## שאלה 2

להלן רשימת המשפטים שנבחרו והדמיון בינם, כאשר הדובר הראשון הינו אדלשטיין והשני ריבלין:

|  |  |
| --- | --- |
| * בבקשה , אדוני . * דבר שני , חשבנו שהנושא שבא לידי דיון ציבורי בכל מקום בעולם ובוודאי בעולם הנאור . | 0.47726664 |
| * חבר הכנסת אברהם מיכאלי , ואחריו - חבר הכנסת אחמד טיבי . * חבר הכנסת פלסנר . | 0.82322127 |
| * תודה . * בבקשה , אדוני השר . | 0.6471806 |
| * חבר הכנסת עבד - מבקש להציע דווקא הצעה שונה מזו של יולי אדלשטיין , אבל למרות זאת אנחנו נאפשר לו לדבר במסגרת דקה . * בשעה תתקיים אזכרה בהר - הרצל , וכל חברי הכנסת כמובן מוזמנים להשתתף זו . | 0.75818926 |
| * בסוף דבריך , כבוד השר , תתייחס גם למה שאתה מציע מבחינה , כי בדרך כלל ביום כזה מתקיים דיון גם בוועדות , אז זה מיותר אולי לשלוח עוד פעם לוועדה . * הוא שמע את השאלה . | 0.672886 |
| * בבקשה , אדוני . * 53 בעד , אין נגד ואין נמנעים . | 0.6942693 |
| * הצעת האי - אמון מטעם סיעת יש עתיד - הצעה אחרת - לכן היא לא צורפה לשלוש הראשונות , ושמה : הרוח מצד שרים בממשלה , אשר מתירה לערער את סמכות הצבא ופוגעת בביטחון המדינה . * השלום זה אלה שרוצים להיפרד ממך , חבר הכנסת . | 0.77346736 |
| * אנחנו אלקטרונית על הסתייגות קבוצת ברכה . * הצעת ועדת הכנסת לתיקון סעיפים 1 , 3 , 6 , 32 ו- 36 , הוספת סעיפים 18 - 24 וביטול סעיפים 24 - 27 , 31 א 34 א 37 , 51 , 58 ו- 58 א לתקנון הכנסת . | 0.48006803 |
| * תודה לחבר הכנסת כבל . * מסיר או מבקש להצביע ? | 0.44333383 |
| * זהו פרויקט משותף של אגף התקציבים , שאינו ידוע גדול , של משרד השיכון ושל משרד הקליטה . * כל אחד אמר את דבריו . | 0.555 |

### הסבר כללי:

דמיון המשפטים מחושב ע"י קוסינוס הזווית בין הווקטורים המשוקללים של כל אחד מהמשפטים. הזווית יכולה להיות אינדיקציה לדמיון שכן ככל שהזווית גדולה ה"מרחק" בין המשפטים גדל ולכן הם דומים פחות. לעיתים, אם המשפט לא מנוסח באופן מתחכם (ראה דוגמה בסוף), הוא מעביר נושא מרכזי וצירוף וקטורי המילים יכל להוות מדד למהו רעיון המשפט. אם הרעיונות של המשפטים קרובים המרחק בינם יהיה קטן כי הם יצביעו לאותו איזור במרחב ה- word embedding.

נשים לב כי הנ"ל נכון במקרים ספציפיים שכן משפטים שמכילים אירוניה יכולים להכין מילים רחוקות והווקטור המשוקלל לא ישקף את רעיון המשפט, לדוגמה, אם נראה לבצע פעולה כגון סיווג סנטימנטים, המילה הבאה עלולה להיות מסווגת כבעלת רגש חיובי: "כישלון מדהים" ו-"כישלון מחריד" שתהיה בעלת רגש שלילי אבל הוקטור המשוקלל שלה כנראה יהיה רחוק משל המילה הקודמת.

### מרחקים עבור הדוברים שלנו:

נשים לב שמשפטים בעלי אורך שונה באופן מובהק הינם רחוקים זה מזה ---