

עיבוד שפות טבעיות – תרגיל בית 4

חלק א' – התנסות ראשונית

השתמשתי ברשימת המילים הבאה:

Word Pairs and Distances:

1. hot - cold: 0.801
2. good - bad: 0.796
3. west - east: 0.954
4. always - never: 0.899
5. yes - no: 0.794
6. apple - banana: 0.561
7. kid - child: 0.556
8. play - act: 0.421
9. little - tiny: 0.584
10. van - car: 0.493

בחרתי את המילים כך שחצי מהן הן מילים דומות או נרדפות, והחצי השני הן מילים הופכיות. ניתן לראות שהמרחק בין המילים ההופכיות (החצי הראשון) הוא מאוד גבוה, ואילו המרחק בין המילים הדומות הוא נמוך. זה הגיוני מכיוון שאם מילים הן דומות, גם הווקטור שלהן אמור להיות דומה, כך שהמרחק קטן. וההפך עבור מילים שהן הפכים.

חלק 2 – יצירת להיט חדש

Said mother , said baby , said baby
What you going ? (What you doing ?)
Where you from ? (Where you at ?)
Oh , you out plans ? (You got plans ?)
Don ' t know that (shut your trap)
I ' m sipping wine (sip , sip) in a headdress (drip , drip)
I look too good (look too good) to be about (woo - woo)
My house water (house clean) , my pool warm (pool warm)
Just shaved (smooth like a baby)
We will be dancing , romancing
In the west wing and the west wing
Of this hotel , what ' s happening ?
I ain ' t got no games
Every word that I want is coming straight from the heart
So if you tryna lay in these weapons
I ' ma leave the window open
(I ' ma leave the window open)
I ' ma leave the window open , girl
(I ' ma leave the window open , hoping)
That you 're the way I feel
And you want me like I want you tomorrow , baby
(ask me that you ' re coming through)
Ooh , you ' re so honey (so sweet) , so tight (so tight)
I won ' t bite (ah - ah) , unless you well (unless you like)
If you smoke (what you smoke ?) I get the haze (Purple Haze)
And if you ' re eat , girl , I got filets (woo - woo)
Ooh , baby , don ' t get me (waiting)
There ' s so much life we could be making (shamone)
I ' m talking strangers , cuddling
Rose petals in the bathroom
Girl , let ' s jump in , it ' s bubbled
I ain ' t got no games
Every word that I want is coming straight from the heart
So (if you) if you (tryna) tryna lay in different arms
I ' ma leave the window open
(I ' ma leave the window open)
I ' ma leave the window open , girl
(I ' ma leave the window open , hoping)
That you 're the way I feel
And you want me like I want you tomorrow , baby
(ask me that you ' re coming through) come on , girl
La - la - la , la - la - la - la (I give you , baby)
La - la - la , la - la - la - la (I gotta do you , baby)
La - la - la , la - la - la - la (girl , I ' m tryna to you this)
nee - ah - ah - ah , ah - ah
Hey - hey , I ' ma leave my doors open , baby (I ' ma leave the door open)

I ' ma leave , I ' ma leave my doors open , girl
 (I ' ma leave the window open , hoping) and I ' m hoping , hoping
 That you 're the way I feel
 And you want me like I want you tomorrow , baby
 (ask me that you ' re coming through) woo !
 La - la - la , la - la - la - la (ask me)
 (ask me that you ' re coming through)
 whahng - woo - woo - woo , woo - woo - woo , woo - woo - woo
 whahng - woo - woo - woo , woo - woo - woo , woo - woo
 de - la - la , la - la - la - la (la - la - la , la - la)
 (ask me that you ' re coming through)
 Girl , I ' m here still waiting for you (oh !)
 Come on over , I ' ll please you (I gotta know !)
 La - la - la , la - la - la - la (I ' m go , waiting , waiting)
 (ask me that you ' re coming through) for you
 Girl , I ' m here just wait for you
 Come on over , I ' ll please you
 de - la - la , la - la - la - la (la - la , la - la)

1. היו משפטים שיצאו הגיוניים והיו כאלה שלא, אבל הרוב הגיוניים. בהרבה משפטים הזמן של המילה השתנה, ולכן המשפט כבר לא היה הגיוני. כמו כן, היו הרבה אובייקטים ששגורים בשפה, ואם מחליפים אותם באובייקט דומה אחר הם כבר לא הגיוניים.

2. לדוגמה:

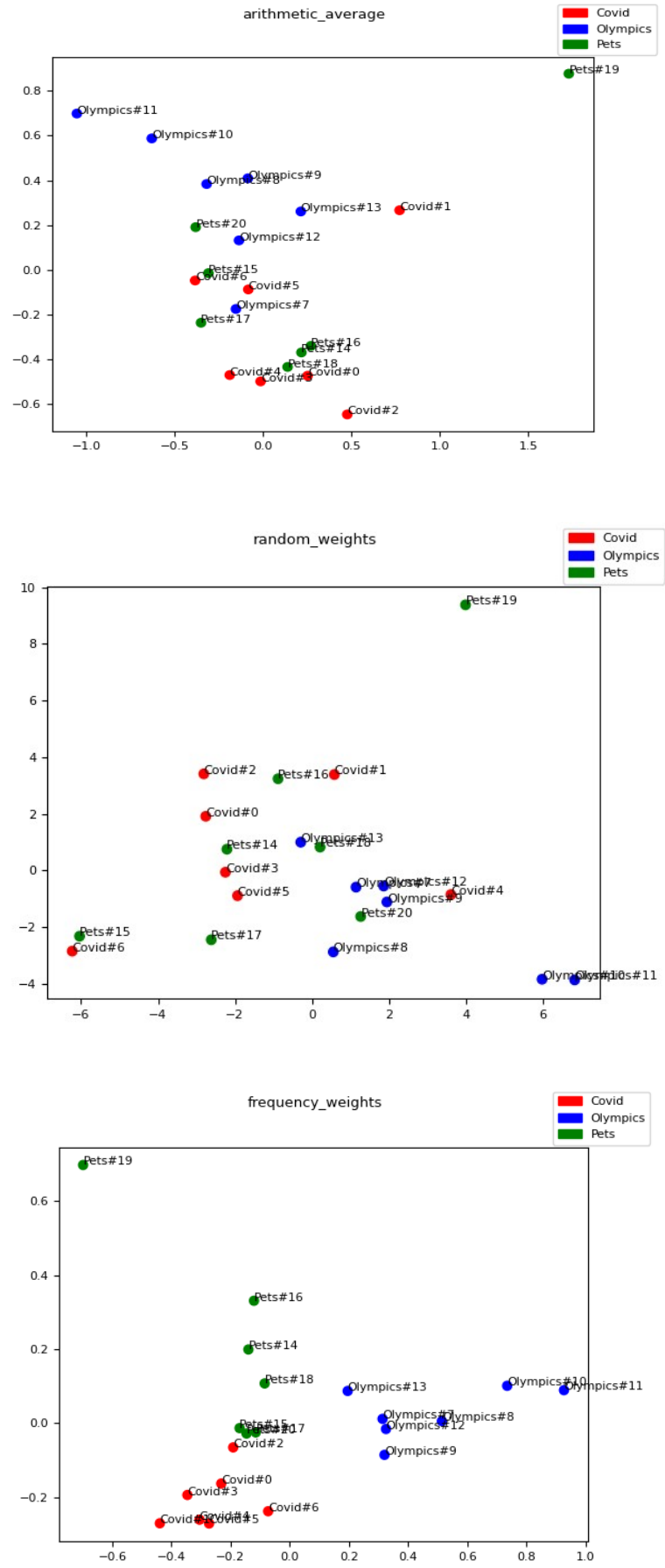
My house water (house clean), my pool warm (pool warm)

במשפט הזה אין הגיון תחבירי בכלל, אבל זה יכול להיות הגיוני ש-clean הוחלף ב-water, כי זה אנלוגית קשור.

והנה נגיד משפט נכון לגמרי:

Where you from ? (Where you at ?)

חלק 3 – מפת הציוצים



בניתי פונקציית משקל שמחזירה כמשקל את ה-similarity בין המילה לקטגוריה של הציוץ, כך בעצם מילים שקשורות לנושא של הציוץ יקבלו משקל גבוה יותר. במקרה של Covid, בגלל שאין מספיק אוצר מילים על הקטגוריה והמילה לא קיימת שם, החזרתי את ה-similarity עם המילה pandemic מכיוון שהיא קשורה ל-Covid.

1. אפשר לראות הפרדה ברורה בפונקציית המשקל שאני בניתי, בממוצע אריתמטי אין הפרדה ברורה, אבל אפשר לראות גושים של מחלקות, הם פשוט לא מופרדים – שוב, לא הפרדה ברורה.
2. הקטגוריות שחולקו הכי טוב הן Olympics ו-Pets, וזה כנראה בגלל שאין הרבה אוצר מילים שקשור ל-Covid. למרות שבפונקציה שאני בניתי, יש הפרדה טובה לכל המחלקות.
3. ניתן להסיק שהמסווג מאוד תלוי בתוכן, ולכן הפונקציה שבניתי מפרידה בצורה הכי טובה, וברורה.