

דו"ח - תרגיל 3

Refael Greenfeld 305030868

Danit Yshaayahu 312434269

חלק א':

1. השתמשנו בממוצע משוקלל (average weighted) כפי שראינו בתירגול. השתמשנו במדד זה היות והוא לוקח בחשבון את ממוצע ההצבעות הכללי של המשתמשים ומנרמל באמצעותו את כלל הדירוגים. דבר זה עוזר למציאת הספרים הכי מומלצים בכך שכל המשתמשים מדרגים באותה הסקאלה וכמו כן נלקח בחשבון מספר ההצבעות לכל ספר.

2.

	title	id	score
0	The Complete Calvin and Hobbes	3628	4.725772
1	Book of Mormon, Doctrine and Covenants, Pearl of Great Price	4154	4.655205
2	Jesus the Christ	4868	4.652004
3	It's a Magical World: A Calvin and Hobbes Collection	4483	4.618368
4	The Book of Mormon: Another Testament of Jesus Christ	1338	4.598453
5	The Essential Calvin and Hobbes: A Calvin and Hobbes Treasury	1010	4.555346
6	Harry Potter Boxset (Harry Potter, #1-7)	422	4.554538
7	The Far Side Gallery 3	4046	4.526230
8	Words of Radiance (The Stormlight Archive, #2)	862	4.502603
9	Calvin and Hobbes	780	4.494695

.3

	title	id	score
0	Dune (Dune Chronicles #1)	126	4.379379
1	All the Light We Cannot See	143	4.327645
2	Unbroken: A World War II Story of Survival, Resilience, and Redemption	144	4.280547
3	Harry Potter and the Goblet of Fire (Harry Potter, #4)	24	4.256506
4	Where the Wild Things Are	102	4.236736
5	Maus I: A Survivor's Tale: My Father Bleeds History (Maus, #1)	490	4.229157
6	The Orphan Master's Son	1462	4.229157
7	Between the World and Me	983	4.229157
8	The Handmaid's Tale	119	4.21041
9	Watchmen	191	4.200513

.4

	title	id	score
0	The Book of Mormon: Another Testament of Jesus Christ	1338	4.511567
1	Harry Potter Boxset (Harry Potter, #1-7)	422	4.443835
2	The Giving Tree	85	4.398191
3	Harry Potter and the Deathly Hallows (Harry Potter, #7)	25	4.38968
4	The Princess Bride	89	4.384318
5	The Runaway Bunny	1984	4.372014
6	Wonder	250	4.371596
7	Oh, The Places You'll Go!	358	4.366132
8	Anne of Green Gables (Anne of Green Gables, #1)	133	4.352603
9	To Kill a Mockingbird	4	4.346225

חלק ג':

8. בחרנו להשתמש בפיצ'רים הבאים: שם הסופר, שפת הספר- כאשר אנו מאחדים שפות שיש להן אותה משמעות (אנגלית ואנגלית מארה"ב), שנת הוצאה – כאשר אנו מאחדים עשורים לאותו פיצ'ר משנת 1850 והילך ולפני כן מפרידים בקפיצות גדולות של כ-1000 שנה. יתר על כן, אנו משתמשים בתגים אשר סופקו לנו על כל ספר, כשאנו משתמשים רק בתגים שקיימים לפחות פעמיים (פעם אחת זה חסר משמעות).

10. רשימת הספרים עבור 'Twilight (Twilight, #1)':

- a. The Great Gatsby
- b. The Hunger Games (The Hunger Games, #1)
- c. The Fault in Our Stars
- d. Pride and Prejudice
- e. Harry Potter and the Sorcerer's Stone (Harry Potter, #1)
- f. The Catcher in the Rye
- g. The Death of Ivan Ilych
- h. Magic Bleeds (Kate Daniels, #4)
- i. The Black Echo (Harry Bosch, #1; Harry Bosch Universe, #1)
- j. Twenty Boy Summer

חלק ד':

11. טבלת מדדים:

	Precision_k	ARHR	RMSE
cosine	0.032	0.084	0.968
euclidean	0.012	0.065	0.967
jaccard	0.036	0.110	0.968

12. ניתן לראות כי קיים פער גדול בין מדד ה-precision וה-ARHR לבין מדד ה-RMSE. פער זה נובע בשל השוני המהותי בין המדדים, בשניים הראשונים אנו בודקים רק את נכונות ה-10 הראשונים בעוד שב-RMSE אנו משווים את התוצאות לכל הרשימה. בשניים הראשונים נקבל ניקוד טוב רק עבור מה שהצלחנו לחזות ב-10 הראשונים בעוד שב-RMSE ירד לנו ניקוד רק על

הפער ב-ranking כך כל חיזוי ילקח בחשבון גם ספר שהוא דורג נמוך וחזינו אותו כנמוך יתמוך בחיזוי, כלומר יעלה את האחוזים.

בנוסף לכך, ב-ARHR כל חיזוי מקבל ניקוד על פי המיקום שלו ברשימת ההמלצות. כך אם הצלחנו לחזות נכון את מקומות 7-10 נקבל ניקוד יותר נמוך מאשר אם היינו חוזים את מקומות 1-3 כלומר המדד יותר מדויק. תוצאת אלו מראות כי גם המיקומים בהמלצות שלנו נכונים כי ARHR גבוה ממדד ה-Precision.