

שאלה 1

לצורך השאלה בחרנו את המילה: Rock.

ה Sense הראשון הוא ד'אנר מוזיקלי, כאשר ה Seed שנבחר הינו: Music ו Band ה Sense השני הוא החומר אבן, כאשר ה Seed שנבחר הינו: Large ו Climbing.

סעיף D

חמשת המשפטים שנבחרו עבור ה Sense הראשון:

- This is called a " blank end " , and the end number usually goes in the farthest column on the right in the row of the team that has the " hammer " (**last rock advantage**) , or on a special spot for blank ends .
- In 1962 and 1963 , with his next band " \ 'Harmoniler (" The Harmonies ") , he recorded cover versions of some of the then popular American twist songs , but also rearrangements of Turkish folk **songs in rock and roll** form , marking the beginning of Anatolian rock movement , a synthesis of Turkish folk music and rock .
- Charles Hardin Holley , known professionally as " \ 'Buddy Holly (September 7 , 1936 February 3 , 1959) was an American singer-songwriter and a **pioneer of rock and roll** .
- While most commercial stations concentrate on a theme , such as 1980s music or classic rock , Radio 1 plays a mix of current songs , including independent/ alternative , Hip **Hop , Rock , House** Music , Electronica , Dance Music , Drum N \ ' Bass and Pop Music Due to restrictions on the amount of commercial music that could be played on radio in the UK until 1988 (the " needle time " limitation) the station has recorded many live performances and studio sessions , many of which have found their way to commercially-available LPs and CDs .
- In the formation of an ordinary agate , it is probable that waters containing silica in solution -- derived , perhaps , from the decomposition of some of the silicates in the lava itself -- percolated **through the rock and deposited** a siliceous coating on the interior of the vapour-vesicles .

חמשת המשפטים שנבחרו עבור ה Sense השני:

- One way that dunes can move is through saltation , where sand particles skip along the ground **like a rock thrown across** a pond might skip across the water 's surface .
- While the Troodos Mountains are a massif formed of **molten igneous rock , the Kyrenia** Range is a narrow limestone ridge that rises suddenly from the plains .
- Elvis Presley , who was known early on as The Hillbilly Cat and was a regular on the radio program " Louisiana Hayride " , went on to become a defining figure in the emerging **genre of rock and roll** .
- Their style has been variously classified as **punk , art rock and post-punk** , but they are most often remembered for their late 1970s and early 1980s New Wave sound .

- With country music record sales in **tatters as rock and roll** took over , Atkins and Bob Ferguson took their cue from Owen Bradley and eliminated fiddles and steel guitar as a means of making country singers appeal to pop fans .

כפי שניתן לראות, עבור ה Sense הראשון, החיזוי צדק ב 3 מתוך 5 משפטים. עבור ה Sense השני, החיזוי צדק ב 2 מתוך 5 משפטים.

סעיף E

במקרה שנרצה להרחיב את האלגוריתם למילים עם 3 Senses, נבצע את האלגוריתם בצורה דומה עד שלב חישוב הדירוג של כל קולוקציה.

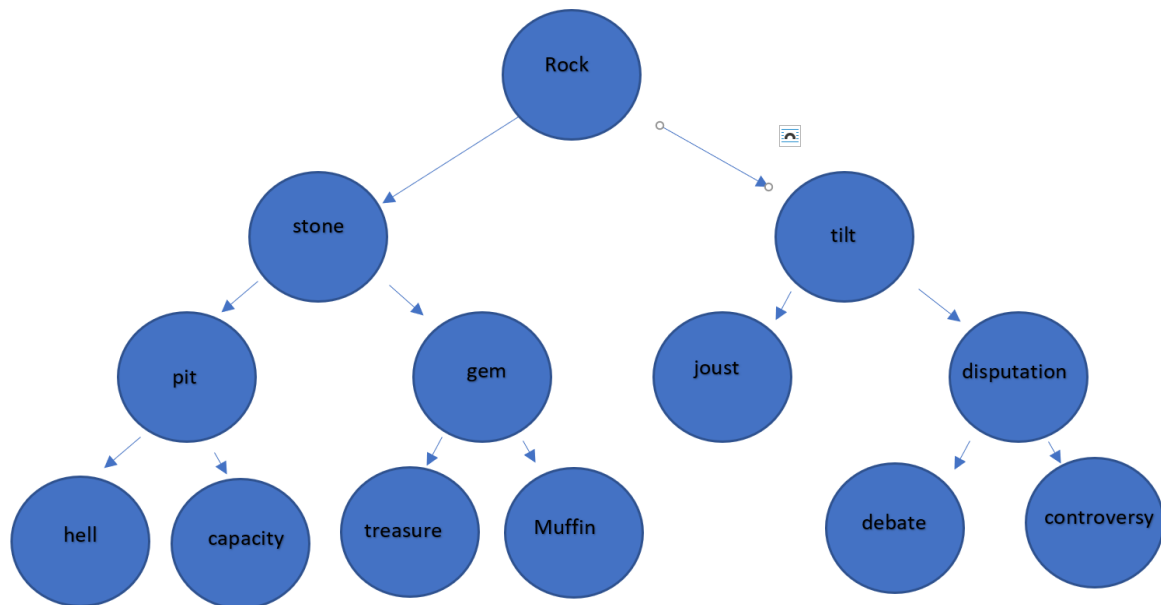
את חישוב הדירוג נבצע באופן הבא:

$$score(collocation_i) = \log \left(\frac{\Pr(sense_A | collocation_i)}{\Pr(sense_B | collocation_i) \cdot \Pr(sense_C | collocation_i)} \right)$$

זאת מכיוון ששני המאורעות במכנה, הינם מאורעות בלתי תלויים ולכן זו ההסתברות ששניהם יתקיימו.

שאלה 2:

א. בשאלה זאת בנינו עץ מהמילה Rock והוא:



	Rock	Stone	Tilt	Pit	Gem	Hell	Capacity	Treasure	Muffin	Joust	Disputation	Debate	Controversy
Rock	0	1	1	2	2	3	3	3	3	2	3	3	3
Stone	1	0	-	1	1	2	2	2	2	-	-	-	-
Tilt	1	-	0	-	-	-	-	-	-	1	1	2	2
Pit	2	1	-	0	-	1	1	-	-	-	-	-	-
Gem	2	1	-	-	0	-	-	1	1	-	-	-	-
Hell	3	2	-	1	-	0	-	-	-	-	-	-	-
Capacity	3	2	-	1	-	-	0	-	-	-	-	-	-
Treasure	3	2	-	-	1	-	-	0	-	-	-	-	-
Muffin	3	2	-	-	1	-	-	-	0	-	-	-	-
Joust	2	-	1	-	-	-	-	-	-	0	-	-	-
Disputation	2	-	1	-	-	-	-	-	-	-	0	1	1
Debate	3	-	2	-	-	-	-	-	-	-	1	0	-
Controversy	3	-	2	-	-	-	-	-	-	-	1	-	0

ב. לאחר הרצת Lin similarity measure נתקלנו במספר דוגמאות בעלות היגיון ברור לנו בהשוואה בין 2 זוגות מילים.

- $treasure \& gem$ בעל דמיון גבוה יותר מ- $treasure \& stone$. נראה לנו הגיוני מאוד, הרי gem הוא לחלוטין $treasure$ אך לא כל $stone$ הוא $treasure$.
- $controversy \& disputation$ בעל דמיון גבוה יותר מ- $controversy \& debate$. הרי הזוג הראשון הינו מילים נרדפות, ו- $debate$ הוא מחלוקת מסוג מאוד ספציפי, ולכן הדבר נראה לנו הגיוני.
- $Hell \& pit$ בעל דמיון גבוה יותר מ- $hell \& stone$. ישנו הביטוי pit of $hell$ ובכל מקרה השאול נחשב שמקום מתחת לאדמה, ולכן מתחבר ל- pit . אין ל- $hell$ קשר נראה לעין עם $stone$ מראות עינינו.
- $Muffin \& gem$ בעל דמיון גבוה יותר מ- $muffin \& hell$. טוב $muffin$ is a gem , וודאי של $hell$.

- Stone & pit בעל דמיון גבוה יותר מ Piti muffin ו Pit יכול להיות מלא באבנים, אך לצערנו לא בmuffins.
- ג. התייעצנו עם גאיה, סטודנטית לספרות ופסיכולוגיה וחובבת פסיכולינגוויסטיקה (ולא מתנגדת למשרת מחקר) שלא לוקחת את הקורס בממצאים שלנו. גאיה הסכימה עם רב הטיעונים שלנו מלבד הטיעון השני. לדעתה, debate אינו מדבר על מחלוקות ספציפיות ולכן מוזר בעינה שהוא בעל דמיון שונה.

שאלה 3

סעיף A

שתי הגדרות מילוניות למילה Rock (ממילון Oxford) על פי ה Senses שלה הן:

- a type of loud popular music, developed in the 1960s, with a strong beat played on electric guitars and drums.
- the hard solid material that forms part of the surface of the earth and some other planets

לפי הגדרות אלו ומהגדרת האלגוריתם, לא יהיה ניתן לסווג אף אחד מהמשפטים שצוינו בסעיף הראשון.

סעיף B

יתרונות האלגוריתם

- מספק אוסף של מילים, שבאמצעות קשורות למילה הנבחרת.
- ה Seed הראשוני רחב (תלוי באורך ההגדרה המילונית) וכך תיאורטית ניתן למצוא יותר התאמות.

חסרונות

- גודל החלון משפיע על יכולת האלגוריתם למצוא התאמה טובה של המילים.
- ה Yarowsky Algorithm ה Seed יכול להיבחר בקפידה כך שצמד המילים כמעט תמיד יופיעו ביחד, דבר המאפשר יכולת התאמה גבוהה, בעיקר בחלונות קטנים.
- האלגוריתם לא גמיש ותלוי בצד שלישי (המילון). לא ניתן לשלוט במילים הכתובות בהגדרה ולפעמים נעשה שימוש במילים כלליות מדי, או פחות נפוצות.