

# Εφαρμογή για υποβοήθηση σύνταξης και υποβολής σύνθετων ερωτημάτων στην τυπική λειτουργία αναζήτησης του NCBI

# **ΙΩΑΝΝΗΣ ΑΝΤΩΝΟΓΙΑΝΝΑΚΗΣ ΑΜ 4104**

Επιβλέπων: Τσατσάκης Νικόλαος

Επόπτης: Τσακαλίδης Παναγιώτης

Μέλος Επιτροπής Αξιολόγησης: Τζίτζικας Ιωάννης

Χρονικό διάστημα: Οκτώβριος 2022 – Ιούνιος 2023

# Περιεχόμενα

1	Εισ	αγωγή	4
	1.1	Γενική περιγραφή	4
2	Avo	ιζήτηση στο NCBI και ανάδειξη προτάσεων βελτίωσης της αναζήτησης του	5
	2.1	Σύντομη αναφορά στο NCBI	5
	2.2	Αναζήτηση στο ΝCBΙ	5
	2.3	ΠΡΟΤΑΣΕΙΣ ΒΕΛΤΙΩΣΗΣ	7
3	ΗО	θόνη της Εφαρμογής – Περιγραφή πεδίων και Λειτουργικότητας	8
	3.1	Πεδίο "Your Query (Preview)"	8
	3.2	Πεδίο "Term x"	8
	3.3	Πτυσσόμενο Μενού "Please choose a field"	9
	3.4	Πεδίο "Connector"	. 11
	3.5	Πτυσσόμενο Μενού "Number of terms"	. 11
4	Λει	τουργία Κουμπιών	. 12
	4.1	Κουμπί "Add parenthesis '(' "	. 12
	4.2	Κουμπί "Add parenthesis ')' "	. 14
	4.3	Κουμπί "Delete parenthesis '(' "	. 15
	4.4	Κουμπί "Delete parenthesis ')' "	. 15
	4.5	Κουμπί "Clean"	. 16
	4.6	Κουμπί "Search"	. 16
5	Επι	λογείς (Checkboxes)	. 17
	5.1	Higher Recall	. 17
	5.2	Unique	. 18
	5.3	Raw data	. 18
	5.4	Sequencing data	. 18
	5.5	Accession	. 19
6	Απο	ιιτήσεις Εφαρμογής στους Browsers	. 20
	6.1	Chrome Browser, Opera Browser και Microsoft Edge	. 20
	6.2	Mozilla Browser	. 20
7	Οθο	όνες της Εφαρμογής	. 21
	7.1	Προσαρμογή για χρήση από διάφορες συσκευές	. 21
8	Αρχ	ιτεκτονική του Συστήματος και Εργαλεία	. 22
	8.1	Αρχιτεκτονική	. 22
	8.2	Τεχνολογίες/frameworks used	. 22

9	Δοκιμή Συστήματος	24
10	Επίλογος	50
Πηγές		51

# 1 Εισαγωγή

Στην παρούσα αναφορά παρουσιάζεται μια εφαρμογή υποβοήθησης σύνταξης και υποβολής σύνθετων ερωτημάτων στη γενική λειτουργία αναζήτησης του NCBI (National Center for Biotechnology Information) και ορίζεται ο βασικός σκοπός για τον οποίο αναπτύχθηκε μαζί με τα χαρακτηριστικά, τις απαιτήσεις, τυχόν περιορισμούς, τις διεπαφές καθώς και τις λειτουργίες που αυτή καλύπτει. Με την ολοκλήρωση του εγγράφου θα υπάρχει μια ολοκληρωμένη εικόνα για το τι είναι και το τι κάνει η εν λόγω εφαρμογή.

Να σημειωθεί ότι η υλοποίηση της εφαρμογής έγινε συμπληρωματικά του συμπεράσματος διπλωματικής εργασίας που πραγματοποιήθηκε από την κ. Δήμητρα Λυδάκη και αφορούσε σε επισκόπηση και κριτική αξιολόγηση των μηχανισμών αναζήτησης και εξόρυξης βιο-ιατρικής πληροφορίας σε διάφορες βάσεις δεδομένων και αποθετηρίων. Έγινε δε στο πλαίσιο εκπόνησης διπλωματικής εργασίας στο Τμήμα Επιστήμης Υπολογιστών του Πανεπιστημίου Κρήτης κατά το διάστημα Οκτώβριος 2022 – Ιούνιος 2023.

# 1.1 Γενική περιγραφή

Η εργασία μας αφορά στην υλοποίηση μιας διαδικτυακής εφαρμογής, η οποία τρέχει σε όλους τους ευρέως χρησιμοποιούμενους φυλλομετρητές. Έρχεται να προσθέσει τη λειτουργικότητα επιστροφής περισσότερων σχετικών αποτελεσμάτων στην περίπτωση που έχουμε ένα σύνθετο ερώτημα, με την επιλογή Higher Recall και τη δυνατότητα επιστροφής καταχωρήσεων μέσα από τις οποίες μπορούμε να βρούμε ακατέργαστα (πρωτογενή) δεδομένα που σχετίζονται με αυτές. Επιπλέον, δίνει τη δυνατότητα επεξεργασίας του ερωτήματος (query) που διαμορφώνουμε, πριν αυτό αποσταλεί στη λειτουργία αναζήτησης του NCBI.

Οι τεχνολογίες που χρησιμοποίησαμε για να δημιουργήσουμε την εφαρμογή μας είναι οι HTML5, CSS, BOOTSTRAP και JavaScript.

Η εφαρμογή μας απευθύνεται στις ακόλουθες κατηγορίες χρηστών:

- ερευνητές που αναζητούν εξειδικευμένη πληροφορία,
- απλούς χρήστες για πληροφόρηση,
- φοιτητές βιολογικών-ιατρικών επιστημών.

Σε επόμενες ενότητες παρουσιάζονται αναλυτικά οι οθόνες τις εφαρμογής με τη λειτουργικότητα που προσφέρουν.

# 2 Αναζήτηση στο NCBI και ανάδειξη προτάσεων βελτίωσης της αναζήτησης του

### 2.1 Σύντομη αναφορά στο NCBI

Το NCBI (National Center for Biotechnology Information) αποτελεί ένα επιστημονικό κέντρο που έχει στόχο να παρέχει μια ολοκληρωμένη πρόσβαση σε πληροφορίες και εργαλεία για την ανάπτυξη, την οργάνωση και την ανάλυση των βιολογικών δεδομένων, ενισχύοντας την επιστημονική έρευνα καθώς και την πρόοδο της βιοϊατρικής.

Το NCBI, φιλοξενεί μια ευρεία γκάμα υπηρεσιών και πόρων που εξυπηρετούν τις ανάγκες της επιστημονικής κοινότητας. Μέσω της ιστοσελίδας του NCBI, δίνεται πρόσβαση σε πλήθος βιολογικών βάσεων δεδομένων όπως είναι η βάση δεδομένων GenBank που περιλαμβάνει ακολουθίες DNA αλλά και πρωτεϊνών, και η βάση PubMed που παρέχει πρόσβαση σε επιστημονικές δημοσιεύσεις στον τομέα της ιατρικής και της βιολογίας. Ακόμα, προσφέρει εξειδικευμένα εργαλεία και αλγόριθμους αναζήτησης, όπως για παράδειγμα το BLAST (Basic Local Alignment Search Tool).

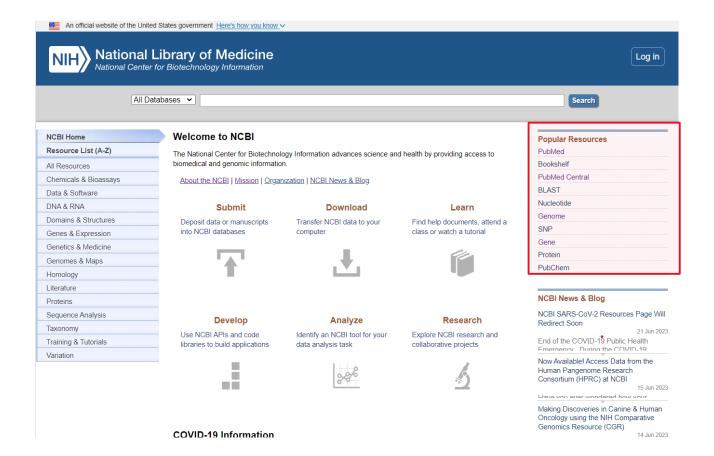
Σημαντικό είναι να αναφέρουμε την δυνατότητα υποβολής δεδομένων από τους χρήστες, ερευνητές αλλά και επιστήμονες, σχετικών με βιοϊατρικά θέματα, με σκοπό την αξιοποίηση τους από την επιστημονική κοινότητα. Τέλος, το NCBI παρέχει μια εκτενή γκάμα ερωτημάτων "How to", τα οποία δημιουργήθηκαν με γνώμονα την διευκόλυνση των νεοεισερχόμενων χρηστών. [2]

# 2.2 Αναζήτηση στο NCBI

Στο NCBI κάθε βάση δεδομένων διαθέτει ξεχωριστό μηχανισμό διατύπωσης ερωτημάτων για αναζήτηση (Query Builder). Ο χρήστης μπορεί να διαλέξει τη βάση δεδομένων που επιθυμεί και να δημιουργήσει το ερώτημα του. Ακόμα, μπορεί να επιλέξει να χρησιμοποιήσει τη λειτουργία της «Προχωρημένης Αναζήτησης» (Advanced Search) στις περιπτώσεις που αυτή έχει υλοποιηθεί.

Τέλος δίνεται η δυνατότητα της αναζήτησης σε όλες τις βάσεις του NCBI από ένα μοναδικό σημείο – one stop. Πρόκειται για την «καθολική αναζήτηση» η οποία προσφέρεται μέσω της ιστοσελίδας <a href="https://www.ncbi.nlm.nih.gov/search/">https://www.ncbi.nlm.nih.gov/search/</a>.

Στη συνέχεια γίνεται αναφορά σε ορισμένες βάσεις στις οποίες παρέχεται πρόσβαση από το NCBI, προκειμένου να διαφανεί το εύρος των πληροφοριών στις οποίες δίδεται πρόσβαση:



Εικόνα 2.2.1

**PubMed** είναι μια βάση δεδομένων με αναφορές και περιλήψεις δημοσιεύσεων από το πεδίο της βιοϊατρικής όπως καταχωρήθηκαν στο MEDLINE και άλλα επιστημονικά περιοδικά βιοεπιστημών. [2]

**Bookshelf** είναι μια συλλογή βιοϊατρικών βιβλίων τα οποία μπορούν να αναζητηθούν απευθείας ή από διασυνδεδεμένα δεδομένα από άλλες βάσεις του NCBI. [2]

**PubMed Central** είναι μια βάση δεδομένων που περιέχει πρόσβαση σε επιστημονικά άρθρα με πλήρες κειμένου στον τομέα της ιατρικής και της βιοϊατρικής. [2]

**Nucleotide** είναι μια συλλογή από αλληλουχίες νουκλεοτιδίων από διάφορες πηγές, συμπεριλαμβανομένων των βάσεων δεδομένων GenBank και RefSeq. [2]

**Genome** είναι μια βάση δεδομένων που περιέχει δεδομένα ακολουθιών από ολόκληρα γονιδιώματα περισσότερων από 1000 οργανισμούς. [2]

**Gene** είναι μια βάση δεδομένων που περιέχει γονίδια και παρέχει δυνατότητα αναζήτησης. Η συγκεκριμένη βάση εστιάζει σε γονίδια που έχουν ολοκληρωθεί οι ακολουθίες τους. [2]

**Protein** είναι μια βάση δεδομένων που περιέχει καταχωρήσεις πρωτεϊνών από διάφορες πηγές όπως τις GenPept, RefSeq, PIR, PRF και PDB. [2]

#### 2.3 ΠΡΟΤΑΣΕΙΣ ΒΕΛΤΙΩΣΗΣ

Προτείνουμε να χρησιμοποιηθεί ο χαρακτήρας μπαλαντέρ (\*), γνωστός και ως "wildcard character \*", σε περιπτώσεις αναζήτησης ενός όρου συνδυάζοντας διάφορες παραλλαγές του. Ο μπαλαντέρ χαρακτήρας αντιπροσωπεύει οποιονδήποτε χαρακτήρα ή ακολουθία χαρακτήρων, προσφέροντας ευελιξία στην αναζήτηση πληροφοριών.

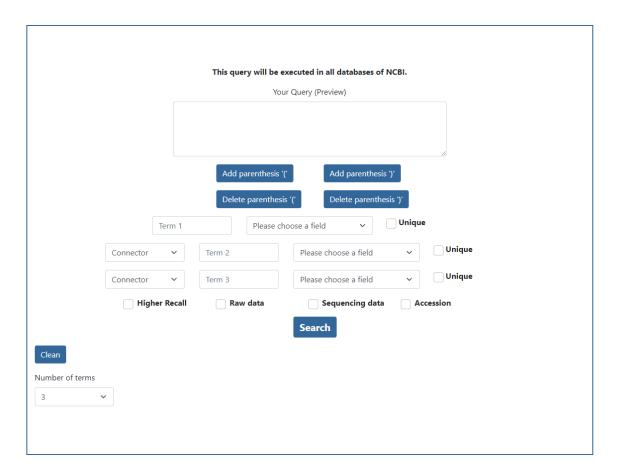
Για παράδειγμα, στην περίπτωση που ένας ερευνητής θέλει να αναζητήσει πληροφορίες για τον όρο melanoma, μπορεί επίσης να τον ενδιαφέρει πληροφορία που σχετίζεται με τον όρο melanomas (πληθυντικός αριθμός) ή melanoma's (γενική κτητική). Με την χρήση του μπαλαντέρ χαρακτήρα, οι αναζητήσεις γίνονται πιο «ανοιχτές» και επιτρέπουν τον εντοπισμό μεγαλύτερου πλήθους αποτελεσμάτων.

Αυτή η ανάγκη για την χρήση του μπαλαντέρ χαρακτήρα, προέκυψε επειδή συχνά, όταν διατυπώνεται ένα σύνθετο ερώτημα αναζήτησης, οι απαντήσεις που επιστρέφονται μπορεί να ειναι περιορισμένες.

Επιπλέον, η μελέτη στο πλαίσιο της πτυχιακής εργασία της κ. Δήμητρας Λυδάκη καταλήγει, μεταξύ άλλων, στο συμπέρασμα πως, «όταν προσθέτουμε τους όρους "raw data" και "sequencing data" στα σωστά σημεία ενός ερωτήματος, μπορούμε να ανακτήσουμε αποτελέσματα με πρόσβαση σε πρωτογενή δεδομένα».

Για αυτόν τον λόγο προσθέσαμε τις επιλογές "Raw data" και "Sequencing data" στην εφαρμογή μας, ώστε οι χρήστες να μπορούν να ανακτούν πρωτογενή δεδομένα τα οποία είναι καίρια για την έρευνα τους.

# 3 Η Οθόνη της Εφαρμογής - Περιγραφή πεδίων και Λειτουργικότητας



Εικόνα 3.1: Η Διεπαφή χρήστη

# 3.1 Πεδίο "Your Query (Preview)"

Καθώς ο χρήστης δημιουργεί το ερώτημα του, στο πεδίο "Your Query (Preview)", φαίνονται οι τροποποιήσεις ή οι προσθήκες που γίνονται. Μόλις ο χρήστης ολοκληρώσει το ερώτημα του, η καθολική αναζήτηση στις βάσεις δεδομένων του NCBI, πραγματοποιείται ακριβώς όπως έχει δημιουργηθεί το ερώτημα στο συγκεκριμένο πεδίο. Πριν υποβάλει ο χρήστης το ερώτημα του, έχει τη δυνατότητα να επεξεργαστεί το ερώτημά του εντός του σχετικού πεδίου.

Αυτό γίνεται με σκοπό να διασφαλιστεί ότι η αναζήτηση με την ακρίβεια που επιθυμεί να καθορίσει ο χρήστης, συνεισφέροντας στη συνάφεια και το πλήθος των επιστρεφόμενων αποτελεσμάτων.

#### 3.2 Πεδίο "Term x"

Στα πεδία "Term 1", "Term 2", …, "Term x", ο χρήστης μπορεί να εισάγει απλούς και σύνθετους όρους αναζήτησης καθώς και φράσεις. Κατά την εισαγωγή ενός σύνθετου όρου, όπως για παράδειγμα το lung cancer, η εφαρμογή θα αναζητήσει την διάζευξη των μοναδιαίων όρων. Από την άλλη πλευρά, όταν ο

χρήστης εισάγει μια φράση, δηλαδή έναν απλό ή σύνθετο όρο που βρίσκεται μεταξύ εισαγωγικών, η εφαρμογή θα αναζητήσει την φράση όπως έχει εισαχθεί, εξασφαλίζοντας την ακρίβεια της αναζήτησης.

#### 3.3 Πτυσσόμενο Μενού "Please choose a field"

Στο πτυσσόμενο μενού "Please choose a field", ο χρήστης αναλαμβάνει να επιλέξει το πεδίο στο οποίο επιθυμεί να πραγματοποιηθεί η αναζήτηση για τον όρο που έχει καταχωρηθεί στο πεδίο "Term x", το όποιο βρίσκεται αριστερά.

Στο πλαίσιο της εφαρμογής μας, προτείνονται, μέσω αναδυόμενου μενού, 31 διαφορετικές επιλογές που διατίθενται στον χρήστη. Ο χρήστης έχει την δυνατότητα να περιηγηθεί μεταξύ αυτών των επιλογών με σκοπό να επιλέξει αυτή που τον εξυπηρετεί καλύτερα στο πλαίσιο των αναγκών και των προτιμήσεών του.

- 1. Το κενό "" που σημαίνει ότι ο όρος θα αναζητηθεί σε όλα τα πεδία.
- 2. Το **Abstract** που σημαίνει ότι ο όρος θα αναζητηθεί στην περίληψη ενός άρθρου εφόσον διαθέτει. [5]
- 3. Το **Accession** αναζητάει τον όρο, για να δει αν αυτός μπορεί να αποτελεί κομμάτι του μοναδικού αναγνωριστικού καταχώρησης της εργασίας σε κάποιο βιοϊατρικό αποθετήριο. [4]
- 4. Το **Acknowledgments** χρησιμοποιείται όταν αναζητούμε έναν όρο στην ενότητα «Ευχαριστίες».
- 5. Το **Affiliation** χρησιμοποιείται για την αναζήτηση δημοσίευσης ή άλλης πληροφορίας από συγκεκριμένα ιδρύματα. [5]
- 6. Το **All Fields** χρησιμοποιείται για να αναζητήσει τον όρο σε όλα τα πεδία που είναι διαθέσιμα.
- 7. Το **Author** χρησιμοποιείται για να αναζητήσει τον συγγραφέα ενός άρθρου. [5]
- 8. Το **Body All Words** χρησιμοποιείται για να αναζητηθεί ένας όρος στο κύριο σώμα του άρθρου. [5]
- 9. Το **Body -Key Terms** χρησιμοποιείται για να βρει όλους τους βασικούς όρους στο σώμα ενός άρθρου εκτός από την περίληψη και τις παραπομπές. [5]
- 10. Το **DOI [Digital Object Identifier]** αναζητάει τον όρο, στο μοναδικό αναγνωριστικό ενός άρθρου. [5]
- 11. Το EC/RN Number[Enzyme Commission / Registry Number ] αναζητάει τον όρο για να εντοπίσει καταχωρήσεις που περιέχουν συγκεκριμένα ένζυμα ή χημικές ενώσεις. Οι αριθμοί EC/RN αποτελούν ένα σύστημα κωδικοποίησης που χρησιμοποιείται για την αναγνώριση και την ταξινόμηση ενζύμων. [5]

- 12. Το **Figure/Table Caption** αναζητάει τον όρο στα κείμενα που περιγράφουν εικόνες ή σχήματα σε ένα άρθρο. [5]
- 13. Το **Filter** αναζητάει τον όρο σύμφωνα με την ετικέτα που έχει προσθέσει ο χρήστης. Για παράδειγμα, ο χρήστης μπορεί να γράψει για Filter το Publication Date, και η εφαρμογή να αναζητήσει τον όρο στο πεδίο που αναγράφει την ημερομηνία που δημοσιεύτηκε σε μια καταχώρηση. [5]
- 14. Το **Full Author Name** αναζητάει τον όρο στο πλήρες όνομα του συγγραφέα ενός άρθρου για να δει αν ταιριάζει. [5]
- 15. Το **Grant Number** αναζητάει τον όρο στον αριθμό επιχορήγησης ή στην πηγή χρηματοδότησης μιας καταχώρησης εφόσον διαθέτει. [5]
- 16. Το **Issue** αναζητάει τον όρο στον αριθμό τεύχους ενός περιοδικού που δημοσιεύθηκε ένα άρθρο σε μια καταχώρηση. [5]
- 17. Το **Journal** αναζητάει τον όρο στον τίτλο ενός περιοδικού ή στον ISSN. [5]
- 18. Το **MeSH (Medical Subject Headings) Major Topic** αναζητάει καταχωρήσεις που έχουν ως κύριο θέμα τον συγκεκριμένο όρο στις Ιατρικές Θεματικές Επικεφαλίδες. [5]
- 19. Το **MeSH (Medical Subject Headings) Subheadings** αναζητάει καταχωρήσεις που περιέχουν ως υπότιτλο τον συγκεκριμένο όρο στις Ιατρικές Θεματικές Επικεφαλίδες. [5]
- 20. Τα **MeSH (Medical Subject Headings) Terms** αναζητάει καταχωρήσεις που ο όρος στην καταχώρηση υπάρχει ως όρος MeSH. [5]
- 21. Το Methods Key Terms αναζητάει τον όρο στην ενότητα «Μέθοδοι» ενός άρθρου. [5]
- 22. Το **Organism** αναζητάει τον όρο στα πεδία όπου αναφέρουν λεπτομέρειες για την γενετική, την φυσιολογία και την συμπεριφορά ζωντανών οργανισμών σε μια καταχώρηση. [5]
- 23. Το **Organism Unsynonymized** αναζητάει άρθρα που περιέχουν τον όρο ως οργανισμό ακριβώς όπως γράφεται και όχι με συνώνυμους όρους. [5]
- 24. Το **Pagination** αναζητάει τον όρο στον αριθμό σελίδας μιας καταχώρησης. [5]
- 25. Το **Reference** αναζητάει τον όρο στους τίτλους των αναφορών μιας καταχώρησης. [5]
- 26. Το **Reference Author** αναζητάει τον όρο στους συγγραφείς που υπάρχουν στην ενότητα αναφορών μιας καταχώρησης. [5]
- 27. Το Section Title αναζητάει τον όρο στους τίτλους ενοτήτων μιας καταχώρησης. [5]

- 28. Το **Supplementary Concept** αναζητάει τον όρο στο συμπληρωματικό μέρος μιας καταχώρησης. [5]
- 29. Το **Text Word** αναζητάει τον όρο στον τίτλο, την περίληψη και το σώμα ενός άρθρου. [5]
- 30. Το **Title** αναζητάει τον όρο στον τίτλο ενός άρθρου. [5]
- 31. Το **Volume** αναζητάει τον όρο στον αριθμό τόμου ενός περιοδικού που έχει δημοσιευθεί ένα άρθρο σε μια καταχώρηση. [5]

Τα πεδία που αναφέραμε, ισχύουν για άρθρα που έχουν ευρετηριαστεί από το PubMed/MEDLINE.

#### 3.4 Πεδίο "Connector"

Με το πεδίο "Connector" ο χρήστης καλείται να συνδέσει τους όρους που έχει καταχωρήσει. Συγκεκριμένα, το πεδίο "Connector" μπορεί να πάρει τιμές AND, OR και NOT. Έστω για παράδειγμα ότι έχουμε 2 όρους, η επιλογή AND υποδηλώνει την σύζευξη των 2 όρων, το OR υποδηλώνει την διάζευξη των 2 όρων και το NOT αποκλείει τα αποτελέσματα που περιέχουν τον όρο που ακολουθεί.

Λόγω της τυπικής λειτουργίας του NCBI, δεν επιτρέπεται η έναρξη ενός ερωτήματος με συνδετικό AND, OR ακόμα και NOT. Συνεπώς, ο χρήστης πρέπει πάντα να ξεκινά το ερώτημά του με έναν όρο και αυτός ο όρος δεν πρέπει να είναι κενός. Για παράδειγμα για να δημιουργηθεί ένα ερώτημα που αναζητά ανοσοθεραπεία και να αποκλείει το μελάνωμα μπορεί να χρησιμοποιηθεί η ακόλουθη σύνταξη: "immunotherapy NOT melanoma".

Αυτός είναι και ο λόγος που η εφαρμογή μας δεν παρέχει "Connector" για τον πρώτο όρο, για να συμμορφώνεται με την τυπική λειτουργία του NCBI, διασφαλίζοντας την έγκυρη αναζήτηση μέσω του μηχανισμού αναζήτησης του NCBI.

#### 3.5 Πτυσσόμενο Μενού "Number of terms"

Αν ο χρήστης επιθυμεί να αναζητήσει περισσότερους ή λιγότερους από 3 όρους, μπορεί να τροποποιήσει τον αριθμό των όρων πηγαίνοντας στο μενού "Number of terms", και να επιλέξει έναν αριθμό από 1 έως 30.

Είναι σημαντικό να αναφερθεί, <u>ότι απαιτείται από τον χρήστη να συμπληρώσει με τιμή κάθε πεδίο που</u> <u>εμφανίζεται</u>, προκειμένου να σταλεί σωστά το ερώτημα στο NCBI. Σε διαφορετική περίπτωση, θα πρέπει να μειώσει τον αριθμό των όρων.

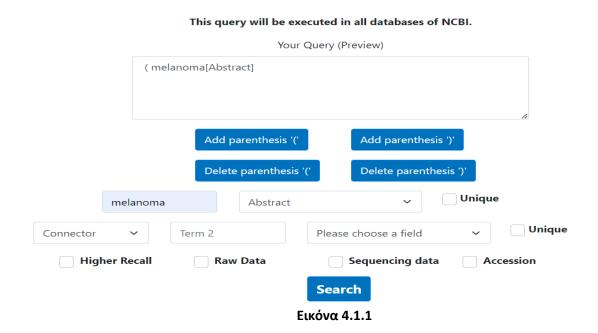
Η επιλογή μας να τοποθετήσουμε πρώτα τα πεδία και έπειτα τα κουμπιά και τις ρυθμίσεις έγινε με σκοπό η διεπαφή να ομοιάζει με τη συνήθη εμφάνιση τέτοιων, προκειμένου να είναι πιο φιλική προς τον χρήστη.

Για αυτόν το λόγο, συνιστάται στον χρήστη να επιλέξει τον αριθμό των όρων που επιθυμεί, και κατόπιν να προβεί στη συμπλήρωση των πεδίων με τιμές.

# 4 Λειτουργία Κουμπιών

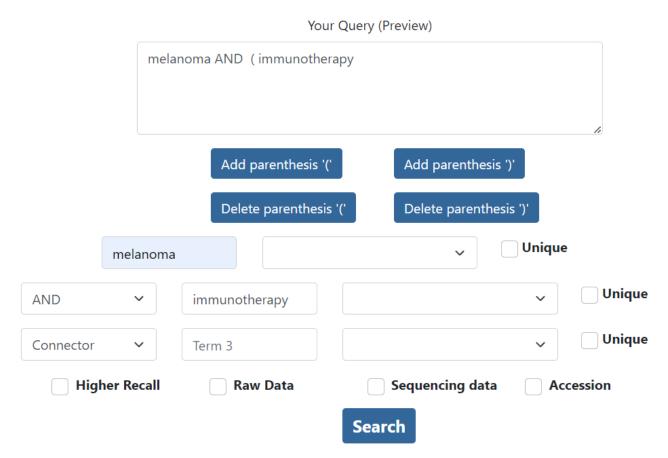
# 4.1 Κουμπί "Add parenthesis '(' "

Το κουμπί "Add parenthesis '('" χρησιμοποιείται για να μπορεί να προσθέσει ο χρήστης αριστερή παρένθεση και να αλλάξει την προκαθορισμένη προτεραιότητα αναζήτησης. Η αριστερή παρένθεση, τοποθετείται πάντα μπροστά από τον τελευταίο όρο που έχει γράψει ο χρήστης. Για παράδειγμα όπως βλέπουμε στην εικόνα 4.1.1 η παρένθεση τοποθετείται αριστερά από την λέξη "melanoma".



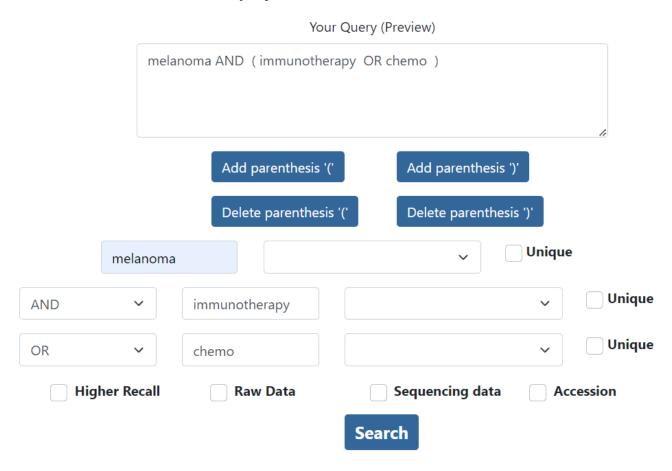
Πώς φτιάχνεται ένα ερώτημα της μορφής: melanoma AND (Immunotherapy OR chemo)?

Όταν όμως ο χρήστης έχει προσθέσει 2 όρους και δεν έχει βάλει καμία παρένθεση, τότε όπως βλέπουμε στην **εικόνα 4.1.2,** η παρένθεση θα μπει μπροστά από τον 2ο όρο.



Εικόνα 4.1.2

Στη συνέχεια,προσθέτουμε τον 3° όρο και την παρένθεση όπως βλέπουμε στην **εικόνα 4.1.3** 

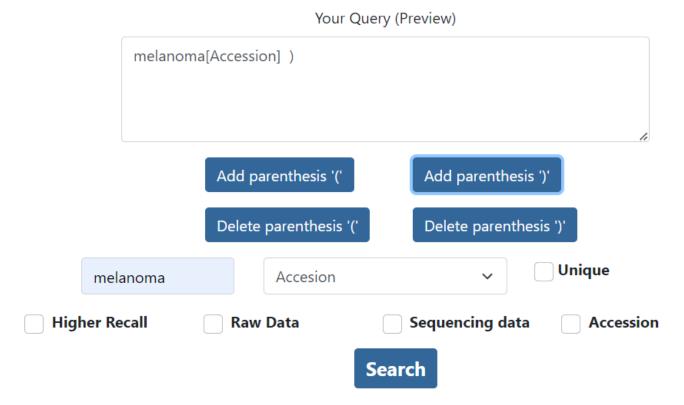


Εικόνα 4.1.3

Υπάρχει πάντα η δυνατότητα να επεξεργαστεί κανείς τις παρενθέσεις από το πεδίο προεπισκόπησης ερωτήματος πριν πατήσει το κουμπί "Search".

# 4.2 Κουμπί "Add parenthesis ')' "

Το κουμπί "Add parenthesis ')'" χρησιμοποιείται για να μπορεί να προσθέσει ο χρήστης δεξιά παρένθεση και να εφαρμόσει προτεραιότητα στους όρους. Η δεξιά παρένθεση, τοποθετείται πάντα μετά τον τελευταίο όρο που έχει εισαγάγει ο χρήστης. Για παράδειγμα όπως βλέπουμε στην **εικόνα 4.2.1** η παρένθεση τοποθετείται δεξιά από το πεδίο "[Accession]".



Εικόνα 4.2.1

Ο χρήστης έχει τη δυνατότητα να προσθέσει και χειροκίνητα τις παρενθέσεις, σε όποιο σημείο επιθυμεί, μόλις έχει τελειώσει την δημιουργία του ερωτήματος του, και πριν πατήσει το κουμπί "Search".

# 4.3 Κουμπί "Delete parenthesis '(' "

Το κουμπί "Delete parenthesis '('" χρησιμοποιείται για να μπορεί ο χρήστης να σβήσει τις αριστερές παρενθέσεις. Κάθε φορά που ο χρήστης πατάει αυτό το κουμπί, η αριστερή παρένθεση που βρίσκεται πιο δεξιά σβήνεται.

# 4.4 Κουμπί "Delete parenthesis ')' "

Το κουμπί "Delete parenthesis ')'" χρησιμοποιείται για να μπορεί ο χρήστης να σβήσει τις δεξιές παρενθέσεις. Κάθε φορά που ο χρήστης πατάει αυτό το κουμπί, η δεξιά παρένθεση που βρίσκεται πιο δεξιά σβήνεται.

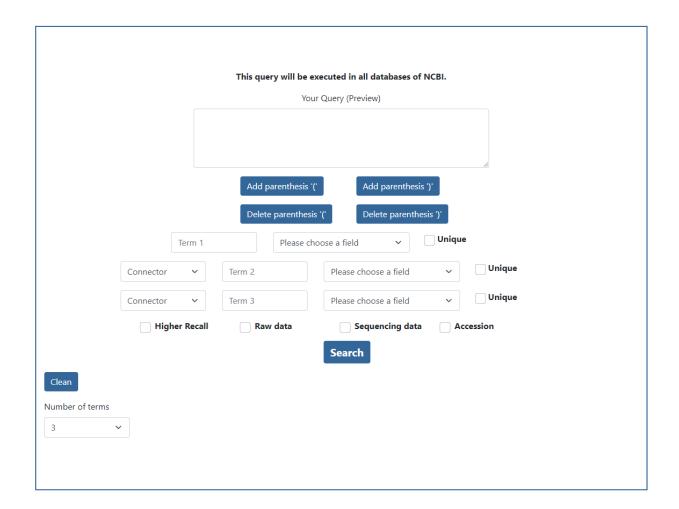
# 4.5 Κουμπί "Clean"

Το κουμπί "Clean" χρησιμοποιείται για να αφαιρεθούν όλες οι προσθήκες που έχει κάνει ο χρήστης και να «καθαρίσει» η φόρμα αναζήτησης. Δηλαδή να επιστρέψει στη μορφή που έχει μόλις ο χρήστης χρησιμοποιήσει την εφαρμογή για πρώτη φορά.

# 4.6 Κουμπί "Search"

Το κουμπί "Search" χρησιμοποιείται για να αναζητήσει ο χρήστης το ερώτημα που δημιούργησε στις βάσεις του NCBI.

# 5 Επιλογείς (Checkboxes)



# 5.1 Higher Recall

Ο επιλογέας "Higher Recall" χρησιμοποιείται όταν ο χρήστης επιθυμεί να πραγματοποιήσει αναζήτηση σε έναν όρο με πολλαπλές παραλλαγές ώστε να ανακτήσει περισσότερα αποτελέσματα. Για να προστεθεί ο χαρακτήρας μπαλαντέρ (\*) στον επιθυμητό όρο, αφαιρείται το τελευταίο γράμμα από αυτόν. Για παράδειγμα ο όρος "melanom\*" θα περιλαμβάνει παραλλαγές όπως το "melanomagnum", "melanommoides" ή "melanomaclinical".

Όταν ο χρήστης επιλέξει τον επιλογέα Higher Recall, ο χαρακτήρας μπαλαντέρ εφαρμόζεται σε όλους τους όρους, ωστόσο αν ο χρήστης επιθυμεί να εξαιρέσει κάποιον συγκεκριμένο όρο από αυτήν την επιλογή, μπορεί να επιλέξει τον επιλογέα Unique ώστε να μην προστεθεί ο χαρακτήρας μπαλαντέρ στον συγκεκριμένο όρο.

Αυτό επιτρέπει στον χρήστη να προσαρμόσει την αναζήτηση και να επιλέξει ποιοι όροι θα υπόκεινται στην εφαρμογή του μπαλαντέρ χαρακτήρα και ποιοι όχι, επιτρέποντάς του έτσι να καθορίσει ακριβέστερα την αναζήτησή του. Όταν ο χρήστης εισάγει μια φράση, δηλαδή έναν απλό ή σύνθετο όρο αναζήτησης μέσα

σε εισαγωγικά, η επιλογή Higher Recall απενεργοποιείται, καθώς η χρήση του μπαλαντέρ χαρακτήρα δεν θα έχει σωστή ερμηνεία και δεν θα επιστρέψει ακριβή αποτελέσματα.

Όταν ο χρήστης επιλέξει τον επιλογέα Higher Recall, μόλις πατήσει το κουμπί "Search" η εφαρμογή στέλνει ένα αίτημα τύπου POST στη λειτουργία αναζήτησης της βάσης PMC (PubMed Central) ώστε να "συνταχθεί" καλύτερα το ερώτημα για να ανακτηθούν περισσότερα αποτελέσματα. Στην συνέχεια, γίνεται ένα αίτημα τύπου GET για να μας επιστραφεί το ανασυνταγμένο ερώτημα και για να τοποθετήσουμε τους χαρακτήρες μπαλαντέρ στα σωστά σημεία των όρων. Τέλος, γίνεται ξανά ένα αίτημα τύπου POST για να σταλεί το νέο ερώτημα στο NCBI.

Ο λόγος που επιλέχθηκε η λειτουργία αναζήτησης της βάσης δεδομένων PMC είναι επειδή η μηχανή αναζήτησης που διαθέτει είναι πιο εξελιγμένη σε σχέση με τις άλλες βάσεις, καθώς αυτή βελτιστοποιεί τη σύνταξη του ερωτήματος και επιτρέπει την ανάκτηση μεγαλύτερου αριθμού αποτελεσμάτων. Επιπλέον, ένας ακόμη σημαντικός λόγος είναι ότι μέσω της αναζήτησης στο PMC είναι πιο εύκολο να ανακτηθεί το ανασυνταγμένο ερώτημα απ' ότι στις άλλες βάσεις.

#### 5.2 Unique

Ο επιλογέας "**Unique**" χρησιμοποιείται, όταν ο χρήστης εισάγει κάποιον ειδικό όρο όπως για παράδειγμα το "NCBI", και δεν υπάρχει παραλλαγή ή επέκταση γι' αυτόν. Επίσης χρησιμοποιείται, σε περίπτωση που ο χρήστης θέλει να ακυρώσει την καθολικά επιβληθείσα αναζήτηση παραλλαγών (μέσω του Higher Recall) για έναν συγκεκριμένο όρο.

#### 5.3 Raw data

Ο επιλογέας "**Raw data**" χρησιμοποιείται όταν ο χρήστης θέλει να γίνεται αναφορά σε πρωτογενή δεδομένα στα αποτελέσματα αναζήτησης που παράγει το ερώτημα το οποίο συντάσσει.

Με τον όρο πρωτογενή δεδομένα εννοούμε πληροφορίες που έχουν συλλεχθεί ή καταγραφεί από πηγές χωρίς να έχουν υποστεί κάποια επεξεργασία ή ανάλυση. Τα πρωτογενή δεδομένα υπάρχουν σε διάφορες μορφές όπως είναι Fasta, Fasta και BAM. Ένας λόγος που ένας ερευνητής θέλει να βρει πρωτογενή δεδομένα είναι ότι ο μεγάλος όγκος διαθέσιμων πρωτογενών δεδομένων μπορεί να χρησιμοποιηθεί συνδυαστικά σε αναλύσεις προκειμένου να προκύψουν νέα σημαντικά συμπεράσματα.

#### **5.4** Sequencing data

Ο επιλογέας "**Sequencing data**" χρησιμοποιείται όταν ο χρήστης θέλει να γίνεται αναφορά σε αλληλουχίες δεδομένων στα αποτελέσματα αναζήτησης που παράγει το ερώτημα το οποίο συντάσσει.

Τα δεδομένα αλληλούχισης αναφέρονται σε πληροφορίες που συλλέγονται και καταγράφονται για να περιγράψουν την αλληλεπίδραση μεταξύ δύο ή περισσότερων στοιχείων. Τα δεδομένα αλληλουχίας αποτελούν υποκλάση των πρωτογενών δεδομένων, αλλά προκύπτουν με ειδική κατεργασία κατάλληλου βιολογικού υλικού από μηχανές αλληλούχισης. Ένας λόγος που ένας ερευνητής μπορεί να ψάξει για τέτοια δεδομένα είναι για να εξερευνήσει τα αίτια μιας ασθένειας.

#### 5.5 Accession

Ο επιλογέας "Accession" χρησιμοποιείται για να εντοπίσει το μοναδικό αναγνωριστικό που συνδέεται με την καταχώρηση μιας μελέτης σε βιοϊατρικά αποθετήρια ή βάσεις. Κάθε εργασία έχει ένα Accession Number για την αναγνώριση της. Αυτός συνήθως αποτελείται από αλφαριθμητικούς χαρακτήρες και έχει μορφή ανάλογη με την βάση στην οποία έχει καταχωρηθεί.

Ο λόγος που προσθέσαμε τον συγκεκριμένο επιλογέα είναι επειδή μερικές φορές δεν γίνεται ρητή αναφορά σε ακατέργαστα δεδομένα ή δεδομένα αλληλούχισης, και ο εντοπισμός τους γίνεται μέσω του Accession Number.

# 6 Απαιτήσεις Εφαρμογής στους Browsers

# 6.1 Chrome Browser, Opera Browser και Microsoft Edge

Για χρήση της εφαρμογής στον Chrome τον Opera ή τον Edge, θα πρέπει ο χρήστης να προσθέσει το extension "Allow CORS: Access-Control-Allow-Origin" και να το ενεργοποιήσει.

#### 6.2 Mozilla Browser

Για χρήση της εφαρμογής στον Mozilla/Firefox, θα πρέπει ο χρήστης να προσθέσει το extension "CORS Everywhere" και να το ενεργοποιήσει.

# 7 Οθόνες της Εφαρμογής

# 7.1 Προσαρμογή για χρήση από διάφορες συσκευές

Η εφαρμογή είναι Responsive και προσαρμόζεται αυτόματα σε όλες τις οθόνες, για να μπορεί ο χρήστης να τη χρησιμοποιήσει από παντού: tablets, κινητά, φορητούς υπολογιστές και σταθερούς υπολογιστές. Αυτό εξασφαλίζεται χάρις στην τεχνολογία BOOTSTRAP που χρησιμοποιήθηκε στην εφαρμογή μας.

# 8 Αρχιτεκτονική του Συστήματος και Εργαλεία

# 8.1 Αρχιτεκτονική

Η αρχιτεκτονική που χρησιμοποιήσαμε στην εφαρμογή μας είναι ο προγραμματισμός χειρισμού γεγονότων. Το συγκεκριμένο μοντέλο εστιάζει στον χειρισμό γεγονότων και την απόκριση το συστήματος σε αυτά.

Ένα γεγονός αναφέρεται σε μια ενέργεια που συμβαίνει στο σύστημα.

Μερικά παραδείγματα γεγονότων είναι όταν ένας χρήστης πατήσει ένα πλήκτρο ή κάποιο κουμπί.

Συνήθως το μοντέλο αυτό λειτουργεί με την χρήση της JavaScript, που μας βοηθάει να ορίσουμε τα γεγονότα που θα συμβούν. Το συγκεκριμένο μοντέλο χρησιμοποιείται ευρέως στην ανάπτυξη ιστοεφαρμογών ή εφαρμογών για κινητές συσκευές.

Τέλος, να αναφέρουμε ότι μπορεί να υπάρξουν διάφορες υλοποιήσεις στο μοντέλο χειρισμού γεγονότων ανάλογα με την γλώσσα προγραμματισμού ή το πλαίσιο ανάπτυξης που χρησιμοποιείται.

# 8.2 Τεχνολογίες/frameworks used

Οι τεχνολογίες που χρησιμοποιήσαμε για την εφαρμογή μας είναι:

#### **HTML5 (Hypertext Markup Language)**

**HTML** είναι ένα σύνολο κανόνων που χρησιμοποιείται για τη διαμόρφωση περιεχομένου σε μια ιστοσελίδα. Δεν πρόκειται για γλώσσα προγραμματισμού, αλλά γλώσσα περιγραφής των στοιχείων που αποτελούν μια ιστοσελίδα.

Η έκδοση 5 είναι η πιο πρόσφατη και πιο πλήρης έκδοση της HTML, και περιλαμβάνει μερικά νέα χαρακτηριστικά όπως για παράδειγμα το <header>, <nav> και το <canvas>. Οι στόχοι της ήταν να βελτίωσει τη γλώσσα, και να υποστηρίξει νέες δυνατότητες, να είναι ευανάγνωστη από τον άνθρωπο και ερμηνεύσιμη από τους υπολογιστές και τις συσκευές.

Επιπροσθέτως, περιλαμβάνει υποστήριξη για πολυμέσα και γραφικά, ενσωματωμένα στοιχεία πολυμέσων και δυνατότητα χρήσης διανυσματικών γραφικών (SVG).

Τέλος, η HTML5 μπορεί να χρησιμοποιηθεί σε εφαρμογές για κινητές συσκευές σε πολλαπλές πλατφόρμες, επειδή περιέχει χαρακτηριστικά για συσκευές χαμηλής κατανάλωσης. [9]

#### **CSS (Cascading Style Sheets)**

CSS είναι μια γλώσσα που χρησιμοποιείται για την παρουσίαση και τη μορφοποίηση των ιστοσελίδων.

Πιο συγκεκριμένα, καθορίζει πως θα εμφανίζεται το περιεχόμενο ενός ιστοτόπου στην οθόνη.

Χρησιμοποιεί κανόνες για να ορίσει τις ιδιότητες των στοιχείων της HTML όπως είναι το χρώμα, η γραμματοσειρά, το μέγεθος του κειμένου καθώς και τις αποστάσεις μεταξύ των στοιχείων που υπάρχουν στην ιστοσελίδα. [10]

#### **JAVASCRIPT**

JAVASCRIPT είναι μια γλώσσα προγραμματισμού που εκτελείται ασύγχρονα και χρησιμοποιείται σε φυλλομετρητές, για να αλλάζει δυναμικά το περιεχόμενο σε μια ιστοσελίδα.

Μπορεί να χρησιμοποιηθεί για την εκτέλεση ενεργειών όπως αλλαγή περιεχομένου, αλλαγή του στυλ μιας ιστοσελίδας καθώς και να διαχειριστεί συμβάντα όπως το πάτημα ενός κουμπιού.

Ένας λόγος που επιλέξαμε την JavaScript για την εφαρμογή μας είναι επειδή μπορεί να εκτελεστεί αμέσως από τον φυλλομετρητή του χρήστη χωρίς να χρειάζεται να κατεβάσει επιπλέον αρχεία ή να προσθέσει περισσότερες ρυθμίσεις.

Η JavaScript, προσφέρει μεγάλη ευελιξία και αρκετές δυνατότητες, και αυτά την καθιστούν ιδανική για την ανάπτυξη ιστοσελίδων. [6]

#### **BOOTSTRAP**

**BOOTSTRAP** είναι ένα framework ανοιχτού κώδικα που χρησιμοποιείται για την ανάπτυξη ιστοσελίδων. Παρέχει ένα σύνολο CSS και JavaScript κανόνων, καθώς και πρότυπα διάταξης που βοηθούν τους προγραμματιστές να δημιουργήσουν πιο ευέλικτες και ελκυστικές αισθητικά ιστοσελίδες ευκολότερα.

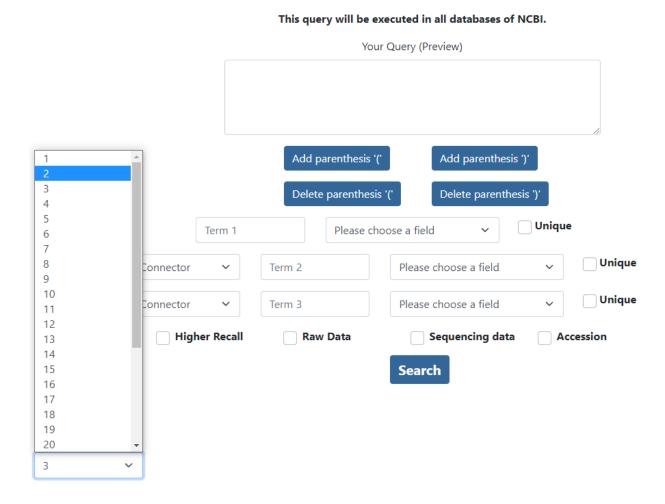
Το BOOTSTRAP παρέχει χαρακτηριστικά όπως τα πλέγματα (Grids) και τα Containers. [7]

# 9 Δοκιμή Συστήματος

#### Δοκιμή 1

Έστω ότι ο χρήστης θέλει να αναζητήσει 2 όρους.

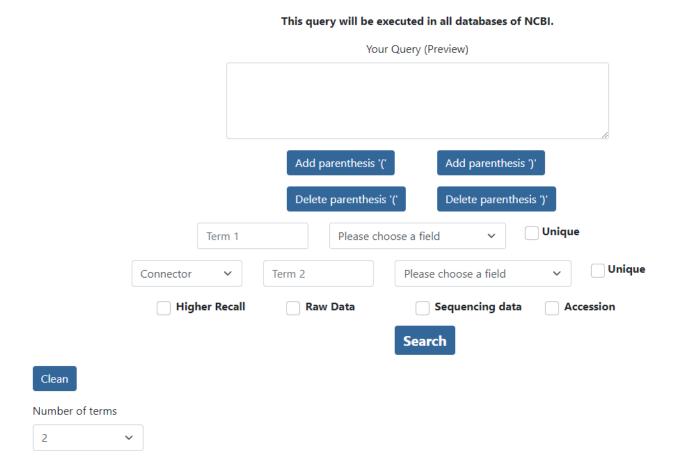
Όπως βλέπουμε στην **εικόνα 9.1.1** η πρώτη ενέργεια του χρήστη θα είναι να αλλάξει τον αριθμό των όρων από το πτυσσόμενο μενού Number of terms επιλέγοντας τον αριθμό 2.



Εικόνα 9.1.1

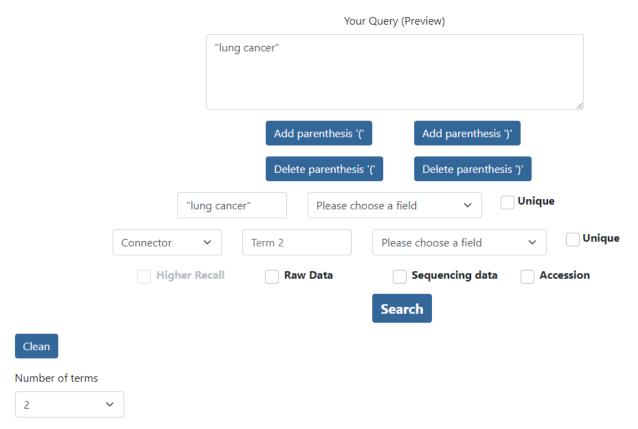
Σελίδα 24 από 51

Στην συνέχεια, όπως βλέπουμε στην εικόνα 9.1.2 ο αριθμός των όρων έχει αλλάξει.



Εικόνα 9.1.2

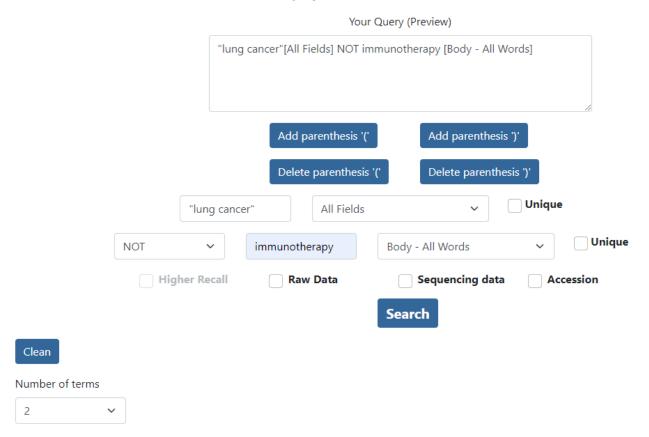
Μετά, ο χρήστης συμπληρώνει τον όρο "lung cancer" όπως βλέπουμε στην **εικόνα 9.1.3**.



Εικόνα 9.1.3

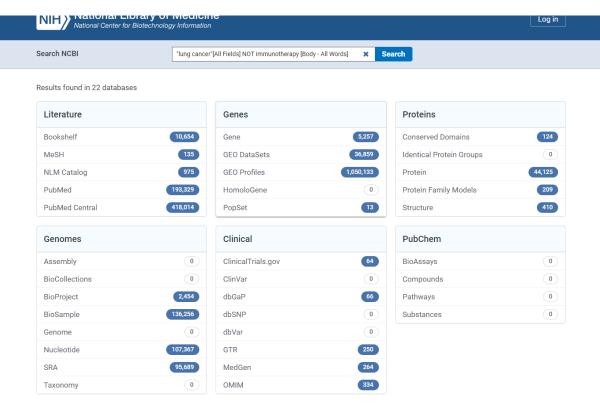
Όπως βλέπουμε στην παραπάνω εικόνα επειδή ο χρήστης έχει εισάγει τον όρο σε εισαγωγικά, η εφαρμογή τον μεταφράζει σαν φράση και για αυτόν τον λόγο ο επιλογέας Higher Recall απενεργοποιείται και δε μπορεί να χρησιμοποιηθεί. Αν θέλει ο χρήστης να τον χρησιμοποιήσει θα πρέπει να μην εισάγει τον όρο σε εισαγωγικά.

Επιπλέον, ο χρήστης εισάγει τον όρο immunotherapy αλλά με τον σύνδεσμο NOT, όπως βλέπουμε στην **εικόνα 9.1.4**.



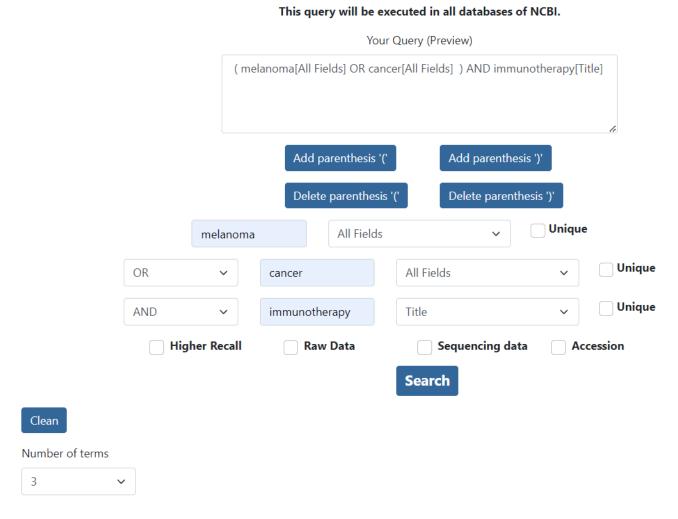
Εικόνα 9.1.4

Μόλις ο χρήστης πατήσει το κουμπί Search το ερώτημα θα σταλθεί στο NCBI όπως βλέπουμε στην **εικόνα 9.1.5.** 



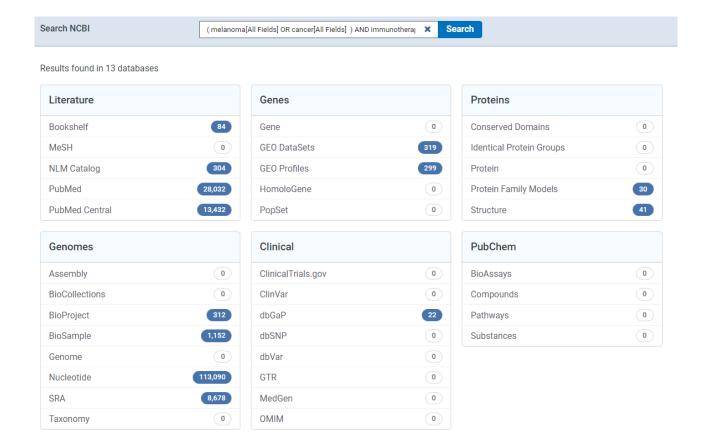
Εικόνα 9.1.5

Έστω ότι θέλουμε να αναζητήσουμε 3 όρους χρησιμοποιώντας παρενθέσεις όπως βλέπουμε στην **εικόνα** 9.2.1



Εικόνα 9.2.1

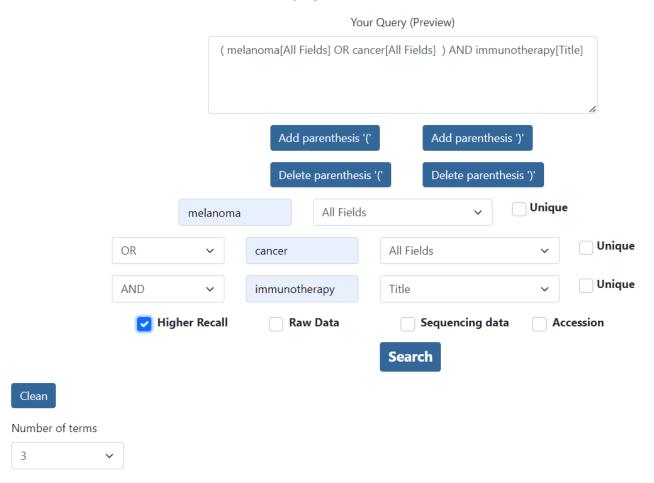
Στην συνέχεια, μόλις πατήσουμε το κουμπί "Search" βλέπουμε στην **εικόνα 9.2.2** ότι μας εμφανίζει κανονικά πάλι τα αποτελέσματα στην σελίδα του NCBI.



Εικόνα 9.2.2

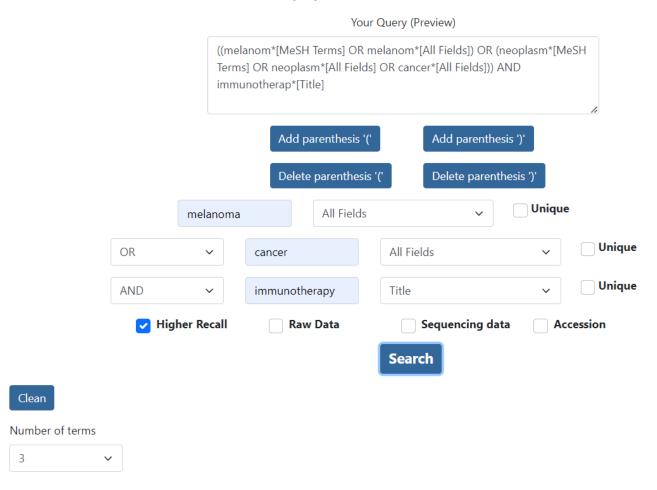
# Δοκιμή 3

Έστω ότι θέλουμε να χρησιμοποιήσουμε την επιλογή Higher Recall. Αρχικά, φτιάχνουμε το ερώτημα μας και στη συνέχεια, επιλέγουμε το σχετικό checkbox όπως βλέπουμε στην εικόνα 9.3.1.



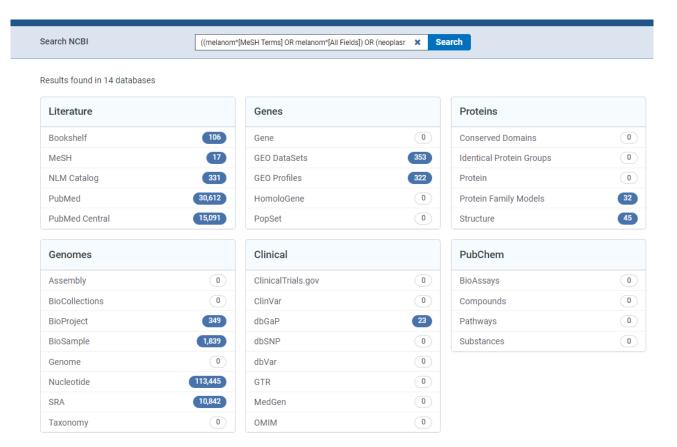
Εικόνα 9.3.1

Όταν έχουμε επιλέξει το checkbox του Higher Recall επειδή όπως έχουμε πει στην ενότητα 6.1 οι όροι που έχουμε εισάγει θα αλλάξουν, βλέπουμε **στην εικόνα 9.3.2** την αλλαγή που γίνεται στους όρους πριν εκείνοι σταλθούν για αναζήτηση στις βάσεις δεδομένων του NCBI.



Εικόνα 9.3.2

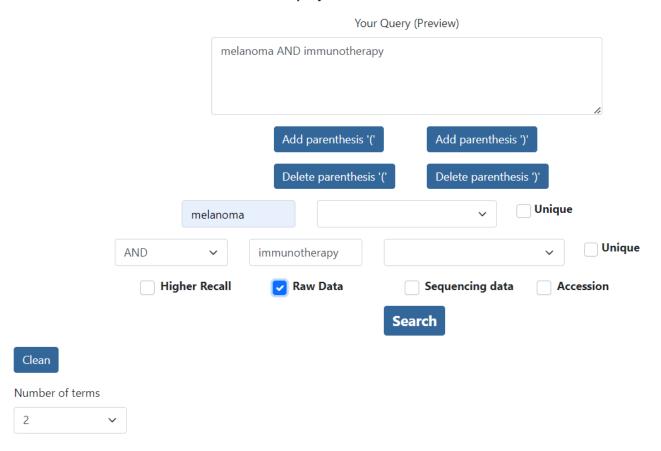
Στην επόμενη εικόνα 9.3.3 μπορούμε να δούμε τα αποτελέσματα του ΝCBI για το ερώτημα που στείλαμε.



Εικόνα 9.3.3

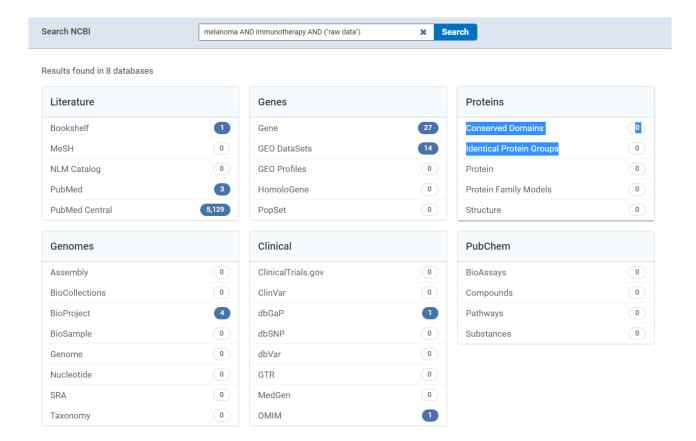
#### Δοκιμή 4

Αν θέλουμε να έχουμε και ακατέργαστα δεδομένα (raw data) στα αποτελέσματα μας, τότε αφού δημιουργήσουμε το ερώτημα , πρέπει να επιλέξουμε το checkbox που γράφει "Raw Data" όπως φαίνεται στην **εικόνα 9.4.1** 



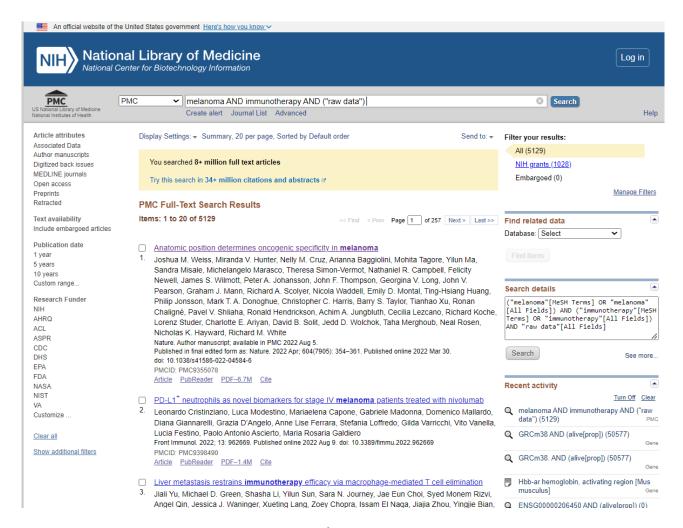
Εικόνα 9.4.1

Ακολουθεί στιγμιότυπο απο τα αποτελέσματα των βάσεων του NCBI για το ερώτημα που δημιουργήσαμε στην **εικόνα 9.4.2.** 



Εικόνα 9.4.2

Για να εξετάσουμε την εγκυρότητα των αποτελεσμάτων μας θα επιλέξουμε την βάση δεδομένων PubMed Central (εικόνα 9.4.3) για να δούμε αν σε κάποιο τυχαίο άρθρο υπάρχουν ακατέργαστα δεδομένα ή όχι .



Εικόνα 9.4.3

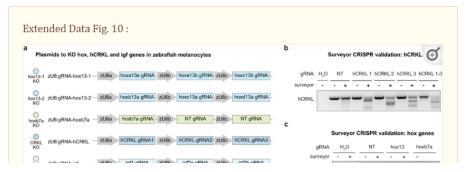
Αφού επιλέξουμε το πρώτο άρθρο, μπορούμε να δούμε στα επόμενα στιγμιότυπα (εικόνα 9.4.4 και εικόνα 9.4.5) ότι υπάρχει η λέξη raw data .Άρα υπάρχουν ακατέργαστα δεδομένα.

#### Data Availability Statement

All plasmids generated in this study will be either deposited with Addgene or available upon request. All fish lines will be made available through Zebrafish International Resource Center (ZIRC) or by request. All cell lines are available by request. Human RNA-seq data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) and are available under study accession EGAS00001001552 with dataset accession EGAD00001006439. Patient sample information and differential expression tables are provided in <u>Supplementary Table 5</u>. The zebrafish fin versus body bulk RNA-seq experiment is available via the NCBI Gene Expression Omnibus (GEO) repository under identifier code <u>GSE158538</u>, with bulk RNA-seq counts and differential expression tables in <u>Supplementary Table 4</u>. The zebrafish fin versus body single cell RNA-seq experiment is available via the NCBI GEO repository under the identifier code <u>GSE181748</u>. Original source data for ChIP–seq analysis is from ref. 31 and can be found at GEO <u>GSE81358</u>. The pathway analysis is made available in <u>Supplementary Table 6</u>. *HOXA13* Cut & Run data are available via the NCBI GEO repository under identifier code <u>GSE181768</u>. MACS2 peak scores and pathway analysis are available in <u>Supplementary Table 7</u>. CRKL immunoprecipitation–mass spectrometry data are available at ProteomeXchange with identifier PXD027968 and in <u>Supplementary Table 8</u>.

#### Εικόνα 9.4.4

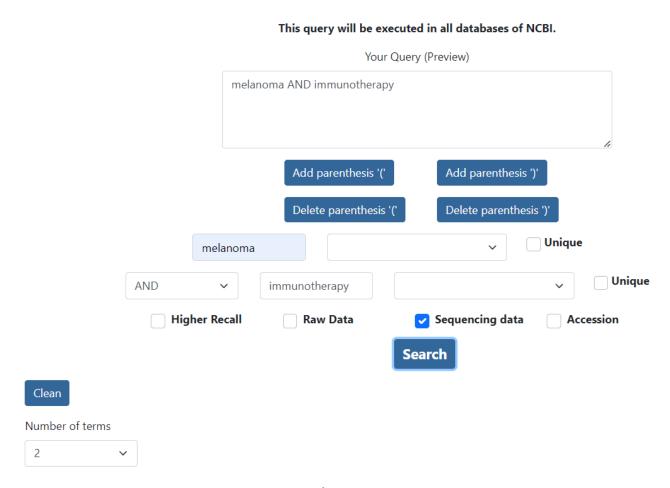
Related to Fig. 5. (a) Phospho-RTK array performed on WM3918 human melanoma cell line with or without overexpression of CRKL. Array tests phosphorylation status of 49 RTKs. For pRTK array source data, see Supplementary Fig. 2. (b) MeWo, SKMEL-1176, and 293T cells were transduced and selected to overexpress CRKL or an empty vector (EV) as a negative control. Western blot for CRKL, pIGF1R, and total IGF1R was then performed to compare basal levels of IGF signaling. Actin was used as a loading control. For western blot source data, see Supplementary Fig. 6. Quantification of western blots shown. Data represents n = 3 biological replicates and was analyzed using a two-sided student's t-test. (c) Schematic representation of CRKL with a V5 tag. CRKL is composed of a SH2 (Src homology) domain that bind proteins with phospho-tyrosines and two SH3 domains that bind to proline-rich proteins. (d) Human melanoma cell line WM3918 overexpressing CRKL with a V5 tag was compared to CRKL overexpressing cells without the V5 tag. N = 5 control replicates and n = 6 CRKL-V5 biological replicates were used for each condition. Lysates were immunoprecipitated using a V5 antibody and then underwent mass spectrometry, yielding 57 significant interactors as defined by logFC > 2 and corrected p-value < 0.05. CRKL itself was among these significant interactors. Significant interactors with CRKL were organized based on whether they Ψ-K), known SH3 binding motif (proline-rich sequence), or no identified binding motif. Genes are color coded by their contribution to PI3K and MAPK signaling. (e) Heatmap showing increased detection of significant interactors in the CRKL-V5 vs control group, expressed as normalized counts. Raw data can be found in Supplementary Table 8. (f) Volcano plot showing log2 fold enrichment of CRKL-V5 IP vs. control IP (CRKL with no V5 tag). \* p-value < 0.05, \*\* p-value < 0.01, \*\*\* p-value < 0.001, \*\*\*\* p-value < 0.0001.



Εικόνα 9.4.5

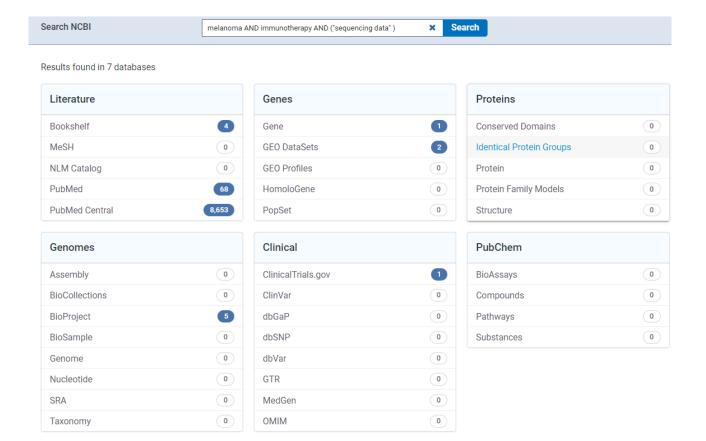
# Δοκιμή 5

Έστω ότι θέλουμε το ερώτημα μας να έχει sequencing data. Φτιάχνουμε το ερώτημα μας και επιλέγουμε την επιλογή "Sequencing data" όπως βλέπουμε στην εικόνα 9.5.1.



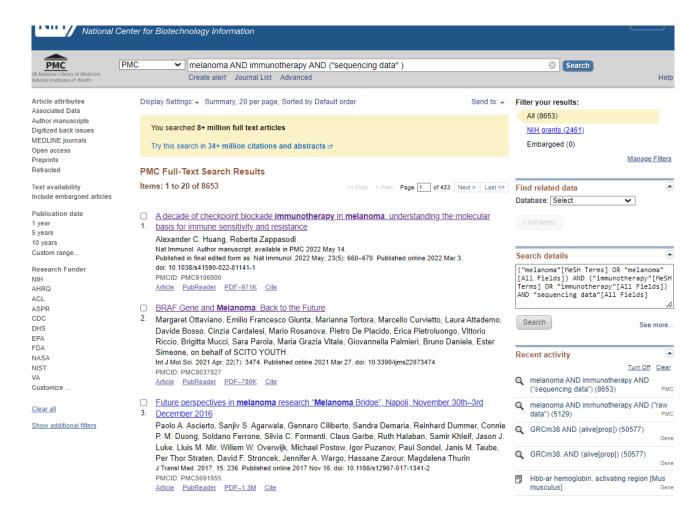
Εικόνα 9.5.1

Στη συνέχεια πατάμε το κουμπί "Search" για να στείλουμε το ερώτημα στις βάσεις δεδομένων του NCBI όπως βλέπουμε στην **εικόνα 9.5.2** .



Εικόνα 9.5.2

Επιλέγουμε την βάση δεδομένων PubMed Central όπως βλέπουμε στην **εικόνα 9.5.3.** 



Εικόνα 9.5.3

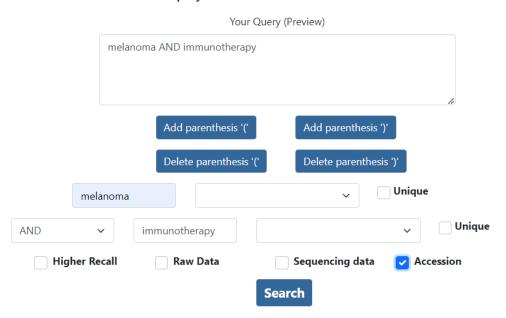
Στη συνέχεια επιλέγουμε το πρώτο άρθρο και όπως βλέπουμε στην **εικόνα 9.5.4** υπάρχει η λέξη sequencing data.

- 35. Rosenberg SA et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin Cancer Res* 17, 4550–4557, doi: 10.1158/1078-0432.CCR-11-0116 (2011). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 36. Goff SL et al. Randomized, Prospective Evaluation Comparing Intensity of Lymphodepletion Before Adoptive Transfer of Tumor-Infiltrating Lymphocytes for Patients With Metastatic Melanoma. *J Clin Oncol* 34, 2389–2397, doi: 10.1200/JC0.2016.66.7220 (2016). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 37. Linnemann C et al. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med* 21, 81–85, doi: 10.1038/nm.3773 (2015). [PubMed] [CrossRef] [Google Scholar]
- 38. Lu YC et al. Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. *J Immunol* 190, 6034–6042, doi: 10.4049/jimmunol.1202830 (2013). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 39. Prickett TD et al. Durable Complete Response from Metastatic Melanoma after Transfer of Autologous T Cells Recognizing 10 Mutated Tumor Antigens. *Cancer immunology research* 4, 669–678, doi: 10.1158/2326-6066.CIR-15-0215 (2016). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 40. Robbins PF et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* 19, 747–752, doi: 10.1038/nm.3161 (2013). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 41. Tran E et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. Science 350, 1387–1390, doi: 10.1126/science.aad1253 (2015). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 42. Krishna S et al. Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer. Science 370, 1328–1334, doi: 10.1126/science.abb9847 (2020). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 43. Gilbert AE et al. Monitoring the systemic human memory B cell compartment of melanoma patients for antitumor IgG antibodies. *PLoS One* 6, e19330, doi: 10.1371/journal.pone.0019330 (2011). [PMC free article] [PubMed] [CrossRef] [Google Scholar]
- 44. Pfreundschuh M The genealogy of SEREX. Cancer Immun 12, 7 (2012). [PMC free article] [PubMed] [Google Scholar]

#### Εικόνα 9.5.4

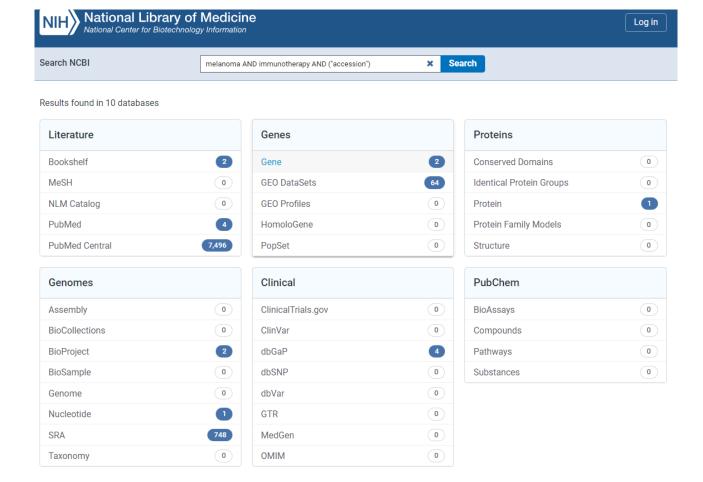
## Δοκιμή 6

Έστω ότι φτιάχνοντας το ερώτημα μας θέλουμε μαζί με τα αποτελέσματα να βρίσκουμε το Accession number του κάθε άρθρου. Μόλις φτιάξουμε το ερώτημα μας θα επιλέξουμε την επιλογή Accession όπως βλέπουμε στην εικόνα 9.6.1.



Εικόνα 9.6.1

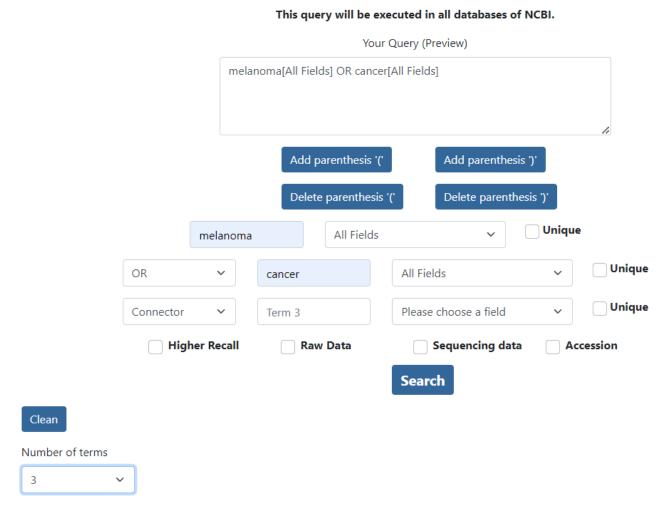
Στη συνέχεια, πατάμε το κουμπί "Search" για να στείλουμε το ερώτημα στις βάσεις δεδομένων του NCBI όπως βλέπουμε στην **εικόνα 9.6.2** .



Εικόνα 9.6.2

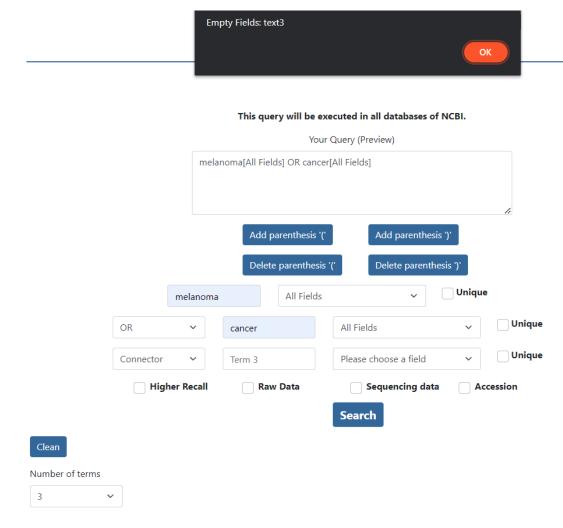
#### ΔΟΚΙΜΗ 7

Έστω ότι ο χρήστης έχει να συμπληρώσει 3 σειρές πεδίων αλλά έχει συμπληρώσει 2 σειρές, όπως βλέπουμε στην **εικόνα 9.7.1** 



Εικόνα 9.7.1

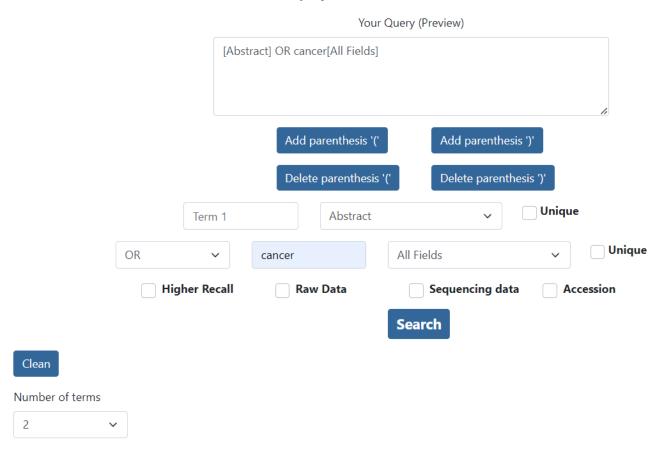
Μόλις ο χρήστης πατήσει το κουμπί "Search" θα εμφανιστεί όπως βλέπουμε στην **εικόνα 9.7.2** το παρακάτω μήνυμα.



Εικόνα 9.7.2

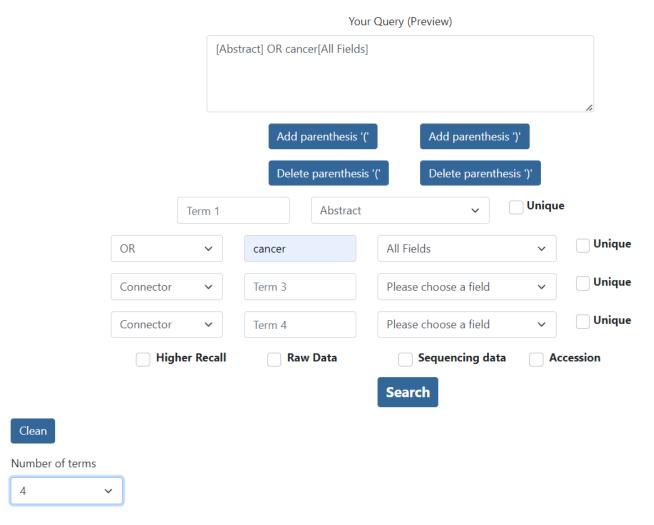
#### ΔΟΚΙΜΗ 8

Έστω ότι ο χρήστης έχει επιλέξει 2 σειρές πεδίων όπως βλέπουμε στην **εικόνα 9.8.1** να συμπληρώσει, αλλά τελικά θέλει να συμπληρώσει άλλες 2 σειρές.



Εικόνα 9.8.1

Μπορεί να αλλάξει τον αριθμό των σειρών από το πτυσσόμενο μενού "**Number of terms**" και να επιλέξει τον αριθμό των σειρών που επιθυμεί χωρίς να σβηστούν τα προηγούμενα που έχει συμπληρώσει όπως βλέπουμε στην **εικόνα 9.8.2** .



Εικόνα 9.8.2

#### ΔΟΚΙΜΗ 9

Έστω ότι ο χρήστης έχει στείλει ένα ερώτημα στο NCBI όπως βλέπουμε στην εικόνα 9.9.1.

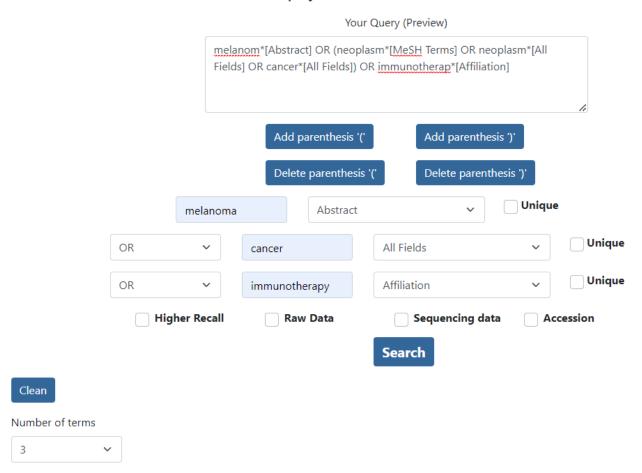
#### Your Query (Preview) melanom\*[Abstract] OR (neoplasm\*[MeSH Terms] OR neoplasm\*[All Fields] OR cancer\*[All Fields]) OR immunotherap\*[Affiliation] Add parenthesis '(' Add parenthesis ')' Delete parenthesis '(' Delete parenthesis ')' Unique Abstract melanoma Unique OR All Fields cancer Unique OR Affiliation immunotherapy Higher Recall Raw Data Sequencing data Accession Search Clean Number of terms

3

This query will be executed in all databases of NCBI.

Εικόνα 9.9.1

Και θέλει για παράδειγμα να αφαιρέσει τον αστερίσκο(\*) από τον όρο cancer και να στείλει ξανά το ερώτημα. Αρχικά θα πρέπει να αντιγράψει από το πεδίο κειμένου Your Query (Preview) το ερώτημα του, και στη συνέχεια να απενεργοποιήσει την επιλογή Higher Recall και να κάνει επικόλληση το ερώτημα που αντέγραψε όπως βλέπουμε στην εικόνα 9.9.2.



Εικόνα 9.9.2

Έπειτα,θα σβήσει από το πεδίο κειμένου τον αστερίσκο από την λέξη cancer και θα πατήσει το κουμπί "Search".

## 10 Επίλογος

Η εφαρμογή που αναπτύξαμε προσφέρει λειτουργικές επεκτάσεις στην καθολική αναζήτηση του NCBI, δίνοντας νέα χαρακτηριστικά που μπορεί να υποστηρίξουν σημαντικά την επιστημονική κοινότητα.

Για παράδειγμα, η επιλογή "Higher Recall" που όπως είδαμε αυξάνει τον αριθμό των επιστρεφόμενων αποτελεσμάτων ειδικά σε σύνθετα ερωτήματα.

Επιπλέον, οι επιλογές "Raw data", "Sequencing data" καθώς και "Accession", οι οποίες προσφέρουν την δυνατότητα ανάκτησης κειμένων που περιέχουν πρωτογενή δεδομένα.

Ακόμα, είναι σημαντικό να αναφερθεί η ημιαυτόματη εισαγωγή παρενθέσεων στα κατάλληλα σημεία του ερωτήματος του χρήστη για την σωστή εφαρμογή προτεραιότητας.

Τέλος, η διεπαφή χρήστη σχεδιάστηκε με γνώμονα την ευκολία και την προσαρμογή χρήσης από κάθε συσκευή καθώς και την άμεση αντίληψη των λειτουργιών της από τους χρήστες.

## Πηγές

- [1] https://www.ncbi.nlm.nih.gov/books/NBK470040/ [27-05-2023]
- [2] https://www.ncbi.nlm.nih.gov/guide/all/ [10-06-2023]
- [3] https://www.nlm.nih.gov/bsd/mms/medlineelements.html [29-05-2023]
- [4] https://www.ncbi.nlm.nih.gov/genbank/samplerecord/ [29-05-2023]
- [5] <a href="https://www.ncbi.nlm.nih.gov/pmc/about/userguide/">https://www.ncbi.nlm.nih.gov/pmc/about/userguide/</a> [30-05-2023]
- [6] https://el.wikipedia.org/wiki/JavaScript [09-06-2023]
- [7] https://el.wikipedia.org/wiki/Bootstrap [09-06-2023]
- [8] <a href="https://www.ncbi.nlm.nih.gov/guide/literature/">https://www.ncbi.nlm.nih.gov/guide/literature/</a> [10-06-2023]
- [9] https://en.wikipedia.org/wiki/HTML5 [22-06-2023]
- [10] https://el.wikipedia.org/wiki/CSS [22-06-2023]

## Κώδικας εφαρμογής

https://drive.google.com/drive/folders/1beWXR\_FfNDMT16K0\_9ML1Q774uhE\_73E?usp=sharing