCG)[18].

Οι επιλεγμένες μετρήσεις αξιολόγησης, nDCG', MAP' και P@10, παρέχουν μια ολοκληρωμένη αξιολόγηση της απόδοσης των συστημάτων ανάκτησης μαθηματικών πληροφοριών. Αυτές οι μετρήσεις επιτρέπουν στους ερευνητές να συγκρίνουν και να αναλύουν διαφορετικά μοντέλα, αλγόριθμους και τεχνικές ανάκτησης στον τομέα του μαθηματικού περιεχομένου. Λαμβάνοντας υπόψη τη συνάφεια, την ποιότητα κατάταξης και την ακρίβεια, αυτές οι μετρήσεις βοηθούν στον εντοπισμό δυνατών σημείων και περιοχών για βελτίωση στα συστήματα ανάκτησης.

Τα nDCG', MAP' και P@10 έχουν αποδειχθεί αποτελεσματικές μετρήσεις στην αξιολόγηση των συστημάτων ανάκτησης. Παρέχουν αντικειμενικά μέτρα απόδοσης, λαμβάνοντας υπόψη τόσο τη συνάφεια όσο και την κατάταξη. Ωστόσο, αυτές οι μετρήσεις έχουν επίσης περιορισμούς, όπως η ευαισθησία στην επιλογή των κρίσεων σχετικών και η υπόθεση της ίσης σημασίας όλων των ανακτημένων εγγράφων. Οι ερευνητές θα πρέπει να γνωρίζουν αυτούς τους περιορισμούς όταν ερμηνεύουν και συγκρίνουν αποτελέσματα.

Στις αξιολογήσεις ARQMath, η συνάφεια καθορίζεται συχνά από ανθρώπινους αξιολογητές που αξιολογούν τη συνάφεια των κορυφαίων αποτελεσμάτων που επιστρέφονται από τα συμμετέχοντα συστήματα για ένα σύνολο ερωτημάτων. Αυτή η διαδικασία αξιολόγησης περιλαμβάνει δύο στάδια: συγκέντρωση και αξιολόγηση. Στο στάδιο της συγκέντρωσης, συλλέγεται ένα σύνολο δυνητικά σχετικών εγγράφων συνδυάζοντας τα κορυφαία αποτελέσματα από όλα τα συμμετέχοντα συστήματα. Στο στάδιο της αξιολόγησης, οι ανθρώπινοι αξιολογητές αξιολογούν τη συνάφεια των εγγράφων στη δεξαμενή σε μια κλίμακα πολλαπλών σημείων. Η κλίμακα περιλαμβάνει συχνά κατηγορίες όπως "πολύ σχετικό", "μερικώς σχετικό" και "μη σχετικό". Αυτές οι κρίσεις σχετικότητας χρησιμοποιούνται στη συνέχεια για τον υπολογισμό των μετρήσεων αξιολόγησης[29].

Ωστόσο, η αξιολόγηση της συνάφειας των μαθηματικών πληροφοριών θέτει μοναδικές προκλήσεις. Για παράδειγμα, ο σημασιολογικός πλούτος των μαθηματικών εκφράσεων, η ποικιλία των τρόπων αναπαράστασης της ίδιας μαθηματικής έννοιας και η ανάγκη για βαθιά κατανόηση για να κριθεί η συνάφεια καθιστούν τη διαδικασία αξιολόγησης περίπλοκη. Επιπλέον, η συνάφεια στο πλαίσιο της μαθηματικής ανάκτησης πληροφοριών δεν αφορά μόνο την επικαιρότητα. εξετάζει επίσης εάν οι μαθηματικές εκφράσεις στο ερώτημα και στο ανακτηθέν έγγραφο είναι ισοδύναμες ή σχετίζονται με κάποιο ουσιαστικό τρόπο. Ως εκ τούτου, η διαδικασία αξιολόγησης για το ARQMath πιθανότατα απαιτεί αξιολογητές με ισχυρό υπόβαθρο στα μαθηματικά[29].

### *Εν κατακλείδι*

Συνολικά, η χρήση των nDCG', MAP' και P@10 ως μετρήσεων αξιολόγησης στο ARQMath παρέχει μια ολοκληρωμένη αξιολόγηση της αποτελεσματικότητας, της ακρίβειας και της ποιότητας κατάταξης των συστημάτων ανάκτησης μαθηματικών πληροφοριών. Αυτές οι μετρήσεις επιτρέπουν στους ερευνητές να συγκρίνουν και να αναλύουν την απόδοση διαφορετικών μοντέλων και αλγορίθμων ανάκτησης στον τομέα του μαθηματικού περιεχομένου.[17]

## *5.5 Task 1*

Ο στόχος του Task1 είναι να βρει και να ταξινομήσει σχετικές απαντήσεις σε μαθηματικές ερωτήσεις.

Τα θέματα δημιουργούνται από ερωτήσεις που δημοσιεύτηκαν στο Math Stack Exchange το 2021 και η συλλογή για αναζήτηση είναι μόνο οι απαντήσεις σε προηγούμενες ερωτήσεις (από το 2010-2018) στη συλλογή ARQMath.

## *5.6 Αποτελέσματα συστήματος των ομάδων*

Τα αποτελέσματα του συστήματος (“runs”) αξιολογούνται χρησιμοποιώντας μέτρα που χαρακτηρίζουν τον βαθμό στον οποίο οι απαντήσεις που κρίνονται από τους αξιολογητές συνάφειας ως με υψηλότερη συνάφεια προηγούνται των απαντήσεων με χαμηλότερη συνάφεια στα αποτελέσματα του συστήματος (π.χ. χρησιμοποιώντας nDCG′).

Η χειροκίνητη κύρια εκτέλεση της ομάδας Approach0 για το task 1 πέτυχε τα καλύτερα αποτελέσματα, με 0,508 nDCG′.

Μεταξύ των αυτόματων εκτελέσεων, το nDCG′, 0,504, επιτεύχθηκε από την ομάδα MSM.

Οι υψηλότερες δυνατές τιμές nDCG′ και MAP′ είναι 1,0, αλλά επειδή λιγότερες από 10 αξιολογημένες σχετικές απαντήσεις βρέθηκαν στις ομάδες για ορισμένα θέματα, η υψηλότερη δυνατή τιμή P′@10 στο Task 1 του ARQMath-3 είναι 0,95 όπως φαίνεται στον παρακάτω πίνακα.

| | | Type | | ARQMath-1 77 Topics | | | ARQMath-2 71 topics | | | ARQMath-3 78 topics | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Run | D | P | M | nDCG' | MAP' | P'@10 | nDCG' | MAP' | P'@10 | nDCG' | MAP' | P'@10 |
| **Baselines** | | | | | | | | | | | | |
| TF-IDF(Terrier) | B | | | 0.204 | 0.049 | 0.073 | 0.185 | 0.046 | 0.063 | 0.272 | 0.064 | 0.124 |
| TF-IDF(PyTerrier) +Tangent-S | B | (✓) | | 0.249 | 0.059 | 0.081 | 0.158 | 0.035 | 0.072 | 0.229 | 0.045 | 0.097 |
| TF-IDF(PyTerrier) | B | | | 0.218 | 0.079 | 0.127 | 0.120 | 0.029 | 0.055 | 0.190 | 0.035 | 0.065 |
| Tangent-S | M | | | 0.158 | 0.033 | 0.051 | 0.111 | 0.027 | 0.052 | 0.159 | 0.039 | 0.086 |
| *Linked MSE posts* | *L | (✓) | | 0.279 | 0.194 | 0.384 | 0.203 | 0.120 | 0.282 | 0.106 | 0.051 | 0.168 |
| **approach0** | | | | | | | | | | | | |
| fusion_alpha05 | B | ✓ | ✓ | 0.462 | 0.244 | 0.321 | 0.460 | 0.226 | 0.296 | **0.508** | **0.216** | **0.345** |
| fusion_alpha03 | B | | ✓ | 0.460 | 0.246 | 0.312 | 0.450 | 0.221 | 0.278 | 0.495 | 0.203 | 0.317 |
| fusion_alpha02 | B | | ✓ | 0.455 | 0.243 | 0.309 | 0.443 | 0.217 | 0.266 | 0.483 | 0.195 | 0.305 |
| rerank_nostemer | B | | ✓ | 0.382 | 0.205 | 0.322 | 0.385 | 0.187 | 0.276 | 0.418 | 0.172 | 0.309 |
| a0porter | B | | ✓ | 0.373 | 0.204 | 0.270 | 0.383 | 0.185 | 0.241 | 0.397 | 0.159 | 0.271 |
| **MSM** | | | | | | | | | | | | |
| Ensemble_RRF | B | ✓ | | 0.422 | 0.172 | 0.197 | 0.381 | 0.119 | 0.152 | 0.504 | 0.157 | 0.241 |
| BM25_system | B | | | 0.332 | 0.123 | 0.168 | 0.285 | 0.082 | 0.116 | 0.396 | 0.122 | 0.194 |
| BM25_Tfidf _system | B | | | 0.332 | 0.123 | 0.168 | 0.286 | 0.083 | 0.116 | 0.396 | 0.122 | 0.194 |
| TF-IDF | B | | | 0.238 | 0.074 | 0.117 | 0.169 | 0.040 | 0.076 | 0.280 | 0.064 | 0.081 |
| CompuBERT22 | B | | | 0.115 | 0.038 | 0.099 | 0.098 | 0.030 | 0.090 | 0.130 | 0.025 | 0.059 |
| **MIRMU** | | | | | | | | | | | | |
| MiniLM+RoBERTa | B | ✓ | | 0.466 | 0.246 | 0.339 | 0.487 | 0.233 | 0.316 | 0.498 | 0.184 | 0.267 |
| MiniLM +MathRoBERTa | B | | | 0.466 | 0.246 | 0.339 | 0.484 | 0.227 | 0.310 | 0.496 | 0.181 | 0.273 |
| MiniLM_tuned +MathRoBERTa | B | | | 0.470 | 0.240 | 0.335 | 0.472 | 0.221 | 0.309 | 0.494 | 0.178 | 0.262 |
| MiniLM_tuned +RoBERTa | B | | | 0.466 | 0.246 | 0.339 | 0.487 | 0.233 | 0.316 | 0.472 | 0.165 | 0.244 |
| MiniLM+RoBERTa | T | | | 0.298 | 0.124 | 0.201 | 0.277 | 0.104 | 0.180 | 0.350 | 0.107 | 0.159 |
| **MathDowsers** | | | | | | | | | | | | |
| L8_a018 | B | ✓ | | 0.511 | 0.261 | 0.307 | 0.510 | 0.223 | 0.265 | 0.474 | 0.164 | 0.247 |
| L8_a014 | B | | | 0.513 | 0.257 | 0.313 | 0.504 | 0.220 | 0.265 | 0.468 | 0.155 | 0.237 |
| L1on8_a030 | B | | | 0.482 | 0.241 | 0.281 | 0.507 | 0.224 | 0.282 | 0.467 | 0.159 | 0.236 |
| **TU_DBS** | | | | | | | | | | | | |
| math_10 | B | ✓ | | 0.446 | 0.268 | 0.392 | 0.454 | 0.228 | 0.321 | 0.436 | 0.158 | 0.263 |
| Khan_SE_10 | B | | | 0.437 | 0.254 | 0.357 | 0.437 | 0.214 | 0.309 | 0.426 | 0.154 | 0.236 |
| base_10 | B | | | 0.438 | 0.252 | 0.369 | 0.434 | 0.209 | 0.299 | 0.423 | 0.154 | 0.228 |
| roberta_10 | B | | | 0.438 | 0.254 | 0.372 | 0.446 | 0.224 | 0.309 | 0.413 | 0.150 | 0.226 |
| math_10_add | B | | | 0.421 | 0.264 | 0.405 | 0.566 | 0.445 | 0.589 | 0.379 | 0.149 | 0.278 |
| **DPRL** | | | | | | | | | | | | |
| SVM-Rank | B | ✓ | | 0.508 | 0.467 | 0.604 | 0.533 | 0.460 | 0.596 | 0.283 | 0.067 | 0.101 |
| RRF-AMR-SVM | B | | | **0.587** | **0.519** | **0.625** | **0.582** | **0.490** | **0.618** | 0.274 | 0.054 | 0.022 |
| QQ-QA-RawText | B | | | 0.511 | 0.467 | 0.604 | 0.532 | 0.460 | 0.597 | 0.245 | 0.054 | 0.099 |
| QQ-QA-AMR | B | | | 0.276 | 0.180 | 0.295 | 0.186 | 0.103 | 0.237 | 0.185 | 0.040 | 0.091 |
| QQ-MathSE-AMR | B | | | 0.231 | 0.114 | 0.218 | 0.187 | 0.069 | 0.138 | 0.178 | 0.039 | 0.081 |
| **SCM** | | | | | | | | | | | | |
| Interpolated_text +positional_word 2vec_tangentl | B | ✓ | | 0.254 | 0.102 | 0.182 | 0.197 | 0.059 | 0.149 | 0.257 | 0.060 | 0.119 |
| Joint_word2vec | B | | | 0.247 | 0.105 | 0.187 | 0.183 | 0.047 | 0.106 | 0.249 | 0.059 | 0.106 |
| Joint_tuned _roberta | B | | | 0.248 | 0.104 | 0.187 | 0.184 | 0.047 | 0.109 | 0.249 | 0.059 | 0.105 |
| Joint_positional _word2vec | B | | | 0.247 | 0.105 | 0.190 | 0.184 | 0.047 | 0.109 | 0.248 | 0.059 | 0.105 |
| Joint_roberta_base | T | | | 0.135 | 0.048 | 0.101 | 0.099 | 0.023 | 0.060 | 0.188 | 0.040 | 0.077 |

Πίνακας 1: Αποτελέσματα ARQMath 2022 Task1 (CQA). P: κύρια εκτέλεση (primary), M: μη αυτόματη εκτέλεση (manual), (✓): baseline συγκεντρωμένη ως κύρια εκτέλεση.

Πηγη: https://www.cs.rit.edu/~rlaz/files/ARQMath_2022_Overview_WorkingNotes.pdf

Πίνακας ομάδων και αποτελέσματα:

Αυτός ο πίνακας είναι μια περίληψη των αποτελεσμάτων από τον διαγωνισμό ARQMath 2022 Task 1 (Community Question Answering, ή CQA).

Κάθε γραμμή στον πίνακα αντιπροσωπεύει ένα διαφορετικό σύστημα ή προσέγγιση που χρησιμοποιείται στον διαγωνισμό, ενώ οι στήλες αντιπροσωπεύουν διάφορα μέτρα αξιολόγησης.

Ορισμένα κοινά μέτρα στον πίνακα περιλαμβάνουν το nDCG', το MAP' και το P'@10 (Ακρίβεια στο 10).

.

Ακολουθεί μια σύντομη εξήγηση για κάθε στήλη:

Run: Υποδεικνύει τα baselines και τις ομάδες που συμμετείχαν ανα τα χρόνια στο διαγωνισμό.

Type: Υποδεικνύει τον τύπο της εκτέλεσης που υποβλήθηκε. Το P για την κύρια εκτέλεση, το M για τη μη αυτόματη εκτέλεση και ένα σημάδι ελέγχου υποδεικνύει ότι μια γραμμή βάσης συγκεντρώθηκε ως κύρια εκτέλεση.

nDCG': Η βαθμολογία στο μέτρο nDCG'. Το nDCG' είναι μια τροποποίηση του nDCG που ενσωματώνει τη συμπεριφορά του χρήστη στο μέτρο. Όσο υψηλότερο είναι το nDCG', τόσο καλύτερη είναι η κατάταξη των σχετικών εγγράφων.

MAP': Η βαθμολογία στο μέτρο του MAP'. Το MAP' είναι μια τροποποίηση του MAP που λαμβάνει επίσης υπόψη τη συμπεριφορά των χρηστών. Όσο υψηλότερο είναι το MAP', τόσο πιο σχετικά έγγραφα κατατάσσονται σε υψηλή κατάταξη.

P'@10: Η βαθμολογία στο μέτρο P'@10. Αυτό μετρά την αναλογία των σχετικών εγγράφων στα 10 κορυφαία ανακτημένα έγγραφα.

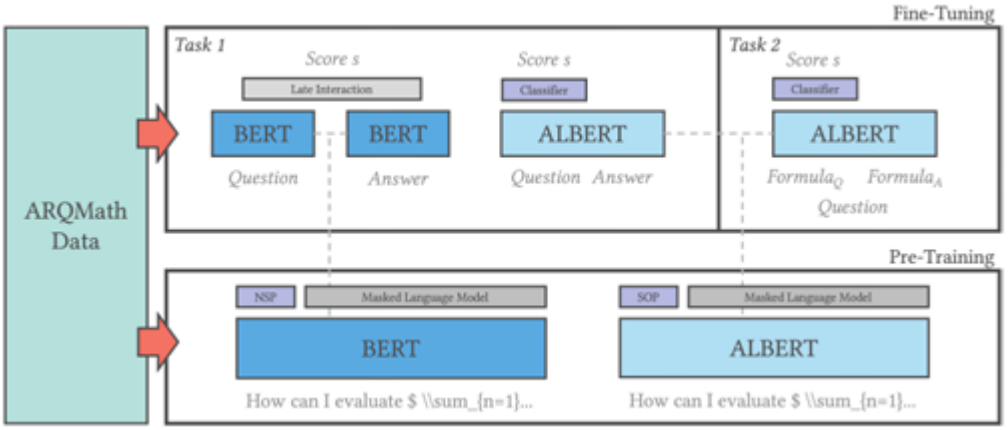
### *5.6.1 Πως προσέγγισαν το πρόβλημα οι ομάδες με τα καλύτερα αποτελέσματα*

#### *5.6.1.1 Η ομάδα TU_DBS*

Η ομάδα *TU_DBS* μελετά την εκπαίδευση και την απόδοση των μοντέλων που βασίζονται στο ColBERT, εστιάζοντας συγκεκριμένα σε μια παραλλαγή που αναφέρεται ως "ColSciBERT". Αυτό το μοντέλο εκπαιδεύτηκε χρησιμοποιώντας τη μέθοδο ColBERT, αλλά αρχικοποιήθηκε με τη βάση του SciBERT, μια παραλλαγή του BERT προεκπαιδευμένη σε ένα μεγάλο σώμα επιστημονικού κειμένου.

Παρά την υπόσχεση του SciBERT για εργασίες που σχετίζονται με επιστημονικό κείμενο, η ομάδα διαπίστωσε ότι το ColSciBERT δεν απέδωσε τη βέλτιστη απόδοση, υποδηλώνοντας πιθανά προβλήματα κατά την εκπαίδευση ή την αξιολόγηση ή ότι το SciBERT μπορεί να μην είναι κατάλληλο για αυτό το συγκεκριμένο σενάριο. Παρόλο που άλλα μοντέλα που βασίζονται στο ColBERT που δεν υποβλήθηκαν

επίσημα στον διαγωνισμό σημείωσαν υψηλότερη βαθμολογία, εξακολουθούν να μην μπορούν να ταιριάζουν με την απόδοση των προσεγγίσεων που βασίζονται στο ALBERT και στις τρεις μετρήσεις αξιολόγησης.




Φιγούρα 1: Επισκόπηση της προ εκπαιδευτικής διαδικασίας

Η ομάδα, ωστόσο, βρήκε ένα σημαντικό πλεονέκτημα όσον αφορά την ταχύτητα αξιολόγησης όταν χρησιμοποιούσε την προσέγγιση που βασίζεται στο ColBERT. Η αξιολόγηση και των 100 θεμάτων από τον διαγωνισμό του 2020 χρειάστηκε μόνο περίπου έξι λεπτά χρησιμοποιώντας δύο GPU NVIDIA GTX 1080, ενώ η αξιολόγηση ενός μεμονωμένου ερωτήματος χρησιμοποιώντας την προσέγγισή τους που βασίζεται στο ALBERT χρειάστηκε από δέκα λεπτά έως μία ώρα σε μία GPU NVIDIA V100.

Δεδομένου αυτού του σημαντικού πλεονεκτήματος ταχύτητας, η ομάδα προτείνει ότι η περαιτέρω έρευνα προς την κατεύθυνση των συστημάτων που βασίζονται στο ColBERT θα μπορούσε να αξίζει τον κόπο, καθώς θα μπορούσαν να παρέχουν ανταγωνιστική απόδοση ενώ παράλληλα μειώνουν σημαντικά τον χρόνο αξιολόγησης. Οι μελλοντικές εργασίες θα πρέπει να στοχεύουν στην καλύτερη κατανόηση των χαρακτηριστικών απόδοσης των συστημάτων που βασίζονται στο ColBERT και στον εντοπισμό των καλύτερων σεναρίων εκπαίδευσης για αυτά τα μοντέλα.

Συμπερασματικά, πρότειναν υποσχόμενες προσεγγίσεις για την Ανάκτηση Μαθηματικών Απαντήσεων και την Αναζήτηση Ομοιότητας Τύπου, εφαρμόζοντας διαφορετικά προεκπαιδευμένα και τελειοποιημένα μοντέλα ALBERT και ένα μοντέλο ColBERT. Παρείχαν επίσης συστάσεις για τη βελτίωση των δυνατοτήτων μοντελοποίησης των μαθηματικών τύπων και κυκλοφόρησαν τον κώδικά τους για προεπεξεργασία δεδομένων και εκπαίδευση μοντέλων. [34]

Τα αποτελέσματα της ομάδας στον διαγωνισμό, παρατηρώντας τα σκορ τα οποία είναι χαμηλότερα σε σύγκριση με την ομάδα Approach0:

Τα αποτελέσματα της ομάδας παρατηρούνται στον Πίνακα 1.

### *5.6.1.2 Η ομάδα MathDowsers*

Η ομάδα MathDowsers από το Πανεπιστήμιο του Waterloo συμμετείχε στον διαγωνισμό ARQMath για την Ανάκτηση Μαθηματικών Πληροφοριών.

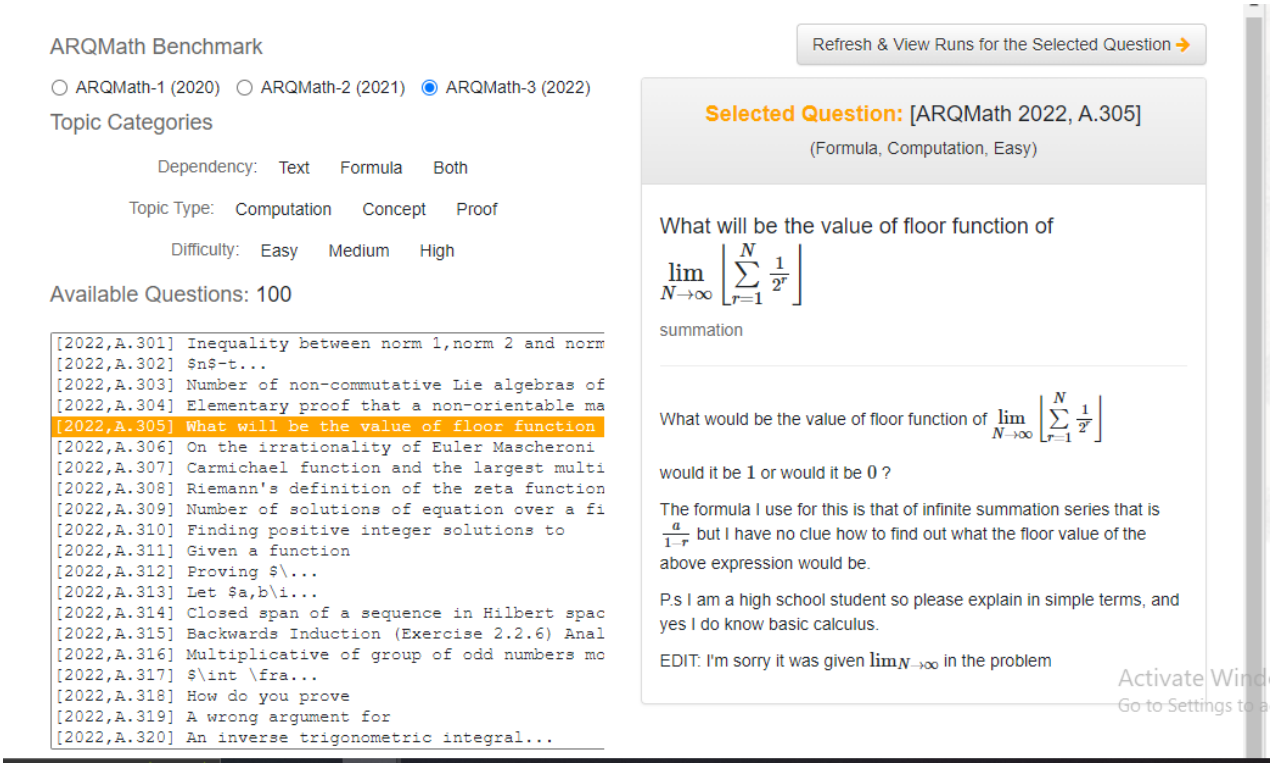
Στους διαγωνισμούς ARQMath-1 και ARQMath-2, η ομάδα υπέβαλε συμμετοχές τόσο για την εργασία Ανάκτησης Απαντήσεων (Εργασία 1) όσο και για την εργασία Ανάκτησης Φόρμουλας (Εργασία 2), επιτυγχάνοντας κορυφαία αποτελέσματα και στις δύο κατηγορίες το 2020 και το 2021. Χρησιμοποίησαν μια μηχανή αναζήτησης μαθηματικών για την εργασία τους η οποία ονομάζεται 'Tangent-L'.

Το Tangent-L είναι μια μηχανή αναζήτησης με γνώση των μαθηματικών που βασίζεται στο Lucene, μια δημοφιλή βιβλιοθήκη αναζήτησης ανοιχτού κώδικα. Σχεδιάστηκε για να κατανοεί και να χειρίζεται μαθηματικούς τύπους, κάτι που το κάνει μετατρέποντας αυτούς τους τύπους σε μορφή παρόμοια με τις λέξεις που χρησιμοποιούνται στις παραδοσιακές μηχανές αναζήτησης κειμένου.

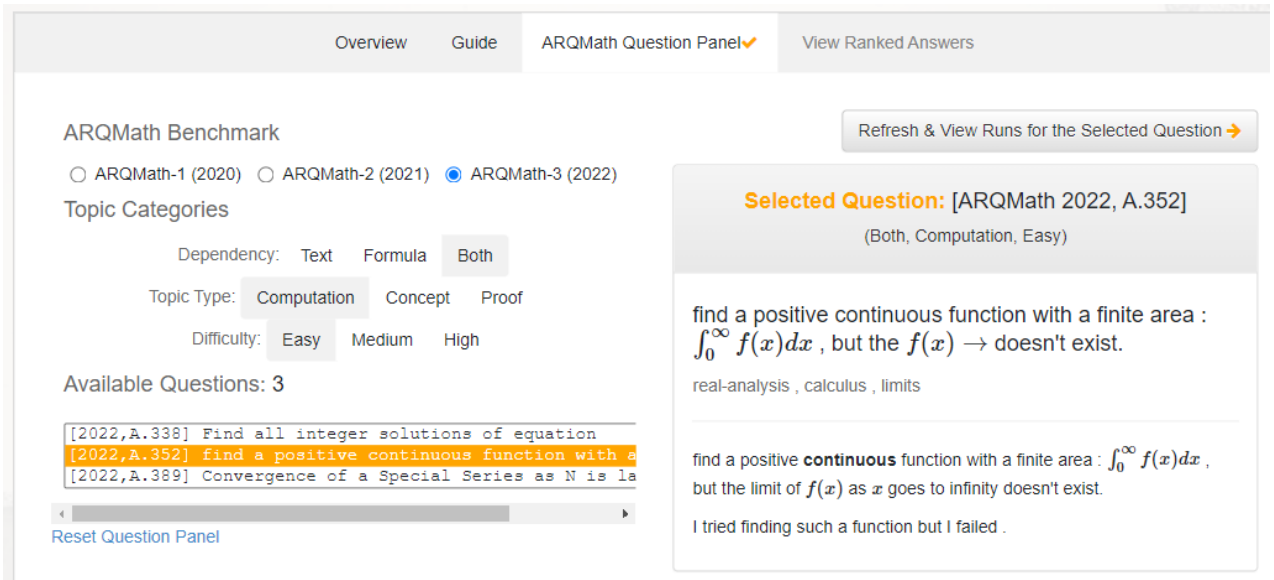
Η λειτουργία του ξεκινάει με ένα τύπο ως είσοδο ο οποίος μετατρέπεται πρώτα σε Δέντρο διάταξης συμβόλων ( Symbol Layout Tree - SLT). Αυτό το δέντρο αντιπροσωπεύει τη δομή του τύπου, με κόμβους που αντιπροσωπεύουν μαθηματικά σύμβολα και άκρες που αντιπροσωπεύουν τις χωρικές σχέσεις μεταξύ αυτών των συμβόλων. Στη συνέχεια, το SLT εξάγει διάφορα σύνολα μαθηματικών συμβολων οποία αποτυπώνουν κάποια χαρακτηριστικά της φόρμουλας. Αυτά τα μαθηματικά σύμβολα αντικαθιστούν τον αρχικό μαθηματικό τύπο στο έγγραφο ή το ερώτημα κατά τη διαδικασία δημιουργίας ευρετηρίου και αναζήτησης.

Το Tangent-L διατηρεί ένα ευρετήριο που συνδυάζει τόσο τα μαθηματικά όσο και τις λέξεις και χρησιμοποιεί μια παραδοσιακή μέθοδο κατάταξης Ανάκτησης Πληροφοριών που βασίζεται σε κείμενο που ονομάζεται BM25+. Υποστηρίζει επίσης αντιστοίχιση χαρακτήρων μπαλαντέρ, επιτρέποντας πιο ευέλικτα ερωτήματα αναζήτησης. Το σύστημα έχει βελτιωθεί με την πάροδο του χρόνου και έχει χρησιμοποιηθεί από την ομάδα MathDowsers για την επίτευξη κορυφαίων επιδόσεων στον διαγωνισμό ARQMath.

Το σύστημα BrushSearch είναι μια διεπαφή χρήστη που επιτρέπει στους χρήστες να εισάγουν αυθαίρετους τύπους και λέξεις-κλειδιά ή φράσεις και να αναζητούν διάφορες βάσεις δεδομένων χρησιμοποιώντας το Tangent-L. Το πρόγραμμα περιήγησης MathDowsers επιτρέπει επίσης στους χρήστες να εξερευνούν τα αποτελέσματα αναζήτησης του Tangent-L για τις εργασίες MathCQA στους διαγωνισμούς ARQMath.




Φιγούρα 2: Η ομάδα MathDowsers χρησιμοποιώντας το BrushSearch για την διεπαφή χρήστη αναζητώντας ένα ερώτημα στα αριστερα και δίνοντας την απάντηση στα δεξια. [36]



Φιγούρα 3: Αλλαγή φίλτρων και ομαδοποίηση των ερωτημάτων. [36]

Παρατηρείται οτι στην Φιγούρα 2 και 3, αλλάζοντας τα φίλτρα, τα ερωτήματα μειώνονται με βάση αυτα.

Για το ARQMath-3, παρουσίασαν μια νέα μηχανή που ονομάζεται mtextsearch. Αυτό το σύστημα απλοποίησε τον πειραματισμό, βελτίωσε την αποτελεσματικότητα ανάκτησης και έδειξε ότι τα παραδοσιακά συστήματα ανάκτησης κειμένου θα μπορούσαν να βελτιωθούν ώστε να γίνουν αποτελεσματικές μηχανές αναζήτησης μαθηματικών.

Εκτός από τις εργασίες του διαγωνισμού, έχουν επίσης αναπτύξει μια Χειρόγραφη Γεννήτρια Μαθηματικών (Handwritten Math Generator) για τη δημιουργία ενός μεγάλου συνόλου δεδομένων διαφορετικών χειρόγραφων εκφράσεων για την εκπαίδευση και τη δοκιμή του μαθηματικού τους συστήματος αναγνώρισης.

Η ομάδα έχει παράσχει πόρους και οδηγίες για τη δημιουργία ενός σώματος εγγράφων για την εργασία Ανάκτησης Απαντήσεων και για τη δημιουργία ερωτημάτων αναζήτησης. Αυτό επιτρέπει σε άλλους να αναπαράγουν τη δουλειά τους ή να την αναπτύξουν. Επιπλέον, παρέχουν μια διεπαφή επίδειξης για τους χρήστες για να εξερευνήσουν τα αποτελέσματα αναζήτησης του Tangent-L για τις εργασίες MathCQA για τα ARQMath-1, ARQMath-2 και ARQMath-3.

Η ερευνητική εργασία και οι υποβολές διαγωνισμών των MathDowsers έχουν τεκμηριωθεί σε αρκετές δημοσιευμένες εργασίες, συμπεριλαμβανομένης μιας διατριβής MMath και πολλαπλών δημοσιεύσεων. Η συνεχιζόμενη έρευνά τους στοχεύει στη βελτίωση της αποτελεσματικότητας και της βιωσιμότητας των παραδοσιακών συστημάτων Ανάκτησης Μαθηματικών Πληροφοριών. [35] [37]

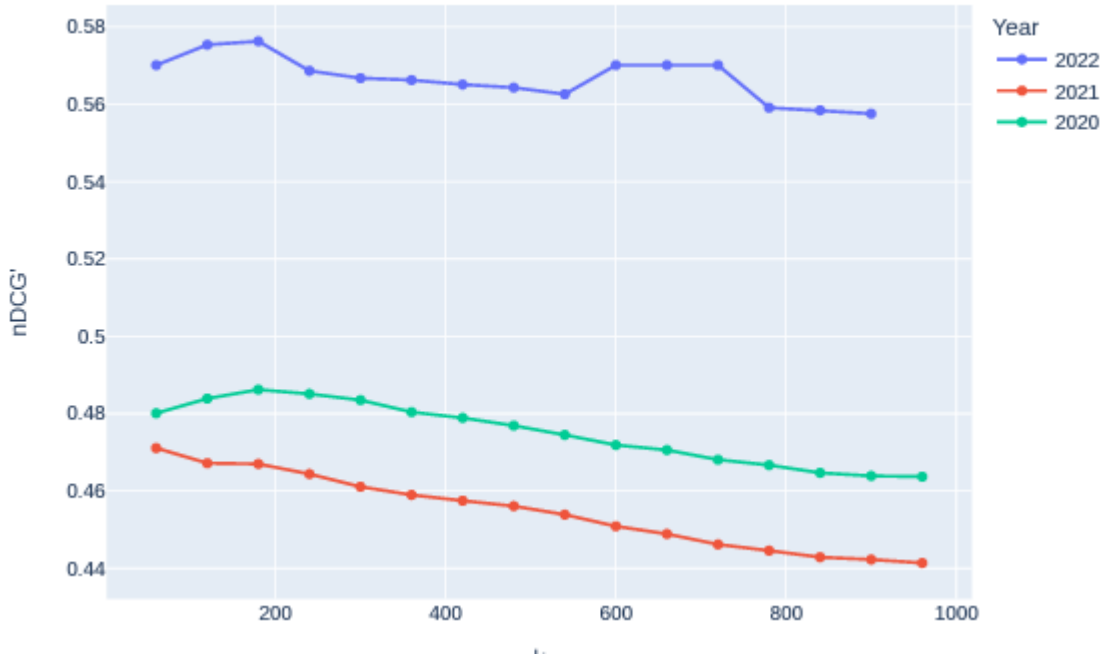
Τα αποτελέσματα της ομάδας παρατηρούνται στον Πίνακα 1.

#### *5.6.1.3 Η ομάδα MSM*

Οι ομάδα του Πανεπιστημίου Masaryk msm ανέπτυξε πολλές διαφορετικές εκδόσεις συστημάτων ανάκτησης πληροφοριών για τον διαγωνισμό ARQMath.

Η ομάδα msm ανέπτυξε πολλά συστήματα, συμπεριλαμβανομένων μοντέλων που βασίζονται στο tf-idf και στο BM25. Κατασκεύασαν επίσης συστήματα συνόλων που συνδυάζουν τις βαθμολογίες μεμονωμένων συστημάτων όπως τα μοντέλα tf-idf και BM25. Τα αποτελέσματα του συνόλου λήφθηκαν χρησιμοποιώντας ένα απλό άθροισμα των βαθμολογιών από τα επιμέρους συστήματα. Χρησιμοποίησαν υλοποιήσεις αυτών των μοντέλων που είναι διαθέσιμες σε βιβλιοθήκες Python όπως το Gensim. Χρησιμοποίησαν τεχνικές για να συνδυάσουν τα δυνατά σημεία διαφορετικών συστημάτων για να λάβουν πιο ακριβή αποτελέσματα. Για αυτό, χρησιμοποίησαν τεχνικές όπως IBC (Inverse Borda Count ), RRF (Reciprocal Rank Fusion) και RBC (Regression Model).

Τα συστήματα εκπαιδεύτηκαν και αναπτύχθηκαν χρησιμοποιώντας διάφορους πόρους, μεθόδους και μαθηματικές αναπαραστάσεις. Δοκίμασαν τα συστήματά τους σε θέματα από προηγούμενους διαγωνισμούς ARQMath και αξιολόγησαν την απόδοσή τους χρησιμοποιώντας μέτρα όπως nDCG , MAP και P@10.




Φιγούρα 4: Απόδοση του συνόλου rrf και των 9 μεμονωμένων συστημάτων που υποβλήθηκαν ανα τα χρόνια.Το καλύτερο *k*=180 που αναφέρθηκε για το arqmath 2022 είναι ndcg′=0,576.

Η ομάδα σημείωσε ότι οι επιδόσεις των συστημάτων τους βελτιώθηκαν με περισσότερα βασικά δεδομένα και εμπειρία από προηγούμενες αξιολογήσεις. Διαπίστωσαν ότι οι μέθοδοι συνόλου, ειδικά εκείνες που χρησιμοποιούν αμοιβαία σύντηξη κατάταξης, έδωσαν τα καλύτερα αποτελέσματα. Για να προωθήσουν την

αναπαραγωγιμότητα, η ομάδα σχεδιάζει να δημοσιεύσει τα σημειωματάρια, τις υλοποιήσεις συνόλων και τα μοντέλα τους στο εργαστήριο και το αποθετήριο μαθημάτων τους. [38]

## *5.7 Το σύστημα "Approach 0"*

Όπως αναφέρθηκε, η ομάδα Approach0 για το task 1 πέτυχε τα καλύτερα αποτελέσματα, με 0,508 nDCG′ score. Η προσέγγιση της ομάδας αυτής έχει το ιδιαίτερο χαρακτηριστικό του "fusion".

Στο πλαίσιο της ανάκτησης πληροφοριών, η λέξη "fusion" αναφέρεται συνήθως σε τεχνικές συγχώνευσης δεδομένων ή αποτελεσμάτων, οι οποίες συνδυάζουν τα αποτελέσματα από πολλαπλά συστήματα ανάκτησης ή πολλαπλές εκτελέσεις του ίδιου συστήματος για τη βελτίωση της συνολικής απόδοσης ανάκτησης. Η τεχνική αυτή είναι γνωστή στην ανάκτηση πληροφοριών διότι ο συνδυασμός στρατηγικών αναζήτησης οδηγεί σε καλύτερα αποτελέσματα.

Στα πλαίσια του fusion η ομάδα χρησιμοποιεί το Approach 0 pass και το Colbert pass χρησιμοποιώντας και στις δύο περιπτώσεις τη βιβλιοθήκη pya0 για το evaluation.

Η προσέγγιση Zero-Pass Retrieval, γνωστή και ως προσέγγιση μηδενικής βολής, είναι μια μέθοδος που χρησιμοποιείται στα συστήματα ανάκτησης πληροφοριών. Στόχος είναι η ανάκτηση σχετικών πληροφοριών χωρίς καμία αλληλεπίδραση ή ανατροφοδότηση από τον χρήστη. Με άλλα λόγια, προσπαθεί να έχει τα καλύτερα δυνατά αποτελέσματα στην πρώτη προσπάθεια, εξ ου και το όνομα 'zero-pass'.

### *5.7.1 Πώς λειτουργεί η Zero-Pass Retrieval:*

Επεξεργασία ερωτήματος (query processing): Το σύστημα ανάκτησης επεξεργάζεται το ερώτημα αναζήτησης του χρήστη για να κατανοήσει ποιες πληροφορίες αναζητούνται.

Εκτέλεση αναζήτησης (search execution): Το σύστημα εκτελεί την αναζήτηση στη βάση δεδομένων ή στο σώμα των εγγράφων χρησιμοποιώντας έναν αλγόριθμο ανάκτησης πληροφοριών. Αυτός ο αλγόριθμος

μπορεί να χρησιμοποιεί τεχνικές όπως ο όρος frequency-inverse document frequency (TF-IDF), η σημασιολογική αντιστοίχιση, η ενσωμάτωση λέξης κ.λπ.

Κατάταξη αποτελεσμάτων (result rankings): Το σύστημα δημιουργεί μια λίστα με πιθανά αποτελέσματα και τα κατατάσσει ανάλογα με τη συνάφειά τους με το ερώτημα αναζήτησης. Η συνάφεια καθορίζεται συνήθως με βάση παράγοντες όπως η αντιστοίχιση λέξεων-κλειδιών, η σημασιολογική ομοιότητα, η δημοτικότητα των εγγράφων και άλλα.

Παρουσίαση Αποτελεσμάτων (result presentation): Τα αποτελέσματα με την κορυφαία κατάταξη παρουσιάζονται στη συνέχεια στον χρήστη.

Κειμενική και σημασιολογική ανάλυση: Το Zero-Pass μετατρέπει τους μαθηματικούς τύπους σε όρους κειμένου. Λαμβάνει υπόψη το κείμενο που περιβάλλει τους μαθηματικούς τύπους, καθώς και το κείμενο μέσα στους ίδιους τους τύπους. Η μηχανή αναζήτησης μπορεί έτσι να χειριστεί ερωτήματα που περιλαμβάνουν τόσο μαθηματικούς τύπους όσο και κανονικό κείμενο.

Ανάλυση μαθηματικών εκφράσεων: Το πιο σημαντικό είναι ότι το Zero-Pass εφαρμόζει δομική αντιστοίχιση μαθηματικών τύπων. Αυτό σημαίνει ότι λαμβάνει υπόψη τη δομή της μαθηματικής έκφρασης στον τύπο, όχι μόνο την κειμενική αναπαράστασή της. Αυτό του επιτρέπει να βρίσκει τύπους που είναι μαθηματικά παρόμοιοι, ακόμα κι αν είναι γραμμένοι διαφορετικά. Με άλλα λόγια, αντιμετωπίζει τη μαθηματική έκφραση ως μια αδόμητη σειρά χαρακτήρων και δεν επιχειρεί να αναγνωρίσει το υποκείμενο μαθηματικό νόημα ή τη σημασιολογία. Αντίθετα, βασίζεται αποκλειστικά σε μοτίβα και χαρακτηριστικά σε επίπεδο επιφάνειας για την ανάκτηση σχετικών πληροφοριών.

Αξίζει να σημειωθεί ότι αυτή η προσέγγιση είναι γρήγορη και λειτουργεί καλά για απλά και συγκεκριμένα ερωτήματα. Ωστόσο, ενδέχεται να μην λειτουργεί τόσο αποτελεσματικά για πολύπλοκα ή διφορούμενα ερωτήματα, όπου η αλληλεπίδραση με τον χρήστη μπορεί να βελτιώσει σημαντικά τα αποτελέσματα ανάκτησης. Για αυτές τις περιπτώσεις, άλλες προσεγγίσεις όπως η Interactive Retrieval ή η Multi-Pass Retrieval μπορεί να είναι πιο κατάλληλες.

Έτσι, το Approach0 χρησιμοποιεί Κειμενική και Σημασιολογική Ανάλυση, αλλά χρησιμοποιεί επίσης και Ανάλυση Μαθηματικής Έκφρασης. Αυτές οι τεχνικές, μαζί με το σχήμα στάθμισης, συνδυάζονται για να δημιουργήσουν μια εξελιγμένη και αποτελεσματική μηχανή μαθηματικής αναζήτησης.

### *5.7.2 Πώς λειτουργεί το ColBERT:*

Το ColBERT (Contextualized Late Interaction over BERT) είναι μια αποτελεσματική παραλλαγή του δημοφιλούς μοντέλου BERT, που εισήχθη για εργασίες ανάκτησης μεγάλης κλίμακας. Το ColBERT εφαρμόζει το BERT ανεξάρτητα σε κάθε έγγραφο σε ένα σώμα και σε κάθε ερώτημα, αντί να εφαρμόζει το BERT στο διασταυρούμενο προϊόν των ζευγών (ερώτημα, έγγραφο), το οποίο θα ήταν υπολογιστικά ανέφικτο για μεγάλα σώματα. Στη συνέχεια, το μοντέλο ταξινομεί τα έγγραφα για ένα δεδομένο ερώτημα με βάση τις μαθημένες αλληλεπιδράσεις μεταξύ του εγγράφου και των ερωτήματων.

Ένα "ColBERT pass" στο πλαίσιο της Ανάκτησης Πληροφοριών αναφέρεται συνήθως σε ένα μόνο βήμα ανάκτησης και κατάταξης εγγράφων με βάση τη συνάφειά τους με ένα ερώτημα χρησιμοποιώντας το μοντέλο ColBERT. Αυτή η διαδικασία αποτελείται από δύο στάδια:

Κωδικοποίηση ερωτήματος (query encoding): Το ερώτημα υποβάλλεται σε επεξεργασία μέσω του BERT για τη λήψη ενσωματώσεων ερωτήματος (query embeddings). Αυτές οι ενσωματώσεις αντιπροσωπεύουν το σημασιολογικό περιεχόμενο του ερωτήματος σε έναν χώρο υψηλών διαστάσεων.

Κωδικοποίηση και κατάταξη εγγράφων (document encoding and ranking): Κάθε έγγραφο επεξεργάζεται ανεξάρτητα μέσω του BERT για να ληφθούν ενσωματώσεις εγγράφων. Στη συνέχεια, οι ενσωματώσεις του ερωτήματος και των εγγράφων συγκρίνονται για να υπολογιστεί μια βαθμολογία ομοιότητας. Τα έγγραφα ταξινομούνται με βάση αυτές τις βαθμολογίες ομοιότητας.

Το μοντέλο ColBERT προσφέρει μια ισορροπία μεταξύ της αποτελεσματικότητας και της αποδοτικότητας ανάκτησης, καθιστώντας το κατάλληλο για εργασίες ανάκτησης μεγάλης κλίμακας.

### *5.7.3 Ένωση (Fusion)*

Το Fusion, που μερικές φορές ονομάζεται επίσης σύνολο ή συνδυασμός, είναι μια μέθοδος που χρησιμοποιείται στην Ανάκτηση Πληροφοριών (IR) για τη βελτίωση της ακρίβειας αναζήτησης. Ο κύριος στόχος είναι να συνδυαστούν ή να συγχωνευτούν τα αποτελέσματα από πολλαπλά συστήματα ή αλγόριθμους ανάκτησης για να παραχθεί καλύτερο αποτέλεσμα.

Συνήθως ένας από τους λόγους χρήσης fusion είναι η βελτίωση της επίδοσης. Διαφορετικά συστήματα ανάκτησης ενδέχεται να έχουν καλή απόδοση σε διαφορετικά είδη ερωτημάτων. Συνδυάζοντας πολλαπλά

συστήματα, μπορούμε να αξιοποιήσουμε τα δυνατά σημεία του καθενός, με αποτέλεσμα τη συνολική καλύτερη απόδοση.Η απόδοση ενός μεμονωμένου συστήματος ανάκτησης μπορεί να ποικίλλει ανάλογα με τη φύση των δεδομένων ή τις ιδιαιτερότητες ενός ερωτήματος. Το Fusion μπορεί να βοηθήσει στην αύξηση της ευρωστίας μειώνοντας την πιθανότητα η κακή απόδοση σε ένα συγκεκριμένο ερώτημα από ένα σύστημα να επηρεάσει σημαντικά το συνολικό αποτέλεσμα.

Ένας ακόμη λόγος χρήσης Fusion είναι η αύξηση της διαφορετικότητας. Μερικές φορές, διαφορετικά συστήματα ενδέχεται να ανακτήσουν διαφορετικά αλλά σχετικά αποτελέσματα για το ίδιο ερώτημα. Το Fusion μπορεί να συγκεντρώσει αυτά τα διαφορετικά αποτελέσματα, προσφέροντας στον χρήστη ένα ευρύτερο φάσμα σχετικών πληροφοριών. Το Fusion μπορεί να βοηθήσει στην αντιμετώπιση της εγγενούς αβεβαιότητας στην ανάκτηση πληροφοριών. Με τη συγκέντρωση αποτελεσμάτων από πολλαπλά συστήματα, ο αντίκτυπος της αβεβαιότητας σε οποιοδήποτε μεμονωμένο σύστημα μπορεί να μειωθεί.

Σύμφωνα με τη θεωρία μάθησης συνόλου, μια ομάδα «αδύναμων» συστημάτων (σε αυτό το πλαίσιο, μεμονωμένα συστήματα ανάκτησης) μπορούν να ενωθούν για να σχηματίσουν ένα «ισχυρό» σύστημα. Αυτό σημαίνει ότι ακόμα κι αν τα μεμονωμένα συστήματα δεν είναι πολύ ακριβή, ο συνδυασμός τους μπορεί να επιτύχει καλή απόδοση.

Στην ουσία, ο στόχος της χρήσης της σύντηξης στο IR είναι να δημιουργηθεί ένα πιο ισχυρό και αξιόπιστο σύστημα ανάκτησης συγκεντρώνοντας τις δυνατότητες πολλαπλών μεμονωμένων συστημάτων.

# 6 *Μελέτη συστήματος "Approach 0"*

Ένα σύστημα ανάκτησης πληροφοριών (όπως μια μηχανή αναζήτησης) επεξεργάζεται ερωτήματα σε ένα σύνολο δεδομένων και παράγει ένα αρχείο αποτελεσμάτων. Το αρχείο αποτελεσμάτων έχει συνήθως μια μορφή όπου κάθε γραμμή αντιπροσωπεύει ένα μεμονωμένο αποτέλεσμα και περιλαμβάνει το αναγνωριστικό ερωτήματος (query id), το αναγνωριστικό του εγγράφου που ανακτήθηκε και την κατάταξη (the id of the document retrieved) ή τη βαθμολογία του εγγράφου (score of the document) για αυτό το ερώτημα.

Ξεχωριστά, υπάρχει ένα αρχείο qrel. Αυτό το αρχείο δημιουργείται από αξιολογητές που έχουν καθορίσει τα σχετικά αποτελέσματα για κάθε ερώτημα. Περιλαμβάνει το αναγνωριστικό ερωτήματος, το αναγνωριστικό του εγγράφου και μια κρίση συνάφειας (συχνά δυαδική, αλλά θα μπορούσε να βαθμολογηθεί).

Ο trec_eval αλγόριθμος λαμβάνει και τα δύο αυτά αρχεία ως είσοδο. Περνάει μέσα από το αρχείο αποτελεσμάτων, ελέγχοντας κάθε επιστρεφόμενο έγγραφο σε σχέση με το αρχείο qrel για να δει αν είναι γνωστό σχετικό αποτέλεσμα. Από αυτό, υπολογίζει μια ποικιλία μετρήσεων αξιολόγησης, όπως ακρίβεια, ανάκληση, βαθμολογία F, MAP, NDCG κ.λπ. Αυτές οι μετρήσεις εξάγονται στην κονσόλα ή σε ένα αρχείο για ανάλυση.

Το εργαλείο trec_eval, είναι ένα εργαλείο που αναπτύχθηκε στα πλαίσια των συνεδριών ανάκτησης κειμένου (TREC) για την αξιολόγηση των μηχανών αναζήτησης σχετικά με τις εργασίες ανάκτησής τους.

Ο κώδικας trec_eval δεν επηρεάζει άμεσα τα αποτελέσματα αναζήτησης. Αντίθετα, παρέχει μετρήσεις και βαθμολογίες με βάση την ποιότητα των αποτελεσμάτων αναζήτησής σε σύγκριση με ένα γνωστό σύνολο σχετικών εγγράφων (QREL).

Εν κατακλείδι, το εργαλείο trec_eval αξιολογεί την αποτελεσματικότητα των συστημάτων ανάκτησης πληροφοριών με βάση τα παρεχόμενα αρχεία αποτελεσμάτων και ένα αρχείο "qrel" (το οποίο είναι ένα σύνολο γνωστών σχετικών αποτελεσμάτων).

## *6.1 Πως λειτουργεί η σύγκριση ερωτήματος-δεδομένων*

### *6.1.1 Η συνάρτηση msearch*

Η συνάρτηση msearch είναι η υπεύθυνη να ορχηστρώσει την διαδικασία. Είναι μια ευέλικτη συνάρτηση αναζήτησης που λειτουργεί με διαφορετικούς τύπους υπηρεσιών ή εργαλείων αναζήτησης, συμπεριλαμβανομένης μιας τοπικής βιβλιοθήκης αναζήτησης (pya0), μιας υπηρεσίας απομακρυσμένης αναζήτησης που εκτελείται σε διακομιστή (tcp) ή μιας διαδικασίας αναζήτησης που διαχωρίζεται ως μια ξεχωριστή διαδικασία λειτουργικού συστήματος (fork_search). Η συνάρτηση παίρνει ένα ερώτημα και ένα ευρετήριο (το οποίο μπορεί να είναι τοπικό ή απομακρυσμένο) και στέλνει το ερώτημα στην υπηρεσία αναζήτησης για να ανακτήσει αποτελέσματα.

Η συνάρτηση cascade_run εκτελεί ένα σύνολο διαφορετικών βημάτων επεξεργασίας στο ερώτημα και τα δεδομένα. Κάθε βήμα ορίζεται από την παράμετρο cascades, η οποία είναι μια λίστα πλειάδων (cascade, args) όπου «cascade» είναι το όνομα του βήματος που πρέπει να εκτελεστεί και «args» είναι τα ορίσματα για αυτό το βήμα. Οι διαφορετικοί cascades περιλαμβάνουν μια αναζήτηση πρώτου σταδίου (πρώτο στάδιο), την εφαρμογή μιας λειτουργίας ανάγνωσης στο αρχείο για τη λήψη αποτελεσμάτων (αναγνώστης), την επέκταση ερωτήματος (rm3) και την εκμάθηση κατάταξης (l2r).

Τα αποτελέσματα από ένα βήμα μεταβιβάζονται στη συνέχεια στο επόμενο βήμα του cascade. Για παράδειγμα, ο cascade πρώτου σταδίου μπορεί να εκτελέσει μια αρχική ανάκτηση από το ευρετήριο, μετά ο cascade rm3 θα εκτελούσε μια επέκταση ερωτήματος με βάση τα αρχικά αποτελέσματα και θα εκτελέσει μια δεύτερη ανάκτηση και, τέλος, ο cascade l2r θα ανακαταταξει τα αποτελέσματα με βάση ένα εκπαιδευμένο μοντέλο κατάταξης.

Κάθε cascade χρησιμοποιεί τη συνάρτηση msearch για να εκτελέσει μια λειτουργία αναζήτησης όταν είναι απαραίτητο, και αφού ολοκληρωθούν όλοι οι cascades, επιστρέφονται τα τελικά αποτελέσματα.

### *6.1.2 Υπηρεσία αναζήτησης που χρησιμοποιεί το msearch:*

Η υπηρεσία αναζήτησης είναι το στοιχείο που λαμβάνει πραγματικά ένα ερώτημα και ανακτά αποτελέσματα από ένα ευρετήριο, το οποίο είναι συνήθως μια δομή δεδομένων που διευκολύνει την αναζήτηση εγγράφων με βάση τους όρους ερωτήματος. Η υπηρεσία αναζήτησης μπορεί να είναι μια τοπική βιβλιοθήκη που ονομάζεται pya0, μια απομακρυσμένη υπηρεσία που εκτελείται σε έναν διακομιστή (tcp) ή μια ξεχωριστή διαδικασία που διαχωρίζεται από την κύρια διαδικασία (fork_search).

Η συνάρτηση msearch είναι η διεπαφή με αυτές τις υπηρεσίες. Ανάλογα με τη διαμόρφωση, στέλνει το ερώτημα σε μία από αυτές τις υπηρεσίες και συλλέγει τα αποτελέσματα.

### *6.1.3 Πως λειτουργεί η βιβλιοθήκη pya0*

Η διαδικασία αναζήτησης PyA0 περιλαμβάνει τα ακόλουθα βήματα:

Ευρετηρίαση: Το PyA0 λαμβάνει ένα μεγάλο σύνολο εγγράφων, συχνά ακαδημαϊκές εργασίες, αναρτήσεις ιστολογίου ή άλλα δεδομένα κειμένου. Επεξεργάζεται αυτά τα έγγραφα για να δημιουργήσει ένα ευρετήριο με δυνατότητα αναζήτησης. Κάθε έγγραφο αναλύεται στα επιμέρους συστατικά του και οι μαθηματικές εκφράσεις μετατρέπονται σε μια δομή δέντρου που αντιπροσωπεύει τη σημασιολογική τους σημασία. Αυτό επιτρέπει στη μηχανή αναζήτησης να αντιστοιχίσει όχι μόνο τα ακριβή σύμβολα σε μια μαθηματική έκφραση, αλλά και τη συνολική δομή και το νόημα της έκφρασης.

Επεξεργασία ερωτήματος: Όταν ένας χρήστης εισάγει ένα ερώτημα αναζήτησης, το PyA0 αναλύει το ερώτημα με τον ίδιο τρόπο που κάνει τα έγγραφα. Αυτό σημαίνει ότι ο χρήστης μπορεί να εισάγει μαθηματικές εκφράσεις στα ερωτήματα αναζήτησής του και το PyA0 θα τις κατανοεί σημασιολογικά.

Αναζήτηση: Μόλις ολοκληρωθεί η επεξεργασία του ερωτήματος, το PyA0 πραγματοποιεί αναζήτηση μέσω του ευρετηρίου για έγγραφα που ταιριάζουν με το ερώτημα. Χρησιμοποιεί έναν συνδυασμό αντιστοίχισης κειμένου και σημασιολογικής μαθηματικής αντιστοίχισης για να βρει σχετικά αποτελέσματα.

Κατάταξη: Αφού εντοπιστούν τα σχετικά έγγραφα, το PyA0 τα κατατάσσει με βάση τη συνάφεια. Οι ακριβείς λεπτομέρειες του αλγόριθμου κατάταξης μπορεί να διαφέρουν, αλλά γενικά περιλαμβάνει έναν συνδυασμό του πόσο καλά ταιριάζει το έγγραφο με το ερώτημα, της συνολικής συνάφειας ή σημασίας του εγγράφου στο σώμα και πιθανώς άλλων παραγόντων.

Επιστροφή αποτελεσμάτων: Τέλος, η μηχανή αναζήτησης επιστρέφει τα αποτελέσματα στον χρήστη, συνήθως ταξινομημένα με βάση την κατάταξή του.

## *6.2 Τροποποίηση της λογικής ή του μοντέλου της μηχανής αναζήτησης*

### *6.2.1 Τροποποίηση της συνάρτησης msearch*

Η συνάρτηση msearch στον κώδικά της ομάδας Approach 0 είναι ένα περιτύλιγμα για μερικές διαφορετικές μεθόδους αναζήτησης, συμπεριλαμβανομένης της αναζήτησης pya0, της αναζήτησης μέσω σύνδεσης TCP και μιας αναζήτησης fork search. Τα αποτελέσματα που επιστρέφονται από το msearch θα καθοριστούν σε μεγάλο βαθμό από τη μέθοδο που θα επιλεγεί και το ερώτημα και το ευρετήριο θα περάσουν σε αυτήν.

Για την τροποποιήση της συνάρτησης msearch, θα πρέπει να αλλάξει ο αλγόριθμος αναζήτησης ή να τροποποιηθεί ο τρόπος διεξαγωγής της αναζήτησης. Μερικοί τρόποι για να τροποποιήσουμε τον αλγόριθμο είναι οι εξής:

#### *6.2.1.1 Αλλαγή του αλγόριθμου αναζήτησης:*

Όπως αναφέρθηκε, το msearch χρησιμοποιεί αυτήν τη στιγμή το pya0 για λειτουργίες αναζήτησης. Μια σημαντική αλλαγή η οποία θα επέστρεφε διαφορετικά αποτελέσματα είναι η αντικατάσταση με διαφορετική βιβλιοθήκη ή αλγόριθμο αναζήτησης.

Αυτή η αλλαγή θα επέστρεφε διαφορετικά αποτελέσματα και ίσως με αρνητικό γνώμονα εφόσον η βιβλιοθήκη pya0 είναι εξειδικευμένη στην αναζήτηση σε μαθηματικό περιεχόμενο.

*6.2.1.2 Τροποποίηση παραμέτρων αναζήτησης:*

Η συμπεριφορά αναζήτησης μπορεί να ελεγχθεί τροποποιώντας τις παραμέτρους που μεταβιβάζονται στη λειτουργία αναζήτησης. Η αλλαγή της παραμέτρου topk που ελέγχει τον αριθμό των αποτελεσμάτων που θα επιστραφούν είναι μία από αυτές.

Η παράμετρος topk στη συνάρτηση msearch καθορίζει τον αριθμό των εγγράφων με την κορυφαία βαθμολογία που θα επιστραφούν μετά τη διεξαγωγή μιας αναζήτησης. Η αλλαγή της παραμέτρου topk μπορεί ενδεχομένως να επηρεάσει την ποιότητα των αποτελεσμάτων που λαμβάνονται, αλλά το αν τα αποτελέσματα γίνονται καλύτερα ή χειρότερα μπορεί να εξαρτάται από τη συγκεκριμένη περίπτωση χρήσης.

Εάν αυξηθεί η παράμετρος topk:

Θα επιστρέφονται περισσότερα αποτελέσματα από κάθε ερώτημα, κάτι που μπορεί να είναι επωφελές εάν οι πληροφορίες που αναζητούνται δεν βρίσκονται στα κορυφαία έγγραφα. Αυτό θα μπορούσε να παρέχει μια μεγαλύτερη δεξαμενή που αποτελείται από σχετικά έγγραφα.

Ωστόσο, μπορεί επίσης να αυξήσει τον θόρυβο στα αποτελέσματά ,δηλαδή, έγγραφα που δεν σχετίζονται με το ερώτημά της αναζήτησης αλλά επιστρέφονται λόγω της βαθμολογίας τους.

Επίσης, μπορεί να αυξήσει τον χρόνο υπολογισμού καθώς θα ανακτηθούν και θα βαθμολογηθούν περισσότερα έγγραφα.

Εάν μειωθεί η παράμετρος topk:

Θα ληφθούν λιγότερα αποτελέσματα, γεγονός που μπορεί να κάνει πιο γρήγορη τη μη αυτόματη επεξεργασία των επιστρεφόμενων εγγράφων.

Ωστόσο, μπορεί να χαθούν σχετικά έγγραφα που δεν μπήκαν στα κορυφαία.

Επίσης, εάν ο αλγόριθμος κατάταξης δεν είναι άρτιος αλγοριθμικά, τα πιο σχετικά αποτελέσματα μπορεί να μην βρίσκονται στα κορυφαία αποτελέσματα, επομένως θα χαθούν.

Συνοπτικά, υπάρχει μια αντιστάθμιση μεταξύ του αριθμού των αποτελεσμάτων, της συνάφειας αυτών των αποτελεσμάτων και του χρόνου υπολογισμού. Η βέλτιστη τιμή του topk μπορεί να ποικίλλει ανάλογα με το τι είναι πιο σημαντικό στη συγκεκριμένη περίπτωση χρήσης: εύρος αναζήτησης ή ακρίβεια.

*6.2.1.3 Αποτελέσματα μετά τη διαδικασία:*

Η τροποποίηση των αποτελέσματατων μπορεί να γίνει και αφού επιστραφούν από τη λειτουργία αναζήτησης, αναβαθμίζοντας τα αποτελέσματα με βάση ορισμένα προσαρμοσμένα κριτήρια που επιθυμεί ο χρήστης.

### *6.2.2 Τροποποίηση της συνάρτησης task1_getresults*

Το αρχείο task1_getresults εκτελεί αξιολόγηση των αποτελεσμάτων ανάκτησης χρησιμοποιώντας τυπικές μετρήσεις Ανάκτησης Πληροφοριών (IR). Το εκτελέσιμο χρησιμοποιεί το εργαλείο trec_eval, ένα τυπικό εργαλείο αξιολόγησης που χρησιμοποιείται στη Διάσκεψη ανάκτησης κειμένου (TREC), για τον υπολογισμό διαφόρων μετρήσεων με βάση τα αποτελέσματα των ερωτημάτων και ένα δεδομένο αρχείο "βασικής αλήθειας" (το αρχείο 'qrel'). Οι μετρήσεις που υπολογίζονται περιλαμβάνουν:

Normalized Discounted Cumulative Gain (nDCG): Μια μέτρηση που μετρά την ποιότητα μιας ταξινομημένης λίστας στοιχείων/εγγράφων, λαμβάνοντας υπόψη τη θέση κάθε σχετικού στοιχείου στη λίστα.

MAP: Ο μέσος όρος της μέσης βαθμολογίας ακρίβειας για κάθε ερώτημα στο σύνολο ερωτημάτων.

Ακρίβεια στο 10 (p@10): Η αναλογία των 10 κορυφαίων ανακτημένων εγγράφων που είναι σχετικά.

Δυαδική προτίμηση (bpref): Μια μέτρηση που μετρά τη συσχέτιση κατάταξης και δεν επηρεάζεται από έγγραφα που δεν έχουν κριθεί.

Ποσοστό κριτών: Μια προσαρμοσμένη μέτρηση που σχετίζεται με την εφαρμογή ή το πείραμα που υπάρχει..

Το εκτελέσιμο δημιουργεί αυτές τις μετρήσεις αξιολόγησης για κάθε αρχείο αποτελεσμάτων ανάκτησης στον δεδομένο κατάλογο (ένα αρχείο ανά σύνολο συστήματος/ερωτημάτων) και εγγράφει τα αποτελέσματα σε ένα καθορισμένο αρχείο εξόδου.

### *6.2.3 Τροποποίηση της συνάρτησης arqmath_to_prim_task1*

Αυτό το σενάριο έχει σχεδιαστεί για να διαμορφώνει ξανά αρχεία σε έναν κατάλογο υποβολής σύμφωνα με τη μορφή TREC, η οποία είναι μια τυπική μορφή για σύνολα δεδομένων στην ανάκτηση πληροφοριών. Πως λειτουργεί το σενάριο:

Ανάλυση arguments: Ξεκινά αναλύοντας τέσσερα ορίσματα της γραμμής εντολών:

Η διαδρομή προς ένα αρχείο qrel (-qre), η οποία είναι μια τυπική μορφή για τον καθορισμό των κρίσεων συνάφειας ενός συνόλου δεδομένων ανάκτησης πληροφοριών.

Η διαδρομή προς έναν κατάλογο που περιέχει αρχεία που έχουν υποβληθεί (-sub), τα οποία είναι σε κάποια άλλη μορφή.

Δύο διαδρομές προς τους καταλόγους όπου θα αποθηκευτούν τα επαναδιαμορφωμένα αρχεία: μία για αρχεία σε τυπική μορφή TREC (-tre) και μία για αρχεία "prime" (-pri).

Ανάγνωση του αρχείου Qrel: Στη συνέχεια, το σενάριο διαβάζει το αρχείο qrel σε ένα dictionary αντιστοιχίζοντας αναγνωριστικά θεμάτων (topic IDs) σε λίστες αναγνωριστικών αναρτήσεων (lists of post IDs), χρησιμοποιώντας τη συνάρτηση read_qrel_to_dictionary.

Επεξεργασία Υποβληθέντων Αρχείων: Μετά από αυτό, επεξεργάζεται κάθε αρχείο στον κατάλογο υποβολής με τη συνάρτηση convert_result_files_to_trec:

Για κάθε αρχείο που υποβάλλεται, ανοίγει δύο αντίστοιχα αρχεία εξόδου στους καθορισμένους καταλόγους εξόδου: ένα τυπικό αρχείο με μορφή TREC και ένα αρχείο "prime".

Στη συνέχεια διαβάζει το υποβληθέν αρχείο γραμμή προς γραμμή και για κάθε γραμμή κάνει τα εξής:

Ελέγχει εάν το αναγνωριστικό θέματος και το αναγνωριστικό ανάρτησης για τη γραμμή έχουν ήδη εμφανιστεί. αν έχουν, παρακάμπτει τη γραμμή (καθώς το TREC απαιτεί όλα τα ανακτημένα αποτελέσματα για να είναι ένα θέμα μοναδικό).

Γράφει τη γραμμή στο αρχείο μορφής TREC.

Εάν το αναγνωριστικό θέματος και το αναγνωριστικό ανάρτησης υπάρχουν στο αρχείο qrel, γράφει τη γραμμή και στο αρχικό αρχείο. (Αυτό φαίνεται να είναι αυτό που διακρίνει τα "prime" αρχεία από τα τυπικά αρχεία σε μορφή TREC.)

Εκτέλεση σεναρίου: Το σενάριο προορίζεται να εκτελεστεί ως αυτόνομο σενάριο Python από τη γραμμή εντολών, όπως υποδεικνύεται από τη γραμμή if __name__ == "__main__". Αξίζει να σημειωθεί ότι ενώ απαιτούνται τα ορίσματα -qre και -sub, τα ορίσματα -tre και -pri δεν απαιτούνται. Εάν δεν παρέχονται, φαίνεται ότι η αντίστοιχη μορφή TREC και τα αρχικά αρχεία δεν θα αποθηκευτούν.

#### *6.2.3.1 Αλλαγή στον αλγόριθμο ωστε να μην αποκλείει τα topic_ids αν υπάρχουν*

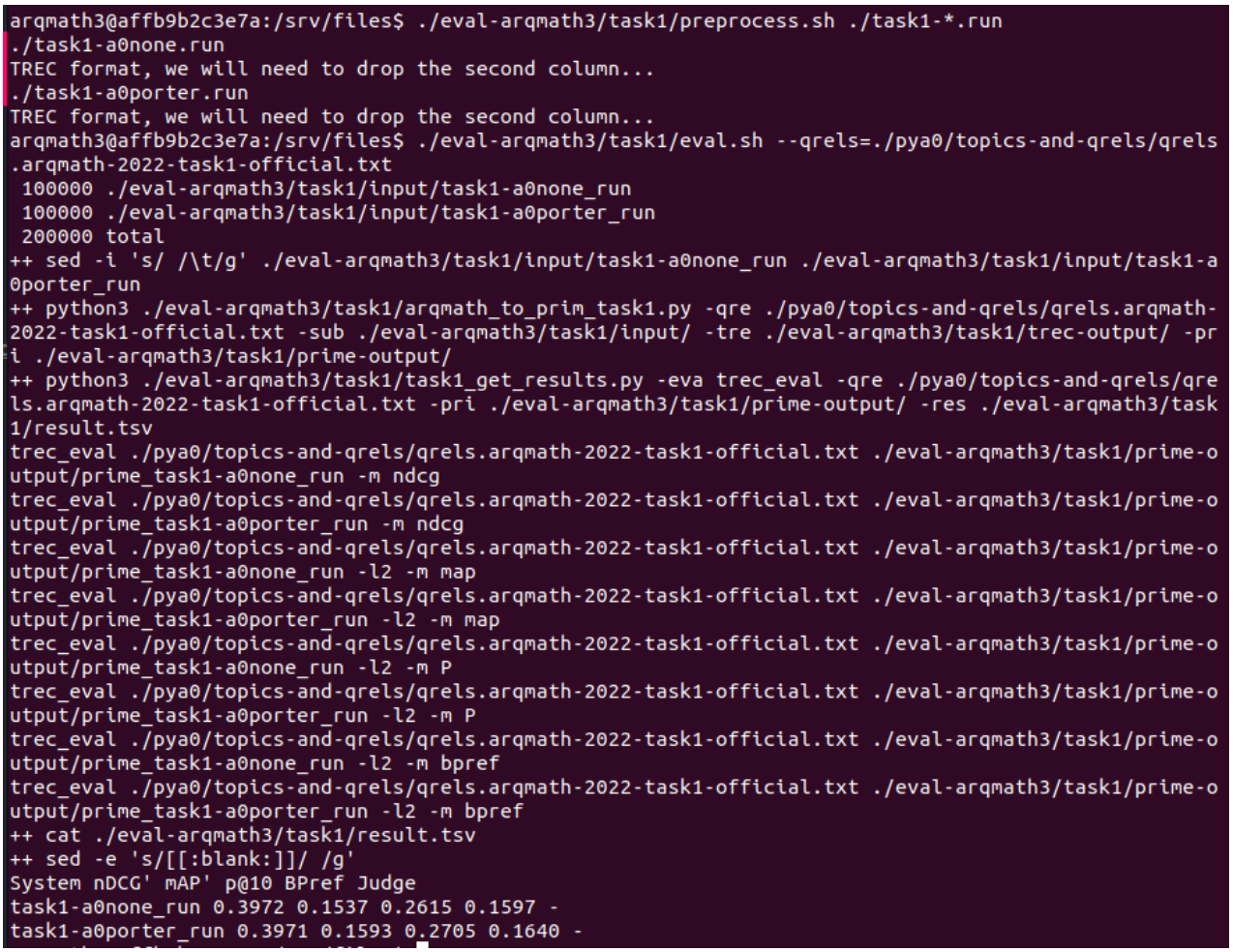
Η κύρια λειτουργία αυτού του script είναι η μετατροπή αρχείων υποβολής σε δύο είδη μορφών: μορφή TREC και μορφή "prime". Η μετατροπή γίνεται κυρίως στη συνάρτηση convert_result_files_to_trec. Η μετατροπή μορφής TREC απλώς μορφοποιεί ξανά τα δεδομένα εισόδου χωρίς να φιλτράρει τίποτα. Η αρχική μορφή είναι ένα υποσύνολο της μορφής TREC που περιλαμβάνει μόνο αναρτήσεις που έχουν κριθεί σχετικές με τα qrels (Κριτικές Σχετικότητας ποιότητας).

Η συνθήκη if ελέγχει εάν το topic_id υπάρχει ήδη ως κλειδί στο λεξικό res_map. Εάν το κάνει, προσθέτει το post_id στην υπάρχουσα λίστα. Εάν δεν το κάνει, δημιουργεί ένα νέο ζεύγος κλειδιού-τιμής στο λεξικό με το topic_id ως κλειδί και μια νέα λίστα που περιέχει το post_id ως τιμή.

```python
def read_qrel_to_dictionary(qrel_file_path):
    """
    Reading the qrel file into a dictionary of topics and annotated answer_ids
    @param qrel_file_path: qrel file path
    @return: dictionart of topics and list of answer_ids
    """
    res_map = {}
    result_file = open(qrel_file_path, newline='', encoding="utf-8")
    csv_reader = csv.reader(result_file, delimiter='\t', quoting=csv.QUOTE_MINIMAL)
    for row in csv_reader:
        topic_id = row[0]
        post_id = row[2]
        if topic_id in res_map:
            res_map[topic_id].append(post_id)
        else:
            res_map[topic_id] = [post_id]
    return res_map
```

```python
def read_qrel_to_dictionary(qrel_file_path):
    """
    Reading the qrel file into a dictionary of topics and annotated answer_ids
    @param qrel_file_path: qrel file path
    @return: dictionart of topics and list of answer_ids
    """
    res_map = {}
    result_file = open(qrel_file_path, newline='', encoding="utf-8")
    csv_reader = csv.reader(result_file, delimiter='\t', quoting=csv.QUOTE_MINIMAL)
    for row in csv_reader:
        topic_id = row[0]
        post_id = row[2]
        # if topic_id in res_map:
        #     res_map[topic_id].append(post_id)
        # else:
        res_map[topic_id] = [post_id]
    return res_map
```

Παρατηρούμε τις βαθμολογίες στο αρχικό πείραμα της ομάδας Approach 0:




System nDCG' mAP' p@10 BPref Judge

task1-a0none_run 0.3972 0.1537 0.2615 0.1597 -

task1-a0porter_run 0.3971 0.1593 0.2705 0.1640 -

Παρατηρούμε τις βαθμολογίες μετά τις αλλαγές στις οποίες προστέθηκαν περισσότερα αποτελέσματα μιας και διαγράφηκε το φίλτρο σύγκρισης σε αντίθεση με το πείραμα της ομάδας Approach 0:

```
arqmath3@affb9b2c3e7a:/srv/files$ ./eval-arqmath3/task1/preprocess.sh ./task1-*.run
./task1-a0none.run
TREC format, we will need to drop the second column...
./task1-a0porter.run
TREC format, we will need to drop the second column...
arqmath3@affb9b2c3e7a:/srv/files$ ./eval-arqmath3/task1/eval.sh --qrels=./pya0/topics-and-qrels/qrels
.arqmath-2022-task1-official.txt
 100000 ./eval-arqmath3/task1/input/task1-a0none_run
 100000 ./eval-arqmath3/task1/input/task1-a0porter_run
 200000 total
++ sed -i 's/ /\t/g' ./eval-arqmath3/task1/input/task1-a0none_run ./eval-arqmath3/task1/input/task1-a
0porter_run
++ python3 ./eval-arqmath3/task1/arqmath_to_prim_task1.py -qre ./pya0/topics-and-qrels/qrels.arqmath-
2022-task1-official.txt -sub ./eval-arqmath3/task1/input/ -tre ./eval-arqmath3/task1/trec-output/ -pr
i ./eval-arqmath3/task1/prime-output/
++ python3 ./eval-arqmath3/task1/task1_get_results.py -eva trec_eval -qre ./pya0/topics-and-qrels/qre
ls.arqmath-2022-task1-official.txt -pri ./eval-arqmath3/task1/prime-output/ -res ./eval-arqmath3/task
1/result.tsv
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -m ndcg
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -m ndcg
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m map
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m map
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m P
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m P
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m bpref
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m bpref
++ cat ./eval-arqmath3/task1/result.tsv
++ sed -e 's/[[:blank:]]/ /g'
System nDCG' mAP' p@10 BPref Judge
task1-a0none_run 0.0494 0.0194 0.0438 0.0194 -
task1-a0porter_run 0.0504 0.0203 0.0414 0.0203 -
arqmath3@affb9b2c3e7a:/srv/files$
```

System nDCG' mAP' p@10 BPref Judge

task1-a0none_run 0.0494 0.0194 0.0438 0.0194 -

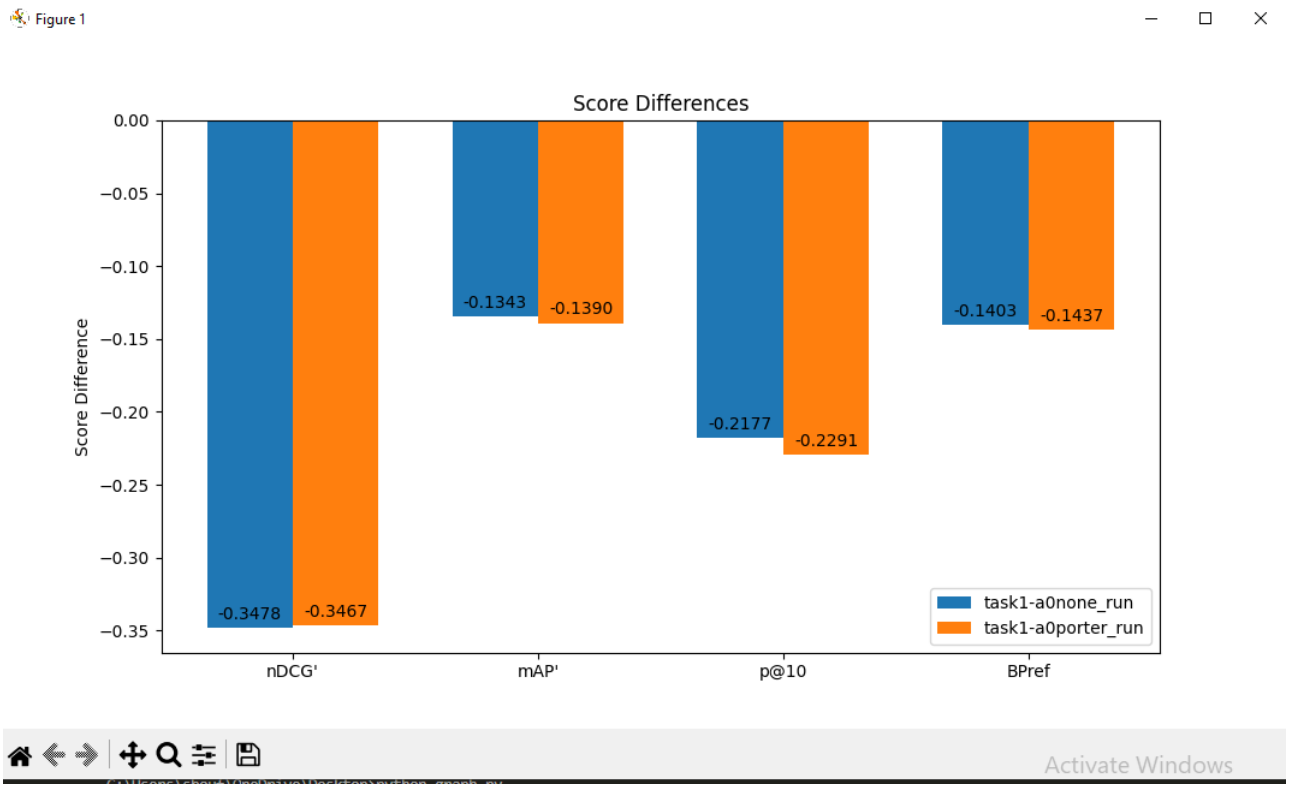
task1-a0porter_run 0.0504 0.0203 0.0414 0.0203 -

Ανάλυση αποτελεσμάτων:

Μετά τις αλλαγές, οι βαθμολογίες για όλες τις μετρήσεις μειώθηκαν σημαντικά, υποδηλώνοντας ότι οι τροποποιήσεις που έγιναν στο σύστημα ανάκτησης μείωσαν την απόδοσή του. Η μείωση αυτών των μετρήσεων υποδηλώνει ότι το σύστημα ανακτά λιγότερο σχετικά έγγραφα, η κατάταξη των σχετικών εγγράφων είναι χειρότερη ή και τα δύο. Θα ήταν απαραίτητο να επανεξεταστούν οι αλλαγές που έγιναν για να προσδιοριστεί γιατί μειώθηκε η απόδοση.

Γράφημα αποτελεσμάτων παρουσιάζοντας τις διαφορές:

(Τα γραφήματα είναι δημιουργημένα με τη βιβλιοθήκη matplotlib της python)




*6.2.3.2Εισαγωγή νέας παραμέτρου. της -pic*

Τα κριτήρια επιλογής για το αρχείο "prime" θα τροποποιηθούν, επί του παρόντος, περιλαμβάνει μόνο αναρτήσεις που κρίνονται σχετικές (relevant) στα qrels. Αλλάζοντας αυτό το κριτήριο, όπως να συμπεριληφθούν αναρτήσεις που κρίνονται μη σχετικές (non-relevant) ή να συμπεριληφθούν όλες οι αναρτήσεις (all) ανεξάρτητα από το καθεστώς κρίσης τους.

Προσθέτοντας την παράμετρο prime_inclusion_criteria

(δηλώνοντας την parser.add_argument('-pic', help='Inclusion criteria for the prime file', choices=['relevant', 'non-relevant', 'all'], default='relevant')

Καλώντας το script από το αρχείο eval.sh με τον εξής τρόπο με νέο argument:

python3 $DIR/arqmath_to_prim_task1.py -qre $QREL -sub "$DIR/input/" -tre $DIR/trec-output/ -pri $DIR/prime-output/ -pic all

Οι αλλαγές στον κώδικα της συνάρτησης είναι οι εξής:

```
if topic_id not in qrel_result_dic:
        continue
    if prime_inclusion_criteria == 'relevant':
        if post_id not in qrel_result_dic[topic_id]:
            continue
    elif prime_inclusion_criteria == 'non-relevant':
        if post_id in qrel_result_dic[topic_id]:
            continue
    csv_writer2.writerow(temp)
```

Με την εισαγωγή της νέας παραμέτρου παρατηρείται οτι επιστρέφονται και αποτελέσματα που έχουν κριθεί και ως σχετικά , μη σχετικά αλλά και όλα τα αποτελέσματα.

Παρατηρούμε τις βαθμολογίες στο αρχικό πείραμα της ομάδας Approach 0:

System nDCG' mAP' p@10 BPref Judge

task1-a0none_run 0.3972 0.1537 0.2615 0.1597 -

task1-a0porter_run 0.3971 0.1593 0.2705 0.1640 -

Παρατηρούμε τις βαθμολογίες μετά τις αλλαγές στις οποίες προστέθηκε νέα παράμετρος σε αντίθεση με το πείραμα της ομάδας Approach 0:

```
++ sed -i 's/ /\t/g' ./eval-arqmath3/task1/input/task1-a0none_run ./eval-arqmath3/task1/input/task1-a
0porter_run
++ python3 ./eval-arqmath3/task1/arqmath_to_prim_task1.py -qre ./pya0/topics-and-qrels/qrels.arqmath-
2022-task1-official.txt -sub ./eval-arqmath3/task1/input/ -tre ./eval-arqmath3/task1/trec-output/ -pr
i ./eval-arqmath3/task1/prime-output/ -pic all
++ python3 ./eval-arqmath3/task1/task1_get_results.py -eva trec_eval -qre ./pya0/topics-and-qrels/qre
ls.arqmath-2022-task1-official.txt -pri ./eval-arqmath3/task1/prime-output/ -res ./eval-arqmath3/task
1/result.tsv
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -m ndcg
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -m ndcg
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m map
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m map
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m P
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m P
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0none_run -l2 -m bpref
trec_eval ./pya0/topics-and-qrels/qrels.arqmath-2022-task1-official.txt ./eval-arqmath3/task1/prime-o
utput/prime_task1-a0porter_run -l2 -m bpref
++ cat ./eval-arqmath3/task1/result.tsv
++ sed -e 's/[[:blank:]]/ /g'
System nDCG' mAP' p@10 BPref Judge
task1-a0none_run 0.3491 0.1036 0.2538 0.1597 -
task1-a0porter_run 0.3525 0.1113 0.2705 0.1640 -
arqmath3@affb9b2c3e7a:/srv/files$ ./eval-arqmath3/task1/preprocess.sh cleanup
```

System nDCG' mAP' p@10 BPref Judge

task1-a0none_run 0.3491 0.1036 0.2538 0.1597 -

task1-a0porter_run 0.3525 0.1113 0.2705 0.1640 -

Ανάλυση αποτελεσμάτων:

nDCG': Το task1-a0none_run μειώθηκε από 0,3972 σε 0,3491, ενώ το task1-a0porter_run μειώθηκε από 0,3971 σε 0,3525. Το nDCG' είναι μια μέτρηση που αξιολογεί την ποιότητα της κατάταξης των ανακτημένων εγγράφων, λαμβάνοντας υπόψη τις θέσεις των εγγράφων στην κατάταξη. Μια μείωση σημαίνει ότι το σύστημα έχει χειρότερη απόδοση όσον αφορά την κατάταξη των εγγράφων μετά την πραγματοποίηση των αλλαγών.

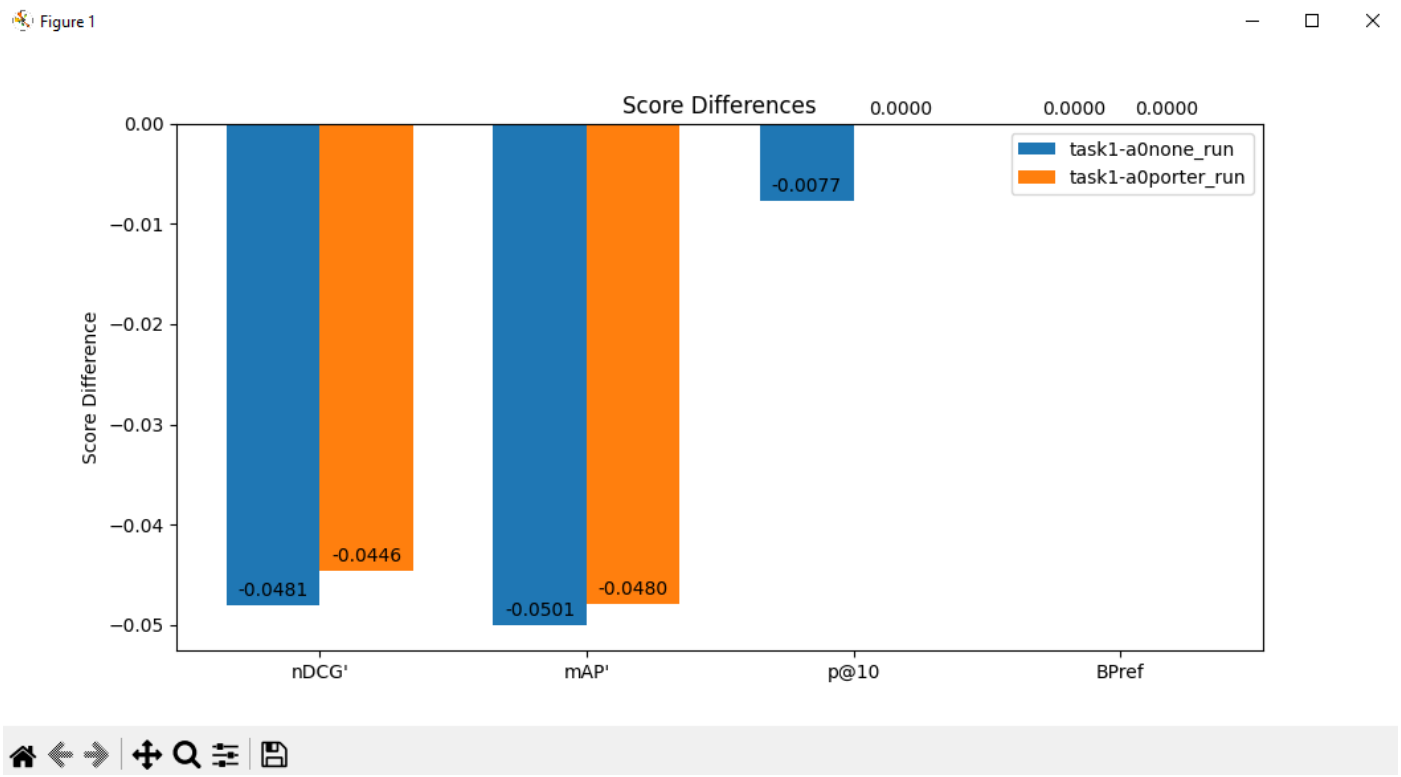
mAP': Το task1-a0none_run μειώθηκε από 0,1537 σε 0,1036 και το task1-a0porter_run μειώθηκε από 0,1593 σε 0,1113. Το mAP' μετρά τη μέση βαθμολογία ακρίβειας στις θέσεις κατάταξης κάθε σχετικού εγγράφου. Παρόμοια με το nDCG', μια μείωση στο mAP' υποδηλώνει χειρότερη κατάταξη των σχετικών εγγράφων στη λίστα αποτελεσμάτων.

Οι άλλες δύο μετρήσεις, p@10 και BPref, δεν άλλαξαν και για τις δύο εκτελέσεις μετά τις τροποποιήσεις. Αυτό σημαίνει ότι η αναλογία των σχετικών εγγράφων στα 10 κορυφαία ανακτηθέντα έγγραφα (p@10) και η μέτρηση δυαδικής προτίμησης (BPref) παρέμεινε η ίδια.

Συνοπτικά, αυτές οι αλλαγές υποδηλώνουν ότι οι τροποποιήσεις που έγιναν στο σύστημα το χειροτέρευσαν όσον αφορά την κατάταξη των σχετικών εγγράφων, αλλά δεν επηρέασαν την αναλογία των σχετικών εγγράφων μεταξύ των κορυφαίων 10 αποτελεσμάτων και τη μέτρηση δυαδικής προτίμησης.

Γράφημα αποτελεσμάτων παρουσιάζοντας τις διαφορές:

(Τα γραφήματα είναι δημιουργημένα με τη βιβλιοθήκη matplotlib της python)



### *6.2.4 Τροποποίηση της συνάρτησης indexer*

Η αλλαγή σε αυτό το αρχείο μπορεί να επηρεάσει τα αποτελέσματα. Το αρχείο περιέχει λειτουργίες που σχετίζονται με την ευρετηρίαση και τη διατήρηση ενός ευρετηρίου αναζήτησης.

Εάν τροποποιηθεί ο κώδικας σε αυτό το αρχείο, μπορεί να επηρεάσει τη συμπεριφορά και τη λειτουργικότητα της διαδικασίας δημιουργίας ευρετηρίου και ανάκτησης. Ανάλογα με τη φύση των αλλαγών, μπορεί να οδηγήσει σε διαφορετικά αποτελέσματα, όπως διαφορετικά αποτελέσματα ευρετηρίασης, ανάκτησης ή κατάταξης.

Πιο συγκεκριμένα, εάν αλλάξει ο αλγόριθμος ευρετηρίασης ή τροποποιηθεί ο τρόπος επεξεργασίας και αποθήκευσης των εγγράφων, μπορεί να επηρεάσει την ποιότητα του ευρετηρίου και, στη συνέχεια, να επηρεάσει τη συνάφεια και την κατάταξη των αποτελεσμάτων αναζήτησης.

Ομοίως, εάν αλλάξει ο αλγόριθμος ανάκτησης ή ο τρόπος αντιστοίχισης των όρων ερωτήματος με το ευρετήριο, μπορεί να επηρεάσει το accuracy και την precision των αποτελεσμάτων αναζήτησης.

Αλλάζοντας τον κώδικα σε :

```
PyObject *add_document(PyObject *self, PyObject *args, PyObject* kwargs)
{
   PyObject *pyindexer;
   const char *content, *url = NULL, *extern_id_str = NULL;
   static char *kwlist[] = {"writer", "content", "url", "extern_id", NULL};
   if (!PyArg_ParseTupleAndKeywords(args, kwargs, "Os|ss", kwlist,
      &pyindexer, &content, &url, &extern_id_str)) {
      PyErr_Format(PyExc_RuntimeError,
         "PyArg_ParseTupleAndKeywords error");
      return NULL;
   }

   // Modified code: Changed the order of field writes

   bool anyfield_written = false;
```

```
if (url) {
    /* for all the other fields */
    (void)indexer_write_field(indexer, 0, "url", url);
    anyfield_written = true;
}

if (content) {
    /* for all the other fields */
    (void)indexer_write_field(indexer, 0, "content", content);
    anyfield_written = true;
}

if (extern_id_str) {
    uint32_t extern_id;
    if (1 == sscanf(extern_id_str, "%lu", &extern_id)) {
        char docIDstr[1024];
        snprintf(docIDstr, 1024, "%d", docID);
        (void)indexer_write_field(indexer, extern_id, "extern_id", docIDstr);
        anyfield_written = true;
    }
}

if (anyfield_written) {
    /* maintain and prepare for the next indexing document */
    docID = indexer_next_doc(indexer);
} else {
    docID = 0;
}
```

Στην αρχική συνάρτηση, το πεδίο περιεχομένου γράφεται πριν από το πεδίο URL. Ωστόσο, στην έκδοση με τις αλλαγές στον κώδικα, η σειρά αντιστρέφεται και το πεδίο URL γράφεται πριν από το πεδίο περιεχομένου. Αυτή η αλλαγή μπορεί να επηρεάσει τη συνάφεια και την κατάταξη των αποτελεσμάτων

αναζήτησης. Εάν η μηχανή αναζήτησης δίνει προτεραιότητα στο πεδίο URL κατά την ανάκτηση ή την κατάταξη, η τροποποιημένη έκδοση ενδέχεται να παράγει ελαφρώς διαφορετικά αποτελέσματα σε σύγκριση με την αρχική έκδοση.

# 7 Συμπεράσματα

## 7.1 Προτεινόμενο εύρος καινοτόμων τεχνικών στην ανάκτηση μαθηματικών πληροφοριών

Το προτεινόμενο εύρος καινοτόμων τεχνικών στην ανάκτηση μαθηματικών πληροφοριών στοχεύει να ξεπεράσει τις προκλήσεις που αφορούν ειδικά το μαθηματικό περιεχόμενο. Μια σημαντική συμβολή αυτής της έρευνας είναι η ανάπτυξη εξειδικευμένων μοντέλων ευρετηρίασης και ανάκτησης που ενσωματώνουν μαθηματικούς τύπους και εξισώσεις ως αναπόσπαστα συστατικά. Αντιμετωπίζοντας τις μαθηματικές εκφράσεις ως βασικές οντότητες στη διαδικασία ανάκτησης, αυτά τα μοντέλα επιτρέπουν μια βαθύτερη κατανόηση της σημασιολογίας και των σχέσεων εντός του μαθηματικού περιεχομένου.

Για να επιτευχθεί αυτό, τα εξειδικευμένα μοντέλα ευρετηρίασης και ανάκτησης αξιοποιούν τεχνικές αναπαράστασης μαθηματικών τύπων όπως MathML, LaTeX ή MathJax. Αυτές οι μορφές αναπαράστασης παρέχουν δομημένους και τυποποιημένους τρόπους κωδικοποίησης μαθηματικών εκφράσεων, διασφαλίζοντας συμβατότητα και διαλειτουργικότητα σε διάφορες πηγές μαθηματικού περιεχομένου.

Με την ενσωμάτωση μαθηματικών τύπων και εξισώσεων ως αναπόσπαστα συστατικά των μοντέλων ανάκτησης, προκύπτουν πολλά οφέλη.

Οι σημασιολογικές σχέσεις (semantic relationships). Οι μαθηματικοί τύποι και εξισώσεις φέρουν πλούσιες σημασιολογικές σχέσεις, συλλαμβάνοντας εξαρτήσεις, πράξεις και μαθηματικές δομές. Με την ενσωμάτωση αυτών των στοιχείων, τα μοντέλα ανάκτησης μπορούν να συλλάβουν τις περίπλοκες σχέσεις εντός των μαθηματικών παραστάσεων, οδηγώντας σε βελτιωμένη ακρίβεια στην ανάκτηση σχετικού μαθηματικού περιεχομένου.

Η κατανόηση συμφραζομένων (contextual understanding). Οι μαθηματικές εκφράσεις συχνά απαιτούν πλαίσιο για την ερμηνεία τους. Η συμπερίληψη μαθηματικών τύπων και εξισώσεων στα μοντέλα ανάκτησης επιτρέπει μια πιο ολοκληρωμένη κατανόηση του πλαισίου που περιβάλλει το μαθηματικό πρόβλημα. Αυτή η συμφραζόμενη κατανόηση επιτρέπει την ακριβέστερη ανάκτηση των σχετικών πόρων, λαμβάνοντας υπόψη τις συγκεκριμένες μαθηματικές έννοιες και σχέσεις μέσα στο πρόβλημα.

Η επέκταση ερωτήματος (query expansion). Η ενσωμάτωση μαθηματικών τύπων και εξισώσεων στα μοντέλα ανάκτησης ανοίγει δυνατότητες για τεχνικές επέκτασης ερωτήματος ειδικά για το μαθηματικό περιεχόμενο. Αναλύοντας τις δομικές και σημασιολογικές ιδιότητες των μαθηματικών παραστάσεων, τα μοντέλα μπορούν να επεκτείνουν το αρχικό ερώτημα χρήστη για να συμπεριλάβουν πρόσθετες σχετικές εκφράσεις ή παραλλαγές, βελτιώνοντας έτσι την κάλυψη των σχετικών πληροφοριών.

Η ακρίβεια και σχετικότητα (precision and relevance). Οι μαθηματικές εκφράσεις συχνά περιέχουν συμβολικές σημειώσεις και εξειδικευμένο λεξιλόγιο που μπορεί να είναι δύσκολο να χειριστεί κανείς με τα παραδοσιακά μοντέλα ανάκτησης που βασίζονται σε κείμενο. Αντιμετωπίζοντας ρητά τους μαθηματικούς τύπους και τις εξισώσεις ως αναπόσπαστα συστατικά, τα μοντέλα ανάκτησης μπορούν να επικεντρωθούν στην αποτύπωση της ακριβούς μαθηματικής σημασιολογίας, οδηγώντας σε βελτιωμένη ακρίβεια και σχετικότητα στα αποτελέσματα ανάκτησης.

Συνολικά, η ανάπτυξη εξειδικευμένων μοντέλων ευρετηρίασης και ανάκτησης που ενσωματώνουν μαθηματικούς τύπους και εξισώσεις ως αναπόσπαστα στοιχεία είναι μια κρίσιμη συμβολή στην ανάκτηση μαθηματικών πληροφοριών. Αυτά τα μοντέλα επιτρέπουν μια πιο εις βάθος κατανόηση του μαθηματικού περιεχομένου, καταγράφουν σημασιολογικές σχέσεις μέσα σε μαθηματικές εκφράσεις και ενισχύουν την ακρίβεια και τη σχετικότητα των ανακτημένων πληροφοριών. Αξιοποιώντας τεχνικές αναπαράστασης μαθηματικών τύπων, όπως MathML, LaTeX ή MathJax, αυτές οι τεχνικές φέρνουν προόδους στη διαδικασία ανάκτησης μαθηματικών προβλημάτων.[13][14][15][16]

## *7.2 Περιοχές που θα βοηθήσουν την εξέλιξη στην ανάκτηση πληροφοριών*

Η ανάκτηση πληροφοριών στα μαθηματικά έχει εξελιχθεί σημαντικά με τα χρόνια, αλλά υπάρχουν ακόμη πολλοί τομείς για περαιτέρω ανάπτυξη και εξερεύνηση. Αυτή η ανάλυση περιγράφει πιθανές προόδους που θα μπορούσαν να ενισχύσουν την αποδοτικότητα και την αποτελεσματικότητα της ανάκτησης μαθηματικών πληροφοριών στο μέλλον.

1. Βελτιωμένη ανάλυση και αναπαράσταση μαθηματικών εκφράσεων

Οι μαθηματικές εκφράσεις μπορεί να είναι δύσκολο να αναλυθούν και να αναπαραστηθούν αποτελεσματικά λόγω του σημασιολογικού τους πλούτου και της ποικιλίας των πιθανών αναπαραστάσεων για την ίδια έννοια. Οι τρέχουσες μέθοδοι βασίζονται συχνά σε ανάλυση LaTeX ή παρόμοιες προσεγγίσεις,

αλλά αυτές οι μέθοδοι μπορεί να είναι υπολογιστικά ακριβές και μπορεί να μην αποτυπώνουν πλήρως τις σημασιολογικές πληροφορίες πολύπλοκων μαθηματικών παραστάσεων [30].

Οι μελλοντικές βελτιώσεις θα μπορούσαν να περιλαμβάνουν την ανάπτυξη πιο αποτελεσματικών και σημασιολογικά συνειδητοποιημένων τεχνικών ανάλυσης που μπορούν να χειριστούν ένα ευρύτερο φάσμα μαθηματικών εκφράσεων. Τεχνικές όπως τα δέντρα διάταξης συμβόλων (SLT) και τα δέντρα χειριστή (OPT) έχουν προταθεί για την ανάλυση και την αναπαράσταση μαθηματικών παραστάσεων και θα μπορούσαν να αναπτυχθούν και να βελτιστοποιηθούν περαιτέρω [30].

## 2. Προηγμένες τεχνικές για τον χειρισμό ερωτημάτων με επίγνωση των μαθηματικών

Η ανάκτηση πληροφοριών στα μαθηματικά αντιμετωπίζει συχνά τις μαθηματικές εκφράσεις ως ξεχωριστές οντότητες από το κειμενικό πλαίσιο, χάνοντας πιθανώς σημαντικές πληροφορίες. Τα μελλοντικά συστήματα ανάκτησης πληροφοριών θα μπορούσαν να επωφεληθούν από πιο εξελιγμένες τεχνικές για το χειρισμό μαθηματικών ερωτημάτων που εξετάζουν την αλληλεπίδραση μεταξύ μαθηματικών εκφράσεων και του κειμενικού τους πλαισίου [29].

Για παράδειγμα, η ενσωμάτωση τεχνικών από την επεξεργασία φυσικής γλώσσας (NLP), όπως μετασχηματιστές ή μηχανισμοί προσοχής, θα μπορούσε να βοηθήσει στην καταγραφή των σημασιολογικών σχέσεων μεταξύ των μαθηματικών εκφράσεων και του περιβάλλοντος κειμένου. Επιπλέον, η ανάπτυξη τεχνικών επέκτασης ερωτημάτων ειδικά για ερωτήματα που γνωρίζουν τα μαθηματικά θα μπορούσε επίσης να βελτιώσει την αποτελεσματικότητα της ανάκτησης [29].

## 3. Μόχλευση μηχανικής μάθησης και προσεγγίσεων βαθιάς μάθησης

Η μηχανική μάθηση και η βαθιά μάθηση έχουν σημαντικές επιπτώσεις στον τομέα της γενικής IR, αλλά η εφαρμογή τους στον τομέα των μαθηματικών IR παραμένει σχετικά ανεξερεύνητη. Οι μελλοντικές εξελίξεις θα μπορούσαν να περιλαμβάνουν τη μόχλευση αυτών των προηγμένων τεχνικών για το μαθηματικό IR.

Για παράδειγμα, τα μοντέλα νευρωνικής κατάταξης θα μπορούσαν να προσαρμοστούν για μαθηματικά IR για να μάθουν πιο αποτελεσματικές συναρτήσεις κατάταξης από δεδομένα εκπαίδευσης. Ομοίως, τα

μοντέλα βαθιάς μάθησης θα μπορούσαν να χρησιμοποιηθούν για την εκμάθηση σημασιολογικών ενσωματώσεων για μαθηματικές εκφράσεις, βελτιώνοντας δυνητικά την ακρίβεια της αντιστοίχισης μαθηματικών ερωτημάτων [29].

4. Διεπαφές χρήστη και σχεδιασμός αλληλεπίδρασης

Το μαθηματικό IR δεν αφορά μόνο τους υποκείμενους αλγόριθμους και μοντέλα, αλλά και τον τρόπο αλληλεπίδρασης των χρηστών με το σύστημα. Οι μελλοντικές βελτιώσεις θα μπορούσαν να περιλαμβάνουν το σχεδιασμό πιο έξυπνων και αποτελεσματικών διεπαφών χρήστη για ανάκτηση πληροφοριών στα μαθηματικά.

Για παράδειγμα, επιτρέποντας στους χρήστες να εισάγουν ερωτήματα χρησιμοποιώντας φυσική μαθηματική σημειογραφία ή ακόμα και χειρόγραφα μαθηματικά σύμβολα θα μπορούσε να κάνει το σύστημα πιο προσιτό και φιλικό προς τον χρήστη. Επιπλέον, η οπτικοποίηση των αποτελεσμάτων ανάκτησης με τρόπο που τονίζει το μαθηματικό περιεχόμενο και τη συνάφειά του με το ερώτημα θα μπορούσε επίσης να βελτιώσει την ικανοποίηση των χρηστών [29].

5. Αξιολόγηση μεγάλης κλίμακας και συγκριτική αξιολόγηση

Τέλος, πιο μεγάλης κλίμακας προσπάθειες αξιολόγησης και συγκριτικής αξιολόγησης, όπως οι εργασίες ARQMath στο CLEF, είναι ζωτικής σημασίας για την προώθηση της προόδου στα μαθηματικά IR. Αυτές οι προσπάθειες παρέχουν μια πλατφόρμα για διαφορετικές μεθόδους και μοντέλα προς δοκιμή και σύγκριση σε ένα κοινό έδαφος.

Οι μελλοντικές προσπάθειες αξιολόγησης θα μπορούσαν να ωφεληθούν από τη συμπερίληψη ενός ευρύτερου φάσματος τύπων εργασιών και πηγών δεδομένων και από τη βελτίωση της διαδικασίας αξιολόγησης συνάφειας για την καλύτερη αποτύπωση των μοναδικών χαρακτηριστικών της μαθηματικής συνάφειας [30].

# *Βιβλιογραφία*


[1] Acquisti, A. (2004), Privacy in electronic commerce and the economics of immediate gratification. In: *Proceedings of the 5th ACM conference on Electronic commerce*, May 17-20, 2004, New York, USA.

[2] B. Mansouri, R. Zanibbi, D. W. Oard, Characterizing Searches for Mathematical Concepts, in: Joint Conference on Digital Libraries (JCDL), 2019.

[3] https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

[4] Precision-Recall versus Accuracy and the Role of Large Data Sets https://ojs.aaai.org/index.php/AAAI/article/view/5193

[5]Kanoulas and Asalam - 2009 - https://dl.acm.org/doi/abs/10.1145/1645953.1646032

[6]Manning, C., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Cambridge University Press. https://nlp.stanford.edu/IR-book/pdf/irbookonlinereading.pdf

[7]Voorhees, E. M., & Harman, D. K. (Eds.). (2005). TREC: Experiment and evaluation in information retrieval. MIT Press.

[8]Kelly, D., & Belkin, N. J. (2015). Evaluation in information retrieval. In Information retrieval: A health and biomedical perspective (pp. 245-280). Springer.

[9]Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.

[10]Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), 1532-1543.

[11]Le, Q., & Mikolov, T. (2014). Distributed representations of sentences and documents. International Conference on Machine Learning (ICML), 1188-1196.

[12]Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 conference of the North American chapter of the Association for Computational Linguistics: Human language technologies, Volume 1 (Long and Short Papers), 4171-4186.

[13]Sojka, P., Langer, J., & Horák, A. (2014). MathML and LaTeX in the ARRS Gold Standard Corpus of Mathematical Formulae. In Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC 2014).

[14]Sojka, P., & Riedl, M. (2018). Digital mathematics libraries and math-aware information retrieval systems. In Proceedings of the 8th Workshop on Mathematical User Interfaces (MathUI 2018), 1-8.

[15]Harispe, S., Ranwez, S., Janaqi, S., & Montmain, J. (2015). Semantic Similarity from Natural Language and Ontology Analysis. Synthesis Lectures on Human Language Technologies, 8(1), 1-254.

[16]Yuan, X., Wei, Z., & Jiang, H. (2020). Incorporating Formula Features into Neural Retrieval Models for Math Word Problem Solving. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020), 5921-5931.

[17]Baeza-Yates, R., & Ribeiro-Neto, B. (2011). Modern Information Retrieval. Addison-Wesley. http://www.dcc.ufmg.br/irbook/

[18]Zuccon, G., & Koopman, B. (2019). Μια κριτική και βελτίωση του κανονικοποιημένου μειωμένου αθροιστικού κέρδους. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2019), 1255-1258.

[19]Oracle. (n.d.). Structured vs. Unstructured Data. Oracle. https://www.oracle.com/big-data/what-is-structured-data/

[20]IBM. (n.d.). What is Unstructured Data? IBM. https://www.ibm.com/cloud/learn/unstructured-data

[21]Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47. https://dl.acm.org/doi/10.1145/505282.505283

[22]Tang, J., Zhang, D., & Yao, L. (2005). Social Network Extraction of Academic Researchers. ICDM '05. Fifth IEEE International Conference on Data Mining. https://ieeexplore.ieee.org/document/1565705

[23]Silla Jr, C. N., & Freitas, A. A. (2011). A survey of hierarchical classification across different application domains. Data Mining and Knowledge Discovery, 22(1-2), 31-72. https://link.springer.com/article/10.1007/s10618-010-0175-9

[24]Kosala, R., & Blockeel, H. (2000). Web mining research: A survey. ACM Sigkdd Explorations Newsletter, 2(1), 1-15. [https://dl.acm.org/doi/10.1145/

[25]Zobel, J., & Moffat, A. (2006). Inverted files for text search engines. ACM Computing Surveys (CSUR), 38(2), 1-56. https://dl.acm.org/doi/10.1145/1132956.1132959

[26]Zamani, H., & Croft, W. B. (2020). Learning a Joint Search and Recommendation Model from User-Item Interactions. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 1-10. https://doi.org/10.1145/3397271.3401113

[27] Youssef, A. (2005). Methods of relevance in mathematical information retrieval. Annals of Mathematics and Artificial Intelligence, 43(1), 5-25. https://doi.org/10.1007/s10472-004-8153-7

[28]Powers, D.M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies. 2 (1): 37–63. http://www.bioinfo.in/contents.php?id=51

[29] Zanibbi, R., Aizawa, A., Kohlhase, M., Ounis, I., & Davila, K. (2020). Overview of ARQMath 2020: CLEF Lab on Answer Retrieval for Questions on Math. In CLEF (Working Notes). https://www.researchgate.net/publication/344337134_Overview_of_ARQMath_2020_CLEF_Lab_on_Answer_Retrieval_for_Questions_on_Math

[30] Youssef, A. (2005). Methods of relevance ranking and hit-content generation in Math Search. In Proceedings of the International Conference on Artificial and Computational Intelligence for Decision, Control and Automation in Engineering and Industrial Applications (ACIDCA'2005). https://arxiv.org/abs/cs/0511086

[31]Miner, R., & Munavalli, R. (2007). MathFind: A Math-Aware Search Engine. In Proceedings of the ACM SIGIR 2007 Workshop on Mathematical Information Retrieval. http://ceur-ws.org/Vol-921/paper8.pdf

[32]Aizawa, A., Kohlhase, M., & Ounis, I. (2014). NTCIR-11 Math-2 Task Overview. In Proceedings of the 11th NTCIR Conference.

http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings11/pdf/NTCIR/Math-2/01-NTCIR11-MATH2-AizawaA.pdf

[33]Overview of ARQMath-3 (2022)
https://www.cs.rit.edu/~rlaz/files/ARQMath_2022_Overview_WorkingNotes.pdf

[34] https://wwwdb.inf.tu-dresden.de/wp-content/uploads/ARQMath_2021_TU_DBS.pdf

[35] https://www.scg.uwaterloo.ca/brushsearch/tangent-l/

[36] https://cs.uwaterloo.ca/~yk2ng/MathDowsers-ARQMath/

[37] https://github.com/kiking0501/MathDowsers-ARQMath

[38] https://ceur-ws.org/Vol-3180/paper-02.pdf