



**ΠΑΠΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ**

**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ  
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ**

**UNIVERSITY OF WEST ATTICA  
DEPARTMENT OF INFORMATICS AND COMPUTER ENGINEERING**

**Πτυχιακή Εργασία**

**Ανάπτυξη συστήματος προτάσεων ταινιών και  
εύχρηστου γραφικού περιβάλλοντος**

**Τσελίκης Δημήτριος (ΑΜ: 711161213)**

**Επιβλέπων: Χρήστος Τρούσσας**

**Αιγάλεω, Ιούλιος 2022**

# Επιτροπή Εξέτασης

1. Χρήστος Τρούσσας  
Επίκουρος Καθηγητής
  
2. Παναγιώτα Τσελέντη  
Μέλος Εργαστηριακού Διδακτικού Προσωπικού
  
3. Γεώργιος Μελετίου  
Μέλος Εργαστηριακού Διδακτικού Προσωπικού

## ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Τσελίκης Δημήτριος, με αριθμό μητρώου 711161213 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Μηχανικών Πληροφορικής και Υπολογιστών, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



## Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Χρήστο Τρούσσα για την επίβλεψη της πτυχιακής και την καθοδήγηση που προσέφερε.

Θα ήθελα επίσης να ευχαριστήσω τους φίλους μου Κωνσταντίνο και Νίκο για τις γνώσεις που παρείχαν στα ανάλογα επιστημονικά πεδία καθώς και τις φίλες μου Εύα και Χριστίνα που χωρίς την αναποφασιστικότητά τους αυτή η ιδέα δε θα είχε γεννηθεί.

## Περίληψη στα ελληνικά

Η έλευση του διαδικτύου έφερε μαζί της μία τεράστια αύξηση στον όγκο της πληροφορίας που καταναλώνεται καθημερινά από τον καθένα. Στην σημερινή, πολυάσχολη και με ταχείς ρυθμούς κοινωνία είναι επιτακτική ανάγκη το φιλτράρισμα της πληροφορίας αυτής προκειμένου να ανακαλυφθούν τα κομμάτια αυτά που αντιστοιχούν στις εκάστοτε προτιμήσεις. Για να βοηθήσουν σε αυτή την προσπάθεια, δημιουργήθηκαν τα συστήματα προτάσεων, τα οποία έχουν αναλάβει όχι μόνο την διαλογή των πληροφοριών, αλλά και να ανακαλύψουν πληροφορίες οι οποίες πιθανώς να ανήκουν στο σύνολο των ενδιαφερόντων ενός ατόμου, δίχως όμως να το έχει αντιληφθεί. Σήμερα, τα συστήματα προτάσεων βρίσκονται παντού, από ένα ηλεκτρονικό εμπορικό κατάστημα όπου αναλαμβάνουν το ρόλο του πωλητή, προτείνοντάς προϊόντα τα οποία πιστεύουν πως ενδιαφέρουν τον πελάτη, μέχρι και στον χώρο της ψυχαγωγίας, με το πλέον αντιπροσωπευτικό παράδειγμα τις πλατφόρμες streaming, όπου προσπαθούν να κρατήσουν τους χρήστες όσο το δυνατόν περισσότερο χρόνο στην πλατφόρμα, προτείνοντάς τους συνεχώς νέο περιεχόμενο προς θέαση, πάντα με βάση τα ενδιαφέροντά τους.

Στην πτυχιακή εργασία αυτή γίνεται αναφορά σε κάποια βασικά ιστορικά στοιχεία των συστημάτων προτάσεων, παρουσιάζοντας κάποια αντιπροσωπευτικά συστήματα και στη συνέχεια αναφέρεται ο τρόπος που τα συστήματα αυτά χρησιμοποιούνται από διαδικτυακούς κολοσσούς προκειμένου να προσφέρουν εξατομικευμένο περιεχόμενο, πολύ προσεκτικά διαλεγμένο για τον κάθε χρήστη χωριστά. Τέλος, έγινε προσπάθεια να αναπτυχθεί ένα, βασικό, σύστημα προτάσεων ταινιών το οποίο λαμβάνει υπόψιν τις προτιμήσεις του χρήστη για τα είδη των ταινιών προκειμένου να παράγει εξατομικευμένες προτάσεις. Μαζί με το σύστημα αναπτύχθηκε, επίσης, κι ένα γραφικό περιβάλλον προκειμένου να το πλαισιώσει και να δημιουργήσει ένα φιλικό, οικείο περιβάλλον για τον χρήστη.

Το σύστημα που αναπτύχθηκε ήταν σε θέση να παράγει, ως επί το πλείστον, αξιόλογες προτάσεις οι οποίες συμφωνούν με τα προτιμητέα είδη του χρήστη. Το κυριότερο πρόβλημα εμφανίζεται στο σύνολο δεδομένων, το οποίο, λόγω της παλαιότητας ενός μεγάλου αριθμού ταινιών, η ομάδα δοκιμής δεν αξιολόγησε τις προτάσεις ως ιδιαίτερα ενδιαφέρουσες. Όσον αφορά το γραφικό περιβάλλον, εκπλήρωσε το σκοπό του, με το σύνολο της ομάδας δοκιμής να το χαρακτηρίζει ως πολύ οικείο και εύχρηστο.

**Λέξεις Κλειδιά:** Σύστημα προτάσεων, εξατομίκευση, ταινίες, προτιμήσεις, κριτήρια, συμπεριφορά χρήστη

## Περίληψη στα αγγλικά

The advent of the internet has brought with it a huge increase in the amount of information consumed by everyone on a daily basis. In today's busy and fast-paced society, it is imperative to filter this information in order to discover the pieces that correspond to each individual's preferences. To help in this effort, recommendation systems were created, which are tasked not only with sorting through information, but also with discovering information that may belong to a person's set of interests without them realizing it. Today, recommendation systems are everywhere, from an e-commerce store where they take on the role of a salesperson, recommending products that they believe are of interest to the customer, even in the entertainment industry, with the most representative example being streaming platforms, where they try to keep users as much time as possible on the platform, constantly suggesting them new content to watch, always based on their interests.

This thesis, reference is made to some basic historical elements of recommendation systems, presenting some representative systems and then the way these systems are used by online giants in order to offer personalized content, very carefully selected for each user separately. Finally, an attempt was made to develop a basic movie recommendation system that takes into account the user's preferences for movie genres in order to produce personalized recommendations. Along with the system, a graphical user interface was also developed in order to frame it and create a friendly, familiar environment for the user.

The developed system was able to produce, for the most part, worthwhile suggestions that agree with the user's preferred genres. The main problem occurs in the data set, which, due to the age of a large number of movies, the test team did not rate the suggestions as particularly interesting. As for the graphical user interface, it served its purpose, with the entire test team describing it as very familiar and easy to use.

**Keywords:** recommendation system, personalization, movies, preferences, criterion, user behavior

# Περιεχόμενα

ΕΠΙΤΡΟΠΗ ΕΞΕΤΑΣΗΣ .....	II
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ .....	III
ΕΥΧΑΡΙΣΤΙΕΣ .....	IV
ΠΕΡΙΛΗΨΗ ΣΤΑ ΕΛΛΗΝΙΚΑ .....	V
ΠΕΡΙΛΗΨΗ ΣΤΑ ΑΓΓΛΙΚΑ .....	VII
<b>ΚΕΦΑΛΑΙΟ 1. ΕΙΣΑΓΩΓΗ.....</b>	<b>1</b>
1.1 ΙΣΤΟΡΙΚΑ ΣΤΟΙΧΕΙΑ.....	2
1.1.1 Σύστημα Grundy.....	2
1.1.2 Το σύστημα συνεργατικού φιλτραρίσματος Tapestry .....	3
1.1.3 Η ερευνητική ομάδα GroupLens .....	5
1.1.4 Η συνεργασία επιστημονικής κοινότητας & βιομηχανίας: Το Βραβείο Netflix ....	7
1.2 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ .....	9
<b>ΚΕΦΑΛΑΙΟ 2. ΑΝΑΣΚΟΠΗΣΗ ΒΙΒΛΙΟΓΡΑΦΙΑΣ .....</b>	<b>10</b>
2.1 ΤΟ ΣΥΣΤΗΜΑ ΠΟΥ ΑΛΛΑΞΕ ΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ: AMAZON.COM .....	10
2.2 ΝΕΤΦΛΙΧ: ΤΑ ΘΕΜΕΛΙΑ ΤΗΣ ΣΥΓΧΡΟΝΗΣ ΟΠΤΙΚΟΑΚΟΥΣΤΙΚΗΣ ΨΥΧΑΓΩΓΙΑΣ .....	13
2.3 ΥΟΥΤΥΒΕ .....	16
2.4 ΤΙΚΤΟΚ (ΔΟΥΙΝ) .....	18
<b>ΚΕΦΑΛΑΙΟ 3. ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ.....</b>	<b>23</b>
3.1 ΔΙΑΡΘΡΩΣΗ ΚΕΦΑΛΑΙΟΥ .....	23
3.2 ΠΑΡΟΜΟΙΕΣ ΕΦΑΡΜΟΓΕΣ .....	23
3.3 ΣΥΝΟΛΟ ΔΕΔΟΜΕΝΩΝ .....	25
<b>ΚΕΦΑΛΑΙΟ 4. ΑΛΓΟΡΙΘΜΙΚΕΣ ΤΕΧΝΙΚΕΣ .....</b>	<b>29</b>
4.1 ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΣΥΣΤΗΜΑΤΟΣ.....	29
4.2 ΑΠΟΘΗΚΕΥΣΗ ΔΕΔΟΜΕΝΩΝ .....	30
4.3 ΤΟ ΣΥΣΤΗΜΑ ΠΡΟΤΑΣΕΩΝ .....	33
4.3.1 Ορισμός.....	33
4.3.2 Ιδιότητες.....	34
4.3.3 Κανονικοποίηση ειδών.....	34
4.3.4 Βάρη.....	36
4.3.5 Παραγωγή προτάσεων.....	38
4.4 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΕΦΑΡΜΟΓΗΣ .....	39
4.4.1 Στρώση γραφικού περιβάλλοντος .....	40
4.4.2 Στρώση δεδομένων.....	41



4.4.3	<i>Android Jetpack</i> .....	42
4.4.4	<i>Single Activity</i> .....	46
<b>ΚΕΦΑΛΑΙΟ 5. ΠΕΡΙΠΤΩΣΕΙΣ ΧΡΗΣΗΣ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ .....</b>		<b>48</b>
5.1	ΠΕΡΙΠΤΩΣΕΙΣ ΧΡΗΣΗΣ .....	48
5.1.1	<i>Πρώτη περίπτωση χρήσης</i> .....	48
5.1.2	<i>Δεύτερη περίπτωση χρήσης</i> .....	49
5.1.3	<i>Τρίτη περίπτωση χρήσης</i> .....	51
5.1.4	<i>Τέταρτη περίπτωση χρήσης</i> .....	52
5.1.5	<i>Πέμπτη περίπτωση χρήσης</i> .....	52
5.2	ΑΞΙΟΛΟΓΗΣΗ.....	53
5.2.1	<i>Αξιολόγηση οικειότητας εφαρμογής</i> .....	54
5.2.2	<i>Αξιολόγηση λειτουργικότητας</i> .....	55
5.2.3	<i>Αξιολόγηση συστήματος προτάσεων</i> .....	56
5.2.4	<i>Αξιολόγηση ταινιών</i> .....	58
5.2.5	<i>Σχόλια χρηστών</i> .....	59
<b>ΚΕΦΑΛΑΙΟ 6. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ .....</b>		<b>60</b>
6.1	ΣΥΜΠΕΡΑΣΜΑΤΑ.....	60
6.2	ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ .....	60
<b>ΚΕΦΑΛΑΙΟ 7. ΒΙΒΛΙΟΓΡΑΦΙΑ .....</b>		<b>62</b>

# Κεφάλαιο 1. Εισαγωγή

Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζει σήμερα η κοινωνία της πληροφορίας είναι η τεράστια και συνεχόμενη ροή -νέων- πληροφοριών. Οι νέες τεχνολογίες που έχουν αναπτυχθεί καθώς και η τεράστια πρόοδος στον τομέα της επικοινωνίας έχουν κάνει την παραγωγή, διαμοιρασμό και κατανάλωση της γνώσης μία τετριμμένη καθημερινή πράξη. Ο ερχομός του Διαδικτύου (Internet) έχει ενώσει όλες τις άκρες του κόσμου και έχει αλλάξει την έννοια της απόστασης, ενώνοντας όλες τις διαφορετικές πηγές πληροφοριών, δημιουργώντας μία τεράστια, απύθμενη πηγή. Μια πληροφορία η οποία δημιουργείται στη μία άκρη του Ατλαντικού Ωκεανού μπορεί ακαριαία (αν όχι ζωντανά όσο συμβαίνει!) να καταναλωθεί από έναν δέκτη στην άλλη άκρη του, γεγονός που λίγες δεκαετίες πριν φάνταζε όνειρο θερινής νυκτός. Σήμερα, περισσότερο από ποτέ άλλοτε, ερχόμαστε καθημερινά αντιμέτωποι με έναν τεράστιο όγκο πληροφοριών· πληροφορίες από τις οποίες χρειάζεται διαλογή προκειμένου να βρεθούν αυτές οι οποίες είναι χρήσιμες για τον καθένα. Μια διαδικασία χρονοβόρα, όπου οι ταχείς ρυθμοί της σημερινής κοινωνίας συχνά δεν το επιτρέπουν.

Η πρώτη προσέγγιση στην εύρεση λύσης για το συγκεκριμένο πρόβλημα ήταν ο ειδήμων. Ένας άνθρωπος, ή ομάδα ατόμων, ο οποίος, χρησιμοποιώντας τις εξειδικευμένες γνώσεις του ήταν σε θέση να αξιολογήσει τις διαθέσιμες πληροφορίες επάνω σε ένα ζήτημα και ύστερα από διαλογή με βάση κάποια πολύ συγκεκριμένα κριτήρια ανά περίπτωση, να **προτείνει** τις πληροφορίες αυτές οι οποίες **πίστευε** πως σχετίζονται καλύτερα με το ζήτημα που του έχει ανατεθεί. Έτσι, ένας άνθρωπος ο οποίος ενδιαφέρονταν για την αγορά ακίνητης περιουσίας, ο οποίος υπό άλλες συνθήκες θα έπρεπε να έχει γνώσεις αρχιτέκτονα, πολιτικού μηχανικού, να γνωρίζει την κατάσταση της αγοράς και πολλές άλλες γνώσεις οι οποίες θα αξιοποιούσε μία ή ίσως δύο φορές στον βίο του, απευθύνονταν σε έναν ειδήμονα, ο οποίος, καθώς επιδίδεται καθημερινά σε τέτοιες δραστηριότητες αποκτώντας εμπειρία, αναλάμβανε να βρει ένα κτίσμα ή γη με βάση τα κριτήρια που έθεσε ο πρώτος. Ή ίσως να ήθελε να παρακολουθήσει μία θεατρική παράσταση και να δυσκολεύονταν να επιλέξει μία μέσα από τις δεκάδες που ανέβαιναν στην σκηνή, οπότε θα απευθύνονταν σε έναν ειδήμονα προκειμένου να του προτείνει μία παράσταση η οποία πιστεύει πως θα του **αρέσει**.

Φυσικά, όσο καλή δουλειά κι αν έκαναν κάποτε οι ειδήμονες, ακόμα κι αυτοί θα ήταν αδύνατο να συμβαδίσουν με την έκρηξη παραγωγής πληροφοριών που έφερε ο ερχομός του

διαδικτύου. Οπότε μία νέα λύση έπρεπε να βρεθεί. Και βρέθηκε. Τα Συστήματα Προτάσεων (Recommendation Systems). Αυτοματοποιημένα συστήματα τα οποία αξιοποιώντας την επανάσταση που έφεραν οι υπολογιστές με τις ταχύτητες υπολογισμών τους καθώς και την αναβάθμιση του παγκόσμιου δικτύου με τεράστιες ταχύτητες μετάδοσης δεδομένων, μπορούν να συλλέξουν έναν πολύ μεγάλο όγκο πληροφοριών, να τον επεξεργαστούν και εφαρμόζοντας διάφορες τεχνικές διαλογής, αξιολόγησης και ταξινόμησης να παράξουν προτάσεις οι οποίες μπορεί να απευθύνονται **προσωπικά** σε ένα άτομο ή σε μία **ομάδα** ατόμων.

## 1.1 Ιστορικά στοιχεία

### 1.1.1 Σύστημα Grundy

Το πρώτο γνωστό σύστημα το οποίο αξιοποιεί τις προτιμήσεις των χρηστών προκειμένου να παράγει προτάσεις για αυτούς είναι ο ηλεκτρονικός βιβλιοθηκάριος Grundy. Το σύστημα αυτό βασιζόταν σε στερεότυπα (stereotypes), δηλαδή μία τεχνική που χρησιμοποιούν οι άνθρωποι προκειμένου γρήγορα να κατασκευάσουν ένα μοντέλο το οποίο πιστεύουν πως ανταποκρίνεται στον άνθρωπο για τον οποίο το κατασκευάζουν. Για να το πετύχουν αυτό, αξιολογούν κάποια χαρακτηριστικά όπως την ηλικία, το μορφωτικό επίπεδο, το επάγγελμα και τα χαρακτηριστικά προσωπικότητας προκειμένου να τον εντάξουν σε μία ομάδα με κοινά χαρακτηριστικά, τα οποία πιστεύουν πως διαθέτει και αυτός, μέχρι αποδείξεως του αντιθέτου. Ο τρόπος με τον οποίο υλοποιείται (implemented) αυτό στο σύστημα είναι μέσω κάποιων ερωτήσεων που θέτει στους χρήστες προκειμένου να τους «γνωρίσει» και να τους κατατάξει σε ένα ή περισσότερα στερεότυπα. (Rich, 1998)

Εκτός από την ικανότητά του να παράγει προτάσεις το Grundy είχε επίσης και την ικανότητα να προσαρμόζει τα στερεότυπά του. Γεγονός το οποίο είναι πολύ σημαντικό διότι με αυτόν τον τρόπο αντισταθμίζει την -σχεδόν- ολοκληρωτική έλλειψη πραγματικών δεδομένων πάνω στα οποία δημιουργήθηκαν τα στερεότυπα. Σύμφωνα με τα λόγια της Elaine Rich:

‘Οι υπολογιστές έχουν ένα σημαντικό πλεονέκτημα έναντι των ανθρώπων όσον αφορά τα στερεότυπα· δεν είναι συναισθηματικά δεσμευμένοι με αυτά και μπορούν να αλλάξουν τα στερεότυπα σύμφωνα με την εμπειρία’

(Rich, 1998)

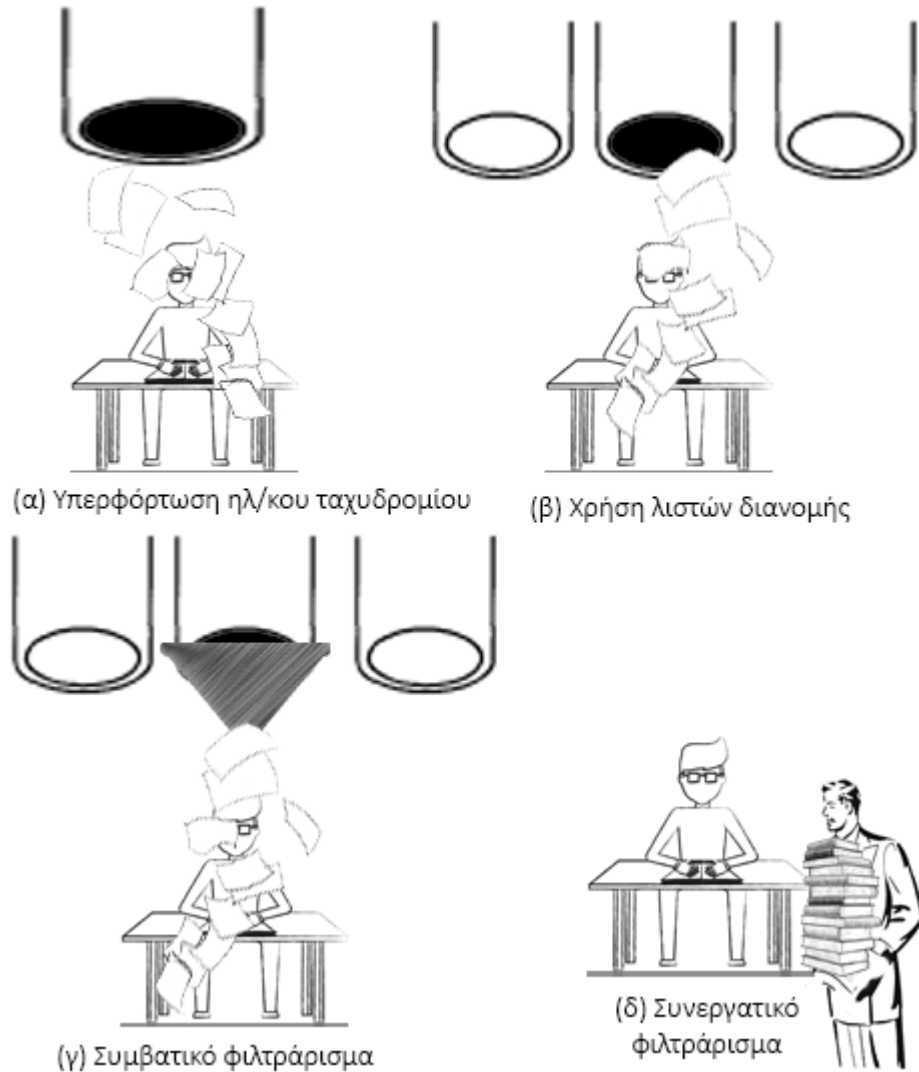
Το σύστημα αξιολογούσε τη συμπεριφορά του χρήστη προκειμένου να αποφανθεί εάν χρειάζεται να αλλάξει το στερεότυπο. Εάν ο χρήστης συμπεριφέρεται όπως ορίζει το στερεότυπο, οδηγεί σε επιβεβαίωση της ορθής πρόβλεψης που οδήγησε σε αυτό το στερεότυπο. Εάν από την άλλη ο χρήστης παρουσιάζει συμπεριφορά η οποία αντικρούει τη συμπεριφορά που θα έπρεπε να είχε με βάση το στερεότυπο, τότε η πρόβλεψη ήταν λανθασμένη.

### Πίνακας 1 Απόσπασμα αλληλεπίδρασης με το σύστημα

```
I'D LIKE TO KNOW WHAT SORT OF PERSON YOU THINK YOU ARE. PLEASE TYPE IN A
FEW SINGLE WORDS THAT YOU THINK CHARACTERIZE THE MOST IMPORTANT ASPECTS
OF YOURSELF
> unconventional open direct honest humomus persistent adventurous
HAVE YOU EVER READ EARTH ANGELS BY SUSAN CAHILL?
> no
YOU MIGHT LIKE IT. IT IS ABOUT A GIRL WHO GOES TO CATHOLIC ELEMENTARY
SCHOOL, HIGH SCHOOL. AND COLLEGE, THEN GOES ON TO BECOME A NUN, BUT FINALLY
BACKS OUT A FEW DAYS BEFORE FINAL VOWS [...] THE MAJOR STRENGTHS OF ONE OF
THE MAIN CHARACTERS IS INDEPENDENCE. IT IS POLITICALLY LIBERAL. IT HAS A
BIT OF ROMANCE. IT HAS A MEDIUM-FAST MOVING PLOT. DOES THAT SOUND GOOD?
> no
WAS IT BECAUSE OF THE DEPTH OF THE CHARACTER DEVELOPMENT? (EITHER TOO MUCH
OR NOT ENOUGH)?
WAS IT BECAUSE OF THE AMOUNT OF ROMANCE? (EITHER TOO MUCH OR NOT ENOUGH)?
```

#### 1.1.2 Το σύστημα συνεργατικού φιλτραρίσματος Tapestry

Το 1992, μέσω της επανάστασης που είχε ήδη φέρει τη ηλεκτρονικό μήνυμα (e-mail), οι εργαζόμενοι σε επιχειρήσεις κατακλύζονταν από έναν μεγάλο αριθμό εισερχομένων μηνυμάτων με αποτέλεσμα να επικρατεί αναρχία στο ηλεκτρονικό γραμματοκιβώτιό τους (mailbox). (Palme, 1984) Μία τάξη σε αυτό το πρόβλημα προσπάθησε να φέρει το σύστημα Tapestry το οποίο αναπτύχθηκε από το Ερευνητικό Κέντρο Xerox στο Πάλο Άλτο της Καλιφόρνια. Παρότι υπήρχαν ήδη κάποια συστήματα τα οποία ήταν ικανά να φιλτράρουν τα μηνύματα με βάση το περιεχόμενό τους (Lutz, Kleist-Retzow, & Hoernig, 1990) (Pollock, 1988) το Tapestry ήθελε να κάνει κάτι διαφορετικό· ήθελε να εντάξει τον άνθρωπο στη διαδικασία του φιλτραρίσματος. Το Tapestry σχεδιάστηκε και αναπτύχθηκε με σκοπό να υποστηρίζει συνεργατικό φιλτράρισμα (**collaborative filtering**) ώστε να μπορεί να φιλτράρει τα εισερχόμενα μηνύματα με βάση θέματα (topics) όπου ο χρήστης έχει ορίσει ως συναφή.



### Εικόνα 1. Συνεργατικό φιλτράρισμα

Το σύστημα επιτυγχάνει τους σκοπούς του επεξεργάζοντας την πληροφορία σε δύο βήματα. Αρχικά, φιλτράρει τον μεγάλο όγκο εισερχόμενων μηνυμάτων με βάση τις προτιμήσεις του χρήστη. Δεύτερον, προτεραιοποιεί τα μηνύματα αυτά με τέτοιο τρόπο ώστε τα περισσότερο σημαντικά να καθίστανται πιο εύκολα προσβάσιμα στον χρήστη. Ο τρόπος με τον οποίο το επιτυγχάνει αυτό είναι μέσω σχολιασμών (annotations). Ενθαρρύνει τους χρήστες να ανατροφοδοτήσουν (feedback) το σύστημα αφήνοντας σχόλια για το συγκεκριμένο μήνυμα. Αυτή η ανατροφοδότηση αποθηκεύεται στη βάση δεδομένων του Tapestry προκειμένου να μπορούν να χρησιμοποιηθεί από τα ερωτήματα (queries) που θέτει ο χρήστης στη διαδικασία φιλτραρίσματος. (Terry, 1993) (Goldberg, Nichols, Oki, & Douglas, 1992)

Το συνεργατικό φιλτράρισμα είναι μία τεχνική προκειμένου να ταιριάζει άτομα με παρόμοια ενδιαφέροντα και στη συνέχεια να παράξει προτάσεις πάνω σε αυτήν τη βάση. Υπάρχουν τρεις βασικοί πυλώνες προκειμένου το εγχείρημα να δουλέψει:

1. Πρέπει να υπάρχει μεγάλη συμμετοχή προκειμένου να αυξηθούν οι πιθανότητες να βρεθούν άτομα με παρόμοιες προτιμήσεις,
2. Πρέπει να υπάρχει ένας απλός και κατανοητός τρόπος οι χρήστες να αναπαριστούν τις προτιμήσεις τους στο σύστημα,
3. Και οι αλγόριθμοι πρέπει να είναι σε θέση να αντιστοιχούν άτομα με παρόμοιες προτιμήσεις.

Με τη χρήση του συνεργατικού φιλτραρίσματος η δουλειά ενός χρήστη απλοποιήθηκε σε μεγάλο βαθμό: εκφράζει τις προτιμήσεις του βαθμολογώντας στοιχεία τα οποία το σύστημα τού παρουσιάζει. Αυτές οι βαθμολογίες λογίζονται ως μια προσεγγιστική αναπαράσταση των προτιμήσεών του για ένα συγκεκριμένο θέμα. Το σύστημα, στη συνέχεια, αντιστοιχεί αυτές τις βαθμολογίες με βαθμολογίες που έχουν εισάγει άλλοι χρήστες. Το αποτέλεσμα αυτής της διαδικασίας είναι οι «κοντινότεροι γείτονες», δηλαδή ένας φορμαλιστικός τρόπος να αναφερθούμε στα άτομα τα οποία έχουν παρόμοιες προτιμήσεις με εμάς. Τέλος, το σύστημα παράγει προτάσεις με βάση τα στοιχεία που οι κοντινότεροι γείτονες έχουν βαθμολογήσει υψηλά και ταυτόχρονα εμείς δεν είμαστε γνώριμοι με αυτά. (Terveen & Hill, 2001)

### 1.1.3 Η ερευνητική ομάδα GroupLens

Λίγο ύστερα από τη δημιουργία του Tapestry, η ερευνητική ομάδα GroupLens με επικεφαλής τον Paul Resnick στο MIT υιοθέτησε τη βασική ιδέα και ανέπτυξε ένα πιο εφαρμόσιμο σύστημα προτάσεων με χρήση στις ομάδες ειδήσεων (newgroups<sup>1</sup>) του usenet.

Η GroupLens επέκτεινε τη βασική ιδέα του Tapestry με δύο τρόπους. Αρχικά, ενώ το Tapestry είναι ένα μονολιθικό σύστημα το οποίο είναι σχεδιασμένο να λειτουργεί εντός ενός εξυπηρετητή (server) (όπως για παράδειγμα στο πλαίσιο μιας επιχείρησης), το σύστημα βαθμολογιών σχεδιάστηκε έχοντας υπόψιν την επεκτασιμότητα. Αξιοποιώντας την

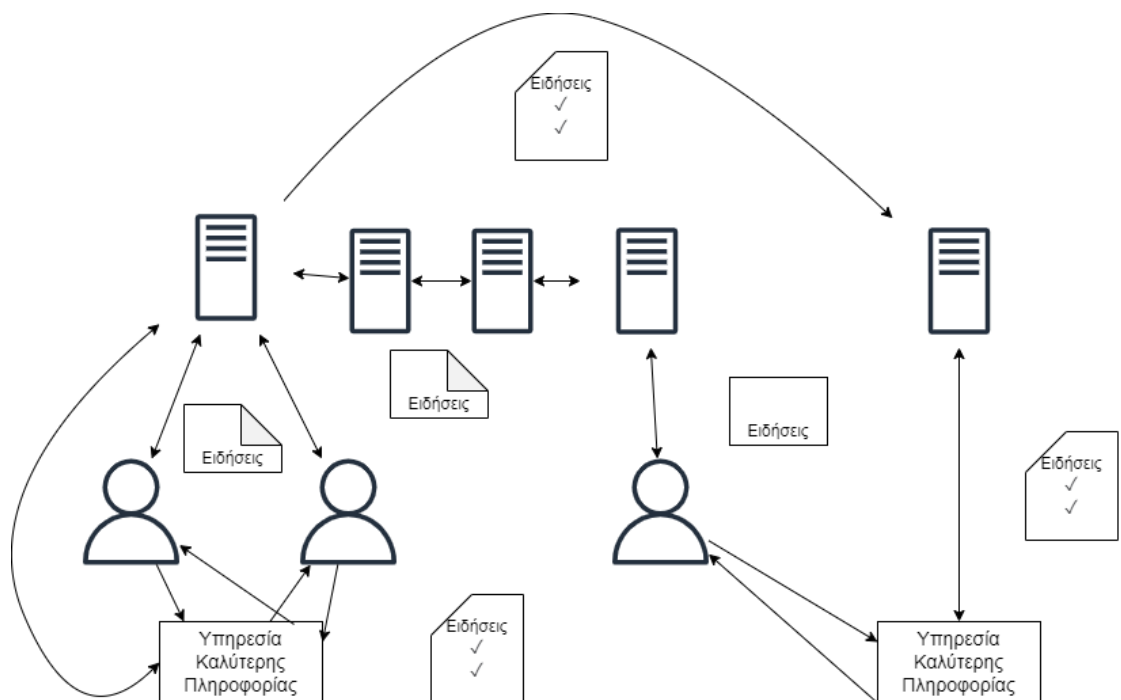
---

<sup>1</sup> Οι ομάδες ειδήσεων του Usenet είναι ομάδες διαλόγου στο διαδίκτυο όπου είναι δυνατός ο διαμοιρασμός πληροφοριών και απόψεων με ανθρώπους από όλον τον κόσμο. Στις ομάδες αυτές, υπάρχει δυνατότητα ανάρτησης απαντήσεων (replies) σε αναγνωσμένα άρθρα καθώς και να η ανάρτηση νέων άρθρων, διαθέσιμα προς ανάγνωση. Υπάρχει μεγάλη ποικιλία όσον αφορά τη θεματολογία και τη σοβαρότητα που διέπει το θέμα το οποίο συζητείται ανά ομάδα. (University Information Technology Services, n.d.)

αρχιτεκτονική αυτή, το σύστημα μπορούσε να περιλαμβάνει πολλές διαφορετικές σελίδες (sites), το οποίο έδωσε στην ομάδα τη δυνατότητα να εντάξει νέες ομάδες ειδήσεων στο ήδη υπάρχον σύνολο δεδομένων (dataset) απόρροια του οποίου ήταν να αυξηθεί το ποσοστό των ορθών προβλέψεων.

Το σύστημα που ανέπτυξε η GroupLens είχε ακόμη ένα θετικό στοιχείο: ήταν πολύ πιο ευέλικτο (flexible) από τους προκατόχους του. Με τους εξυπηρετητές που ανέπτυξαν προκειμένου να αποθηκεύουν τις βαθμολογίες, έδωσαν στο σύστημα τη δυνατότητα να συγκεντρώνει βαθμολογίες από διαφορετικές πηγές, λαμβάνοντας υπόψιν τις προηγούμενες βαθμολογίες ενός ατόμου καθώς και να αναγνωρίζει το μοτίβο βαθμολογιών της πηγής και βάση αυτού δημιουργούσε ένα νέο σύστημα βαθμολογιών. (Konstan, Riedl, Borchers, & Herlocker, 1998)

Η GroupLens εισήγαγε, επίσης, μία νέα οντότητα στην αρχιτεκτονική των διαδικτυακών ειδήσεων. Η οντότητα η οποία ανέπτυξε ονομάστηκε Υπηρεσία Καλύτερης Πληροφορίας (Better Bit Bureaus). Η Υπηρεσία παρείχε βαθμολογίες (scores) οι οποίες προέβλεπαν το ποσοστό αρεσκείας του χρήστη για το άρθρο που θα διάβαζε. Χρησιμοποιούσε ακόμη ειδικές ομάδες ειδήσεων προκειμένου να διαμοιράζονται οι βαθμολογίες (ratings) μεταξύ τους, ώστε να είναι εφικτό το συνεργατικό φιλτράρισμα μεταξύ χρηστών διαφορετικών σελίδων.



**Εικόνα 2. Η αρχιτεκτονική της GroupLens**

Οι βαθμολογίες στο σύστημα της GroupLens ήταν ένας αριθμός μεταξύ του ενός (1) και του πέντε (5). Οι χρήστες ενθαρρύνονταν να βαθμολογήσουν με βάση το επίπεδο αρεσκείας τους προς το άρθρο που διάβασαν, με το 5 να είναι η υψηλότερη βαθμολογία και να υποδηλώνει απόλυτη αρέσκεια και το 1 να είναι η χαμηλότερη βαθμολογία και να υποδηλώνει την απόλυτη δυσαρέσκεια.

Συνολικά, η θεμελιώδης λογική πάνω στην οποία βασίζεται το σύστημα της GroupLens είναι πολύ παρόμοιο με αυτήν που χρησιμοποιούν τα σύγχρονα συστήματα προτάσεων και αυτός είναι και ο λόγος που θεωρείται από πολλούς ως ο πρωτοπόρος των σύγχρονων συστημάτων προτάσεων.

#### **1.1.4 Η συνεργασία επιστημονικής κοινότητας & βιομηχανίας: Το Βραβείο Netflix**

Η ευρεία υιοθέτηση του διαδικτύου στις αρχές της δεύτερης χιλιετίας και η μεγάλη επιτυχία των διαδικτυακών αγορών (online retailing) έφερε πρόσφορο έδαφος στην αγορά για εμπορικές εφαρμογές μεγάλης κλίμακας. Με την τεράστια αύξηση των προϊόντων που προσφέρονταν προς πώληση διαδικτυακά και τον αριθμό των θαμώνων του διαδικτύου να μεγαλώνει, τα συστήματα προτάσεων έπαιξαν καθοριστικό ρόλο στον στρατηγικό σχεδιασμό εμπορικών σελίδων, όπως το Amazon.

Ο σκοπός τους [των επιχειρήσεων] ήταν να αντιστοιχήσουν τις προτιμήσεις των χρηστών με τα προϊόντα που προσέφεραν προκειμένου να αυξήσουν τις πωλήσεις τους. Για το λόγο αυτό χρησιμοποιήθηκαν συστήματα προτάσεων προκειμένου να γίνονται στοχευμένες προτάσεις στους πελάτες των διαδικτυακών καταστημάτων, αυξάνοντας έτσι την πιθανότητα μίας πώλησης.

Αν και η ευρεία διάδοση των συστημάτων προτάσεων οφείλεται στα εμπορικά συμφέροντα, η βελτίωσή τους προήλθε από την ακαδημαϊκή κοινότητα. Κανείς από τους δύο τομείς δεν ήταν σε θέση να σημειώσει αξιόλογη πρόοδο μόνος του χωρίς την αρωγή του άλλου. Ο λόγος ήταν πως, ενώ οι ερευνητές ανακάλυπταν νέους τρόπους προκειμένου να βελτιώσουν τα συστήματα προτάσεων, προκειμένου να δοκιμάσουν τις μεθόδους τους και να αξιολογήσουν τα αποτελέσματα, χρειάζονταν ογκώδη σύνολα δεδομένων, παραγόμενα από τον πραγματικό κόσμο και όχι το εργαστήριο· δεδομένα τα οποία δε διέθεταν και μόνον η βιομηχανία θα μπορούσε να τους παρέχει.

Μία από τις σημαντικότερες προσπάθειες που έγιναν για τη βελτίωση των συστημάτων προτάσεων προέκυψε από τη συνεργασία της βιομηχανίας, μίας ιδιωτικής επιχείρησης και της ακαδημαϊκής κοινότητας. Το 2016, το Netflix διοργάνωσε έναν



διαγωνισμό με όνομα «Το Βραβείο Netflix» (Netflix Prize). Το έπαθλο του διαγωνισμού ανέρχονταν στα ένα εκατομμύριο αμερικάνικα δολάρια και νικητής θα ήταν το άτομο ή η ομάδα η οποία θα κατάφερνε να αναπτύξει ένα σύστημα το οποίο θα είχε μεγαλύτερη ακρίβεια στις προτάσεις σε σχέση με τον αλγόριθμο του Netflix. Οι νικητές ήταν υποχρεωμένοι να καταγράψουν και να δημοσιεύσουν τα ευρήματα και τις μεθόδους που χρησιμοποίησαν, δίνοντας έτσι την ευκαιρία στον καθένα να τα μελετήσει και να ωφεληθεί από τις γνώσεις και τις τεχνικές που απαιτούνται προκειμένου να επιτευχθεί αυτό το επίπεδο ακρίβειας. (Bennett & Lanning, 2007)

Για τους σκοπούς του διαγωνισμού, το Netflix παρείχε ένα πολύ μεγάλο σύνολο δεδομένων αποτελούμενο από περισσότερες από εκατό εκατομμύρια βαθμολογίες (ratings) από περισσότερους από τετρακόσιους ογδόντα χιλιάδες τυχαία επιλεγμένους συνδρομητές για πάνω από δεκαοχτώ χιλιάδες ταινίες. Το μεγάλο μέγεθος του συνόλου δεδομένων, έδωσε μία μοναδική -για τότε- ευκαιρία στην ακαδημαϊκή κοινότητα να δοκιμάσουν τα εργαστηριακά ανεπτυγμένα συστήματά τους με δεδομένα του πραγματικού κόσμου, κάτι το οποίο δεν ήταν πριν εφικτό.

Ο διαγωνισμός διήρκησε τρία χρόνια και οι συμμετέχοντες μπορούσαν να δοκιμάσουν την αποτελεσματικότητα των συστημάτων που ανέπτυξαν χρησιμοποιώντας το σύνολο δεδομένων που παρείχε το Netflix. Τον Σεπτέμβριο του 2019 το Netflix ανακοίνωσε πως η ομάδα «Bellkor's Pragmatic Chaos» στέφθηκε νικήτρια, με το σύστημά της να πετυχαίνει μία βελτίωση της τάξης του 10%. Ωστόσο, η εταιρία τελικά δεν αξιοποίησε το νικητήριο σύστημα καθώς ύστερα από δοκιμές που πραγματοποίησαν εσωτερικά, κατέληξαν στο συμπέρασμα που οι βελτιώσεις στην ακρίβεια πρόβλεψης που θα επιτύγχαναν δεν δικαιολογούσαν την τεράστια προσπάθεια που θα απαιτούνταν προκειμένου να εντάξουν το σύστημα στην πλατφόρμα. Κύριο ρόλο κατείχε επίσης και το γεγονός πως ύστερα από τρία χρόνια που χρειάστηκαν προκειμένου να ολοκληρωθεί ο διαγωνισμός, η εταιρία είχε αποφασίσει να προχωρήσει στο «επόμενο επίπεδο», το οποίο ήταν η μετάβαση από την ταχυδρόμηση DVD στο σημερινό επιχειρηματικό μοντέλο της streaming πλατφόρμας. (Amatriain, 2012)

Το 2010 ανακοινώθηκε ο διάδοχος του διαγωνισμού, ο οποίος, ωστόσο, ακυρώθηκε, ύστερα από μία μαζική αγωγή κατά του Netflix βασιζόμενη σε μία δημοσίευση από το Πανεπιστήμιο του Τέξας, όπου οι συγγραφείς έδειξαν πως το ανωνυμοποιημένο σύνολο δεδομένων που είχε δοθεί από το Netflix μπορούσε τελικά να ταυτοποιήσει τους χρήστες. (Narayanan & Shmatikov, 2006)

## 1.2 Σκοπός της εργασίας

Παρότι τα συστήματα προτάσεων βρίσκονται σχεδόν πίσω από κάθε εμπορική δραστηριότητα στο διαδίκτυο η οποία προσφέρει ποικιλία προϊόντων κάθε μορφής σε χρήστες, συνήθως δίνεται έμφαση μόνο στις ατομικές προτάσεις και όχι σε ομαδικές. Για παράδειγμα, το Netflix, το Amazon Prime και το Disney Plus, τρεις από τις δημοφιλέστερες πλατφόρμες streaming περιεχομένου, διαθέτουν όπως είναι αναμενόμενο συστήματα προτάσεων για εξατομικευμένα προφίλ ενός ατόμου, δεν παρέχουν ωστόσο κάποια μέθοδο ομαδικών προτάσεων. Η παρούσα εργασία αποπειράθηκε να βρει μία προσεγγιστική λύση στο πρόβλημα της χρονοβόρας αναζήτησης επιθυμητού περιεχομένου προς θέαση όταν αποδέκτες είναι παραπάνω του ενός ατόμου. Έγινε προσπάθεια ανάπτυξης ενός συστήματος το οποίο, αφού πρώτα παράξει εξατομικευμένες προτάσεις ταινιών, τις συνδυάζει και ακολούθως προτείνει τις ταινίες αυτές οι οποίες έχουν τις περισσότερες πιθανότητες να αρέσουν και στα δύο άτομα που συμμετέχουν στην διαδικασία.

## Κεφάλαιο 2. Ανασκόπηση βιβλιογραφίας

Στο κεφάλαιο αυτό παρουσιάζονται παραδείγματα συστημάτων προτάσεων που χρησιμοποιούνται από κολοσσούς στη βιομηχανία και που η έρευνά τους στον τομέα αυτόν έφερε μεγάλες αλλαγές στην ποιότητα των παραγόμενων προτάσεων.

### 2.1 Το σύστημα που άλλαξε τις ηλεκτρονικές αγορές:

#### Amazon.com

Είναι αδύνατο για μία εργασία η οποία τοποθετεί τον εαυτό της στον τομέα των συστημάτων προτάσεων να μην κάνει ιδιαίτερη μνεία στο, ίσως μεγαλύτερο, ηλεκτρονικό κατάστημα αγορών που η έρευνά του για την επίτευξη προτάσεων μεγάλης ακρίβειας διεύρυνε τους ορίζοντες και χρησιμοποιήθηκε από πολλούς.

Σχεδόν δύο δεκαετίες έχουν περάσει από όταν το Amazon ξεκίνησε να χρησιμοποιεί συστήματα προτάσεων, εξυπηρετώντας εκατομμύρια ανθρώπους, βοηθώντας τους να ανακαλύψουν προϊόντα τα οποία, πιθανότατα, δε θα είχαν βρει με προσωπική αναζήτηση. Κάθε επισκέπτης της σελίδας έχει μία μοναδική εμπειρία καθότι του παρουσιάζονται αντικείμενα που αντιστοιχούν στα ενδιαφέροντά του. Από την βιτρίνα (αρχική σελίδα) του ηλεκτρονικού καταστήματος μέχρι την ολοκλήρωση της αγοράς, ο πελάτης βιώνει μία εξατομικευμένη εμπειρία. Από έναν κατάλογο εκατοντάδων εκατομμυρίων προϊόντων, το σύστημα προτάσεων του Amazon επιλέγει έναν μικρό αριθμό προϊόντων που κρίνει ότι **μάλλον** αντιστοιχούν στα ενδιαφέροντα του πελάτη, συνδυάζοντάς τα με προϊόντα που άλλοι έχουν ήδη ανακαλύψει.

Θα ήταν αναμενόμενο ο αλγόριθμος πίσω από το σύστημα του ηλεκτρονικού καταστήματος να προσπαθεί να προτείνει προϊόντα σε ένα πελάτη προσπαθώντας να αντιστοιχήσει το ατομικό του προφίλ με προφίλ άλλων πελατών που μοιάζουν μεταξύ τους. Αντ' αυτού, προσπαθεί να βρει συσχετίσεις μεταξύ των ίδιων των προϊόντων. Μία διαίσθηση του αλγορίθμου είναι η ακόλουθη: (Linden, Smith, & York, 2003)

Για κάθε προϊόν στον κατάλογο προϊόντων  $I_1$

Για κάθε πελάτη  $C$  που αγόρασε το προϊόν  $I_1$

Για κάθε προϊόν  $I_2$  που αγοράστηκε από τον πελάτη  $C$

Κατέγραψε πως ένας πελάτης αγόρασε τα προϊόντα  $I_1$

και  $I_2$

Για κάθε προϊόν  $I_2$

Υπολόγισε την ομοιότητα μεταξύ των προϊόντων  $I_1$  και

$I_2$

Για πολύ μεγάλα σύνολα δεδομένων, ένας επεκτάσιμος (scalable) αλγόριθμος προτάσεων πρέπει να εκτελεί τους πιο χρονοβόρους υπολογισμούς σε μη πραγματικό χρόνο (offline), συνθήκες τις οποίες ο αλγόριθμος του Amazon ικανοποιεί. Ιδιαίτερα, καταφέρνει να διαχειριστεί εκατοντάδες εκατομμύρια χρήστες και δεκάδες χιλιάδες εκατομμύρια προϊόντων χωρίς να εφαρμόζει τεχνικές μείωσης του συνόλου δεδομένων οι οποίες θα επέφεραν μείωση στην ποιότητα των προτάσεων.

### Συχνά αγοράζονται μαζί



**Εικόνα 3. Το Amazon μας προτείνει προϊόντα που άλλοι πελάτες έχουν αγοράσει μαζί με την αγορά του πρώτου**

Ωστόσο το Amazon μέσω της πολυετούς του έρευνας έχει παρατηρήσει πως, για κάθε ζεύγος προϊόντων  $X$  και  $Y$ , οι πελάτες που έχουν ήδη προβεί στην αγορά του προϊόντος  $X$  είναι πολύ πιθανότερο να αγοράσουν το προϊόν  $Y$  παρότι ο γενικός πληθυσμός. Ο λόγος είναι πως ένας πελάτης με πολύ μεγάλο ιστορικό αγορών είναι πολύ πιθανότερο να έχει ήδη αγοράσει το προϊόν  $X$  οπότε κάποια στιγμή στο μέλλον θα προβεί και στην αγορά του προϊόντος  $Y$ . Για παράδειγμα, ένας πελάτης που έχει πραγματοποιήσει 1000 αγορές είναι

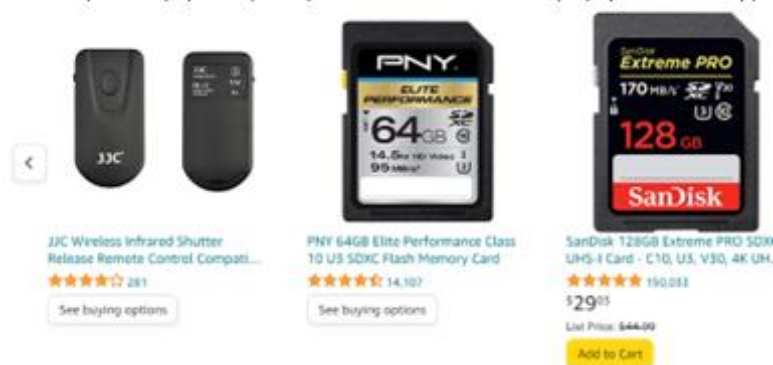
πολύ πιθανότερο να επιλεγθεί από έναν με 20 αγορές, κατά έναν συντελεστή x50. Για τον λόγο αυτό, το δείγμα που χρησιμοποιείται δεν είναι αμερόληπτο και αυτή η ανομοιόμορφη κατανομή οδηγεί στο συμπέρασμα πως κάθε αγορασμένο προϊόν X είναι και μία ξεχωριστή ευκαιρία να έχει αγοραστεί και το προϊόν Y. (Smith & Linden, 2017)

Σύμφωνα με τους Brent Smith και Greg Linden:

‘Ένα σύστημα προτάσεων ουσιαστικά είναι μία εφαρμογή της στατιστικής επιστήμης. Η ανθρώπινη συμπεριφορά είναι δύσκολο να καταγραφεί με ακρίβεια (noisy) και η πρόκληση σε αυτό το εγχείρημα είναι η ανακάλυψη χρήσιμων μοτίβων μέσα στην τυχαιότητα [της ανθρώπινης συμπεριφοράς].’

Μια ακόμη παρατήρηση που έγινε είναι πως δεν υπάρχει μία καθολική βαθμολογία (score) η οποία δουλεύει σωστά σε όλες τις συνθήκες. Με χρήση της μηχανικής μάθησης, ο αλγόριθμος του Amazon είναι σε θέση να μετρήσει, όχι μόνο το είδος των προτάσεων που είναι αποτελεσματικές, αλλά και να λάβει υπόψιν του ποιες προτάσεις άρεσαν στους πελάτες και ποιες από αυτές επέλεξαν (clicked), στοιχεία τα οποία ανατροφοδοτούν τον αλγόριθμο και τον βελτιώνουν. Βασιζόμενοι σε αυτό, οι μηχανικοί του Amazon κατέληξαν στο συμπέρασμα πως παρέχοντας αρκετά δεδομένα και μία σωστή μετρική για τη μέτρηση της συσχέτισης των προϊόντων, μπορεί να προκύψει συμβατότητα μεταξύ των ατόμων με βάση τη συμπεριφορά τους, με τα λανθάνοντα σήματα να εξαφανίζονται και τα κατάλληλα προϊόντα να αναδύονται.

Πελάτες που αγόρασαν τα προϊόντα στο καλάθι σας αγόρασαν επίσης



**Εικόνα 4. Έχοντας στο καλάθι μας μία φωτογραφική μηχανή το Amazon μας προτείνει προϊόντα τα οποία είναι συναφή με αυτή.**

Ο κατάλογος του Amazon συνεχώς επεκτείνεται με νέα προϊόντα, τα οποία λόγω ελλιπών δεδομένων δεν μπορούν να συσχετισθούν με άλλα. Το πρόβλημα αυτό είναι γνωστό ως cold-start problem. Το ίδιο πρόβλημα αντιμετωπίζουν και οι νέοι πελάτες του ηλεκτρονικού καταστήματος καθώς το σύστημα δεν διαθέτει αρκετές πληροφορίες για

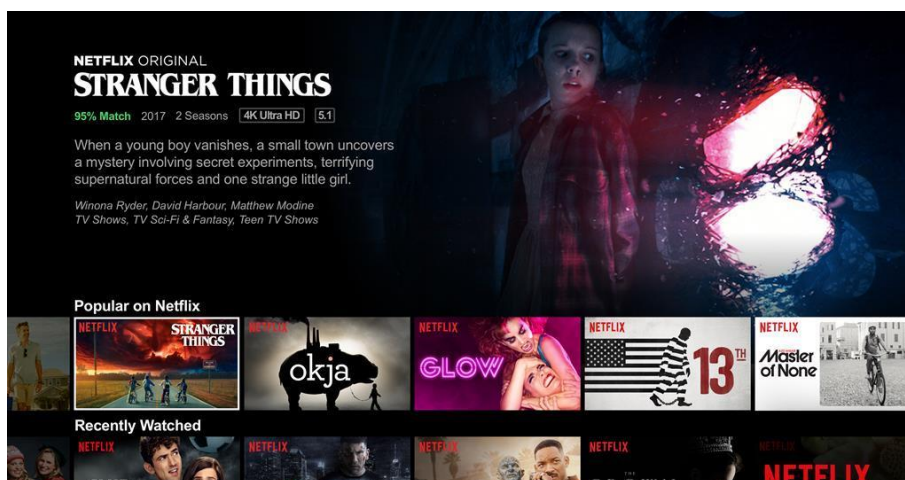
αυτούς προκειμένου να παράξει ακριβής προτάσεις που να ανταποκρίνονται στις προτιμήσεις τους. Σε αυτήν την περίπτωση χρειάζεται πολύ μεγάλη προσπάθεια για να βρεθεί η σωστή ισορροπία μεταξύ εξατομικευμένων προτάσεων και των ευρέως δημοφιλών προϊόντων.

Λόγο της απλότητας, της επεκτασιμότητας, της εξηγηματικότητας του λόγου επιλογής των προτάσεων, της προσαρμοστικότητας και της σχετικά υψηλής ποιότητας προτάσεων, ο αλγόριθμος προτάσεων βασισμένων σε αντικείμενα (item-based collaborative filtering) παραμένει ένας από τους πιο δημοφιλείς αλγορίθμους.

## **2.2 Netflix: Τα θεμέλια της σύγχρονης οπτικοακουστικής ψυχαγωγίας**

Η εφεύρεση της τηλεόρασης έφερε μαζί της μία τεράστια αλλαγή στον τρόπο που οι ιστορίες διηγούνται: τη χρήση του βίντεο. Η διανομή του οπτικοακουστικού υλικού πέρασε από πολλά στάδια μέχρι να φτάσει στη σημερινή μορφή του, όπου η χρήση του διαδικτύου είναι προτιμητέα μέθοδος ψυχαγωγίας του κοινού. (Richter, 2019) Η τεράστια επανάσταση που προκάλεσε η χρήση του διαδικτύου είναι η επιλογή που έδωσε στους χρήστες να διαλέγουν οι ίδιοι το υλικό που επιθυμούν να παρακολουθήσουν, σε αντίθεση με τη συμβατική τηλεόραση όπου καθένας μπορεί να παρακολουθήσει μονάχα ότι προβάλλεται εκείνη τη στιγμή στο πρόγραμμα του εκάστοτε καναλιού. Όμως οι άνθρωποι δυσκολεύονται πολύ να επιλέξουν όταν πολλές διαφορετικές επιλογές είναι διαθέσιμες. Ένας από τους βασικούς πυλώνες του Netflix είναι να βοηθά τους συνδρομητές του να βρουν υλικό που θέλουν να παρακολουθήσουν. Ο τρόπος που η πλατφόρμα το πετυχαίνει αυτό είναι με τη χρήση ενός συστήματος προτάσεων, το οποίο δεν αποτελείται από ένα μοναδικό αλγόριθμο, παρά από μία συλλογή αλγορίθμων, καθένας από τους οποίους εξυπηρετεί διαφορετικό σκοπό.

Ιστορικά για το Netflix το πρόβλημα παραγωγής ακριβών προτάσεων ήταν ισοδύναμο με την πρόβλεψη της βαθμολογίας που θα έδινε ένας χρήστης στο υλικό που παρακολούθησε. Αξιοποιώντας τη σημερινή τεχνολογία, είναι σε θέση να γνωρίζει πολλές παραπάνω πληροφορίες πέραν της βαθμολογίας, όπως για παράδειγμα το υλικό που παρακολουθεί κάθε χρήστης της πλατφόρμας, τις συνθήκες που επικρατούν (σε ποια συσκευή παρακολουθεί, ποια μέρα της εβδομάδας και ποια ώρα της ημέρας), τον τρόπο με τον οποίο ο χρήστης ανακάλυψε το υλικό, μέχρι και προτάσεις οι οποίες εμφανίστηκαν αλλά δεν επιλέχθηκαν.



**Εικόνα 5. Η αρχική σελίδα του Netflix. Το περιεχόμενο της πρώτης γραμμής προέρχεται λαμβάνοντας υπόψιν την γενική δημοτικότητα ενός υλικού αλλά και τις προσωπικές προτιμήσεις του χρήστη**

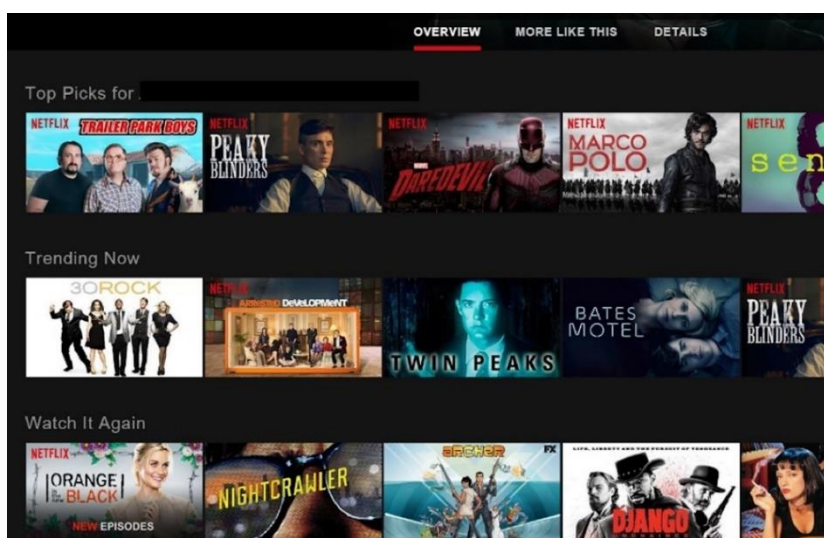
Το γραφικό περιβάλλον του Netflix αποτελείται από γραμμές (rows), κάθε μία από τις οποίες αναπαριστά μία έννοια (π.χ. δημοφιλείς ταινίες), ένα ίδιο (genre) υλικού (π.χ. δράση) ή μία κατηγορία (π.χ. πρωτότυπες παραγωγές του Netflix). Το υλικό που παρουσιάζεται σε κάθε γραμμή προέρχεται από τον αλγόριθμο «Ταξινομητής Προτεινόμενων Βίντεο» (Personalized Video Ranker), ο οποίος ορίζει τη σειρά (order) παρουσίασης του υλικού, εξατομικευμένα για κάθε προφίλ χρήστη και είναι ο λόγος που το ίδιο είδος υλικού μπορεί να έχει εντελώς διαφορετικές προτάσεις ανά χρήστη (Gomez-Uribe & Hunt, 2015).

Πέραν των διαφόρων ειδών, το Netflix αφιερώνει και μία γραμμή για την παρουσίαση υλικού για τις κορυφαίες προτάσεις (Top Picks) ανά χρήστη. Ο αλγόριθμος «Top N» που η πλατφόρμα χρησιμοποιεί είναι βελτιστοποιημένος προκειμένου να συνδυάζει την εξατομίκευση κάθε χρήστη με τη δημοτικότητα κάθε υλικού καθώς και λαμβάνει υπόψιν τις τάσεις του υλικού που παρακολουθείται μέσα από μία πληθώρα χρονικών περιόδων, από μία μέρα μέχρι ένα χρόνο, προκειμένου να ανακαλύψει ένα περιορισμένο πλήθος προτάσεων, που είναι οι βέλτιστες για κάθε χρήστη.

Εκτός από τις εξατομικευμένες προτάσεις κάθε χρήστη παρουσιάζεται και υλικό το οποίο είναι γενικά δημοφιλές αυτήν την χρονική περίοδο. Το Netflix κάνει διάκριση μεταξύ δύο κατηγοριών δημοφιλούς υλικού:

- Επαναλαμβανόμενα γεγονότα,
- Μοναδικά γεγονότα

Στην πρώτη κατηγορία ανήκουν γεγονότα όπως ετήσιες γιορτές και επέτειοι. Για παράδειγμα, τη χρονική περίοδο της γιορτής του Αγίου Βαλεντίνου, παρουσιάζεται μία άνοδος στη θέαση ρομαντικού υλικού. Η δεύτερη κατηγορία συμπεριλαμβάνει γεγονότα όπως η επικείμενη καταστροφή από έναν μεγάλο τυφώνα, γεγονός το οποίο καλύπτεται σε μεγάλη κλίμακα από όλα τα μέσα ενημέρωσης. Ένα τέτοιο γεγονός θα προκαλούσε τη βραχυπρόθεσμη άνοδο ενδιαφέροντος για υλικό σχετιζόμενο με φυσικές καταστροφές. Το Netflix λαμβάνει υπόψιν γεγονότα σαν κι αυτά, με τη βοήθεια του «Ταξινομητή Δημοφιλών Γεγονότων» (Trending ranker) και με τη σωστή δόση εξατομικευμένων προτάσεων.



**Εικόνα 6. Επιπλέον γραμμές της αρχικής σελίδας. Σε αυτήν την περίπτωση η γραμμή «Κορυφαίες επιλογές» είναι μια άκρως εξατομικευμένη γραμμή που περιλαμβάνει προτάσεις που το σύστημα πιστεύει πως θα αρέσουν στο χρήστη**

Μέχρι στιγμής, όλοι οι αλγόριθμοι χρησιμοποιούν συναγόμενες πληροφορίες προκειμένου να προτείνουν υλικό που ο θεατής δεν έχει παρακολουθήσει. Υπάρχει όμως και το ενδεχόμενο ο θεατής να ξεκινήσει τη θέαση ενός υλικού και να μην την ολοκληρώσει ποτέ, γεγονός που μπορεί να οφείλεται σε πλειάδα λόγων. Το Netflix διαθέτει έναν μηχανισμό του οποίου στόχος είναι να ταξινομήσει μόνο το υποσύνολο αυτό, βασισμένο στις προβλέψεις του αλγόριθμου για το αν ο θεατής έχει σκοπό να συνεχίσει τη θέαση του συγκεκριμένου υλικού ή όχι, γεγονός που μάλλον υποδηλώνει πως το συγκεκριμένο υλικό δεν ήταν τόσο ενδιαφέρον για το χρήστη όσο αναμένονταν. Για να πραγματοποιηθεί αυτή η πρόβλεψη, ο μηχανισμός χρησιμοποιεί διάφορες πληροφορίες, όπως ο χρόνος που έχει μεσολαβήσει από την θέαση του υλικού, το σημείο που η θέαση σταμάτησε (αρχή, μέση του υλικού κτλ.), εάν έχει γίνει θέαση παρόμοιου υλικού από εκείνη τη χρονική στιγμή μέχρι τώρα, καθώς και η συσκευή στην οποία πραγματοποιήθηκε η θέαση.



Όπως το Amazon, έτσι και το Netflix χρησιμοποιεί έναν αλγόριθμο που σχετίζεται με τα προϊόντα (στην προκειμένη περίπτωση, το οπτικοακουστικό υλικό). Ο συγκεκριμένος αλγόριθμος λαμβάνει υπόψιν προηγούμενες θεάσεις προκειμένου να δημιουργήσει προτάσεις για μια γραμμή με τον τίτλο «Επειδή είδατε» (Because You Watched). Παρότι ο αλγόριθμος αυτός καθ' αυτός δεν δημιουργεί εξατομικευμένες προτάσεις, το ποιο σύνολο υλικού θα υπάρχει εν τέλει στη γραμμή είναι εξατομικευμένο.

Γνωρίζοντας πως ένας λογαριασμός μπορεί να χρησιμοποιηθεί από πολλά μέλη μιας οικίας, το Netflix χρησιμοποιεί όλους τους αλγόριθμους που αναφέρθηκαν, μαζί με άλλες τεχνικές και αλγορίθμους, προκειμένου να παράξει μία εξατομικευμένη αρχική σελίδα για τον καθένα. Λόγο του τεράστιου όγκου υπολογισμών που χρειάζεται ένα τέτοιο εγχείρημα, η εταιρία χρησιμοποιεί έναν προσωποποιημένο (personalized) και βασισμένο στην επιστήμη των μαθηματικών, αλγόριθμο, ο οποίος είναι σε θέση να επιλέξει και να ταξινομήσει πολλές παραγόμενες γραμμές (rows) από μία μεγάλη δεξαμενή υποψήφιων γραμμών, προκειμένου να κατασκευάσει μία σειρά (order) βελτιστοποιημένη ως προς την σχετικότητα (relevance) με το προφίλ του χρήστη αλλά ταυτόχρονα να προσφέρει και μία ποικιλία.

## 2.3 YouTube

Ως η μεγαλύτερη και γνωστότερη κοινότητα παροχής βίντεο παρεχόμενο από τους ίδιους τους χρήστες, το YouTube παρουσιάζει κάποιες μοναδικές ευκαιρίες και δυσκολίες όσον αφορά την ανακάλυψη νέου περιεχομένου. Με πάνω από 500 ώρες νέου περιεχομένου να ανεβαίνει στην πλατφόρμα κάθε λεπτό (YouTube Press, n.d.) είναι φανερό πως δίχως ένα σύστημα προτάσεων, μεγάλο μέρος από αυτά τα βίντεο δε θα ανακαλύπτονταν ποτέ.

Το σύστημα προτάσεων του YouTube έχει περάσει από διάφορα στάδια και συνεχώς εξελίσσεται. Ο σκοπός του, ωστόσο, παραμένει ο ίδιος: να παρέχει εξατομικευμένες προτάσεις που βοηθούν τους χρήστες να βρουν βίντεο υψηλής ποιότητας που να ανταποκρίνονται στις προτιμήσεις τους.

Το σύστημα διακρίνει δύο μεγάλες κατηγορίες δεδομένων που δέχεται ως είσοδο:

- Δεδομένα σχετικά με το περιεχόμενο
- Δεδομένα σχετικά με τη συμπεριφορά του χρήστη

Στην πρώτη κατηγορία ανήκουν δεδομένα όπως το ίδιο το οπτικοακουστικό υλικό, ο τίτλος, η περιγραφή και η διάρκεια. Η δεύτερη κατηγορία περιλαμβάνει ενέργειες που εκτελεί ο χρήστης ενόσω βρίσκεται στην πλατφόρμα, όπως το να δηλώσει ότι του αρέσει ένα βίντεο πατώντας το κουμπί «Μου αρέσει», να μοιραστεί στο βίντεο, να κάνει εγγραφή στο

κανάλι του δημιουργού, ακόμη και ότι ο χρήστης επέλεξε να δει ένα βίντεο και είδε ένα μεγάλο μέρος του, αλλά όχι ολόκληρο.

Θα περίμενε κανείς πως το γεγονός πως ένας χρήστης παρακολούθησε ολόκληρη τη διάρκεια του βίντεο θα σήμαινε πως του άρεσε ή τουλάχιστον ότι δεν αποτέλεσε δυσάρεστη εμπειρία. Ωστόσο η διάρκεια θέασης δεν είναι από μόνη της ικανή συνθήκη προκειμένου να εξαχθεί το συμπέρασμα εάν το περιεχόμενο άρεσε στον χρήστη ή όχι.

Προτού συμπεριλάβει αλγόριθμους μηχανικής μάθησης στο σύστημά του, το YouTube κατασκεύαζε ένα σύνολο βίντεο  $R_i$  όπου δεδομένου ενός βίντεο  $u_i$ , ο χρήστης ήταν πιθανό να δει. Προκειμένου να το επιτύχει αυτό, το σύστημα, για κάθε ζεύγος βίντεο  $(u_i, v_j)$  μετρούσε, μέσα σε μία ορισμένη χρονική διάρκεια (συνήθως 24 ωρών), πόσο συχνά έγινε θέαση και των δύο μελών του ζεύγους σε μία συνεδρία. Στην εξίσωση υπολογισμού χρησιμοποιήθηκε και μία μέθοδος κανονικοποίησης προκειμένου βίντεο με λιγότερη απήχηση να είναι περισσότερο ευνοημένα. (Davidson, Liebald, Liu, Nandy, & Van Vleet, 2010). Από το σύνολο δεδομένων που προκύπτει, γίνεται διαλογή των  $N$  κορυφαίων (top  $N$ ) βίντεο, εφαρμόζοντας μία κατώτατη βαθμολογία που το ζεύγος θα πρέπει να συγκεντρώνει.

Η λίστα που προκύπτει ύστερα από τη διαλογή συνδυάζεται με την δραστηριότητα του κάθε χρήστη, τόσο άμεση (όπως το πάτημα του κουμπιού «Μου αρέσει») όσο και έμμεση (όπως η διάρκεια θέασης ενός βίντεο). Στο τελικό αυτό σύνολο, εφαρμόζεται μία κατάταξη τριών σταδίων:

1. Ποιότητα του βίντεο
2. Ακρίβεια πρότασης για τον χρήστη
3. Διαφοροποίηση

Η ποιότητα, στην οποία συγκαταλέγονται στοιχεία όπως ο αριθμός προβολών και το κατά πόσο οι θεατές αφήνουν σχόλια είναι μία γενική μετρική που δεν επηρεάζεται από το προφίλ κάθε χρήστη. Η δεύτερη κατηγορία χρησιμοποιείται προκειμένου να κατατάξει ψηλότερα τα βίντεο εκείνα που ανταποκρίνονται στις ατομικές προτιμήσεις του χρήστη. Η τελική λίστα με τις προτάσεις που προκύπτει είναι ισορροπημένη μεταξύ συνάφειας με τις προτιμήσεις του χρήστη και ποικιλίας.

Με την εξέλιξη της τεχνολογίας επήλθαν αλλαγές και στο σύστημα προτάσεων του YouTube, όπου νευρωνικά δίκτυα χρησιμοποιήθηκαν για την παραγωγή προτάσεων υψηλής ακρίβειας. Υλοποιήθηκαν δύο νευρωνικά δίκτυα, ένα για την παραγωγή υποψήφιων (candidate generation) και ένα για κατάταξη (ranking). Το πρώτο νευρωνικό δίκτυο δέχεται ως είσοδο τις ενέργειες του χρήστη στην πλατφόρμα και ανακτά ένα μικρό υποσύνολο εκατοντάδων βίντεο, με την πρόθεση να έχουν ιδιαίτερα υψηλή συνάφεια με τις προτιμήσεις

του χρήστη. Για την παραγωγή των προτάσεων το δίκτυο κάνει χρήση του συνεργατικού φιλτραρίσματος. Η ομοιότητα μεταξύ των χρηστών μετράτε συναρτήσεως συγκεκριμένων χαρακτηριστικών, όπως τα βίντεο που παρακολούθησαν, τις αναζητήσεις τους καθώς και δημογραφικά στοιχεία.

Το δεύτερο νευρωνικό δίκτυο επιτυγχάνει τον σκοπό του αναθέτοντας μία βαθμολογία (score) σε κάθε βίντεο λαμβάνοντας υπόψη ένα πλούσιο σύνολο από χαρακτηριστικά τα οποία περιγράφουν το βίντεο και τον χρήστη που τα παρακολουθεί. Τα βίντεο με την υψηλότερη βαθμολογία παρουσιάζονται στον χρήστη. (Covington, Adams, & Emre, 2016)

Η εξέλιξη της τεχνολογίας δεν ήταν ο μόνος παράγοντας που ώθησε το YouTube σε αλλαγές· οι ίδιοι οι χρήστες της πλατφόρμας έπαιξαν σημαντικό ρόλο. Οι λόγοι για τους οποίους επισκέπτεται κάποιος τη σελίδα έχουν αλλάξει πολύ από τις πρώτες ημέρες ζωής του. Σήμερα, μέσα από την πληθώρα βίντεο που διαθέτει η πλατφόρμα, ένας επισκέπτης μπορεί όχι μόνο να ψυχαγωγηθεί αλλά και να ενημερωθεί. Το οποίο έφερε νέες προκλήσεις στο σύστημα προτάσεων. Η εύκολη πρόσβαση σε τεχνολογία καταγραφής και δημιουργίας βίντεο οδήγησε σε άνοδο του οπτικοακουστικού υλικού με ψευδές περιεχόμενο (fake news). Οι αλγόριθμοι πλέον δεν είχαν μόνο την ευθύνη να προτείνουν υλικό το οποίο να είναι συναφές προς τον χρήστη, αλλά και το ίδιο το υλικό θα πρέπει να ανταποκρίνονταν στην πραγματικότητα.

Η προσέγγιση που ακολουθήθηκε για τη λύση αυτή του προβλήματος ήταν να δημιουργηθεί μία μεγάλη, διεθνής ομάδα η οποία να εποπτεύει το υλικό και να κρίνει την καταλληλότητά του καθώς και να βελτιωθούν οι ταξινομητές (classifiers) του περιεχομένου των βίντεο. Η ομάδα επιτήρησης αναλαμβάνει να απαντήσει σε κάποιες ερωτήσεις σχετικές με το περιεχόμενο, και μέσω αυτής της διαδικασίας ένας βαθμός (score) ανατίθεται σε κάθε βίντεο. Για ενημερωτικό και πληροφοριακό περιεχόμενο, όσο υψηλότερος είναι ο βαθμός τόσο περισσότερο το συγκεκριμένο βίντεο προωθείται. Ο βαθμός αυτός συνδυάζεται με το αποτέλεσμα των ταξινομητών και ένας ενιαίος βαθμός προκύπτει, ο οποίος δείχνει πόσο πιθανό είναι το βίντεο να περιέχει επιβλαβές υλικό. Το σύστημα, στη συνέχεια, εκπαιδεύεται από την ομάδα επιτήρησης και επεκτείνεται σε όλα τα βίντεο της πλατφόρμας. (Goodrow, 2021)

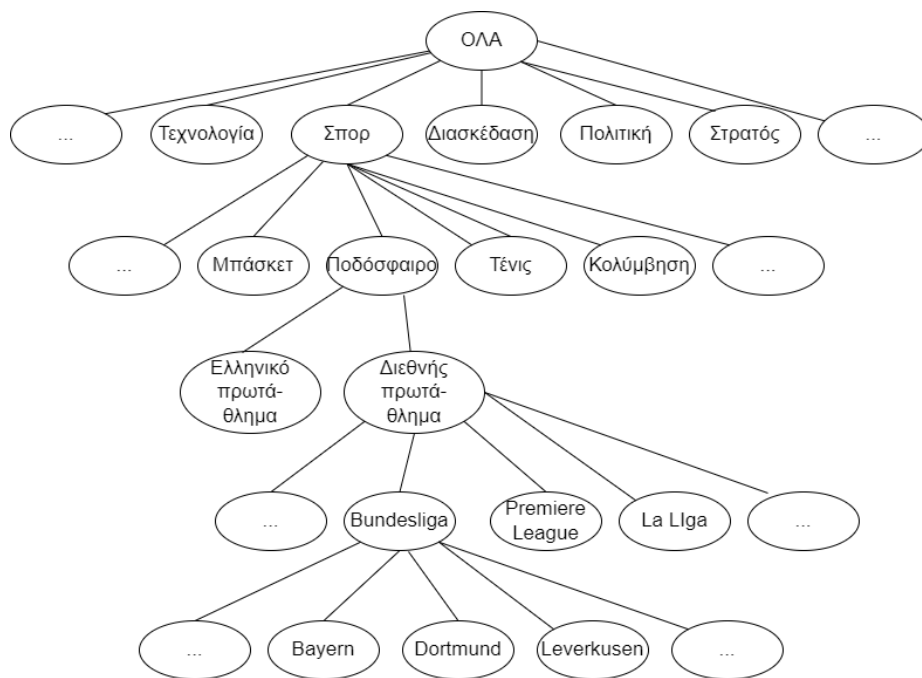
## 2.4 TikTok (Douyin)

Ανάμεσα στις εφαρμογές παροχής μικρής διάρκειας βίντεο που υπάρχουν, το TikTok έχει καταφέρει να ξεχωρίσει και να ηγηθεί της κατηγορίας αλλά και να διχάσει την κοινωνία

σε πολλές περιπτώσεις (Bonfiglio, 2022) (Fung, 2022). Το κύριο χαρακτηριστικό που έχει διαφοροποιήσει το TikTok από τους ανταγωνιστές του είναι πως το επιχειρηματικό του μοντέλο βασίζεται στο περιεχόμενο, το οποίο πολύ προσεχτικά επιλέγεται από τους αλγόριθμους που χρησιμοποιεί.

Παρότι ένα σύντομο βίντεο μπορεί να διαρκεί μόνο δεκαπέντε δευτερόλεπτα, είναι πολύ σύννηθες φαινόμενο οι χρήστες να καταναλώνουν πολύ περισσότερο χρόνο από αυτόν χωρίς καν το αντιλαμβάνονται. Αυτό επιτυγχάνεται μέσω του συστήματος προτάσεων που η ίδια η πλατφόρμα έχει αναπτύξει. Κύριο χαρακτηριστικό του συστήματος είναι πως επιτρέπει στους χρήστες να παρακολουθούν περιεχόμενο που ταιριάζει στα ενδιαφέροντά τους, χωρίς ωστόσο να χρειαστεί οι ίδιοι να ψάξουν να το βρουν. Σύμφωνα με μελέτη που διεξήχθη από βασικό μέλος της ομάδας ανάπτυξης του αλγόριθμου της ByteDance (κατασκευάστρια του TikTok), φαίνεται πως υπάρχει μία ξεκάθαρη συσχέτιση με την ευχαρίστηση του χρήστη και τη επιθυμία του να συνεχίσει να αλληλοεπιδρά με την πλατφόρμα. Παρατηρήθηκε, επίσης, πως η ικανοποίηση (satisfaction) είναι η σημαντικότερη μεταβλητή που επηρεάζει την επιθυμία του χρήστη να συνεχίσει την αλληλεπίδραση. (Li, Cao, Jiang, & Oh, 2013)

Το 2019 η μητρική εταιρία πατεντάρισε μία μέθοδο κατηγοριοποίησης μεγάλου συνόλου υλικού προκειμένου να ταιριάζει στα ενδιαφέροντα των χρηστών. Η υλοποίηση της μεθόδου αυτή περιλαμβάνει την κατασκευή ενός δέντρου τίτλων (labels). Με τη βοήθεια του δέντρου μπορεί εύκολα να φανεί η ιεραρχία και η σχέση των δεδομένων. Η ρίζα του δέντρου περιλαμβάνει όλο το διαθέσιμο υλικό της πλατφόρμας. Κάθε νέο επίπεδο του δέντρου αποτελεί εξειδίκευση ενός πατέρα-κόμβου. Για παράδειγμα, από τη ρίζα του δέντρου συγκαταλέγουμε μία μερίδα του περιεχομένου με ένα γενικό τίτλο «Αθλήματα». Στη συνέχεια αυτός ο κόμβος μπορεί να εξειδικευτεί περαιτέρω, περιλαμβάνοντας ήδη αθλημάτων (ποδόσφαιρο, μπάσκετ κτλ) και κάθε ένα είδος μπορεί να εξειδικευτεί στα ανάλογα πρωταθλήματα κ.ο.κ. (Zhao, 2021) Στο πλαίσιο αυτό, οι δημιουργοί της πλατφόρμας ενθαρρύνονται να εξειδικεύονται σε έναν μόνο τομέα καθώς ο αλγόριθμος θα μπορέσει να πραγματοποιήσει ακριβέστερες προτάσεις στους χρήστες που όντως ενδιαφέρονται για τον συγκεκριμένο τομέα, πετυχαίνοντας έτσι το υλικό να καταναλωθεί από το σωστό κοινό.



**Εικόνα 7. Το δένδρο ιεραρχιών του TikTok**

Με τη χρήση του συστήματος προτάσεων, κάθε χρήστης θα λάβει μία εντελώς εξατομικευμένη εμπειρία. Λόγου αυτού του γεγονότος, καθώς και ότι σκοπός της πλατφόρμας είναι να κρατά τον χρήστη ευχαριστημένο προτείνοντάς του περιεχόμενο με βάση τα ενδιαφέροντα και τις ανάγκες του, ο ίδιος καταναλώνει παθητικά το περιεχόμενο, ίσως και για ώρες.

Ο αλγόριθμος χρησιμοποιεί διάφορα χαρακτηριστικά προκειμένου να περιγράψει έναν χρήστη (Cao, 2018). Αρχικά, όταν κάποιος συνδέεται (log in) στην πλατφόρμα, του δίνεται η δυνατότητα της σύνδεσης μέσω κάποιου άλλου κοινωνικού δικτύου. Εφόσον ο χρήστης κάνει χρήση αυτής της επιλογής, ο αλγόριθμος θα κάνει χρήση των στοιχείων που θα βρει από το συγκεκριμένο κοινωνικό δίκτυο προκειμένου να επιταχύνει την ανάλυση των συνήθειών του χρήστη. Επειδή όμως η σύνδεση μέσω άλλου κοινωνικού δικτύου δεν είναι υποχρεωτική, ο αλγόριθμος δίνει μεγαλύτερη βαρύτητα στη συμπεριφορά του χρήστη εντός της πλατφόρμας.

Άλλα χαρακτηριστικά που λαμβάνονται υπόψιν είναι οι κατηγορίες, οι δημιουργοί και τίτλοι (labels) όπου οι χρήστες δείχνουν ενδιαφέρον. Χρήστες με κοινά ενδιαφέροντα δημιουργούν μία συστάδα (cluster) και διαμοιράζονται τα χαρακτηριστικά τους. Σημαντικά είναι επίσης τα στοιχεία που σχετίζονται με τη συμπεριφορά του χρήστη, όπως για παράδειγμα πότε ο χρήστης χρησιμοποιεί την εφαρμογή. Αν, για παράδειγμα, η συχνότητα

χρήσης είναι πολύ μεγάλη αλλά η διάρκεια μικρή, ο αλγόριθμος ενδέχεται να προτείνει βίντεο μικρής διάρκειας.

Το TikTok χρησιμοποιεί μία μέθοδο υπολογιστικής ροής (stream computing<sup>2</sup>) προκειμένου να επεξεργάζεται δυναμικά τα προφίλ των χρηστών. Η σύγχρονη (synchronous) αυτή τεχνική έχει ως αποτέλεσμα η συμπεριφορά του χρήστη να αντανακλά τα βίντεο που επιλέγει ο αλγόριθμος για αυτόν. Τη στιγμή που ο χρήστης αποφασίζει πως ένα βίντεο του άρεσε ή όχι, ο αλγόριθμος το λαμβάνει υπόψιν του για το επερχόμενο περιεχόμενο.

Προκειμένου το περιεχόμενο να παρουσιαστεί σε ένα ευρύ κοινό, η πλατφόρμα χρησιμοποιεί μία μέθοδο δεξαμενών δεδομένων (data buckets). Αρχικά, οι χρήστες μοιράζονται σε μικρές δεξαμενές με βάση μία σχετική τυχαιότητα (όπως το τελευταίο ψηφίο του ID χρήση). Όταν νέο περιεχόμενο εισάγεται στην πλατφόρμα, θα διαμοιραστεί πρώτα στους χρήστες των μικρών δεξαμενών και κάποιες μετρήσεις θα λάβουν μέρος. Συγκεκριμένα, θα μετρηθεί ο αριθμός προβολών, ο αριθμός των likes, το πλήθος των επαναλήψεων και των διαμοιρασμών, καθένα με διαφορετικό βάρος. Όταν η προκύπτουσα βαθμολογία γίνει ίση με ένα δεδομένο όριο, σημαίνει ότι το συγκεκριμένο περιεχόμενο έχει προοπτικές να γίνει δημοφιλές. Οπότε το βίντεο μεταφέρεται -καθότι «υπερχειλίσσε» (overflow)- σε μία δεξαμενή μεσαίου μεγέθους και η διαδικασία επαναλαμβάνεται. Το περιεχόμενο κάθε δεξαμενής δέχεται διαφορετικά επίπεδα προβολών, με τα βίντεο που εμφανίζουν τη μεγαλύτερη απόδοση θα μεταφέρονται συνεχώς σε μεγαλύτερες δεξαμενές μέχρις ότου να καταλήξουν να προβάλλονται σε ολόκληρη την πλατφόρμα.

Το TikTok πατεντάρισε μία δική του μέθοδο προκειμένου να υπολογίζει τη συσχέτιση μεταξύ τίτλων (labels) και ενδιαφερόντων του χρήστη. Αρχικά, το μοντέλο θέτει ένα κατώφλι (threshold) συσχέτισης μεταξύ δύο τίτλων με βάση τη μικρότερη απόσταση μεταξύ τους. Στη συνέχεια, συγκεντρώνει όλους τους γνωστούς (known) τίτλους που αντιστοιχούν στα ενδιαφέροντα του χρήστη και στη συνέχεια τα τροφοδοτεί σε μία συνάρτηση η οποία λαμβάνει υπόψιν τη συχνότητα με την οποία δύο τίτλοι εμφανίζονται αμφότεροι στα ενδιαφέροντα ενός χρήστη καθώς και την ατομική συχνότητα του καθενός. Εάν το αποτέλεσμα της συνάρτησης είναι μικρότερο από το κατώφλι, σημαίνει πως το περιεχόμενο των δύο αυτών τίτλων είναι τελείως άσχετο οπότε και δε θα προταθούν.

---

<sup>2</sup> Ο όρος “Stream computing” αναφέρεται σε ένα πρότυπο υπολογισμού όπου τα δεδομένα προέρχονται από συλλογές συνεχούς ροής πληροφοριών και τα αποτελέσματα παράγονται σε πραγματικό χρόνο, κι αυτά με συνεχή ροή. (Sun, Zhang, Zheng, & Li, 2015)

Αφότου υπολογίσει την ομοιότητα μεταξύ δύο τίτλων, το μοντέλο, στη συνέχεια, αναθέτει μία βαθμολογία (score) η οποία καθορίζεται από την δημοτικότητα ενός τίτλου και τον βαθμό στον οποίο έχει εξερευνηθεί. Στόχος είναι ο αλγόριθμος να μπορεί να προτείνει στο χρήστη, όχι μόνο περιεχόμενο με ακριβή ταύτιση στα ενδιαφέροντά του, αλλά και περιεχόμενο νέο και απρόσμενο για αυτόν. (Zhao, 2021)

## Κεφάλαιο 3. Μεθοδολογία έρευνας

### 3.1 Διάρθρωση κεφαλαίου

Στο κεφάλαιο αυτό παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για τη συγκέντρωση των απαραίτητων πληροφοριών, τεχνικών και δεδομένων τα οποία χρησιμοποιούνται για την ανάπτυξη του συστήματος προτάσεων.

Αρχικά, γίνεται μία έρευνα προκειμένου να βρεθούν εφαρμογές με παρόμοιο σκοπό με αυτή της παρούσας πτυχιακής εργασίας και μελετάται το γραφικό τους περιβάλλον, προκειμένου να αναπτυχθεί ένα εφάμιλλο το οποίο να είναι οικείο για τους χρήστες. Στη συνέχεια αναφέρεται ο τρόπος με τον οποίο επιλέγηκε το σύνολο δεδομένων που χρησιμοποιήθηκε και αναλύεται η δομή αυτού.

### 3.2 Παρόμοιες εφαρμογές

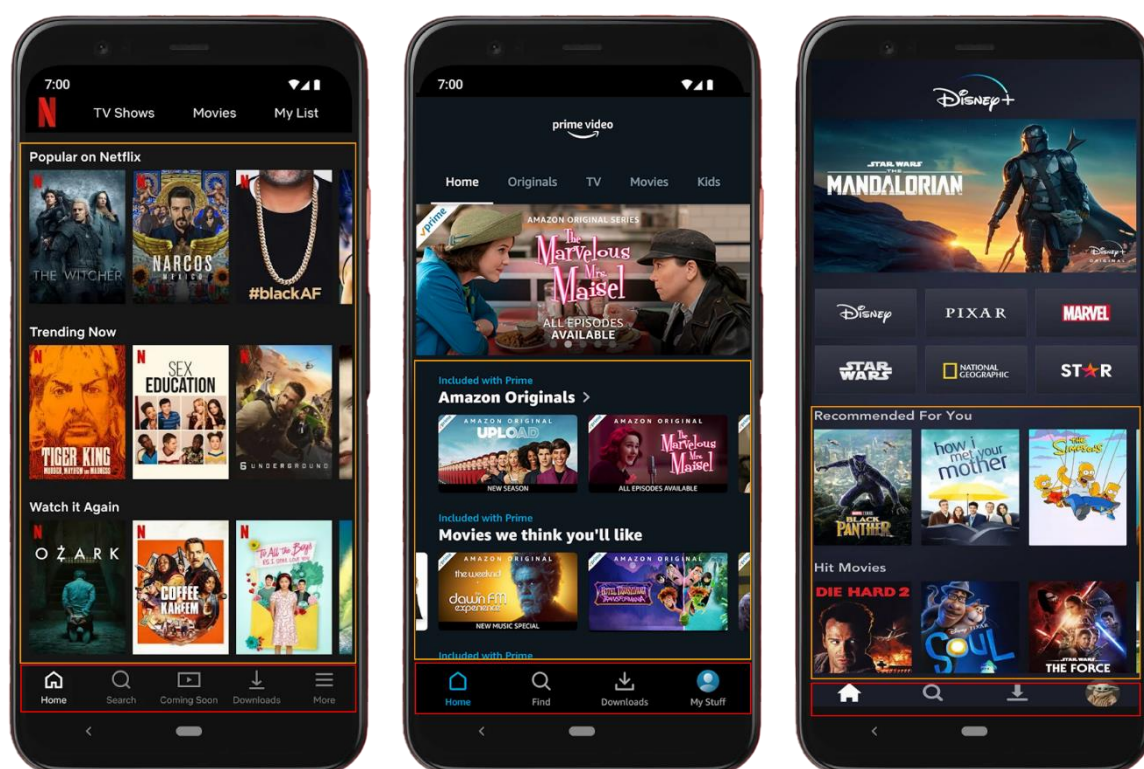
Όταν πρόκειται για τη δημιουργία μίας νέας εφαρμογής / προγράμματος είναι σημαντικό να λαμβάνεται υπόψιν και η προσέγγιση των ήδη υπαρχόντων. Ο λόγος είναι προκειμένου η διαίσθηση που έχουν οι χρήστες στην χρήση παρόμοιων εφαρμογών να μεταφερθεί και στην υπό ανάπτυξη εφαρμογή, προκειμένου να αλληλοεπιδράσουν με ένα φιλόξενο και οικείο περιβάλλον. Συνεπώς, το πρώτο βήμα ήταν η εύρεση των πιο γνωστών από τις εφαρμογές που περιλαμβάνουν ένα σύστημα προτάσεων ταινιών.



Εικόνα 8. Στατιστικά δημοτικότητας streaming υπηρεσιών στις ΗΠΑ την περίοδο Ιουλίου 2020 και Ιουνίου 2021



Το επόμενο βήμα ήταν η αξιολόγηση του γραφικού τους περιβάλλοντος και η μεταξύ τους σύγκριση προκειμένου να βρεθούν κοινά στοιχεία – μεθοδολογίες και ίσως κάποιο μοτίβο που ακολουθείται. Παρατηρήθηκε πως όλες οι εφαρμογές ακολουθούν ένα συγκεκριμένο μοτίβο παρουσιάσεις των ταινιών. Χρησιμοποιούν γραμμές (rows) κάθε μία από τις οποίες περιέχει προτάσεις ενός συγκεκριμένου είδους (genre) -όπως κωμωδίες- ή μίας συγκεκριμένης κατηγορίας -όπως «Δημοφιλή τώρα»-. Παρατηρήθηκε επίσης πως όλα τα γραφικά περιβάλλοντα διέθεταν μία μπάρα πλοήγησης στο κάτω μέρος της οθόνης.



**Εικόνα 9. Γραφικό περιβάλλον εφαρμογών με σύστημα προτάσεων ταινιών. Και οι τρεις εφαρμογές χρησιμοποιούν την παρουσία σε γραμμές (rows) και μία μπάρα πλοήγησης στο κάτω μέρος.**

### 3.3 Σύνολο δεδομένων

Προκειμένου μία εφαρμογή προτάσεων να είναι λειτουργική χρειάζεται δεδομένα. Συγκεκριμένα, χρειάζεται δύο είδους δεδομένα:

1. Πληροφορίες για το περιεχόμενο
2. και κρητικές

Στην προκειμένη περίπτωση που το περιεχόμενο αποτελείται από ταινίες είναι απαραίτητα κάποια βασικά δεδομένα όπως ο τίτλος, η περιγραφή, το έτος κυκλοφορίας, η δημοτικότητα και τα είδη (genres). Ως στόχος τέθηκε επίσης τα δεδομένα να βασίζονται στον πραγματικό κόσμο και να μην είναι ψεύτικα (dummy).

Η πρώτη σκέψη ήταν να χρήση του συνόλου δεδομένων<sup>3</sup> που παρήχε το Netflix για τους σκοπούς του [διαγωνισμού](#). Ωστόσο κρίθηκε πως ο όγκος της πληροφορίας είναι αρκετά μεγάλος για τα πλαίσια της εργασίας. Τελικά επιλέχθηκε το σύνολο δεδομένων MovieLens που παρέχεται από την GroupLens, ερευνητική ομάδα με αντικείμενο έρευνας στα συστήματα προτάσεων. Το σύνολο δεδομένων παρέχεται σε πολλές εκδόσεις, κάθε μία με περισσότερα δεδομένα<sup>4</sup>. Επιλέχθηκε το σύνολο δεδομένων MovieLens 1M το οποίο περιέχει 100.000 κριτικές ταινιών από 1000 χρήστες για 1700 ταινίες. (Harper & Konstan, 2015) Ως χρήστες ορίζονται άτομα που διατηρούν ατομικό προφίλ στον διαδικτυακό τόπο του MovieLens<sup>5</sup>.

Το σύνολο δεδομένων αποτελείται από τρία αρχεία:

- users.dat
- ratings.dat
- movies.dat

Ακολουθεί επεξήγηση του συνόλου δεδομένων:

---

<sup>3</sup> <https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data>

<sup>4</sup> <https://grouplens.org/datasets/movielens/>

<sup>5</sup> <https://movielens.org/>

**Πίνακας 2. Περιγραφή ιδιοτήτων αρχείου users.dat**  
 Οι αγνές καταχωρήσεις δεν χρησιμοποιούνται στα πλαίσια της εργασίας

Πεδίο	Περιγραφή
UserID	Μοναδικός αναγνωριστικός κωδικός κάθε χρήστη
Gender	Φύλλο χρήστη. Μ για άρρεν F για θήλυ
Age	Ηλικιακή ομάδα χρήστη
Occupation	Απασχόληση
Zip-code	Ταχυδρομικός κώδικας
<b>Διάρθρωση αρχείου</b>	
UserID::Gender::Age::Occupation::Zip-code	

**Πίνακας 3. Περιγραφή ιδιοτήτων αρχείου ratings.dat**  
 Οι αγνές καταχωρήσεις δεν χρησιμοποιούνται στα πλαίσια της εργασίας

Πεδίο	Περιγραφή
UserID	Μοναδικός αναγνωριστικός κωδικός κάθε χρήστη
MovieID	Μοναδικός αναγνωριστικός κωδικός κάθε ταινίας
Rating	Βαθμολογία χρήστη για κάθε ταινία (1-10)
Timestamp	Χρονική στιγμή που καταχωρήθηκε η βαθμολογία
<b>Διάρθρωση αρχείου</b>	
UserID::MovieID::Rating::Timestamp	

**Πίνακας 4. Περιγραφή ιδιοτήτων αρχείου movies.dat**  
 Οι αγνές καταχωρήσεις δεν χρησιμοποιούνται στα πλαίσια της εργασίας

Πεδίο	Περιγραφή
MovieID	Μοναδικός αναγνωριστικός κωδικός κάθε ταινίας
Title	Τίτλος ταινίας
Genres	Λίστα ειδών
<b>Διάρθρωση αρχείου</b>	
MovieID::Title::Genres	

Όπως προκύπτει από την επεξήγηση, το σύνολο δεδομένων του MovieLens μπορεί να παρέχει τις πληροφορίες που χρειάζονται μόνο για το ένα σκέλος· αυτό των βαθμολογιών, καθώς οι πληροφορίες που διαθέτει για τις ταινίες είναι οι απολύτως βασικές. Προκειμένου να βρεθούν οι απαραίτητες πληροφορίες για τις ταινίες χρησιμοποιήθηκε η σελίδα TMDb. Πρόκειται για μία διαδικτυακή βάση δεδομένων που περιλαμβάνει πληροφορίες για συνολικά πάνω από 800.000 ταινίες και σειρές. Όλες οι καταχωρήσεις προέρχονται από τα μέλη της κοινότητας. Η επιλογή της σελίδας έγινε καθώς διαθέτει δημόσιο, εύκολο στη χρήση API<sup>6</sup>, παρέχει δωρεάν πρόσβαση και χρησιμοποιείται και από πολλές γνωστές σελίδες ανάλογου περιεχομένου.

Προκειμένου να πραγματοποιηθεί η ανάκτηση των δεδομένων των ταινιών που από το API, δημιουργήθηκε ένα python script το οποίο εκτελούσε τα εξής βήματα:

- Αναζήτηση στο API με βάση τον τίτλο της ταινίας
- Εύρεση της κατάλληλης αντιστοιχίας (σε περίπτωση συνωνυμίας)
- Εύρεση των δεδομένων της ταινίας μέσω του API
- Αποθήκευση των δεδομένων σε ένα αρχείο .dat

Τελικά, το σύνολο δεδομένων για τις πληροφορίες ταινιών κατέληξε στην εξής μορφή:

**Πίνακας 5. Περιγραφή ιδιοτήτων συνόλου δεδομένων**

Πεδίο	Περιγραφή
_id	Μοναδικός αναγνωριστικός αριθμός καταχώρησης στη βάση δεδομένων
adult	Ταινία ενήλικου περιεχομένου (true/false)
backdrop_path	Στιγμιότυπο ταινίας
genre_ids	Πίνακας που περιέχει τα ids των ειδών, με βάση τη βάση του TMDb
original_language	Πρωτότυπη γλώσσα παραγωγής
original_title	Πρωτότυπος τίτλος παραγωγής
overview	Επισκόπηση ταινίας
popularity	Δημοτικότητα
poster_path	Μερική διαδρομή (path) στη οποία βρίσκεται η αφίσα της ταινίας

<sup>6</sup> Για την επικοινωνία με το API χρησιμοποιήθηκε μία τρίτη (third-party) βιβλιοθήκη ανοιχτού κώδικα με όνομα tmdbsimple: <https://github.com/celiao/tmdbsimple>

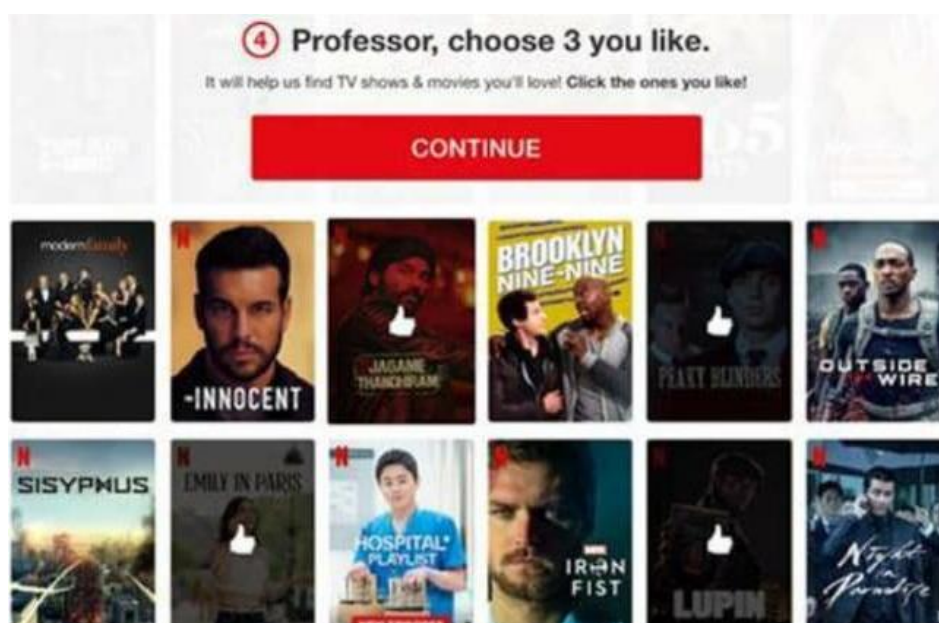
release_date	Ημερομηνία κυκλοφορίας
title	Τίτλος (μπορεί να διαφέρει από γλώσσα σε γλώσσα)
video	Διαθέτει τρέιλερ (true/false)
vote_average	Μέσος όρος βαθμολογιών
vote_count	Σύνολο ψήφων
movielens_id	Μοναδικό αναγνωριστικό ταινίας στο σύνολο δεδομένων μας
imdb_id	Μοναδικό αναγνωριστικό ταινίας στη σελίδα IMDB

Όπως είδαμε, οι παρόμοιες εφαρμογές που εξετάστηκαν είχαν ως κύριο χαρακτηριστικό παρουσίαση την αφίσα της εκάστοτε ταινίας. Συνεπώς το σύνολο δεδομένων καθαρίστηκε από τυχών καταχωρήσεις οι οποίες για τον οποιονδήποτε λόγο δεν διέθεταν αφίσα, καθώς θα δημιουργούσε πρόβλημα μετέπειτα. Το τελικό σύνολο δεδομένων αποτελούνταν από ~3.400 ταινίες.

## Κεφάλαιο 4. Αλγοριθμικές τεχνικές

### 4.1 Διαχείριση δεδομένων συστήματος

Προκειμένου ένα σύστημα προτάσεων να μπορέσει να παράξει εξατομικευμένες προτάσεις, χρειάζεται να γνωρίζει κάποια βασικά στοιχεία για τον χρήστη. Στην προκειμένη περίπτωση, χρειάζεται να γνωρίζει ποια είδη (genres) ταινιών προτιμάει και ποιες ταινίες έχει ήδη δει. Μία προσέγγιση που μπορεί να ακολουθηθεί προκειμένου να συλλεχθούν αυτές οι πληροφορίες είναι η παρουσίαση στον χρήστη μία λίστας από αντιπροσωπευτικές ταινίες κάθε είδους και η προτροπή να τις βαθμολογήσει. Επειδή οι προτιμήσεις του στα είδη είναι αρκετά σημαντική πληροφορία, η διαδικασία αυτή θα γίνει κατά την δημιουργία ενός νέου λογαριασμού. Μία αντίστοιχη μέθοδο χρησιμοποιεί και το Netflix, όπου του παρουσιάζει συγκεκριμένους τίτλους ταινιών (συνήθως με μεγάλη δημοτικότητα προκειμένου να αυξηθούν οι πιθανότητες ο χρήστης να έχει δει κάποια) και του ζητάει να επιλέξει ποιες από τις ταινίες αυτές του αρέσουν. Ύστερα, θα προσαρμόσει τις προτάσεις του με βάση τις επιλογές του χρήστη.



**Εικόνα 10** Το Netflix ζητάει από τον χρήστη να επιλέξει τουλάχιστον τρεις ταινίες ή σειρές που του αρέσουν προκειμένου να λάβει κάποιες αρχικές πληροφορίες από τον χρήστη και να αρχίσει να του προτείνει συναφές περιεχόμενο

Εκτός από ταινίες που αρέσουν στον χρήστη παρέχεται η δυνατότητα να επιλέξει και ρητά ποια είδη προτιμά, καθώς οι ταινίες μπορεί να είναι κατηγοριοποιημένες διαφορετικά από την αντίληψη του. Έτσι, για παράδειγμα, ο χρήστης μπορεί να βαθμολογήσει υψηλά τρεις ταινίες που, κατά τον ίδιο (και κατά τα φαινόμενα), ανήκουν στο Sci-Fi είδος, ωστόσο είναι πολύ πιθανό να περιέχουν πάνω από ένα είδη, όπως για παράδειγμα να ανήκουν και στο είδος του δράματος, κι έτσι να βρεθεί να του προτείνονται δράματα, χωρίς ωστόσο ο ίδιος να ενδιαφέρεται για τέτοιου είδους περιεχόμενο.

Δίνεται στον χρήστη την δυνατότητα να επιλέξει όσα είδη θέλει, ακόμη και κανένα, όπου στην περίπτωση αυτό το σύστημα, όπως είναι προφανές, δημιουργεί προτάσεις χωρίς να λαμβάνει υπόψιν του τα είδη των ταινιών.

Τα είδη τα οποία μπορεί να επιλέξει προέρχονται από τη βάση δεδομένων του TMDb και είναι τα εξής 19:

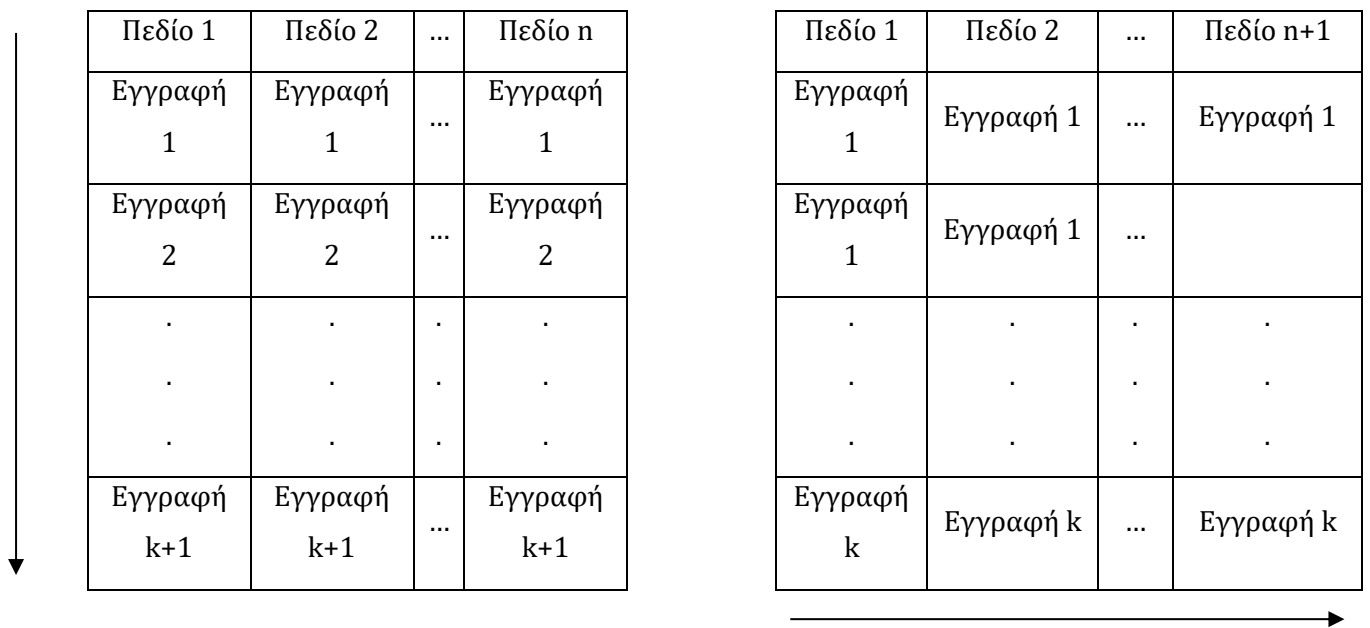
Action	Drama	Mystery	Western
Adventure	Family	Romance	
Animation	Fantasy	Sci-Fi	
Comedy	History	TV Movie	
Crime	Horror	Thriller	
Documentary	Music	War	

Μετά την επιλογή του, οι προτιμήσεις του αποθηκεύονται στη βάση δεδομένων και συνδέονται με το ατομικό προφίλ του.

## 4.2 Αποθήκευση δεδομένων

Τα δεδομένα, ή αλλιώς η πληροφορία, αποτελεί αναπόσπαστο κομμάτι ακόμη και της πιο απλής εφαρμογής. Συνεπώς χρειάζεται ένας τρόπος προκειμένου τα δεδομένα να αποθηκευτούν, να διατηρηθούν και, κυριότερα, να ανακτηθούν. Όταν πρόκειται για δεδομένα με σαφή δομή αλλά και δεδομένα μεταξύ των οποίων υπάρχει λογική σχέση εξάρτησης, η τετριμμένη λύση είναι η χρήση μιας βάσης δεδομένων.

Οι βάσεις δεδομένων χωρίζονται σε δύο μεγάλες κατηγορίες, τις SQL και NoSQL βάσεις με κάθε κατηγορία να έχει τα προτερήματα και τα μειονεκτήματά της. Μία ειδοποιός διαφορά που υπάρχει ανάμεσα σε αυτές τις δύο κατηγορίες είναι ο τρόπος με τον οποίο επεκτείνονται.



**Σχήμα 10. διαφορετικός τρόπος επέκτασης των δύο κατηγοριών βάσεων δεδομένων. Αριστερά η SQL βάση και δεξιά η NoSQL**

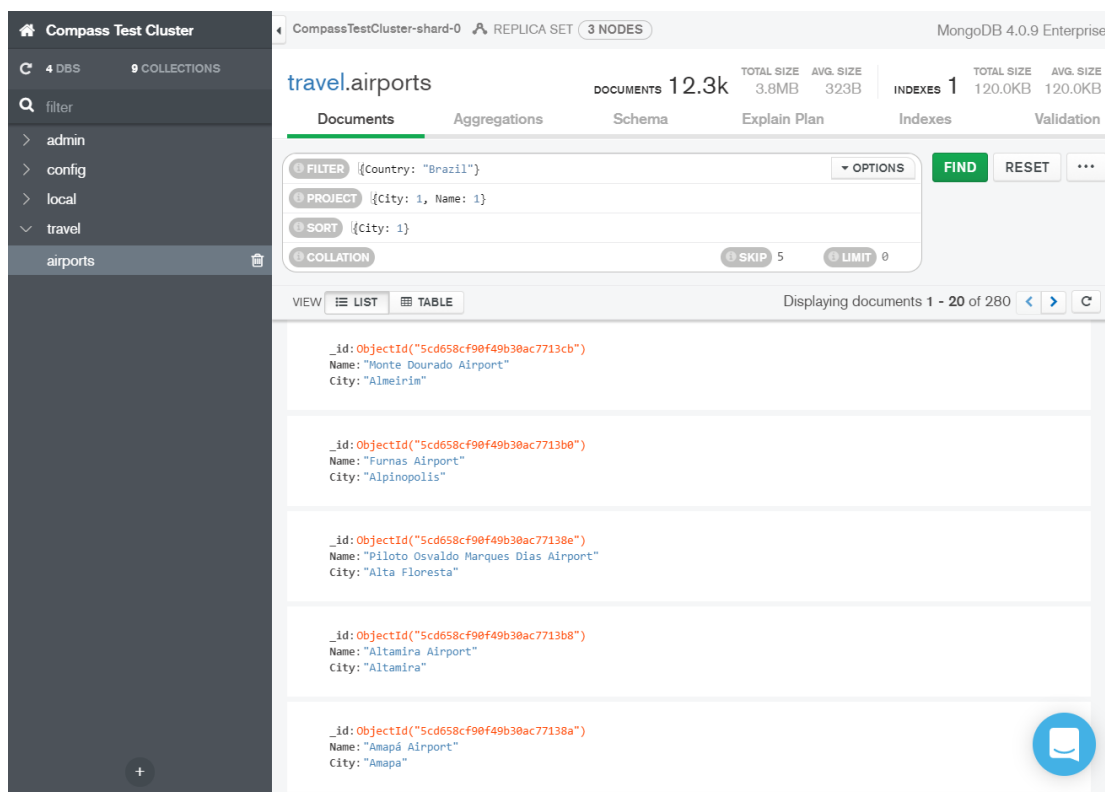
Οι SQL βάσεις επεκτείνονται ευκολότερα κάθετα προσθέτοντας νέες γραμμές (rows), δηλαδή δίνουν τη δυνατότητα πολύ εύκολης προσθήκης νέων καταχωρήσεων (records). Ωστόσο δεν είναι το ίδιο εύκολη και η πρόσθεση μίας νέας στήλης (column), δηλαδή μίας νέας ιδιότητας. Ο λόγος είναι ότι κάθε καταχώρηση πρέπει οπωσδήποτε να διαθέτει όλες τις ιδιότητες του πίνακα, ακόμη κι αυτή είναι κενή (null). Το θετικό αντίκτυπο αυτής της προσέγγισης είναι ότι επιτρέπει την ύπαρξη σύνθετων σχέσεων μεταξύ των δεδομένων σε επίπεδο βάσης δεδομένων.

Σε αντίθεση με τις SQL βάσεις, οι NoSQL επεκτείνονται ευκολότερα οριζόντια προσθέτοντας νέες στήλες, δηλαδή νέες ιδιότητες στις καταχωρήσεις μας. Πράγματι, σε μία NoSQL βάση οι καταχωρήσεις που βρίσκονται στο ίδιο έγγραφο (document) δεν είναι απαραίτητο να διαθέτουν τις ίδιες ιδιότητες, χαρακτηριστικό το οποίο διευκολύνει πάρα πολύ την περίπτωση που δεν είναι γνωστές εξ αρχής όλες οι πληροφορίες που θα χρειαστούν ή στην περίπτωση πρόσθεσης νέων ιδιοτήτων μόνο σε συγκεκριμένες καταχωρήσεις.

Καθώς οι ανάγκες τις εργασίας όσον αφορά το κομμάτι της βάσεις δεδομένων δεν είναι απαιτητικές, και οι δύο προσεγγίσεις θα μπορούσαν να ανταποκριθούν. Παρόλα αυτά, έγινε επιλογή της NoSQL τεχνολογίας, κυρίως για την δυνατότητα που παρέχει οι καταχωρήσεις ενός εγγράφου να είναι ανομοιομορφες. Πιο συγκεκριμένα χρησιμοποιήθηκε



η βάση δεδομένων MongoDB<sup>7</sup>, για την οποία παρέχεται κι ένα άκρως βοηθητικό γραφικό περιβάλλον<sup>8</sup>, του οποίου η εγκατάσταση είναι προαιρετική καθότι ολόκληρη η διαχείριση της βάσης μπορεί να γίνει από γραμμή εντολών. Η βάση διατίθεται τόσο σε επί πληρωμή όσο και σε ελεύθερη έκδοση.



**Εικόνα 11. Το γραφικό περιβάλλον MongoDB Compass.**

Στα αριστερά φαίνεται το σύνολο των συλλογών (collections), δηλαδή των διαθέσιμων βάσεων. Στο επάνω μέρος φαίνεται πως έχει γίνει επιλογή του εγγράφου “airports” της συλλογής “travel” και δεξιά εξ αυτών κάποια βασικά στοιχεία. Αμέσως από κάτω φαίνονται τα φίλτρα που έχουν χρησιμοποιηθεί για μία αναζήτηση και στο κάτω μέρος τα αποτελέσματα της αναζήτησης

Η επικοινωνία και διαχείριση της βάσης γίνεται μέσω της βιβλιοθήκης ανοιχτού κώδικα PyMongo<sup>9</sup>, η οποία συντηρείται από τους δημιουργούς της MongoDB. Η βάση διαθέτει μία ευρεία γκάμα από φίλτρα τα οποία μπορούν να χρησιμοποιηθούν για την ανάκτηση των δεδομένων.

<sup>7</sup> Διαθέσιμη στην επίσημη σελίδα <https://www.mongodb.com/>

<sup>8</sup> Διαθέσιμο στην επίσημη σελίδα <https://www.mongodb.com/products/compass>

<sup>9</sup> Διαθέσιμη στο αποθετήριο <https://github.com/mongodb/mongo-python-driver>

## 4.3 Το σύστημα προτάσεων

### 4.3.1 Ορισμός

Όπως έγινε φανερό στα πρώτα δύο κεφάλαια υπάρχουν διάφορες μέθοδοι για να προσεγγίσει κανείς ένα σύστημα προτάσεων. Η προσέγγιση που ακολουθήθηκε στα πλαίσια αυτής της εργασίας θα βασιστεί στον αλγόριθμο της μεθόδου Σταθμισμένων Βαρών (Weighted Sum Method, WSM). Η μέθοδος αυτή χρησιμοποιείται ευρέως σε πολυκριτήρια προβλήματα βελτιστοποίησης. Η βασική ιδέα πίσω από τη μέθοδο είναι η διάσπαση ενός πολυσύνθετου προβλήματος σε μικρότερα επιμέρους προβλήματα. (Chagas & Wagner, 2020) Η επιλογή του αλγορίθμου έγινε βασιζόμενοι στο γεγονός πως παρέχει ελαστικότητα όσον αφορά τον αριθμό των κριτηρίων απόφασης. (Trousas, Giannakas, Sgouroroulou, & Voyiatzis, 2020)

Στην Μέθοδο Σταθμισμένων Βαρών στόχος είναι η μετατροπή του συνόλου των στόχων σε ένα, πολλαπλασιάζοντας το κάθε στόχο με ένα βάρος. Τα βάρη που δίνονται είθισται να είναι αναλογικά με την σπουδαιότητα που έχει κάθε στόχος για το πρόβλημά που καλείται η μέθοδος να λύσει. Όμως είναι εξίσου σημαντικό να ληφθεί υπόψιν και το μέγεθος του εκάστοτε στόχου. Για τον λόγο αυτό, γίνεται κλιμάκωση κάθε ιδιότητας προκειμένου όλες να έχουν το ίδιο μέγεθος· με άλλα λόγια, κανονικοποιούνται.

Οι ιδιότητες της μεθόδου χωρίζονται σε δύο κατηγορίες. Τις ωφέλιμες ιδιότητες, όπου είναι επιθυμητές οι μέγιστες τιμές και τις ανώφελες ιδιότητες όπου είναι επιθυμητές οι ελάχιστες τιμές. Για την κανονικοποίηση των ιδιοτήτων χρησιμοποιείται διαφορετική μέθοδος ανά κατηγορία:

$$\text{Για τις ωφέλιμες ιδιότητες, } X = \frac{x}{x_{\max}}$$

$$\text{Για τις ανώφελες ιδιότητες, } X = \frac{x_{\min}}{x}$$

Όπου  $x$  = στοιχείο ιδιότητας,  $x_{\max}$  = μέγιστη τιμή ιδιότητας,  $x_{\min}$  = ελάχιστη τιμή ιδιότητας.

Μαθηματικά, ο αλγόριθμος εκφράζεται ως:

$$F(x) = \sum_{m=1}^M w_m f_m(x)$$

όπου:

- $M$  ο αριθμός των ιδιοτήτων
- $w_m$  το βάρος που ανατίθεται σε κάθε ιδιότητα
- $f_m$  η ιδιότητα και
- $x$  η μεταβλητή απόφασης (decision variable)

### 4.3.2 Ιδιότητες

Όπως αναφέρθηκε, ο αλγόριθμος WSM αποτελείται από έναν αριθμό ιδιοτήτων, κάθε μία από τις οποίες εκφράζει ένα διαφορετικό κριτήριο στην διαδικασία παραγωγής προτάσεων.

Σε αυτήν την περίπτωση, στόχος είναι το σύστημά μας να παράγει προτάσεις για ταινίες, ένα ευρέως διαδεδομένο μέσω ψυχαγωγίας, συνεπώς, εκτός από τις προτιμήσεις του ίδιου του χρήστη θα ληφθεί υπόψιν και η άποψη του συνόλου.

Στο [κεφάλαιο 4.1](#) εξηγήθηκε πώς γίνεται η συλλογή των δεδομένων που είναι απαραίτητα για την παραγωγή των προτάσεων. Στο σημείο αυτό θα παρουσιαστεί ο τρόπος με τον οποίο θα αξιοποιηθούν. Μέσω της διαδικασίας που περιεγράφηκε νωρίτερα είναι γνωστό ποιες είναι οι προτιμήσεις του χρήστη όσον αφορά τα είδη των ταινιών. Είναι επίσης γνωστό, ως συναγόμενη πληροφορία, είτε από την αρχική βαθμολόγηση των ταινιών είτε ύστερα από χρήση της εφαρμογής, οι ταινίες που έχει ήδη παρακολουθήσει ο χρήστης. Λαμβάνεται ως παραδοχή πως το σύστημα προτάσεων δε θα προτείνει ταινίες που ο χρήστης έχει ήδη δει. Χρησιμοποιούνται τέσσερα κριτήρια επιλογής:

1. Το είδος της ταινίας
2. Ο μέσος όρος βαθμολογίας της ταινίας
3. Το πλήθος των ψήφων της ταινίας
4. Η δημοτικότητα της ταινίας

### 4.3.3 Κανονικοποίηση ειδών

Αφού ορίστηκαν τα κριτήρια σύμφωνα με τα οποία θα παραχθούν οι προτάσεις, ήρθε η ώρα να γίνει αναφορά στην ίδια την διαδικασία και τυχών ιδιαιτερότητες που έχει.

Όπως έχει ήδη αναφερθεί, ως παραδοχή του συστήματός έχει οριστεί πως δε θα προτείνει ταινίες τις οποίες ο χρήστης έχει ήδη παρακολουθήσει, δηλαδή ταινίες για τις οποίες υπάρχει καταχωρημένη βαθμολογία στη βάση για τον χρήστη αυτόν. Το πρώτο μέρος της διαδικασίας είναι να ανακτηθούν οι υποψίες ταινίες. Συνεπώς, το σύστημα κάθε φορά

ανακτά ολόκληρη τη λίστα με τις ταινίες από τη βάση δεδομένων και στη συνέχεια αφαιρεί από τη λίστα τις ταινίες που ο χρήστης έχει ήδη παρακολουθήσει.

Αφού η λίστα με τις ταινίες ανακτηθεί, πρέπει να υπάρξει μέριμνα ώστε να συμπεριληφθούν τα είδη ταινιών που προτιμά ο χρήστης. Όπως είναι λογικό, μία ταινία μπορεί να ανήκει σε παραπάνω από ένα είδη ταυτόχρονα, το οποίο σημαίνει πως διακρίνονται τρεις περιπτώσεις:

- Τα είδη μιας ταινίας να μην αντιστοιχούν με κανένα από τα προτιμητέα είδη του χρήστη
- Τα είδη μιας ταινίας να ανήκουν εν μέρη στα προτιμητέα είδη του χρήστη
- Όλα τα είδη μιας ταινίας να ανήκουν στα προτιμητέα είδη του χρήστη

Όπως είναι λογικό, είναι προτιμότερο μία ταινία της οποίας όλα τα είδη αποτελούν και προτιμήσεις του χρήστη να αποτελεί καλύτερη πρόταση από μία που τα είδη ταιριάζουν εν μέρη. Ωστόσο, προκύπτει το εξής πρόβλημα: το πλήθος των ειδών που ανήκει μία ταινία. Για παράδειγμα, ένας χρήστης προτιμά τα είδη Δράση, Περιπέτεια, Μυστήριο, Δράμα, συνεπώς:

$$u = \{Action, Adventure, Mystery, Fantasy\}$$

και δύο ταινίες, μία που να ανήκει στις κατηγορίες Δράση και Περιπέτεια και μία που ανήκει στις κατηγορίες Δράση, Μυστήριο, Επιστημονική φαντασία, Φαντασία, Κινούμενα σχέδια:

$$m_1 = \{Action, Adventure\}$$

$$m_2 = \{Action, Mystery, Sci - fi, Fantasy, Animation\}$$

Όπως είναι προφανές, η πρώτη ταινία έχει πλήρη ταύτιση με τις προτιμήσεις του χρήστη, όπου και τα δύο είδη της συμπεριλαμβάνονται στα προτιμητέα είδη του χρήστη, ενώ η δεύτερη έχει μερική ταύτιση, με τρία από τα πέντε είδη να ανήκουν στις προτιμήσεις του χρήστη. Σύμφωνα λοιπόν με όσα αναφέρθηκαν, η πρώτη ταινία που έχει ολική ταύτιση, θα έπρεπε να αποτελεί καλύτερη πρόταση από τη δεύτερη που έχει μερική ταύτιση, παρόλο που η δεύτερη διαθέτει περισσότερα είδη που ανήκουν στα προτιμητέα του χρήστη.

$$m_1 \cap u = \{Action, Adventure\}$$

$$m_2 \cap u = \{Action, Mystery, Fantasy\}$$

Συνεπώς, έπρεπε να βρεθεί μία μέθοδος προκειμένου το πλήθος των ειδών να λαμβάνεται υπόψιν. Η απλούστερη λύση στο πρόβλημα αυτό είναι η χρήση της

κανονικοποίησης (normalization). Ως μέθοδος κανονικοποίησης χρησιμοποιήθηκε η διαίρεση μεταξύ του πλήθους των προτιμητέων ειδών του χρήστη και του πλήθους των ειδών μίας ταινίας. Η προκύπτουσα τιμή ανατίθεται στην εκάστοτε ταινία.

#### 4.3.4 Βάρη

Αφού ολοκληρώθηκε και το στάδιο κανονικοποίησης των ειδών, ήρθε η ώρα να οριστεί ο βαθμός στον οποίο κάθε κριτήριο λαμβάνεται υπόψιν στην παραγωγή των προτάσεων. Με άλλο λόγια, να αποφασιστεί ποια θα είναι τα βάρη των ιδιοτήτων.

Τα κριτήρια που αναφέρθηκαν χωρίζονται σε δύο κατηγορίες, άμεσα και τα έμμεσα. Ως άμεσα νοούνται τα κριτήρια τα οποία προήλθαν από τον χρήστη και γνωρίζει την ύπαρξή τους και έμμεσα ως αυτά που δεν σχετίζονται με κανέναν τρόπο με τον χρήστη και που ο ίδιος ο χρήστης δεν γνωρίζει πως συμμετέχουν στη διαδικασία (ίσως, όμως, το υποψιάζεται). Στην πρώτη κατηγορία ανήκουν τα είδη της ταινίας και στην δεύτερη κατηγορία ανήκουν ο μέσος όρος βαθμολογίας, το πλήθος των ψήφων και η δημοτικότητα της ταινίας.

Στόχος στην υλοποίησή αυτού του συστήματος είναι τα άμεσα κριτήρια, δηλαδή αυτά που ο ίδιος ο χρήστης ορίζει και που, εν δυνάμει, είναι διαφορετικά για κάθε χρήστη, να έχουν σημαντικότερο ρόλο από τα έμμεσα. Συνεπώς στο κριτήριο αυτό ανατέθηκε και το μεγαλύτερο βάρος. Απόρροια αυτού είναι οι ταινίες που θα προταθούν να σχετίζονται περισσότερο με τις προτιμήσεις του χρήστη και λιγότερο με την γενική αντιμετώπισή τους από το κοινό. Τα βάρη έχουν ως εξής:

$$w_{\text{είδους}} = 0.4$$

$$w_{\text{μ.ο.βαθμολογίας}} = 0.22$$

$$w_{\text{πλήθος ψήφων}} = 0.18$$

$$w_{\text{δημοτικότητας}} = 0.20$$

Εξίσου σημαντικός στόχος αποτελεί προτάσεις που θα παράγει το σύστημα να είναι και αντικειμενικά καλές ταινίες. Αυτός είναι και ο λόγος που προτάθηκαν και τα υπόλοιπα βάρη. Οπότε το αμέσως επόμενο σημαντικό βάρος είναι αυτό του μέσου όρου της βαθμολογίας. Στην αρχική έκδοσή του συστήματός, όσον αφορά τη βαθμολογία μίας ταινίας, λαμβάνονταν υπόψιν σαν κριτήριο μονάχα ο μέσος όρος αυτός καθ' αυτός και όχι το πώς προέκυψε. Δηλαδή δεν συνυπολογίζονταν το πλήθος των ψήφων. Αυτό όμως δημιούργησε το εξής πρόβλημα: Έστω μία ταινία με μέσο όρο βαθμολογιών 4 και πλήθος ψήφων 1 και μία άλλη ταινία με μέσο όρο βαθμολογιών 3.2 και πλήθος ψήφων 2000. Κοιτώντας την βαθμολογία και μόνο, θα έβγαινε το συμπέρασμα πως η πρώτη πρόκειται για πολύ καλύτερη ταινία από τη δεύτερη, κάτι το οποίο αναιρείται όταν στην εξίσωση συμπεριλαμβάνεται και

το πλήθος των ψήφων, το οποίο φανερώνει, αφενός πως για την πρώτη ταινία δεν μπορεί να εξαχθεί ασφαλές συμπέρασμα όσον αφορά την ποιότητά της, καθώς το δείγμα θεατών είναι πρακτικά ανύπαρκτο, και αφετέρου πως η δεύτερη ταινία πρόκειται για μία, αντικειμενικά, μέτριας ποιότητας ταινίας. Οπότε, τα δύο βάρη ρυθμίσθηκαν ώστε να αντικατοπτρίζουν αυτήν την κατάσταση.

Τέλος, ένας ακόμη στόχος είναι μία ταινία εκτός από ποιοτική να είναι και δημοφιλής. Παρότι διαισθητικά μία αντικειμενικά ποιοτική ταινία είναι και δημοφιλής, στην πράξη αυτό δεν ισχύει πάντα. Υπάρχουν ταινίες οι οποίες στα διεθνή φεστιβάλ κινηματογράφου κάνουν θραύση, ταινίες που είναι κινηματογραφικά αριστουργήματα, με πολλά βραβεία στην κατοχή τους, που παρόλα αυτά απευθύνονται σε πολύ μικρή μερίδα κοινού. Στην υλοποίηση αυτή τέθηκε ως στόχος η δημοτικότητα μίας ταινίας να λαμβάνει μέρος στην διαδικασία παραγωγής προτάσεων ώστε να αυξηθούν οι πιθανότητες να αρέσει στον χρήστη.

Η δημοτικότητα είναι ένα μετά δεδομένο το οποίο ανακτήθηκε από την [βάση του TMDB](#). Το TMDB χρησιμοποιεί διαφορετικά μοντέλα προκειμένου να υπολογίσει τη δημοτικότητα ταινιών, τηλεοπτικών προγραμμάτων και ανθρώπων. Για τον υπολογισμό της δημοτικότητας των ταινιών λαμβάνει υπόψιν τα παρακάτω<sup>10</sup>:

- Πλήθος ψήφων ημέρας
- Πλήθος προβολών ημέρας
- Πλήθος χρηστών που έβαλαν την ταινία στη λίστα των αγαπημένων τους σήμερα
- Πλήθος χρηστών που έβαλαν σήμερα την ταινία στη λίστα των ταινιών που έχουν παρακολουθήσει
- Ημερομηνία κυκλοφορίας
- Πλήθος ψήφων
- Σκορ (δημοτικότητας) προηγούμενων ημερών

Μία παρατήρηση που μπορεί να γίνει είναι πως ενώ η βαθμολογία των ταινιών από τη βάση του TMDB δεν χρησιμοποιείται άμεσα, αυτή συμπεριλαμβάνεται στον υπολογισμό της δημοτικότητας της κάθε ταινίας. Λαμβάνοντας υπόψιν ότι οι βαθμολογίες από το σύνολο δεδομένων του MovieLens είναι αρκετά παλιές και η κουλτούρα της κοινωνίας έχει αλλάξει με την πάροδο των χρόνων, το γεγονός πως, έστω και με αυτόν τον έμμεσο τρόπο,

---

<sup>10</sup> Ανάκτηση από: <https://developers.themoviedb.org/3/getting-started/popularity>

συμπεριλαμβάνονται στη συνάρτηση και πιο σύγχρονες, εν δυνάμει, βαθμολογίες, αυξάνει την αντικειμενικότητα των προτάσεων.

#### 4.3.5 Παραγωγή προτάσεων

Έχοντας ορίσει τα επιμέρους στοιχεία που χρειάζονται προκειμένου το σύστημα να παράξει τις προτάσεις, ήρθε η ώρα να συνδυαστούν όλα μαζί για την παραγωγή προτάσεων.

Μία διαίσθηση του αλγόριθμου σε ψευδοκώδικα είναι η εξής:

```
Προσκόμισε τις ταινίες τις οποίες δεν έχει δει ο χρήστης
Προσκόμισε τα προτιμητέα είδη του χρήστη
Κανονικοποίησε τα είδη με βάση το πλήθος των ειδών ανά ταινία
Για κάθε κριτήριο:
    Όρισε το βάρος του
    Βρες τη μέγιστη τιμή στο σύνολο των ταινιών
Για κάθε ταινία
    Κανονικοποίησε την τιμή κάθε κριτηρίου
    Πολλαπλασίασε με το βάρος κάθε κριτηρίου
    Πρόσθεσε όλες τις τιμές για να βρεις την βαθμολογία
Ταξινόμησε σε αύξουσα σειρά τη λίστα με βάση την βαθμολογία
```

Έχει ήδη αναφερθεί τι συμβαίνει στις πρώτες 7 γραμμές του ψευδοκώδικα, οπότε θα αναλυθούν οι υπόλοιπες.

Όπως έχει ήδη [αναφερθεί](#) ο αλγόριθμος WSM χρησιμοποιεί δύο ειδών ιδιότητες, τις ωφέλιμες και τις ανώφελες, με κάθε μία να χρησιμοποιεί διαφορετική μέθοδο υπολογισμού. Αυτές οι μέθοδοι θα χρησιμοποιηθούν προκειμένου να κανονικοποιηθούν οι τιμές των ιδιοτήτων. Στην περίπτωση αυτού του συστήματος, όλες οι ιδιότητες είναι ωφέλιμες, που σημαίνει ότι η σύγκριση θα γίνει με την μέγιστη τιμή της εκάστοτε ιδιότητας και πως όλες οι ιδιότητες δρουν θετικά ως προς το αποτέλεσμα. Συνεπώς, για κάθε στοιχείο κάθε ιδιότητας – κριτηρίου, διαιρείται η τιμή του με τη μέγιστη τιμή που λαμβάνει αυτή η ιδιότητα και στη συνέχεια το πηλίκο πολλαπλασιάζεται με το εκάστοτε βάρος προκειμένου να της ανατεθεί ο βαθμός (score) κατά τον οποίο συμμετέχει στο συνολικό βαθμό καταλληλότητας πρότασης της συγκεκριμένης ταινίας. Τέλος, αρκεί να αθροιστούν όλοι οι επιμέρους βαθμοί των ιδιοτήτων προκειμένου να υπολογιστεί ο γενικός βαθμός πρότασης της ταινίας.

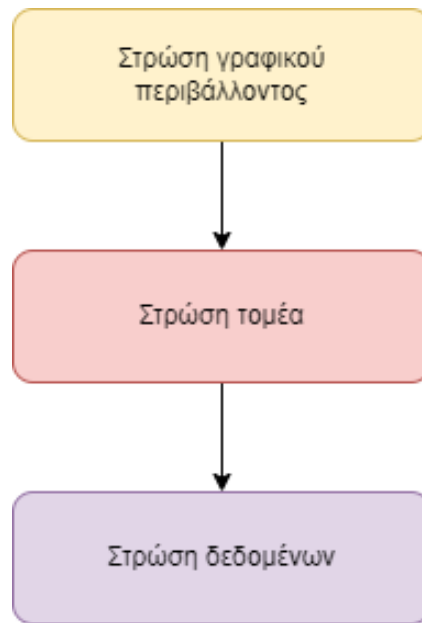
## 4.4 Αρχιτεκτονική εφαρμογής

Οι σημερινές απαιτήσεις που έχει ένας χρήστης από μία εφαρμογή κινητής συσκευής είναι πολύ περισσότερες απ' ό,τι όταν τα πρώτα έξυπνα τηλέφωνα δημιουργήθηκαν. Προκειμένου να μπορέσουν να καλύψουν αυτές τις απαιτήσεις, οι εφαρμογές μεγάλωσαν σε μέγεθος και έγιναν πιο περίπλοκες. Μία τυπική, βασική, εφαρμογή χρειάζεται να αλληλοεπιδρά με τον χρήστη, να αποθηκεύει και να διαχειρίζεται δεδομένα, να έχει πρόσβαση στο διαδίκτυο και να χρησιμοποιεί διάφορους από τους διαθέσιμους αισθητήρες μιας κινητής συσκευής. Όπως είναι φανερό, κάθε μία από αυτές τις λειτουργίες απαιτεί διαφορετική λογική και νοηματικά είναι ανεξάρτητη η μία από την άλλη.

Προκειμένου να δώσει μία λύση σε αυτό το πρόβλημα η Google πρότεινε μία αρχιτεκτονική καθώς κι ένα σύνολο καλών πρακτικών τα οποία συνιστά έντονα να χρησιμοποιούνται από τους προγραμματιστές. Μια αρχιτεκτονική μπορεί να οριστεί ως μία αφαιρετική (abstraction) περιγραφή ενός συστήματος. Αποτελείται από δομές και οι δομές με τη σειρά τους αποτελούνται από στοιχεία (components) και σχέσεις, συνεπώς μπορεί να ειπωθεί πως μία αρχιτεκτονική αποτελείται από προγραμματιστικά στοιχεία και πώς αυτά αλληλοεπιδρούν μεταξύ τους.

Σύμφωνα με την αρχιτεκτονική που πρότεινε η Google, κάθε εφαρμογή πρέπει να αποτελείται από τουλάχιστον δύο στρώσεις (layers): τη στρώση του γραφικού περιβάλλοντος (UI layer) και τη στρώση των δεδομένων (data layer). Προαιρετικά μπορεί να υπάρχει και μία ακόμη στρώση, η οποία τοποθετείται ανάμεσα στις δύο προαναφερθείσες, το domain layer, όπου η δουλειά του είναι να απλοποιήσει και να βοηθήσει στην επαναχρησιμοποίηση των διαδικασιών μεταξύ του γραφικού περιβάλλοντος και των δεδομένων. (Kaseb, 2022)



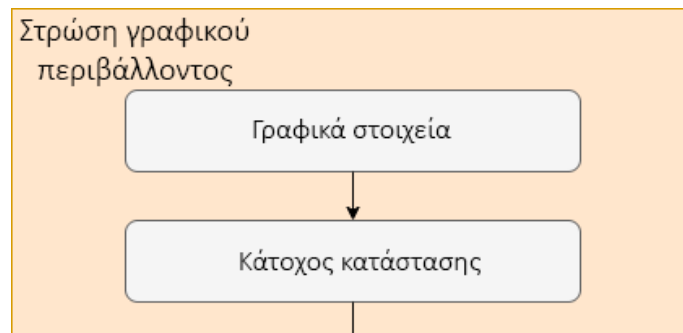


**Εικόνα 12. Γενικό διάγραμμα της αρχιτεκτονικής μιας τυπικής εφαρμογής. Η στρώση του γραφικού περιβάλλοντος λαμβάνει τα δεδομένα της εφαρμογής από την προαιρετική στρώση τομέα ή τη στρώση δεδομένων**

#### **4.4.1 Στρώση γραφικού περιβάλλοντος**

Ο σκοπός της στρώσης αυτής (που έχει και την εναλλακτική ονομασία «στρώση παρουσίασης» (presentation layer)) είναι να εμφανίζει στην οθόνη τα δεδομένα της εφαρμογής. Το γραφικό περιβάλλον οφείλει να αντικατοπτρίζει τις αλλαγές των δεδομένων, είτε αυτά προέρχονται ως αποτέλεσμα αλληλεπίδρασης της εφαρμογής με τον χρήστη, είτε από εξωτερικές πηγές (όπως το διαδίκτυο ή κάποιος αισθητήρας). Ουσιαστικά, το γραφικό περιβάλλον είναι μία οπτική αναπαράσταση της κατάστασης της εφαρμογής, σύμφωνα με τα δεδομένα της στρώσης δεδομένων.

Η στρώση αυτή αποτελείται από δύο ομάδες στοιχείων: τα γραφικά στοιχεία (UI elements) τα οποία προβάλλουν τα γραφικά στην οθόνη, και τα State holders, στοιχεία τα οποία περιέχουν δεδομένα τα οποία εκθέτουν στο γραφικό περιβάλλον καθώς και εμπεριέχουν τη λογική αυτών.



**Εικόνα 13. Η στρώση του γραφικού περιβάλλοντος περιλαμβάνει κατηγορίες στοιχείων, καθένα από τα οποία έχει τον δικό του σκοπό.**

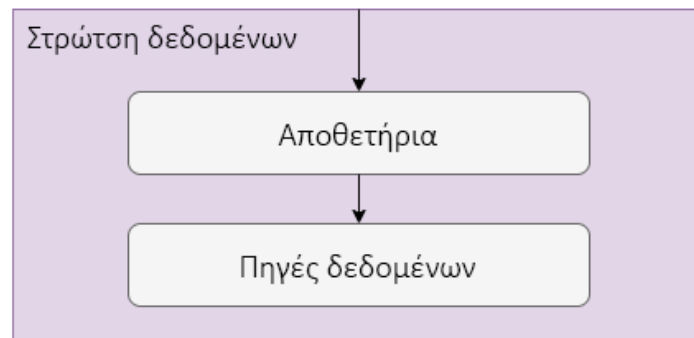
Η αρμοδιότητες που περιλαμβάνει η στρώση αυτή είναι:

- Η μετατροπή των δεδομένων σε μορφή που ανταποκρίνεται στο γραφικό περιβάλλον
- Η ενημέρωση του γραφικού περιβάλλοντος ώστε να αντικατοπτρίζει αυτά τα δεδομένα
- Η διαχείριση της συμπεριφοράς του χρήστη που οδηγεί σε μεταβολές του γραφικού περιβάλλοντος
- Να επαναλαμβάνει τα προαναφερθέντα βήματα για όσο χρειαστεί

#### **4.4.2 Στρώση δεδομένων**

Η στρώση των δεδομένων σε μία εφαρμογή περιέχει το business logic της, δηλαδή ένα σύνολο κανόνων που ορίζουν τον τρόπο με τον οποίο η εφαρμογή δημιουργεί, αποθηκεύει και μεταβάλλει δεδομένα.

Σύμφωνα με την αρχιτεκτονική που έχει προτείνει η Google, η στρώση αυτή πρέπει να αποτελείτε από αποθετήρια (repositories), ξεχωριστά για κάθε τύπο δεδομένων που διαχειρίζεται η εφαρμογή, καθένα από τα οποία μπορεί να περιλαμβάνει περισσότερες από μία πηγές δεδομένων (όπως ένα API ή μία τοπική βάση δεδομένων).



**Εικόνα 14. Τα μέρη από τα οποία αποτελείται η στρώση δεδομένων. Τα αποθετήρια αποτελούν ειδικές κλάσεις που περιέχουν πηγές δεδομένων από τις οποίες λαμβάνουν τα, σχετικά με την εφαρμογή, δεδομένα.**

Τα αποθετήρια σαν οντότητες έχουν ορισμένες ευθύνες:

- Να εκθέτουν τα δεδομένα της εφαρμογής στα υπόλοιπα μέρη της
- Να είναι το μοναδικό μέρος στο οποίο πραγματοποιούνται μεταβολές στα δεδομένα
- Να επιλύουν τυχών συγκρούσεις μεταξύ πολλαπλών πηγών δεδομένων
- Να κρύβουν τις πηγές δεδομένων από την υπόλοιπη εφαρμογή
- Να περιέχουν το business logic

Κάθε κλάση πηγής δεδομένων πρέπει να είναι υπεύθυνη μονάχα για μία πηγή, όπως ένα αρχείο, μία δικτυακή πηγή ή μια τοπική βάση δεδομένων. Οι κλάσεις αυτές αποτελούν τη γέφυρα μεταξύ την εφαρμογής και του συστήματος για εργασίες δεδομένων.

Καμία άλλη στρώση πέραν αυτής δεν πρέπει να έχει απευθείας πρόσβαση στα δεδομένα· το σημείο εισόδου στη στρώση των δεδομένων είναι πάντοτε τα αποθετήρια. Με τον τρόπο αυτόν, είναι εφικτή η ανεξάρτητη κλιμάκωση (scaling) των διάφορων στρώσεων.

Είναι σημαντικό κάθε αποθετήριο να ορίζει μία μοναδική πηγή αλήθειας (single source of truth). Η πηγή αυτή περιέχει πάντα δεδομένα τα οποία είναι συνεπή, ορθά και ενήμερα. Πιο συγκεκριμένα, τα δεδομένα που εκτίθενται από το αποθετήριο πρέπει να προέρχονται απευθείας από αυτήν την πηγή.

#### **4.4.3 Android Jetpack**

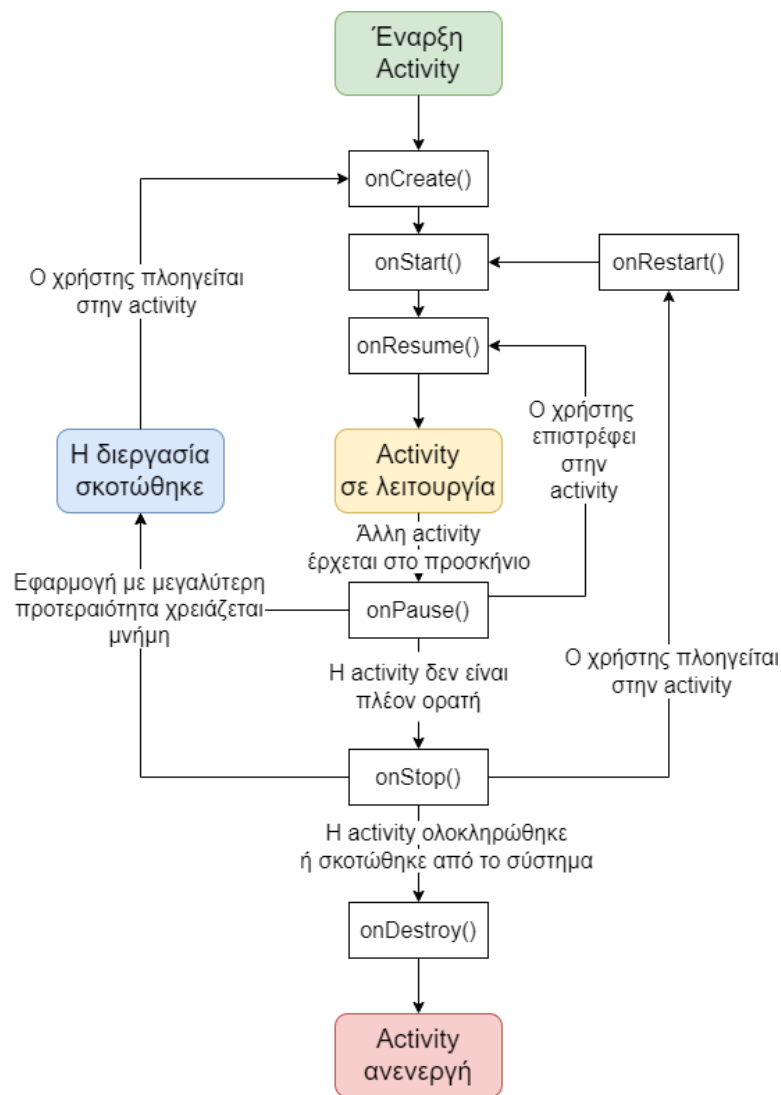
Το Jetpack είναι μία σουίτα βιβλιοθηκών που σκοπό έχουν να βοηθήσουν τους προγραμματιστές να ακολουθήσουν καλές πρακτικές, να μειώσουν τον τετριμμένο (boilerplate) κώδικα, καθώς και να γράψουν κώδικα ο οποίος δουλεύει με συνέπεια τόσο σε διαφορετικές εκδόσεις του Android όσο και σε διαφορετικές συσκευές.

Μέρος του Jetpack αποτελεί η κλάση ViewModel η οποία έχει σχεδιαστεί προκειμένου να φυλάσσει (store) και να διαχειρίζεται δεδομένα που αφορούν το γραφικό περιβάλλον, λαμβάνοντας υπόψιν τον κύκλο ζωής (lifecycle). Με τη χρήση της κλάσης αυτής τα δεδομένα μας επιβιώνουν από μεταβολές του περιβάλλοντος εκτέλεσης (configuration changes), όπως η περιστροφή του κινητού.

Στο λειτουργικό σύστημα Android υπάρχει η έννοια του κύκλου ζωής. Ο κύκλος ζωής περιγράφει τις διάφορες καταστάσεις στις οποίες μπορεί να βρεθεί κάποιο αντικείμενο, ύστερα από μεταβολές στο περιβάλλον εκτέλεσης. Παραδείγματα τέτοιων μεταβολών είναι:

- Η περιστροφή του κινητού
- Το κλείσιμο της εφαρμογής
- Το άνοιγμα της εφαρμογής
- Μετάβαση σε διαφορετική οθόνη (screen)
- Αναγκαστική διακοπή της εφαρμογής από κάποια άλλη

Το Android Framework διαχειρίζεται τον κύκλο ζωής των δραστηριοτήτων και των κομματιών (fragments). Το λειτουργικό ενδέχεται να αποφασίσει να τα καταστρέψει ή να τα επαναδημιουργήσει ως αντίδραση σε πράξεις του χρήστη ή γεγονότα (events) σχετιζόμενα με την ίδια τη συσκευή, στα οποία ο χρήστης δεν έχει κανέναν έλεγχο.

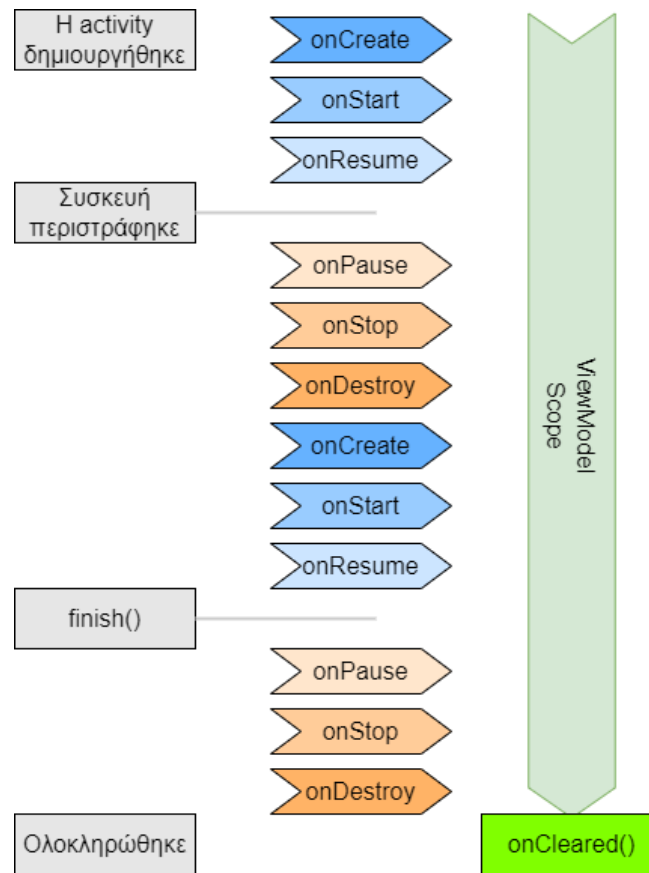


**Εικόνα 15 Απλοποιημένη εικονογράφηση του κύκλου ζωής μίας δραστηριότητας**

Όταν αυτό συμβαίνει, τα δεδομένα που είναι αποθηκευμένα σε αυτά τα αντικείμενα χάνονται επίσης. Αντικείμενα όπως οι δραστηριότητες και τα κομμάτια έχουν πρωτεύοντα σκοπό να εμφανίζουν γραφικά στοιχεία, να ανταποκρίνονται σε ενέργειες του χρήστη ή να διαχειρίζονται την επικοινωνία με το λειτουργικό σύστημα, όπως τα αιτήματα δικαιωμάτων. Η ανάθεση περισσότερων ευθυνών όπως η ανάκτηση δεδομένων από μία βάση ή από μία διαδικτυακή πηγή αποτελεί περιττό (bloat) κώδικα για τον σκοπό της κλάσης. Αυτό θα έχει ως αποτέλεσμα μόνο μία κλάση να αναλαμβάνει να εκτελέσει ολόκληρη τη δουλειά της εφαρμογής μόνη της. Είναι ευκολότερο και αποτελεσματικότερο να γίνει διαχωρισμός των δεδομένων από το γραφικό περιβάλλον.

Η αρχιτεκτονική που ανέπτυξε η Google περιλαμβάνει τα ViewModels για αυτόν τον σκοπό. Αποτελούν βοηθητικές κλάσεις του γραφικού περιβάλλοντος με την αρμοδιότητα να

του προετοιμάζουν τα δεδομένα. Τα αντικείμενα αυτά αυτόματα διατηρούνται από αλλαγές του περιβάλλοντος εκτέλεσης και είναι άμεσα ξανά διαθέσιμα στο επόμενο στιγμιότυπο της δραστηριότητας ή του κομματιού. Η κλάση αυτή είναι έτσι σχεδιασμένη ώστε να μην υπακούει ακριβώς σε κάθε συμβάν του κύκλου ζωής του γραφικού περιβάλλοντος.



**Εικόνα 16.** Αριστερά οι διάφορες καταστάσεις μία δραστηριότητας ύστερα από περιστροφή της συσκευής. Δεξιά φαίνεται η διάρκεια ζωής ενός στιγμιότυπου της κλάσης ViewModel που έχει συνδεθεί με αυτήν την δραστηριότητα

Στην υλοποίησή του συστήματος αντικείμενα της κλάσης αυτής έχουν χρησιμοποιηθεί κατά κόρων, σύμφωνα με την προαναφερθείσα αρχιτεκτονική και ακολουθώντας τις καλές πρακτικές που προτείνει η Google.

#### 4.4.4 Single Activity

Στο Google I/O<sup>11</sup> του 2018 αναφέρθηκε για πρώτη φορά από την Google η αρχιτεκτονική της Ατομικής Δραστηριότητας (Single Activity Architecture), η οποία περιλαμβάνει είτε μία και μοναδική δραστηριότητα είτε έναν μικρό αριθμό αυτών. Στην αρχιτεκτονική αυτή οι δραστηριότητες, αντί να αντιπροσωπεύουν μία ξεχωριστή οθόνη, νοούνται περισσότερο σαν ένα μεγάλο δοχείο (container) με τα κομμάτια (fragments) να έχουν το ρόλο της οθόνης.

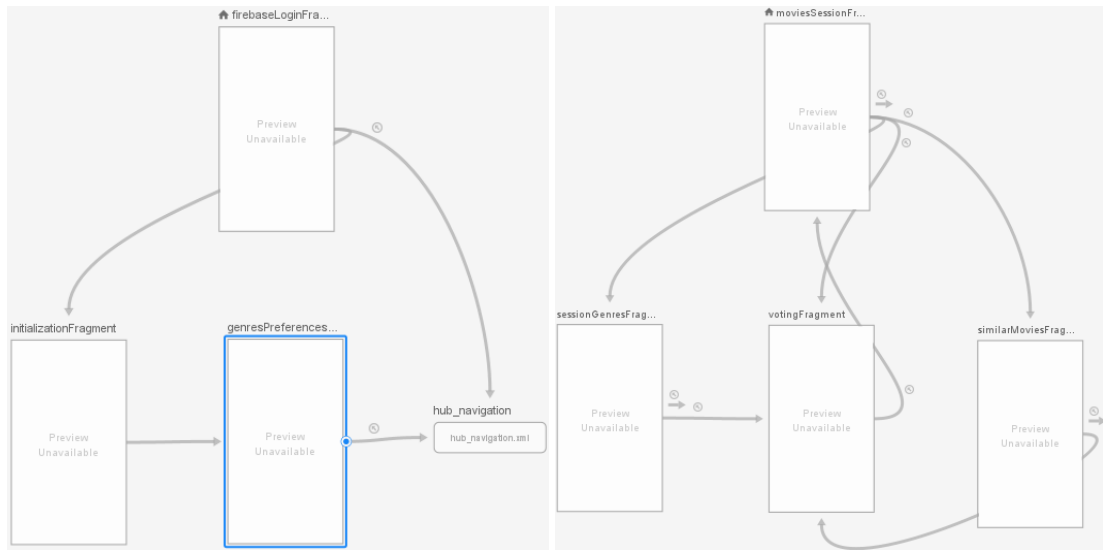
Η προσέγγιση αυτή συνυπάρχει άψογα μαζί με ένα από τα στοιχεία του Jetpack, το οποίο αναλαμβάνει την πλοήγηση (navigation) της εφαρμογής. Ο όρος πλοήγηση αναφέρεται στις αλληλεπιδράσεις που επιτρέπουν στον χρήστη να μα περιηγηθεί σε διάφορα μέρη της εφαρμογής. Το Navigation Component είναι μια βιβλιοθήκη όπου απλοποιεί αυτήν την διαδικασία και εξασφαλίζει μία συνεπή και προβλέψιμη εμπειρία για τον χρήστη, ακολουθώντας κάποιους καθιερωμένους βασικούς κανόνες.

Το navigation component περιλαμβάνει τα γραφήματα πλοήγησης (navigation graph) τα οποία είναι ειδικά xml αρχεία που περιλαμβάνουν πληροφορίες για τους διάφορους προορισμούς. Στην ανάπτυξη του συστήματος, έγινε χρήση αυτής της αρχιτεκτονικής, έχοντας μόνο μία δραστηριότητα και αναπαριστώντας όλες τις οθόνες με fragments. Ωστόσο έχει γίνει μία λογική ομαδοποίηση, χρησιμοποιώντας ξεχωριστά γραφήματα περιήγησης για την κάθε ομάδα.



Εικόνα 17 Γράφημα περιήγησης κεντρικού hub

<sup>11</sup> Ετήσιο συνέδριο προγραμματιστών που διοργανώνεται από την Google



**Εικόνα 18** Αριστερά το γράφημα περιήγησης της διαδικασίας σύνδεσης. Δεξιά το γράφημα περιήγησης μια συνεδρίας



## Κεφάλαιο 5. Περιπτώσεις χρήσης και Αξιολόγηση

Στο κεφάλαιο αυτό θα παρουσιαστούν κάποιες αντιπροσωπευτικές περιπτώσεις χρήσης παραγωγής προτάσεων και τη σχετική αξιολόγηση που λήφθηκε από μία ομάδα που δοκίμασε την εφαρμογή.

### 5.1 Περιπτώσεις χρήσης

Στην παρουσίαση των περιπτώσεων χρήσης θα ακολουθηθεί η εξής μορφή:

- Οι προτιμήσεις του χρήστη θα αναπαρασταθούν ως ένα σύνολο από τις αγγλικές ονομασίες των ειδών
- Θα παρουσιάζονται οι τρεις πρώτες και οι 3 τελευταίες προτάσεις. Η κάθε παραχθείσα πρόταση θα αναπαρασταθεί ως ένα σύνολο το οποίο θα περιλαμβάνει:
  - Τον αγγλικό τίτλο της ταινίας
  - Ο βαθμός που της αποδόθηκε ως προτίμηση από το σύστημα προτάσεων
  - Ο βαθμός αντιστοίχισης με τα προτιμητέα είδη του χρήστη
  - Τα είδη στα οποία ανήκει
  - Η δημοτικότητα της
  - Ο μέσος όρος βαθμολογίας της
  - Το πλήθος ψήφων

#### 5.1.1 Πρώτη περίπτωση χρήσης

Ο χρήστης Χ έχει τις εξής προτιμήσεις σε είδη:

$$p_X = \{Fantasy, Action\}$$

και το σύστημα προτάσεων παράγαγε τις εξής προτάσεις:

$$r_1 = \{Batman Returns, 0.7018, 1.0, \{Fantasy, Action\}, 149.467, 6.9, 5183\}$$

$$r_2 = \{The Matrix, 0.6173, 0.5, \{Action, Sci - fi\}, 88.288, 8.2, 21405\}$$

$$r_3 = \{Death Wish, 0.5834, 1.0, \{Action\}, 16.553, 6.9, 580\}$$

$$r_{18} = \{Unforgettable, 0.5421, 1.0, \{Action\}, 7.375, 5.6, 76\}$$

$$r_{19} = \{The Mummy, 0.5407, 0.66, \{Action, Adventure, Fantasy\}, 79.723, 6.9, 7327\}$$

$$r_{20} = \{Ben - Hur, 0.5395, 0.33, \{Action, Drama, History\}, 317.662, 7.8, 2069\}$$

Κοιτώντας τα αποτελέσματα κάποιες ενδιαφέρουσες παρατηρήσεις προκύπτουν. Αρχικά, συγκρίνοντας τις πρώτες δύο προτάσεις, είναι φανερό πως ενώ η ταινία «The Matrix» έχει αρκετά υψηλότερο μέσο όρο βαθμολογίας και από πολύ περισσότερες ψήφους, επιλέχθηκε ως δεύτερη πρόταση από το σύστημα. Ο λόγος είναι πως το «Batman Returns», αν και λιγότερο ποιοτική (σύμφωνα με το μέσο όρο βαθμολογίας) είναι πολύ πιο δημοφιλής από το «The Matrix» και κυρίως, έχει απόλυτη ταύτιση με τα προτιμητέα είδη του χρήστη. Οπότε η συζήτηση που έγινε στο κεφάλαιο [4.3.4](#) σχετικά με τα βάρη επαληθεύεται και στην πράξη, με το σύστημα προτάσεων να έχει δώσει μεγαλύτερο βάρος στην ταύτιση των ειδών, δηλαδή την άμεση ιδιότητα, σε σχέση με τις έμμεσες ιδιότητες των ταινιών.

Συγκρίνοντας την τρίτη πρόταση με την τελευταία φαίνεται πως η ταινία «Ben-Hur» είναι πολύ πιο δημοφιλής από το «Death Wish», για την ακρίβεια είναι πιο δημοφιλής από όλες τις ταινίες στο σύνολο προτάσεων, με μεγαλύτερο μέσο όρο βαθμολογίας και περισσότερες ψήφους. Παρ' όλα αυτά βρίσκεται στο τέλος του συνόλου προτάσεων. Ο λόγος κι εδώ είναι τα είδη. Το «Ben-Hur» έχει αντιστοιχία κατά 1/3 με τα προτιμητέα είδη του χρήστη, σε αντίθεση με το «Death Wish» του οποίου το ένα και μοναδικό είδος στο οποίο ανήκει, συγκαταλέγεται στα προτιμητέα είδη του χρήστη.

### 5.1.2 Δεύτερη περίπτωση χρήσης

Ο χρήστης Υ έχει τις εξής προτιμήσεις σε είδη:

$$p_Y = \{Action, Adventure, Fantasy, Drama\}$$

και το σύστημα προτάσεων παράγαγε τις εξής προτάσεις:

$$r_1 = \{The Matrix, 0.6173, 0.5, \{Action, Sci - Fi\}, 88.288, 8.2, 21405\}$$

$$r_2 = \{Death Wish, 0.5834, 1.0, \{Action\}, 16.553, 6.9, 580\}$$

$$r_3 = \{Hard Target, 0.5833, 1.0, \{Action\}, 36.617, 6.3, 837\}$$

$$r_{18} = \{Best of the Best 3, 0.5326, 1.0, \{Action\}, 8.283, 5.2, 41\}$$

$$r_{19} = \{Let's Get Hurry, 0.5168, 1.0, \{Action\}, 2.965, 4.7, 12\}$$

$$r_{20} = \{Batman Returns, 0.5018, 0.5, \{Action, Fantasy\}, 149.467, 6.9, 5183\}$$

Η συγκεκριμένη περίπτωση χρήσης παρουσιάζει ιδιαίτερο ενδιαφέρον επειδή εμφανίζονται ξανά ταινίες οι οποίες είχαν εμφανιστεί και στις προτάσεις για τον χρήστη Χ.

Αρχικά, παρατηρείται η χαοτική διαφορά στην κατάταξη μεταξύ των ταινιών «The Matrix» και «Batman Returns». Στις προτάσεις της προηγούμενης περίπτωσης χρήσης, οι δύο ταινίες όχι μόνο απείχαν μεταξύ τους στην κατάταξη μονάχα μία θέση, αλλά και το «Batman Returns» είχε εμφανιστεί ως καλύτερη πρόταση από το «The Matrix». Θα μπορούσε να

υποτεθεί πως αυτό οφείλεται στην μεγαλύτερη ταύτιση της πρώτης πρότασης με τα προτιμητέα είδη του χρήστη, ωστόσο αυτό δεν επαληθεύεται καθότι και οι δύο ταινίες έχουν ίδιο βαθμό αντιστοίχισης ειδών (2/4). Η αιτία είναι οι μεγάλες διαφορές που παρατηρούνται στις έμμεσες ιδιότητες. Πιο συγκεκριμένα, μπορεί το «Batman Returns» να είναι πολύ πιο δημοφιλής σαν ταινία, ωστόσο το «The Matrix» έχει αρκετά υψηλότερη βαθμολογία και με πολλές περισσότερες ψήφους, γεγονός που την καθιστά μία, κατά τα δεδομένα, αντικειμενικά καλύτερη ταινία.

Αξίζει επίσης να αναφερθεί πως, ενώ πράγματι η ιδιότητα των ειδών έχει το μεγαλύτερο ατομικό βάρος, συλλογικά, το άθροισμα όλων των υπόλοιπων ιδιοτήτων είναι μεγαλύτερο (0.6 έναντι 0.4) γεγονός που επαληθεύεται και στην πράξη μιας και τόσο η πρώτη όσο και η τελευταία πρόταση έχουν βαθμό ειδών 0.2, ωστόσο το άθροισμα των υπόλοιπων ειδών είναι  $\sim 0.41728$  και  $\sim 0.30175$  αντίστοιχα<sup>12</sup>.

Μια ακόμη παρατήρηση που μπορεί να γίνει είναι μεταξύ της δεύτερης και τρίτης πρότασης. Και οι δύο ταινίες ανήκουν στο ίδιο είδος και για τον λόγο αυτό θα περίμενε κανείς πως οι θέσεις θα έπρεπε να ήταν αντίστροφες, και η τρίτη πρόταση να ήταν στη δεύτερη θέση καθότι είναι περισσότερο δημοφιλής και με περισσότερες ψήφους. Αυτό που συμβαίνει ωστόσο είναι πως επειδή ο μέσος όρος βαθμολογίας έχει μεγαλύτερο βάρος από τις δύο προαναφερθείσες ιδιότητες και επειδή η δεύτερη πρόταση έχει ελαφρός καλύτερη βαθμολογία από την τρίτη, καταλήγει να έχει αυτό το απειροελάχιστο προβάδισμα.

---

<sup>12</sup> Πιο συγκεκριμένα, για το “The Matrix”:

- Βαθμός μέσου όρου βαθμολογίας = 0.20044
- Βαθμός δημοτικότητας = 0.05555
- Βαθμός πλήθους ψήφων = 0.16129

Και για το «Batman Returns»:

- Βαθμός μέσου όρου βαθμολογίας = 0.16866
- Βαθμός δημοτικότητας = 0.09410
- Βαθμός πλήθους ψήφων = 0.0390

### 5.1.3 Τρίτη περίπτωση χρήσης

Ο χρήστης Κ έχει τις εξής προτιμήσεις σε είδη:

$$p_K = \{Animation, Family, Fantasy, Romance, Sci - fi, Drama\}$$

και το σύστημα προτάσεων παρήγαγε τις εξής προτάσεις:

$$r_1 = \{Titanic, 0.8675, 1.0, \{Drama, Romance\}, 186.34, 7.9, 20855\}$$

$$r_2 = \{The Lion King, 0.8637, 1.0, \{Family, Animation, Drama\}, 232.504, 8.3, 15197\}$$

$$r_3 = \{Fight Club, 0.8228, 1.0, \{Drama\}, 69.349, 8.4, 23887\}$$

$$r_{18} = \{Bambi, 0.6768, 1.0, \{Animation, Drama, Family\}, 111.992, 7.0, 4678\}$$

$$r_{19} = \{12 Angry Men, 0.6751, 1.0, \{Drama\}, 31.001, 8.5, 6344\}$$

$$r_{20} = \{Dumbo, 0.6688, 1.0, \{Animation, Family\}, 106.074, 7.0, 4103\}$$

Στην περίπτωση αυτή, όπου υπάρχουν και οι περισσότερες επιλογές προτιμήσεων από πλευράς χρήστη, φαίνεται πως η διαφορά του βαθμού πρότασης είναι πολύ μικρή μεταξύ των ταινιών στην αμέσως επόμενη θέση κατάταξης. Παρατηρείται ωστόσο ότι, όπως είναι αναμενόμενο λόγω των περισσότερων διαθέσιμων επιλογών, υπάρχει μεγαλύτερη ποικιλία στις προτάσεις. Για παράδειγμα, αν και οι δύο πρώτες προτάσεις έχουν πρακτικά απειροελάχιστη διαφορά στον βαθμό καταλληλότητας πρότασης, πρόκειται για δύο εντελώς διαφορετικές ταινίες. Παρατηρείται, επίσης, ότι αν και η ταινία «The Lion King» είναι αρκετά πιο δημοφιλής και με λίγο μεγαλύτερο μέσο όρο βαθμολογίας (κριτήριο με μεγαλύτερο βάρος σε σχέση με τη δημοτικότητα και το πλήθος των ψήφων) καταλήγει με ελάχιστα μικρότερη βαθμολογία σε σχέση με την ταινία «Titanic», η οποία, αν και λιγότερο δημοφιλής και λιγότερο ποιοτική (με βάση το μέσο όρο της βαθμολογίας), κατέχει περισσότερους ψήφους. Η ίδια ακριβώς κατάσταση επικρατεί και στο τέλος της λίστας προτάσεων, με τις δύο τελευταίες ταινίες να έχουν μικρή διαφορά στον βαθμό καταλληλότητας, αλλά να ανήκουν σε εντελώς διαφορετικά είδη.

Μία ακόμη μικρή παρατήρηση που μπορεί να γίνει είναι ότι, λόγω των περισσότερων επιλογών του χρήστη ως προς τα είδη, όλες οι προτάσεις έχουν είδη που ταυτίζονται εντελώς με τις προτιμήσεις του χρήστη, σε αντίθεση με τις προηγούμενες περιπτώσεις χρήσης όπου υπήρξαν και προτάσεις με μερική ταύτιση.

### 5.1.4 Τέταρτη περίπτωση χρήσης

Ο χρήστης R έχει τις εξής προτιμήσεις σε είδη:

$$p_R = \{Comedy, Family, Horror, Thriller\}$$

και το σύστημα προτάσεων παρήγαγε τις εξής προτάσεις:

$$r_1 = \{The\ Shining, 0.7543, 1.0, \{Horror, Thriller\}, 73.364, 8.2, 14294\}$$

$$r_2 = \{The\ Exorcist, 0.7219, 1.0, \{Horror\}, 137.727, 7.7, 6235\}$$

$$r_3 = \{The\ Parent\ Trap, 0.7158, 1.0, \{Comedy, Family\}, 187.187, 7.1, 3247\}$$

$$r_{18} = \{Young\ Frankenstein, 0.6257, 1.0, \{Comedy\}, 21.469, 7.9, 2535\}$$

$$r_{19} = \{Carrie, 0.6219, 1.0, \{Horror, Thriller\}, 34.599, 7.3, 2885\}$$

$$r_{20} = \{Airplane!, 0.6213, 1.0, \{Comedy\}, 26.255, 7.3, 3503\}$$

Σε αυτήν την περίπτωση χρήσης εσκεμμένα συμπεριλαμβάνονται είδη ταινιών τα οποία αντιπροσωπεύουν τα δύο άκρα (αλλά παρ' όλα αυτά δεν είναι απίθανο να αντιστοιχούν στις προτιμήσεις ενός πραγματικού ατόμου) προκειμένου να εξεταστεί πώς θα συμπεριφερθεί ο αλγόριθμος.

Δεν υπάρχει κάποια αξιολογή παρατήρηση που μπορεί να γίνει παρατηρώντας το συγκεκριμένο σύνολο προτάσεων, πέραν του γεγονότος πως το σύστημα προτάσεων κατάφερε επιτυχώς να εντοπίσει και να προτείνει ταινίες και για τις δύο, λογικές, ομάδες των ειδών.

### 5.1.5 Πέμπτη περίπτωση χρήσης

Ο χρήστης T έχει της εξής προτίμηση σε είδος:

$$p_T = \{Animation\}$$

και το σύστημα προτάσεων παρήγαγε τις εξής προτάσεις:

$$r_1 = \{The\ Lion\ King, 0.7304, 0.33, \{Family, Animation, Drama\}, 232.504, 8.3, 15197\}$$

$$r_2 = \{Pinocchio, 0.6894, 0.5, \{Animation, Family\}, 126.448, 7.1, 4809\}$$

$$r_3 = \{Dumbo, 0.6688, 0.5, \{Animation, Family\}, 106.074, 7.0, 4103\}$$

$$r_{18}$$

$$= \{Snow\ White\ and\ the\ Seven\ Dwarfs, 0.5728, 0.33, \{Fantasy, Animation, Family\}, 137.26, 7.1, 6133\}$$

$$r_{19}$$

$$= \{The\ Secret\ Adventures\ of\ Tom\ Thumb, 0.5656, 0.5, \{Animation, Scifi\}, 2.549, 6.7, 29\}$$

$$r_{20}$$

$$= \{The\ Jungle\ Book, 0.5499, 0.33, \{Family, Animation, Adventure\}, 103.825, 7.3, 5241\}$$

Σε αυτήν την περίπτωση χρήσης σκοπός ήταν να εξεταστεί πώς θα ανταποκριθεί το σύστημα προτάσεων έχοντας ως περιορισμό ότι ο χρήστης έχει μία και μοναδική προτίμηση.

Αρχικά παρατηρείται ότι καμία από τις προτάσεις δεν βρίσκει πλήρη ταύτιση με την μία και μοναδική προτίμηση του χρήστη. Για την ακρίβεια, φαίνεται πως το είδος «Family» σχετίζεται με το είδος «Animation» καθ' όσον 4/5 φορές εμφανίζονται μαζί.

Παρατηρείται επίσης πως η ταινία «The Lion King» έχει το μικρότερο βαθμό ταύτισης ειδών από την πρώτη τριάδα, αλλά παρ' όλα αυτά είναι στην κορυφή των προτάσεων. Κοιτώντας τα υπόλοιπα κριτήρια, και λαμβάνοντας υπόψιν τη συζήτηση που έγινε στην [δεύτερη περίπτωση χρήσης](#), φαίνεται η χαοτική διαφορά που υπάρχει στα άλλα τρία κριτήρια, δηλαδή τη δημοτικότητα, το μέσο όρο βαθμολογίας και το πλήθος των ψήφων, που ακόμη κι αν η δεύτερη και η τρίτη πρόταση έχουν μεγαλύτερο βαθμό ταύτισης των ειδών, η ταινία «The Lion King» είναι πολύ πιο δημοφιλής, πολύ πιο ποιοτική (με βάση τη βαθμολογία) και πολύ πιο αντικειμενικά ποιοτική.

Αξίζει επίσης να αναφερθεί πως, συγκρίνοντας τις δύο τελευταίες προτάσεις, παρ' όσον η ταινία «The Secret Adventures of Tom Thumb» είναι σχετικά άγνωστη και, κατά πάσα πιθανότητα, αντικειμενικά κακή ταινία<sup>13</sup>, έχει βρεθεί υψηλότερα στην κατάταξη από την ταινία «The Jungle Book» διότι έχει μεγαλύτερο βαθμό ταύτισης με τα προτιμητέα είδη του χρήστη, ενώ φαίνεται η μεγάλη διαφορά που υπάρχει μεταξύ των δύο ταινιών στα υπόλοιπα τρία κριτήρια.

## 5.2 Αξιολόγηση

Η εφαρμογή διαμοιράστηκε σε μία ομάδα δοκιμής αποτελούμενη από 19 άτομα προκειμένου να τη χρησιμοποιήσουν και στη συνέχεια να την αξιολογήσουν. Η ομάδα δοκιμής αποτελούνταν από άτομα 14 έως 40 ετών από όλες τις βαθμίδες της εκπαίδευσης. Για την αξιολόγηση, τους δόθηκε ένα ερωτηματολόγιο 10 ερωτήσεων, οι οποίες χωρίζονται σε 5 θεματικές ενότητες:

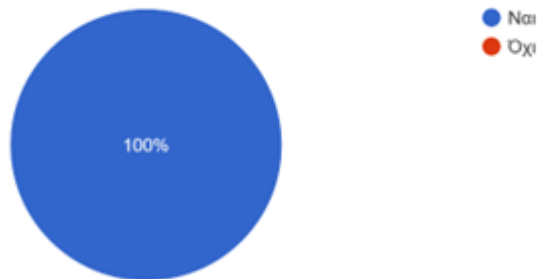
- την οικειότητα με την εφαρμογή και το γραφικό της περιβάλλον
- την ευκολία χρήσης της εφαρμογής
- την ποιότητα των προτάσεων του συστήματος προτάσεων
- την ποιότητα του συνόλου δεδομένων
- γενικά σχόλια χρηστών

---

<sup>13</sup> Με βάση τα, λιγοστά, στοιχεία που διατίθενται. Υπάρχει το ενδεχόμενο εάν η ταινία βαθμολογηθεί από ευρύτερο κοινό τα αποτελέσματα να είναι διαφορετικά

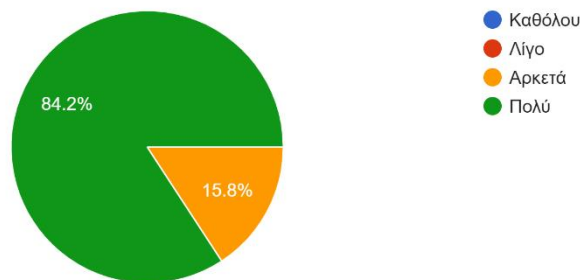
### 5.2.1 Αξιολόγηση οικειότητας εφαρμογής

Έχετε χρησιμοποιήσει παρόμοιες εφαρμογές στο παρελθόν;  
19 responses



Εικόνα 19 Αποτελέσματα πρώτης ερώτησης αξιολόγησης

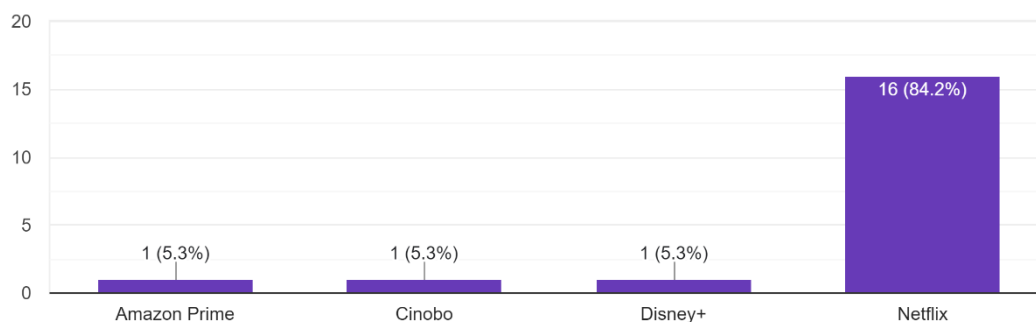
Πόσο οικείο σας ήταν το γραφικό περιβάλλον;  
19 responses



Εικόνα 20 Αποτελέσματα δεύτερης ερώτησης αξιολόγησης

Με ποια εφαρμογή θα την προσομοιάζατε; (Μόνο το όνομα)

19 responses



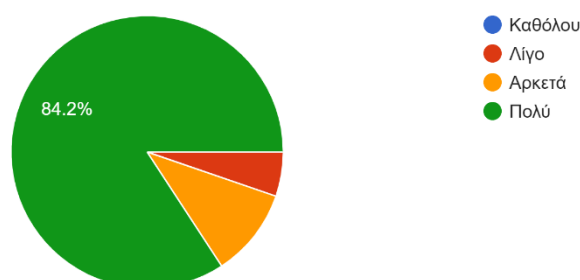
### Εικόνα 21 Αποτελέσματα τρίτης ερώτησης αξιολόγησης

Όπως είναι ξεκάθαρο –και αναμενόμενο- από τα αποτελέσματα των τριών πρώτων απαντήσεων, όλοι οι συμμετέχοντες είχαν πρότερη εμπειρία με παρόμοιες εφαρμογές, και συγκεκριμένα κυρίως με το Netflix σε ποσοστό 84.2%. Το αποτέλεσμα της δεύτερης ερώτησης ήταν επίσης αναμενόμενο, μιας και όπως έχει ήδη αναφέρει στο [κεφάλαιο 3.1](#), η εφαρμογή του Netflix ήταν μέσα στο σύνολο εφαρμογών που εξετάστηκε το γραφικό τους περιβάλλον προκειμένου να προσφερθεί μια οικειότητα.

### 5.2.2 Αξιολόγηση λειτουργικότητας

Πόσο εύκολη ήταν για εσάς η διαδικασία δημιουργίας λογαριασμού;

19 responses

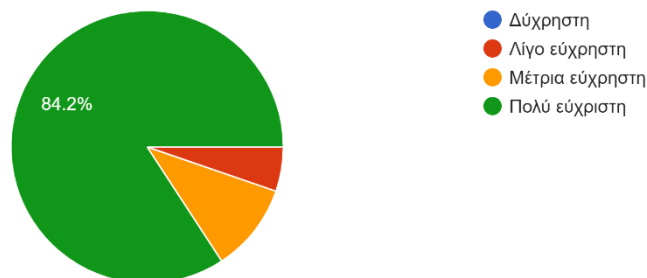


### Εικόνα 22 Αποτελέσματα τέταρτης ερώτησης αξιολόγησης



Πώς θα αξιολογούσατε την ευχρηστία της εφαρμογής;

19 responses



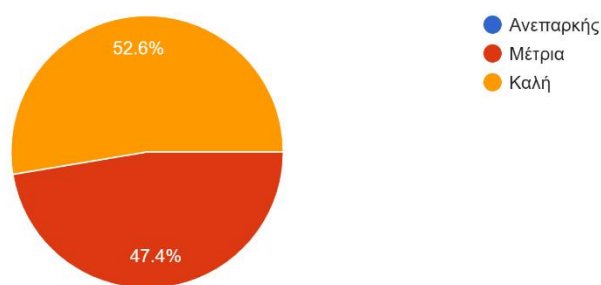
**Εικόνα 23** Αποτελέσματα πέμπτης ερώτησης αξιολόγησης

Κοιτώντας τα αποτελέσματα φαίνεται πως, στην πλειοψηφία τους, τα άτομα τα οποία χρησιμοποίησαν την εφαρμογή δεν αντιμετώπισαν κάποιο ιδιαίτερο πρόβλημα στη χρήση της. Αποτέλεσμα το οποίο είναι εξίσου αναμενόμενο αφού δόθηκε ιδιαίτερο βάρος προκειμένου η εφαρμογή να ανταποκρίνεται στη διαίσθηση που θα έχουν αναπτύξει οι χρήστες χρησιμοποιώντας παρόμοιες εφαρμογές.

### 5.2.3 Αξιολόγηση συστήματος προτάσεων

Πώς θα βαθμολογούσατε την ποικιλία των ταινιών;

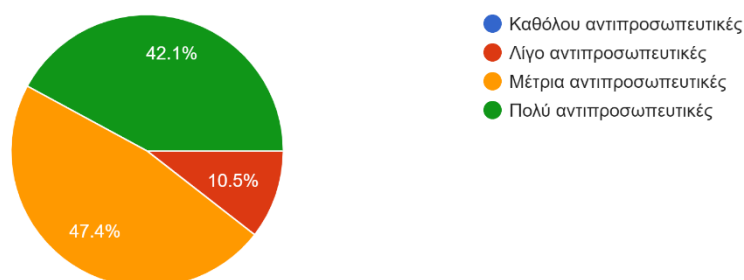
19 responses



**Εικόνα 24** Αποτελέσματα έκτης ερώτησης αξιολόγησης

Πώς θα αξιολογούσατε τις προτάσεις ταινιών που παρήγαγε η εφαρμογή για εσας;

19 responses



### Εικόνα 25 Αποτελέσματα έβδομης ερώτησης αξιολόγησης

Παρατηρώντας τα αποτελέσματα της αξιολόγησης που αφορούν το σύστημα προτάσεων, οι απόψεις είναι διχασμένες. Αρχικά είναι φανερό πως σχεδόν οι μισοί από τους ερωτηθέντες δεν βρήκαν ιδιαίτερη ποικιλία ανάμεσα στις προτάσεις που τους παρήγαγε το σύστημα προτάσεων, με τους άλλους μισούς να θεωρούν πως η ποικιλία ήταν επαρκής. Μία αιτία στην οποία μπορεί να οφείλεται το γεγονός αυτό είναι η πως το σύνολο δεδομένων διαθέτει, κατά γενική ομολογία, παλιές, έως πολύ παλιές, ταινίες. Συγκεκριμένα, η πιο πρόσφατη ταινία που διαθέτει κυκλοφόρησε στα τέλη του 2000, ενώ η παλαιότερη ταινία που διαθέτει κυκλοφόρησε στα μέσα του 1919! Δεδομένου ότι τα περισσότερα άτομα ήταν νεαρής ηλικίας, θα ήταν λογικό κάποιος να υποθέσει πως, με εξαίρεση τις πολύ γνωστές ταινίες, μάλλον οι υπόλοιπες έχουν παλιώσει αρκετά ώστε να μην απευθύνονται στο κοινό που αποτελούσαν το δείγμα των ερωτηθέντων.

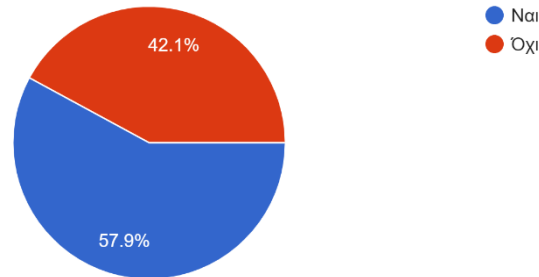
Αυτό θα εξηγούσε και το γεγονός πως το μεγαλύτερο ποσοστό έκρινε πως οι προτάσεις που δέχθηκε ήταν μέτριες. Ξανά, αν ληφθεί υπόψιν η ηλικία των ταινιών καθώς και οι διαφορές της τότε κοινωνίας με τη σημερινή, είναι πιθανό, οι ταινίες αυτές καθ' αυτές να είναι όντως αντιπροσωπευτικές, απλά για μία άλλη εποχή. Η συγκεκριμένη υπόθεση χρειάζεται παραπάνω διερεύνηση.

Επίσης, όπως έγινε φανερό από το [κεφάλαιο 2](#), οι γνωστότερες εφαρμογές με συστήματα προτάσεων διαθέτουν πολυσύνθετα και πολυεπίπεδα συστήματα τα οποία βασίζονται μέχρι και σε τεχνικές τεχνητής νοημοσύνης και μηχανικής μάθησης. Θα ήταν μάλλον αφελής η προσδοκία μία τόσο απλοϊκή μέθοδος να παράγει το ίδιο ποιοτικές προτάσεις.

## 5.2.4 Αξιολόγηση ταινιών

Σας πρότεινε η εφαρμογή κάποια ταινία την οποία να έχετε παρακολουθήσει και πράγματι να σας άρεσε;

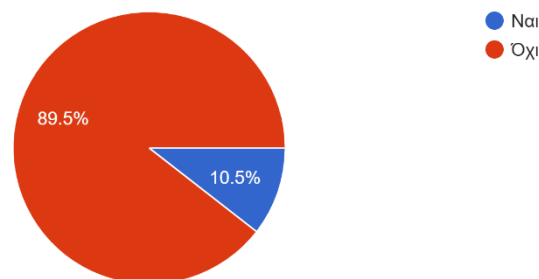
19 responses



Εικόνα 26 Αποτελέσματα όγδοης ερώτησης αξιολόγησης

Υπήρξε κάποια ταινία που η εφαρμογή να σας πρότεινε η οποία να σας κέντρισε το ενδιαφέρον;

19 responses



Εικόνα 27 Αποτελέσματα ένατης ερώτησης αξιολόγησης

Από το πρώτο από δύο προηγούμενα γραφήματα βλέπουμε πως, για ένα μεγάλο ποσοστό του δείγματος, τα άτομα δεν είχαν παρακολουθήσει καμία από τις ταινίες που το σύστημα πρότεινε, ενώ, βλέποντας και το δεύτερο διάγραμμα συμπεραίνεται πως δεν ήταν και ιδιαίτερα διατεθειμένοι να παρακολουθήσουν τις εν λόγω ταινίες. Εδώ θα αναφερθεί ξανά πως είναι πιθανών η παλαιότητα των ταινιών να φέρει μέρος της ευθύνης. Ωστόσο σε αυτό το σημείο θα εισαχθεί και μία ακόμη συνιστώσα. Το γεγονός πως το σύστημα προτάσεων λαμβάνει υπόψιν μόνο τις ρητές (explicit) των χρηστών. Όπως παρατηρήθηκε κατά την ανασκόπηση της βιβλιογραφίας, υπάρχουν πολυσύνθετα συστήματα που όχι μόνο λαμβάνουν υπόψιν τις συναγόμενες (implicit) προτιμήσεις των χρηστών, αλλά τους δίνουν

και πολύ μεγαλύτερο βάρος. Για παράδειγμα, μπορεί ένας χρήστης να έχει δηλώσει ότι του αρέσουν οι ταινίες δράσεις, ωστόσο αυτή η πληροφορία από μόνη δίνει ελάχιστη γνώση. Είναι πιθανό ο συγκεκριμένος χρήστης να μην προτιμά κωμωδίες οι οποίες να έχουν ταυτόχρονα και δράση ή έναν συγκεκριμένο πρωταγωνιστή. Πληροφορίες τις οποίες το σύστημα που αναπτύχθηκε δεν διαθέτει.

### 5.2.5 Σχόλια χρηστών

Υπάρχει κάποιο σχόλιο που θα θέλατε να κάνετε όσον αφορά γενικά την εφαρμογή ή τις προτάσεις αυτής;

9 responses

Παλιές ταινίες

Αρκετά παλιές ταινίες

Οι ταινίες είναι κυρίως παλιές

Παλιές ταινίες γενικά

Οι ταινίες ήταν αρκετά παλιές

Οι ταινίες που προτεινόταν αφοτούσαν κυρίως τον mainstream κινηματογράφο

Θα προτιμούσα να είχε πιο καινούργιες ταινίες

Θα ήθελα οι ταινίες να ήταν πιο σύγχρονες

**Εικόνα 28 Σχόλια που άφησαν οι συμμετέχοντες στην έρευνα ανεξάρτητα από τις προκαθορισμένες ερωτήσεις**

Όπως είναι πλέον σαφές, η υπόθεσή που έγινε σχετικά με την παλαιότητα των ταινιών και το κατά πόσον επηρέαζε τα αποτελέσματα της έρευνας επαληθεύεται. Αν και λιγότερα από τα μισά άτομα προτίμησαν να αφήσουν κάποιο σχόλιο, φαίνεται πως στο σύνολο του δείγματος υπάρχει 100% ομοφωνία για το πρόβλημα της ηλικίας των ταινιών.

## Κεφάλαιο 6. Συμπεράσματα – Μελλοντικές επεκτάσεις

### 6.1 Συμπεράσματα

Στην παρούσα πτυχιακή εργασία έγινε μία προσπάθεια ανάπτυξης ενός συστήματος προτάσεων και ενός γραφικού περιβάλλοντος που να το πλαισιώνει. Όπως έγινε γνωστό μέσα από την βιβλιογραφική ανασκόπηση και την εξέταση παρόμοιων εφαρμογών, τα συστήματα προτάσεων έχουν μετατραπεί πλέον σε πολυσύνθετους μηχανισμούς με πάρα πολλές εφαρμογές στον κόσμο του διαδικτύου. Όσο περνούν τα χρόνια και η τεχνολογία εξελίσσεται, τόσο εξελίσσονται και τα συστήματα προτάσεων, αξιοποιώντας τεχνικές τεχνητής νοημοσύνης και μηχανικής μάθησης. Με τις τεχνικές αυτές οι αλγόριθμοι των συστημάτων έχουν καταφέρει να λαμβάνουν υπόψιν, πέραν τα ρητών προτιμήσεων του χρήστη (αυτά δηλαδή που ο ίδιος ο χρήστης έχει δηλώσει πως του αρέσουν) και τις έμμεσες προτιμήσεις του, λαμβάνοντας υπόψιν τη συμπεριφορά του.

Ωστόσο, το γεγονός ότι τα συστήματα αυτά μπορούν και μαθαίνουν τόσο καλά έναν χρήστη, μπορεί τελικά να ενέχει και κινδύνους. Όπως αναφέρει ο (Zhao, 2021) σχετικά με την περίπτωση του TikTok, παρότι ο αλγόριθμος αρχικά κατασκευάστηκε προκειμένου να βελτιώσει την αποδοτικότητα της διαμοίρασης της πληροφορίας, δηλαδή να μειώσει τον χρόνο που οι χρήστες δαπανούν στην διαλογή πληροφοριών, το αποτέλεσμα ήταν πως, και μεν οι χρήστες όντως σπαταλούν λιγότερο χρόνο σε αναζήτηση περιεχομένου λόγω της ακριβούς (accurate) μεθόδου προτάσεων, ωστόσο καταλήγουν να δαπανούν περισσότερο χρόνο βλέποντας παθητικά υλικό που τους προτάθηκε. Το γεγονός αυτό ανήκει περισσότερο στη σφαίρα της φιλοσοφίας παρά σε μία τεχνική συζήτηση.

### 6.2 Μελλοντικές επεκτάσεις

Όπως έγινε φανερό από την [αξιολόγηση](#) της εφαρμογής, ένα πρόβλημα που θα πρέπει μελλοντικά να λυθεί αφορά το σύνολο δεδομένων, το οποίο θα πρέπει να εμπλουτιστεί με νεότερες (και ίσως να καθαριστεί από σχετικά άγνωστες και πολύ παλιές ταινίες). Ένας τρόπος για να γίνει αυτό θα ήταν να η χρήση ενός άλλου συνόλου δεδομένων, όπως αυτό που παρείχε το Netflix στα πλαίσια του [διαγωνισμού](#) ή ίσως το MovieTweatings<sup>14</sup>,

---

<sup>14</sup> <https://github.com/sidooms/MovieTweatings>

ένα σύνολο δεδομένων που αντλεί κριτικές από το Tweeter. Μία άλλη προσέγγιση θα ήταν η χρήση του TMBD και της τεράστιας βάσης δεδομένων του προκειμένου να γίνει διαλογή ταινιών.

Το σύστημα προτάσεων θα μπορούσε επίσης να βελτιωθεί. Εκτός από έναν απλοϊκό αλγόριθμο με βάρη, θα μπορούσε να γίνει πιο σύνθετο, εφαρμόζοντας κάποιες από τις τεχνικές που αναφέρθηκαν στην παρούσα. Ο περιορισμός που υπάρχει εδώ βέβαια είναι πως, σε αντίθεση με πλατφόρμες όπως το Netflix ή το Amazon οι οποίες διαθέτουν οι ίδιες τα προϊόντα και είναι σε θέση να γνωρίζουν πώς ο χρήστης αλληλοεπιδρά μεταξύ τους, κάτι τέτοιο δε θα ήταν εφικτό σε ένα τρίτο σύστημα όπου θα μπορεί μόνο να γνωρίσει τα στοιχεία αυτά που ο ίδιος ο χρήστης επιλέγει, όπως για παράδειγμα, η βαθμολογία που θα απέδιδε σε μία ταινία που παρακολούθησε.

Μία ακόμη πιθανή επέκταση θα ήταν η εφαρμογή να αποκτήσει έναν πιο κοινωνικό χαρακτήρα. Θα μπορούσε, για παράδειγμα, να δίνει τη δυνατότητα στους χρήστες να δημιουργούν λίστες αναπαραγωγής (playlists) που θα συγκεντρώνουν ταινίες με ένα δικό τους κριτήριο ομαδοποίησης και να τη μοιράζονται με φίλους.

## Κεφάλαιο 7. Βιβλιογραφία

- Amatriain, X. (2012, Απρίλιος 6). Netflix Recommendations: Beyond the 5 stars (Part 1).
- Bennett, J., & Lanning, S. (2007). The Netflix Prize. Καλιφόρνια.
- Bonfiglio, N. (2022, Ιούνιος 16). The 10 most dangerous TikTok trends of all time. Ανάκτηση από <https://wegotthiscovered.com/social-media/the-10-most-dangerous-tiktok-trends-of-all-time/>
- Brothy, P. (2007). The library in the twenty-first century. London: Facet Publishing.
- Cao, H. (2018, Ιανουάριος 16). Today's headline algorithm principle. Ανάκτηση από <https://baijiahao.baidu.com/s?id=1589714822818540008>
- Chagas, J., & Wagner, M. (2020). A weighted-sum method for solving the bi-objective traveling thief. doi:<https://doi.org/10.48550/arXiv.2011.05081>
- Covington, P., Adams, J., & Emre, S. (2016). Deep Neural Networks for YouTube Recommendations. RecSys. doi:10.1145/2959100.2959190
- Davidson, J., Liebald, B., Liu, J., Nandy, P., & Van Vleet, T. (2010). The YouTube Video Recommendation System. RecSys. doi:10.1145/1864708.1864770
- Fennelle, E. (1998, September 22). The harsh law of averages. The Times, σ. 41.
- Fung, B. (2022, Ιούνιος 29). FCC commissioner calls on Apple and Google to remove TikTok from their app stores. Ανάκτηση από <https://edition.cnn.com/2022/06/29/tech/fcc-google-apple-tiktok-block/index.html>
- Goldberg, D., Nichols, D., Oki, B. M., & Douglas, T. (1992, Δεκέμβριος --). ACM Digital Livbrary. doi:10.1145/138859.138867
- Gomez-Uribe, C. A., & Hunt, N. (2015). The Netflix Recommender System: Algorithms, Business Value, and Innovation. doi:10.1145/2843948
- Goodrow, C. (2021, Σεπτέμβριος 15). On YouTube's recommendation system. Ανάκτηση από <https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/#:~:text=What%20is%20a%20recommendation%20system,the%20%E2%80%9CUp%20Next%E2%80%9D%20panel.>
- Harper, F. M., & Konstan, J. A. (2015, Δεκέμβριος 22). The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems. doi:10.1145/2827872

- Harrar, J. H. (1975). Photographs, pictures and prints. Στο S. P. Grove (Επιμ.), *Non-print media in academic libraries* (σσ. 173-192). Chicago: American Library Association.
- Kaseb, K. (2022, Απρίλιος 20). The Implementation of UI Layer in Android App Architecture.
- Kirkendall, K. (1998). Teaching the online catalogue users. *Library Review*, 19(4), 27-28.
- Konstan, J. A., Riedl, J., Borchers, A., & Herlocker, J. L. (1998, Ιανουάριος --). Research Gate. Ανάκτηση από Research Gate:  
[https://www.researchgate.net/publication/247356333\\_Recommender\\_Systems\\_A\\_GroupLens\\_Perspective](https://www.researchgate.net/publication/247356333_Recommender_Systems_A_GroupLens_Perspective)
- Li, H., Cao, N., Jiang, J., & Oh, L.-B. (2013, Απρίλιος). A Maslow's hierarchy of needs analysis of social networking services continuance. *Journal of Service Management*, σ. 20. doi:10.1108/09564231311323953
- Linden, G., Smith, B., & York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. *IEEE*. doi:10.1109/MIC.2003.1167344
- Lutz, E., Kleist-Retzow, H. V., & Hoernig, K. (1990, 12 2). Semantic scholar. doi:10.1145/122220.122222
- Narayanan, A., & Shmatikov, V. (2006). Robust De-anonymization of Large Sparse Datasets. Όστιν. doi:arXiv:cs/0610105
- Nicholas, D. (1998). Hacking the net. Ανάκτηση 9 22, 1998, από Ariadne:  
<http://www.ariadne.ac.uk/issue16/cover>
- Palme, J. (1984). *You Have 134 Unread Mail!* Stockholm: Elsevier Science Publshers B.V. (North-Holland).
- Pollock, S. (1988, 7 1). Semantic Scholar. doi:10.1145/45945.214327
- Rich, E. (1998, 6 1). [www.semanticscholar.org](http://www.semanticscholar.org). doi:10.1207/S15516709COG0304\_3
- Richter, F. (2019, Ιανουάριος 24). Streaming Dominates U.S. Home Entertainment Spending. Ανάκτηση από <https://www.statista.com/chart/7654/home-entertainment-spending-in-the-us/>
- Smith, B., & Linden, G. (2017). Two Decades of Recommender. *IEEE*. doi:10.1109/MIC.2017.72
- Sun, D., Zhang, G., Zheng, W., & Li, K. (2015). Key Technologies for Big Data Stream Computing. Στο *Big Data*. Chapman and Hall/CRC. doi:10.1201/b18050-15
- Terry, D. (1993, Δεκέμβριος 1). ACM DIgital Library. doi:10.1145/168555.168558



- Terveen, L., & Hill, W. (2001, -- --). Sematic Scholar. Ανάκτηση από Sematic Scholar: <https://www.semanticscholar.org/paper/Beyond-Recommender-Systems%3A-Helping-People-Help-Terveen-Hill/9fed1483fd60a19372d082d7ac0e6719d6e7fd9a>
- Troussas, C., Giannakas, F., Sgouropoulou, C., & Voyiatzis, I. (2020). Collaborative activities recommendation based on. doi:10.1080/10494820.2020.1761835
- University Information Technology Services. (χ.χ.). Ανάκτηση από Indiana University: <https://kb.iu.edu/d/aebt>
- YouTube Press. (χ.χ.). Ανάκτηση από YouTube Press: <https://blog.youtube/press/>
- Zhao, Z. (2021, Φεβρουάριος 3). Analysis on the “Douyin (Tiktok) Mania” Phenomenon Based on. σ. 10. doi:10.1051/e3sconf/202123503029
- Christos Troussas, Akrivi Krouska, Cleo Sgouropoulou: Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education, *Computers & Education*, Volume 144, 2020, 103698, <https://doi.org/10.1016/j.compedu.2019.103698>.
- Troussas C, Krouska A, Sgouropoulou C, Voyiatzis I. Ensemble Learning Using Fuzzy Weights to Improve Learning Style Identification for Adapted Instructional Routines. *Entropy*. 2020; 22(7):735. <https://doi.org/10.3390/e22070735>.
- Krouska A., Troussas C., Sgouropoulou C. (2020) A Personalized Brain-Based Quiz Game for Improving Students’ Cognitive Functions. In: Frasson C., Bamidis P., Vlamos P. (eds) *Brain Function Assessment in Learning*. BFAL 2020. Lecture Notes in Computer Science, vol 12462. Springer, Cham. [https://doi.org/10.1007/978-3-030-60735-7\\_11](https://doi.org/10.1007/978-3-030-60735-7_11)
- Papakostas, C., Troussas, C., Krouska, A. et al. User acceptance of augmented reality welding simulator in engineering training. *Educ Inf Technol* (2021). <https://doi.org/10.1007/s10639-020-10418-7>.
- Troussas, C., Krouska, A. & Sgouropoulou, C. Impact of social networking for advancing learners’ knowledge in E-learning environments. *Educ Inf Technol* 26, 4285–4305 (2021). <https://doi.org/10.1007/s10639-021-10483-6>.
- Krouska A., Troussas C., Sgouropoulou C. (2020) Applying Genetic Algorithms for Recommending Adequate Competitors in Mobile Game-Based Learning Environments. In: Kumar V., Troussas C. (eds) *Intelligent Tutoring Systems*. ITS 2020. Lecture Notes in Computer Science, vol 12149. Springer, Cham. [https://doi.org/10.1007/978-3-030-49663-0\\_23](https://doi.org/10.1007/978-3-030-49663-0_23)

- Christos Troussas, Cleo Sgouropoulou: Innovative Trends in Personalized Software Engineering and Information Systems - The Case of Intelligent and Adaptive E-learning Systems. *Frontiers in Artificial Intelligence and Applications* 324, IOS Press 2020, ISBN 978-1-64368-096-5, pp. 1-96
- Troussas C., Krouska A., Sgouropoulou C. (2020) Towards a Reference Model to Ensure the Quality of Massive Open Online Courses and E-Learning. In: Frasson C., Bamidis P., Vlamos P. (eds) *Brain Function Assessment in Learning*. BFAL 2020. *Lecture Notes in Computer Science*, vol 12462. Springer, Cham. [https://doi.org/10.1007/978-3-030-60735-7\\_18](https://doi.org/10.1007/978-3-030-60735-7_18)
- Troussas C., Krouska A., Sgouropoulou C. (2020) Dynamic Detection of Learning Modalities Using Fuzzy Logic in Students' Interaction Activities. In: Kumar V., Troussas C. (eds) *Intelligent Tutoring Systems*. ITS 2020. *Lecture Notes in Computer Science*, vol 12149. Springer, Cham. [https://doi.org/10.1007/978-3-030-49663-0\\_24](https://doi.org/10.1007/978-3-030-49663-0_24)
- Christos Troussas, Akrivi Krouska, Filippos Giannakas, Cleo Sgouropoulou, Ioannis Voyiatzis: Automated reasoning of learners' cognitive states using classification analysis. *PCI 2020-24th Pan-Hellenic Conference on Informatics*, November 2020, Pages 103–106. <https://doi.org/10.1145/3437120.3437285>
- Christos Troussas, Akrivi Krouska, Filippos Giannakas, Cleo Sgouropoulou, Ioannis Voyiatzis: Redesigning teaching strategies through an information filtering system. *PCI 2020-24th Pan-Hellenic Conference on Informatics*, November 2020, Pages 111-114. <https://doi.org/10.1145/3437120.3437287>
- Akrivi Krouska, Christos Troussas, Cleo Sgouropoulou: Usability and Educational Affordance of Web 2.0 tools from Teachers' Perspectives. *PCI 2020-24th Pan-Hellenic Conference on Informatics*, November 2020, Pages 107-110. <https://doi.org/10.1145/3437120.3437286>
- F. Giannakas, C. Troussas, I. Voyiatzis, C. Sgouropoulou, A deep learning classification framework for early prediction of team-based academic performance, *Applied Soft Computing*, Volume 106, 2021, 107355. <https://doi.org/10.1016/j.asoc.2021.107355>.
- Christos PAPAΚOSTAS, Christos TROUSSAS, Akrivi KROUSKA, Cleo SGOUROPOULOU, Exploration of Augmented Reality in Spatial Abilities Training: A Systematic Literature Review for the Last Decade, *Informatics in Education* 20 (2021), no. 1, 107-130, DOI 10.15388/infedu.2021.06

- C. Troussas, A. Krouska and C. Sgouropoulou, "A Novel Teaching Strategy Through Adaptive Learning Activities for Computer Programming," in *IEEE Transactions on Education*, vol. 64, no. 2, pp. 103-109, May 2021, doi: 10.1109/TE.2020.3012744.
- Krouska, A., Troussas, C. and Sgouropoulou, C. 2019. Fuzzy Logic for Refining the Evaluation of Learners' Performance in Online Engineering Education. *European Journal of Engineering and Technology Research*. 4, 6 (Jun. 2019), 50-56. DOI:<https://doi.org/10.24018/ejers.2019.4.6.1369>.
- Troussas C, Krouska A, Sgouropoulou C. Improving Learner-Computer Interaction through Intelligent Learning Material Delivery Using Instructional Design Modeling. *Entropy*. 2021; 23(6):668. <https://doi.org/10.3390/e23060668>
- Papakostas C, Troussas C, Krouska A, Sgouropoulou C. Measuring User Experience, Usability and Interactivity of a Personalized Mobile Augmented Reality Training System. *Sensors*. 2021; 21(11):3888. <https://doi.org/10.3390/s21113888>
- Giannakas F., Troussas C., Krouska A., Sgouropoulou C., Voyiatzis I. (2021) XGBoost and Deep Neural Network Comparison: The Case of Teams' Performance. In: Cristea A.I., Troussas C. (eds) *Intelligent Tutoring Systems. ITS 2021. Lecture Notes in Computer Science*, vol 12677. Springer, Cham. [https://doi.org/10.1007/978-3-030-80421-3\\_37](https://doi.org/10.1007/978-3-030-80421-3_37)
- Krouska, A., Troussas, C. & Sgouropoulou, C. Mobile game-based learning as a solution in COVID-19 era: Modeling the pedagogical affordance and student interactions. *Educ Inf Technol* (2021). <https://doi.org/10.1007/s10639-021-10672-3>
- C. Troussas, A. Krouska, C. Sgouropoulou, "Enhancing Human-Computer Interaction in Digital Repositories through a MCDA-Based Recommender System", *Advances in Human-Computer Interaction*, vol. 2021, Article ID 7213246, 7 pages, 2021. <https://doi.org/10.1155/2021/7213246>
- Krouska A, Troussas C, Sgouropoulou C. A Cognitive Diagnostic Module Based on the Repair Theory for a Personalized User Experience in E-Learning Software. *Computers*. 2021; 10(11):140. <https://doi.org/10.3390/computers10110140>
- Troussas, C., Krouska, A. and Sgouropoulou, C., A User-centric System for Improving Human-Computer Interaction through Fuzzy Logic-based Assistive Messages, *Proceedings of the 17th International Conference on Web Information Systems and Technologies (WEBIST 2021)*, pages 365-370. DOI: 10.5220/0010702800003058.

- Akrivi KROUSKA, Christos TROUSSAS, Filippos GIANNAKAS, Cleo SGOUROPOULOU, and Ioannis VOYIATZIS, Enhancing the Effectiveness of Intelligent Tutoring Systems Using Adaptation and Cognitive Diagnosis Modeling, *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 40-45). IOS Press, doi:10.3233/FAIA210073.
- Kanetaki, Z., Stergiou, C., Bekas, G., Troussas, C., & Sgouropoulou, C. (2021). Data Mining for Improving Online Higher Education Amidst COVID-19 Pandemic: A Case Study in the Assessment of Engineering Students. *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 157-165). doi:10.3233/FAIA210088.
- Kanetaki, Z., Stergiou, C., Troussas, C., & Sgouropoulou, C. (2021, October). Development of an Innovative Learning Methodology Aiming to Optimise Learners' Spatial Conception in an Online Mechanical CAD Module During COVID-19 Pandemic. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 31-39). IOS Press. doi:10.3233/FAIA210072.
- Kanetaki, Z., Stergiou, C., Bekas, G., Troussas, C., & Sgouropoulou, C. (2021). Creating a Metamodel for Predicting Learners' Satisfaction by Utilizing an Educational Information System During COVID-19 Pandemic. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 127-136). IOS Press. doi:10.3233/FAIA210085.
- Kapetanaki, A., Krouska, A., Troussas, C., & Sgouropoulou, C. (2021). A Novel Framework Incorporating Augmented Reality and Pedagogy for Improving Reading Comprehension in Special Education. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 105-110). IOS Press. doi:10.3233/FAIA210081.
- C. Papakostas, C. Troussas, A. Krouska, C. Sgouropoulou (2021, October). On the Development of a Personalized Augmented Reality Spatial Ability Training Mobile Application. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 75-83). IOS Press. doi:10.3233/FAIA210078.

- Kontellis, E., Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). Real-time face mask detector using convolutional neural networks amidst COVID-19 pandemic. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 247-255). IOS Press. doi:10.3233/FAIA210102
- Troussas, C., Krouska, A., Giannakas, F., Sgouropoulou, C., & Voyiatzis, I. (2021). An Alternative Educational Tool Through Interactive Software over Facebook in the Era of COVID-19. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 3-11). IOS Press. doi:10.3233/FAIA210069.
- Troussas C., Krouska A., Giannakas F., Sgouropoulou C., Voyiatzis I. (2021) Representation of Generalized Human Cognitive Abilities in a Sophisticated Student Leaderboard. In: Cristea A.I., Troussas C. (eds) *Intelligent Tutoring Systems. ITS 2021. Lecture Notes in Computer Science*, vol 12677. Springer, Cham. [https://doi.org/10.1007/978-3-030-80421-3\\_44](https://doi.org/10.1007/978-3-030-80421-3_44)
- A. Marougkas, C. Troussas, A. Krouska, C. Sgouropoulou (2021, October). A Framework for Personalized Fully Immersive Virtual Reality Learning Environments with Gamified Design in Education. In *Novelties in Intelligent Digital Systems: Proceedings of the 1st International Conference (NIDS 2021)*, Athens, Greece, September 30-October 1, 2021 (Vol. 338, p. 95-104). IOS Press. doi:10.3233/FAIA210080.
- Kanetaki Z, Stergiou C, Bekas G, Troussas C, Sgouropoulou C. Evaluating Remote Task Assignment of an Online Engineering Module through Data Mining in a Virtual Communication Platform Environment. *Electronics*. 2022; 11(1):158. <https://doi.org/10.3390/electronics11010158>
- Kanetaki, Z., Stergiou, C., Bekas, G., Troussas, C., & Sgouropoulou, C. (2021). Analysis of Engineering Student Data in Online Higher Education During the COVID-19 Pandemic. *International Journal of Engineering Pedagogy (ijEP)*, 11(6), pp. 27-49. <https://doi.org/10.3991/ijep.v11i6.23259>.
- Kanetaki, Z.; Stergiou, C.; Bekas, G.; Troussas, C.; Sgouropoulou, C. The Impact of Different Learning Approaches Based on MS Teams and Moodle on Students' Performance in an on-Line Mechanical CAD Module. *Global Journal of Engineering Education (GJEE)*. 2021, 23, 185-190.