

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ



ΔΙΟΙΚΗΣΗ ΤΗΣ ΨΗΦΙΑΚΗΣ ΕΠΙΧΕΙΡΗΣΗΣ
Εργαστήριο Διοίκησης Πληροφοριακών Συστημάτων
(2020-2021)

4^η Εργασία
ΑΝΑΖΗΤΗΣΗ & RECOMMENDER SYSTEMS

Ονοματεπώνυμο:

- Χρήστος Τσούφης

Αριθμός Μητρώου:

- 03117176

Στοιχεία Επικοινωνίας:

- el17176@mail.ntua.gr

1. Αναζήτηση – TFIDF

Έστω ότι πραγματοποιείται μια αναζήτηση σε μια μηχανή αναζήτησης με τους όρους “thriller, mission, planet”. Επίσης, έστω ότι τα μοναδικά δεδομένα της μηχανής είναι οι παρακάτω περιγραφές ταινιών.

α) Ζητείται η χρήση της μετρικής TF-IDF (χωρίς TF normalization) για να βρεθεί ποια θα είναι η κατάταξη των αποτελεσμάτων. Να χρησιμοποιηθεί Cosine Similarity και να ληφθούν υπόψη το κείμενο των τίτλων και οι παράγωγες λέξεις (Stemming). Να σχολιαστούν τα αποτελέσματα.

Για την αναζήτηση με χρήση TF/IDF ακολουθείται η εξής διαδικασία:

Για κάθε λέξη σε ένα query [term_A, term_B, term_C], όπου:

term_A = thriller term_B = mission term_C = planet

Αρχικά, αναζητείται και υπολογίζεται η συχνότητα εμφάνισής της (TF: Term Frequency) σε κάθε έγγραφο που δίνεται. Συνεπώς:

	Sunshine	Gattaca	Gravity	Event Horizon	The Arrival
thriller	2	1	2	0	1
mission	2	0	0	1	1
planet	2	0	0	1	0

(Σημείωση: στην ταινία The Arrival, έχει ληφθεί υπόψη ο όρος “transmissions” για τον όρο “mission”).

Στη συνέχεια, υπολογίζεται το IDF από τον τύπο: $IDF = \log\left(\frac{|D|}{|d:t_i \in d|}\right) = \log\left(\frac{\text{total number of documents}}{\# \text{ documents with word } x \text{ in it}}\right)$

Οπότε, όπου D: είναι το πλήθος των κειμένων, προκύπτει:

$$IDF(\text{thriller}) = \log\left(\frac{5}{4}\right) = 0.0969$$

$$IDF(\text{mission}) = \log\left(\frac{5}{3}\right) = 0.2218$$

$$IDF(\text{planet}) = \log\left(\frac{5}{2}\right) = 0.3979$$

Έπειτα, πολλαπλασιάζεται το TF με το IDF οπότε προκύπτει:

	Sunshine	Gattaca	Gravity	Event Horizon	The Arrival
thriller	0.1938	0.0969	0.1938	0	0.0969
mission	0.4436	0	0	0.2218	0.2218
planet	0.7958	0	0	0.3979	0

Υστερα, χρησιμοποιείται Cosine Similarity

$$\text{Cosine Similarity (Query, Document)} = \frac{\text{Query} * \text{Document}}{||\text{Query}|| * ||\text{Document}||}$$

Όπου για όλα τα Documents, θα ισχύει: $||\text{Document}|| = \sqrt{0.0969^2 + 0.2218^2 + 0.3979^2} = 0.4657$

Sunshine:

$$\text{Query} * \text{Sunshine} = (0.1938 * 0.0969) + (0.4436 * 0.2218) + (0.7958 * 0.3979) = 0.4338$$

$$\|\text{Query}\| = \sqrt{0.1938^2 + 0.4436^2 + 0.7958^2} = 0.9314$$

$$\text{Cosine Similarity (Query, Sunshine)} = \frac{\text{Query} * \text{Sunshine}}{\|\text{Query}\| * \|\text{Sunshine}\|} = \frac{0.4338}{0.9314 * 0.4657} = \mathbf{1.000}$$

Gattaca:

$$\text{Query} * \text{Gattaca} = (0.0969 * 0.0969) + (0 * 0.2218) + (0 * 0.3979) = 0.0093$$

$$\|\text{Query}\| = \sqrt{0.0969^2 + 0^2 + 0^2} = 0.0969$$

$$\text{Cosine Similarity (Query, Gattaca)} = \frac{\text{Query} * \text{Gattaca}}{\|\text{Query}\| * \|\text{Gattaca}\|} = \frac{0.0093}{0.0969 * 0.4657} = \mathbf{0.2060}$$

Gravity:

$$\text{Query} * \text{Gravity} = (0.1938 * 0.0969) + (0 * 0.2218) + (0 * 0.3979) = 0.0187$$

$$\|\text{Query}\| = \sqrt{0.1938^2 + 0^2 + 0^2} = 0.1938$$

$$\text{Cosine Similarity (Query, Gravity)} = \frac{\text{Query} * \text{Gravity}}{\|\text{Query}\| * \|\text{Gravity}\|} = \frac{0.0187}{0.1938 * 0.4657} = \mathbf{0.2071}$$

Event Horizon:

$$\text{Query} * \text{Event Horizon} = (0 * 0.0969) + (0.2218 * 0.2218) + (0.3979 * 0.3979) = 0.2075$$

$$\|\text{Query}\| = \sqrt{0^2 + 0.2218^2 + 0.3979^2} = 0.4555$$

$$\text{Cosine Similarity (Query, Event Horizon)} = \frac{\text{Query} * \text{Event Horizon}}{\|\text{Query}\| * \|\text{Event Horizon}\|} = \frac{0.2075}{0.4555 * 0.4657} = \mathbf{0.9781}$$

The Arrival:

$$\text{Query} * \text{The Arrival} = (0.0969 * 0.0969) + (0.2218 * 0.2218) + (0 * 0.3979) = 0.0585$$

$$\|\text{Query}\| = \sqrt{0.0969^2 + 0.2218^2 + 0^2} = 0.2420$$

$$\text{Cosine Similarity (Query, The Arrival)} = \frac{\text{Query} * \text{The Arrival}}{\|\text{Query}\| * \|\text{The Arrival}\|} = \frac{0.0585}{0.2420 * 0.4657} = \mathbf{0.5190}$$

Με βάση τα αποτελέσματα, η μηχανή αναζήτησης θα εμφανίσει τα έγγραφα με τον εξής τρόπο:

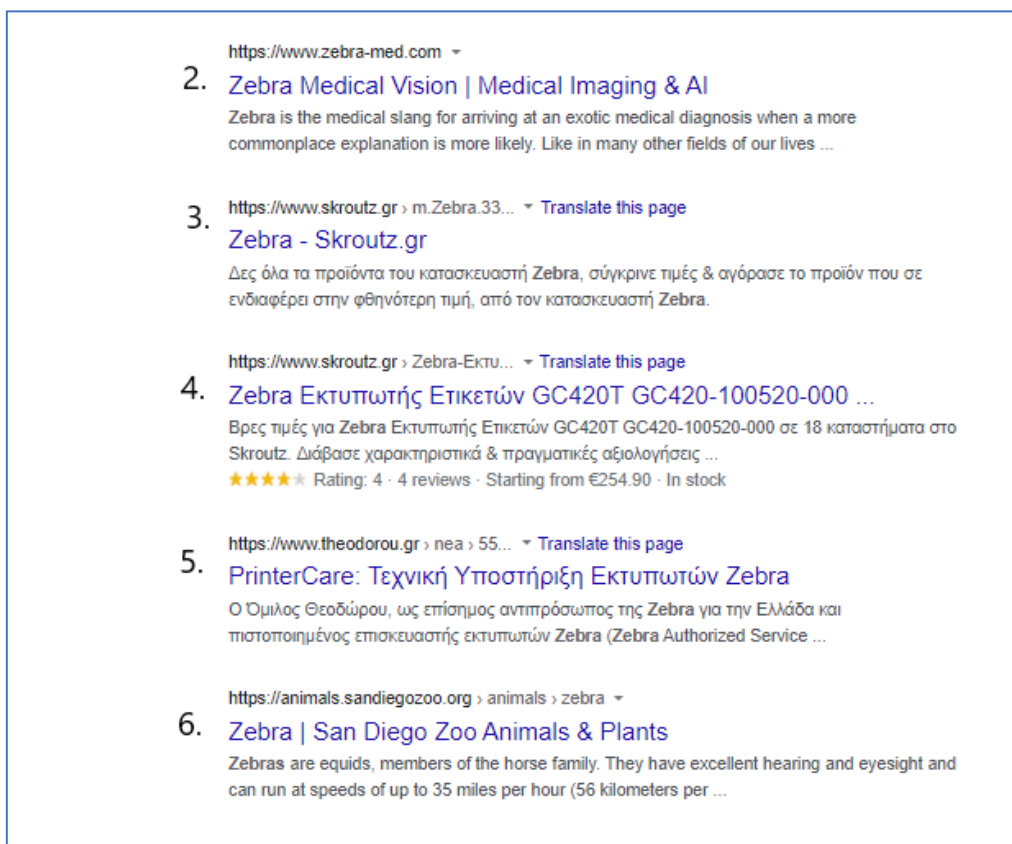
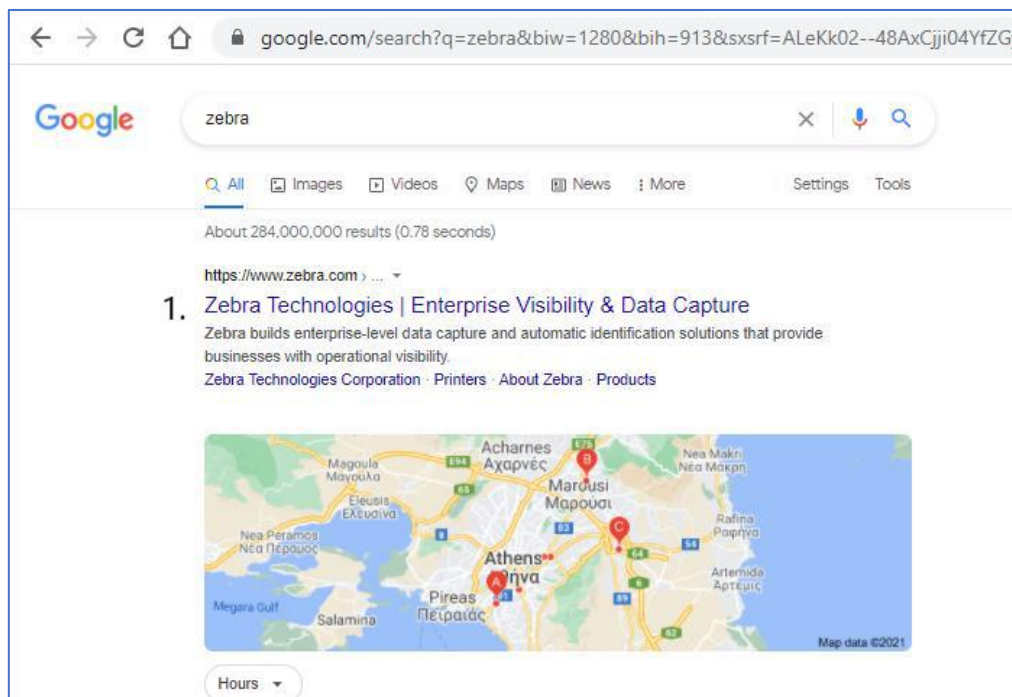
Sunshine, Event Horizon, The Arrival, Gravity, Gattaca.


Σχολιασμός:

Η σειρά των αποτελεσμάτων είναι λογική και αναμενόμενη αν παρατηρηθεί το πόσες κοινές λέξεις περιέχουν από τους όρους αναζήτησης και το Cosine Similarity απλώς το επιβεβαιώνει.

2. Αναζήτηση – PRECISION/RECALL

Ενώ αναζητούνται στοιχεία σχετικά με το ζώο «ζέβρα», η μηχανή αναζήτησης “Google” επιστρέφει τα αποτελέσματα που φαίνονται παρακάτω.



7. <https://www.youtube.com/watch> ▾
Zebra - YouTube
 One is that they serve as camouflage from the lion, its greatest predator. To a human, the black-and-white ...
Jun 12, 2007 · Uploaded by Explore Documentary Films
8. <https://www.awf.org/wildlife-conservation/zebra> ▾
Zebra | African Wildlife Foundation
Where do plains zebras live? They have a wide range in east and southern Africa. They usually live in treeless grasslands and savanna woodlands and are ...
9. <https://en.wiktionary.org/wiki/zebra> ▾
zebra - Wiktionary
NounEdit. zebra (plural zebra or zebras). Any of three species of genus Equus: E. grevyi, E. quagga, or E. ...
10. <https://www.thebarcodewarehouse.co.uk/shop/zebra> ▾
Zebra Printer | Buy Zebra Label Printers, Mobile Label ...
Buy Zebra Label Printers, Mobile Printers, Thermal Transfer Ribbons, Printheads & Labels from Zebra Technologies.
11. <https://zebrasandlibras.com> ▾
Zebras & Libras – ZEBRAS&LIBRAS / SWIMWEAR
Zebras and Libras is a brand inspired by the balance between earth and space. Z&L creates swimwear with high quality textile and new era design cuts.
12. <https://www.multisystems.gr/zebra> ▾ [Translate this page](#)
Zebra εκτυπωτές - Ταμειακές μηχανές
Zebra LP2824 Plus, 8 dots/mm (203 dpi), peeler, RTC, EPL, ZPL, USB, print server (ethernet) (282P-201521-040). Χωρίς ΦΠΑ: 366,89 € Με ΦΠΑ: 454,94 €.
13. <https://www.nationalgeographic.com/mammals/facts> ▾
Plains Zebra | National Geographic
Population and Herd Behavior. Zebras are social animals that spend time in herds. They graze together, primarily on grass, and even groom one another. Plains ...
14. <https://www.linkedin.com/company/zebra-technologies> ▾
Zebra Technologies | LinkedIn
Zebra Technologies | 194845 followers on LinkedIn. Zebra empowers those on the front line in retail, health care, transportation and logistics, manufacturing and ...
15. http://www.albanis.gr/synergates/zebra_technologies ▾
Zebra Technologies Service & Support | Partners :: Albanis.gr
Zebra is committed to offering our customers the highest quality customer care in the industry, delivering outstanding customer service and providing a global ...

16. <https://www.britannica.com/.../Hoofed-Mammals> ▾
zebra | Size, Diet, & Facts | Britannica
Feb 12, 2021 — Zebra, any of three species of strikingly black-and-white striped mammals of the horse family Equidae (genus Equus). All zebras are ...
17. <https://www.sigma-hellas.gr/6-zebra> ▾ [Translate this page](#)
Zebra - Sigma-Hellas.gr
Scanner Zebra (Motorola) LS1203 με βάση. Κωδικός: 001034. 74,00 € +φπα. 91,76 € με ΦΠΑ. Αναγνώστης 1D Laser, Ταχύτητα 100 scan/sec, Ενσύρματο (USB) ...
18. <https://www.novatron.gr/zebra-zt410> ▾ [Translate this page](#)
Zebra ZT410 Industrial Barcode/Label Printer | Novatron
Zebra ZT410. Βιομηχανικός θερμικός εκτυπωτής επικετών-barcode με μέγιστο πλάτος εκτύπωσης 104 mm.

α) Ποια αποτελέσματα αναφέρονται στην αναζήτηση που έγινε και ποια όχι; Να εξηγηθεί με συντομία. Ποια από τα αποτελέσματα είναι σωστά (*true positive*) και ποια λάθος (*false positive*);

β) Αν είναι γνωστό ότι υπάρχουν ακόμη 400 αποτελέσματα που σχετίζονται με την αναζήτηση και δεν βρέθηκαν – *false negative* – να υπολογιστούν τα παρακάτω:

- i) *Precision*
- ii) *Recall*
- iii) *F-Measure*

Να σχολιαστούν τα αποτελέσματα.

(α) Η μηχανή αναζήτησης επιστρέφει 18 αποτελέσματα από τα οποία τα 6 αφορούν το ζώο που αναζητείται. Συγκεκριμένα τα αποτελέσματα 6, 7, 8, 9, 13, 16. Άρα, είναι 6 σωστά (*true positive*) αποτελέσματα και 12 λανθασμένα (*false positive*). Είναι λογικό να προκύπτουν και αποτελέσματα που είναι αδιάφορα για την συγκεκριμένη αναζήτηση αφού δεν προσδιορίστηκε τι ακριβώς αναζητείται, δηλαδή το ζώο. Έτσι, ένας διαφορετικός χρήστης, θα μπορούσε να ψάχνει πληροφορίες για τους εκτυπωτές Zebra ή για *swimwear* που έχουν το ίδιο όνομα. Επομένως, τα αποτελέσματα που προκύπτουν είναι αναμενόμενα.

(β) Από το προηγούμενο ερώτημα, θα ισχύει ότι $TP = 6$, $FP = 12$ & $FN = 400$.

(i) Precision: $Precision = \frac{TP}{TP+FP} = 0.3$

Το precision αφορά την ακρίβεια ή πιστότητα των αποτελεσμάτων αφού υπολογίζεται ο αριθμός των σωστών αποτελεσμάτων προς το σύνολο των σωστών και λανθασμένων αποτελεσμάτων. Στα 18 αποτελέσματα που προέκυψαν, τα 6 είναι σωστά που είναι λιγότερα από τα μισά.

(ii) Recall: $Recall = \frac{TP}{TP+FN} = 0.0147$

Το Recall αφορά την πληρότητα των αποτελεσμάτων που προέκυψαν. Εφόσον υπήρξαν 6 σωστά και υπάρχουν ακόμα 400 αποτελέσματα για το ζώο ‘ζέβρα’ που δεν εμφανίσθηκαν, το αποτέλεσμα είναι πολύ χαμηλό.

(iii) F-Measure: $F-Measure = \frac{2 \cdot (Precision \cdot Recall)}{Precision + Recall} = 0.0281$

Το F-Measure είναι ένας σταθμισμένος αρμονικός μέσος όρος μεταξύ Precision & Recall.

3. RECOMMENDER SYSTEMS

Έξι άνθρωποι αξιολόγησαν τις ταινίες του ερωτήματος 1 με βαθμολογία από 1 (καθόλου καλή) έως 10 (εξαιρετική) και τα αποτελέσματα φαίνονται στον παρακάτω πίνακα. Να αντικατασταθεί όπου X τον τελευταίο ψηφίο του Αριθμού Μητρώου σας, αφού προστεθεί 1.

	Sunshine	Gattaca	Gravity	Event Horizon	Arrival
Χρήστης 1	X	2		5	8
Χρήστης 2	3	3		5	7
Χρήστης 3	9	2	6	3	X
Χρήστης 4	6	4	4	7	3
Χρήστης 5		4	X	8	3
Χρήστης 6	8	8	7	5	6

α) Να υπολογιστεί η ομοιότητα (similarity) μεταξύ των 6 χρηστών χρησιμοποιώντας δυο μεθόδους: Ευκλείδεια απόσταση και Pearson Correlation.

β) Χρησιμοποιώντας K-Nearest Neighbors με $k=2$ και weighted average και με τις δύο μετρικές του ερωτήματος (α) να υπολογιστεί το πως αναμένεται να αξιολογήσει την ταινία Gravity ο Χρήστης 2. Να σχολιαστούν τα αποτελέσματα.

γ) Έστω ότι χρησιμοποιούνται οι προτιμήσεις των χρηστών στις ταινίες για να προτείνουμε φίλους, τότε ποιες σχέσεις φαίνονται πιο πιθανές; Να εξηγηθεί ο τρόπος σκέψης.

Ο πίνακας μετά την αντικατάσταση του X με 7:

	Sunshine	Gattaca	Gravity	Event Horizon	Arrival
Χρήστης 1	7	2		5	8
Χρήστης 2	3	3		5	7
Χρήστης 3	9	2	6	3	7
Χρήστης 4	6	4	4	7	3
Χρήστης 5		4	7	8	3
Χρήστης 6	8	8	7	5	6

(α) Ακολουθεί ο υπολογισμός της ομοιότητας (similarity) μεταξύ των 6 χρηστών χρησιμοποιώντας:

Ευκλείδεια Απόσταση – 1^{ος} Τρόπος:

Στις περιπτώσεις που υπάρχει κενό σε τουλάχιστον ένα κελί, το ζευγάρι δεν υπολογίζεται.

Ακολουθώντας τον τύπο που φαίνεται παρακάτω από τις διαφάνειες, προκύπτει ο πίνακας.

$$\text{sum} = \text{sum} + ((\text{rating}(\text{User1}, \text{item } j) - \text{rating}(\text{User2}, \text{item } j))^2$$

$$\text{Δηλαδή, } \text{dist}((x, y), (a, b)) = (x - a)^2 + (y - b)^2$$

$$\text{Και similarity: } \text{similarity}(0, 1) = \frac{1}{1 + \sqrt{\text{sum}}}$$

Οπότε, ο πίνακας των Ευκλείδειων Αποστάσεων, θα είναι:

	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	18	9	34	38	41
Χρήστης 2		-	41	30	26	51
Χρήστης 3			-	49	46	43
Χρήστης 4				-	10.0	42
Χρήστης 5					-	34
Χρήστης 6						-

Και ο πίνακας ομοιότητας θα είναι:

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	0.1907	0.25	0.1463	0.1395	0.1350
Χρήστης 2		-	0.1350	0.1543	0.1639	0.1228
Χρήστης 3			-	0.1252	0.1284	0.1323
Χρήστης 4				-	0.2402	0.1336
Χρήστης 5					-	0.1463
Χρήστης 6						-

Και ο ενδεικτικός κώδικας για την περίπτωση (Χρήστης 3, Χρήστης 4) είναι:

```
from math import*
d1 = 9-6
d2 = 2-4
d3 = 6-4
d4 = 3-7
d5 = 7-3
d11 = d1**2
d22 = d2**2
d33 = d3**2
d44 = d4**2
d55 = d5**2
sum1 = d11+d22+d33+d44+d55
sim1 = 1/(1+math.sqrt(sum1))
print("sum is", sum1)
print("sim is", sim1)
```


Ευκλείδεια Απόσταση – 2^{ος} Τρόπος:

Στις περιπτώσεις που υπάρχει κενό σε τουλάχιστον ένα κελί, το ζευγάρι υπολογίζεται (αν και είναι λάθος έτσι με βάση τις διαλέξεις).

Ακολουθώντας ένα διαφορετικό τύπο για την Ευκλείδεια Απόσταση, ο οποίος είναι της μορφής

$$\text{Δηλαδή, } \text{dist}((x, y), (a, b)) = \sqrt{(x - a)^2 + (y - b)^2}$$

$$\text{Και similarity: } \text{similarity}(0, 1) = \frac{1}{1 + \sqrt{\text{sum}}}$$

Οπότε, ο πίνακας των Ευκλείδειων Αποστάσεων, θα είναι:

	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	4.2426	6.7082	7.0710	11.6619	9.4868
Χρήστης 2		-	8.7749	6.7823	9.1651	10.0
Χρήστης 3			-	7.0	11.2694	6.5574
Χρήστης 4				-	6.7823	6.4807
Χρήστης 5					-	9.8994
Χρήστης 6						-

Και ο πίνακας ομοιότητας θα είναι:

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	0.3268	0.2785	0.2732	0.2265	0.2450
Χρήστης 2		-	0.2523	0.2774	0.2482	0.2402
Χρήστης 3			-	0.2742	0.2295	0.2808
Χρήστης 4				-	0.2774	0.2820
Χρήστης 5					-	0.2411
Χρήστης 6						-

Και ο ενδεικτικός κώδικας για την περίπτωση (Χρήστης 3, Χρήστης 4) είναι:

```
from math import*
d1 = 9-6
d2 = 2-4
d3 = 6-4
d4 = 3-7
d5 = 7-3
d11 = d1**2
d22 = d2**2
d33 = d3**2
d44 = d4**2
d55 = d5**2
sum1 = d11+d22+d33+d44+d55
sum2 = math.sqrt(sum1)
sim2 = 1/(1+math.sqrt(sum2))
print("sum is", sum2)
print("sim is", sim2)
```

Pearson Correlation:

Αρχικά, υπολογίζεται το άθροισμα των ratings και του τετραγώνου των ratings για κάθε χρήστη.

- $\text{Sum1} = \text{sum}(\text{ratings user1}), \text{Sum2} = \text{sum}(\text{ratings user2})$
- $\text{Sum1Sq} = \text{sum}[(\text{ratings user1})^2], \text{Sum2Sq} = \text{sum}[(\text{ratings user2})^2]$

Έπειτα, υπολογίζεται το άθροισμα των γινομένων των ratings για τα n αντικείμενα που έχουν αξιολογηθεί και από τους δύο.

- $\text{pSum} = \text{sum}((\text{rating user1 for item i}) * (\text{rating user2 for item i}))$
- $\text{num} = \text{pSum} - \frac{\text{sum1} \cdot \text{sum2}}{n}$
- $\text{den} = \sqrt{(\text{sum1Sq} - \frac{\text{sum1}^2}{n}) \cdot (\text{sum2Sq} - \frac{\text{sum2}^2}{n})}$
- $\text{correlation} = \frac{\text{num}}{\text{den}}$
- $\text{similarity} = \frac{1 + \text{correlation}}{2}$

Οπότε, από τους παραπάνω κανόνες, κατασκευάζεται ο παρακάτω πίνακας.

Sum	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	22	22	22	22	15	22
Χρήστης 2	18	18	18	18	15	18
Χρήστης 3	21	21	27	27	18	27
Χρήστης 4	20	20	24	24	18	24
Χρήστης 5	15	15	22	22	22	22
Χρήστης 6	27	27	34	34	26	34

SumSq	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	142	142	142	142	93	142
Χρήστης 2	92	92	92	92	83	92
Χρήστης 3	143	143	179	179	98	179
Χρήστης 4	110	110	126	126	90	126
Χρήστης 5	89	89	138	138	138	138
Χρήστης 6	189	189	238	238	174	238

Psum	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	108	-				
Χρήστης 3	138	97	-			
Χρήστης 4	141	86	128	-		
Χρήστης 5	72	73	95	109	-	
Χρήστης 6	145	115	187	161	139	-

Num	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	9	-				
Χρήστης 3	22.5	2.3	-			
Χρήστης 4	31	-4	-1.6	-		
Χρήστης 5	-3	-2	-4	10	-	
Χρήστης 6	-3.5	-10.5	3.4	-2.2	-4	-

Den	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	18.198	-				
Χρήστης 3	26.224	18.980	-			
Χρήστης 4	14.491	10.488	18.935	-		
Χρήστης 5	15.874	10.583	17	12.369	-	
Χρήστης 6	11.905	8.616	15.025	8.569	9.219	-

Correlation	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	0.592	-				
Χρήστης 3	0.857	0.131	-			
Χρήστης 4	2.139	-0.381	-0.084	-		
Χρήστης 5	-0.188	-0.001	-0.235	0.808	-	
Χρήστης 6	-0.293	-1.218	0.226	-0.256	-0.433	-

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	0.796	-				
Χρήστης 3	0.928	0.565	-			
Χρήστης 4	1.569	0.309	0.458	-		
Χρήστης 5	0.406	0.499	0.382	0.904	-	
Χρήστης 6	0.353	-0.109	0.613	0.372	0.283	-

(β) Χρησιμοποιώντας K-Nearest Neighbors με $k = 2$ και weighted average προκύπτει ότι:

- Από τον Similarity πίνακα της Ευκλείδειας Απόστασης του 1^{ου} Τρόπου που παρατίθεται και παρακάτω για ευκολία, προκύπτει ο πίνακας γειτνίασης για τον Χρήστη 2.

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	0.1907	0.25	0.1463	0.1395	0.1350
Χρήστης 2		-	0.1350	0.1543	0.1639	0.1228
Χρήστης 3			-	0.1252	0.1284	0.1323
Χρήστης 4				-	0.2402	0.1336
Χρήστης 5					-	0.1463
Χρήστης 6						-

Πίνακας Γειτνίασης του Χρήστη 2:

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 2	0.1907	-	0.1350	0.1543	0.1639	0.1228

Οπότε, προκύπτει ότι οι δύο πιο κοντινοί γείτονες του Χρήστη 2 είναι ο Χρήστης 1 και ο Χρήστης 5. Όμως, ο Χρήστης 1 δεν έχει βαθμολογήσει την ταινία Gravity. Έτσι, θα υπολογιστεί ο αμέσως πιο κοντινός γείτονας που είναι ο Χρήστης 4. Έτσι,

$$\text{PredictedRating}_{u2} = \frac{\text{sim}(u2,u4) \cdot \text{rating}(u4) + \text{sim}(u2,u5) \cdot \text{rating}(u5)}{\text{sim}(u2,u4) + \text{sim}(u2,u5)} = \frac{0.1543 \cdot 4 + 0.1639 \cdot 7}{0.1543 + 0.1639} = \frac{1.7645}{0.3182} = 5.5452$$

Δηλαδή, η αναμενόμενη βαθμολογία του Χρήστη 2 για το Gravity θα είναι 5.5.

- Από τον Similarity πίνακα της Pearson Correlation που παρατίθεται και παρακάτω για ευκολία, προκύπτει ο πίνακας γειτνίασης για τον Χρήστη 2.

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	0.796	-				
Χρήστης 3	0.928	0.565	-			
Χρήστης 4	1.569	0.309	0.458	-		
Χρήστης 5	0.406	0.499	0.382	0.904	-	
Χρήστης 6	0.353	-0.109	0.613	0.372	0.283	-

Πίνακας Γειτνίασης του Χρήστη 2:

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 2	0.796	-	0.565	0.309	0.499	-0.109

Οπότε, προκύπτει ότι οι δύο πιο κοντινοί γείτονες του Χρήστη 2 είναι ο Χρήστης 1 και ο Χρήστης 3. Όμως, ο Χρήστης 1 δεν έχει βαθμολογήσει την ταινία Gravity. Έτσι, θα υπολογιστεί ο αμέσως πιο κοντινός γείτονας που είναι ο Χρήστης 5. Έτσι,

$$\text{PredictedRating}_{u2} = \frac{\text{sim}(u2,u3) \cdot \text{rating}(u3) + \text{sim}(u2,u5) \cdot \text{rating}(u5)}{\text{sim}(u2,u3) + \text{sim}(u2,u5)} = \frac{0.565 \cdot 6 + 0.499 \cdot 7}{0.565 + 0.499} = \frac{6.883}{1.064} = 6.468$$

Δηλαδή, η αναμενόμενη βαθμολογία του Χρήστη 2 για το Gravity θα είναι 6.5.

Σχολιασμός:

Παρατηρείται ότι η μέθοδος της Ευκλείδειας Απόστασης και η μέθοδος Pearson Correlation διαφέρουν κατά μια μονάδα γεγονός που δικαιολογείται από τον τρόπο που υπολογίζεται η κάθε μέθοδος καθώς στην πρώτη μέθοδο υπολογίζεται το διάνυσμα της σχετικής διαφοράς ενώ στην δεύτερη υπολογίζεται και η weighted average των βαθμολογιών.

(γ) Το καταλληλότερο κριτήριο είναι το Similarity. Αναλυτικά, για τις δύο μεθόδους θα ισχύει:

Για τη μέθοδο της Ευκλείδειας Απόστασης:

	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-	0.1907	0.25	0.1463	0.1395	0.1350
Χρήστης 2		-	0.1350	0.1543	0.1639	0.1228
Χρήστης 3			-	0.1252	0.1284	0.1323
Χρήστης 4				-	0.2402	0.1336
Χρήστης 5					-	0.1463
Χρήστης 6						-

Σχολιασμός:

Έτσι, προτείνονται οι σχέσεις για τους χρήστες (1, 3), (4, 5), (1, 2), αφού έχουν το μεγαλύτερο similarity (0.25, 0.2402 και 0.1907).

Για τη μέθοδο του Pearson Correlation:

Similarity	Χρήστης 1	Χρήστης 2	Χρήστης 3	Χρήστης 4	Χρήστης 5	Χρήστης 6
Χρήστης 1	-					
Χρήστης 2	0.796	-				
Χρήστης 3	0.928	0.565	-			
Χρήστης 4	1.569	0.309	0.458	-		
Χρήστης 5	0.406	0.499	0.382	0.904	-	
Χρήστης 6	0.353	-0.109	0.613	0.372	0.283	-

Σχολιασμός:

Έτσι, προτείνονται οι σχέσεις για τους χρήστες (4, 1), (3, 1), (5, 4), αφού έχουν το μεγαλύτερο similarity (1.569, 0.928, 0.904).