דוח חלק א'

1. **פירוט השיטות והסבר על מימושן:**

Wordnet

במימוש שיטה זו ייבאנו מחבילת nltk את wordnet ואת המילון שלו.   
אנחנו רצים על המילים בשאילתא , לכל מילה חיפשנו את המילים שקרובות אליה מבחינת מילים נרדפות בלבד ולקחנו את כל המילים שמשקל הדימיון בינהן מהמילה שמצאנו למילה המקורית בשאילתא עולה על 0.6. למילים האלה נתנו משקל יותר נמוך מהמילים המקוריות בשאילתא לכן המילים בשאילתא קיבלו משקל 1 וכל שאר המילים שחזרו משיטה זו משקלן היה משקל הדימון מוכפל ב0.8 .

את המילים הללו הכנסנו למטריצת tfidf ועשינו cosSim בשביל למצוא את המסמכים הרלנווטיים ביותר.

thesaurus

במימוש שיטה זו ייבאנו מחבילת nltk את lin\_thesaurus ואת המילון שלו.   
אנחנו רצים על המילים בשאילתא , לכל מילה חיפשנו את המילים שקרובות אליה מבחינת מילים נרדפות בלבד ולקחנו את כל המילים שמשקל הדימיון בינהן מהמילה שמצאנו למילה המקורית בשאילתא עולה על 0.175. למילים האלה נתנו משקל יותר נמוך מהמילים המקוריות בשאילתא לכן המילים בשאילתא קיבלו משקל 1 וכל שאר המילים שחזרו משיטה זו משקלן היה 0.8.

את המילים הללו הכנסנו למטריצת tfidf ועשינו cosSim בשביל למצוא את המסמכים הרלנווטיים ביותר.

Spelling correction

במימוש שיטה זו ייבאנו מחבילת pyspellchecker את spellchecke ואת המילון שלו.   
אנחנו רצים על המילים בשאילתא , לכל מילה בדקנו האם היא לא ידועה , כלומר יש שם שגיאת כתיב כלשהי , במידה וכן בעזרת שיטה זו אנחנו מוצאים את המילה הקרובה ביותר למילה "השגויה" ומוסיפים אותה לשאילתא. את המילים הללו הכנסנו למטריצת tfidf ועשינו cosSim בשביל למצוא את המסמכים הרלנווטיים ביותר.

Advanced parser

במימוש שיטה זו הוספנו חוקים נוספים לפארסר שלנו :

* coronaNames = ["covid", "cov", "corona","coronavirus","covid19","cov19","coronavir","coronovirus"]

הוספנו את מילון זה בשביל להפוך את כל הצורות שהם קורונה יכולה להיות רשומה , לצורה אחת , דבר שעזר מאוד גם במדד הMAP.

* self.curse\_word=["shit","fuck","fucking","bull","bullshit","asshole","dickhead","dick","wtf","bastard","nigger","bitch","cunt","cock","pussy","motherfuker","motherfuck","dickhole","crap","nigaa","twat","slut","whore","dork","thot","simp","scumbag","pimp","boomer","alm","master","slave"]

הוספנו מילון זה גם בשביל החלק היצירתי , על מנת להפוך את המנוע שלנו למנוע ידידותי למשתמש החלטנו להוריד את רוב המילים הלא נאותות בשפה ובנוסף, גם החלטנו שהמילים הללו לא מוסיפות לנו מידע רלוונטי אודות הנושא שעליו עוסק המנוע.

* הורדנו אימוג'ים וכל דבר שלא נמצא בטבלת ההאסקי הבסיסית (כמו דברים שהם בbold)
* הורדנו כל מה שהוא לא אנגלית
* מצאנו תאריכים בעזרת regex
* הפכנו כמה נקודות לנק' אחת – "...." יהפוך ל "."
* הפרדנו מילים אשר בינהן יש נקודה ולא רווח
* הורדנו תווי רווח שהם נראים בפורמט של nbsm

LSI

יצרנו iterable corpus שזוהי מחלקה ששמה my Corpus על מנת שנוכל להסטרים את המידע (את כל הטוויטים) למודל lsi מבלי לשמור את כל הקרופוס על הזיכרון.   
אימנו מודל על 100,000 מילים השכיחות ביותר מהקורפוס שקיבלנו בחלק א'.

אימנו את המודל בעזרת זה שנתנו לו את הקורפוס, את המילון שלנו. שמרנו את המודל כקובץ כחלק מהpre-processing.

לאחר שיש לנו את המודל שמור, יצרנו מטריצת similarity העובדת עם cosine .   
לאחר מכן , קיבלנו שאילתא והפכנו אותה לוקטור וכך יכלנו לעשות את הדירוג עם המטריצת similarity.

נעזרנו בקוד של ניר על מנת לאמן את המודל ובחבילה של genism.

2.**הפארסר:**

יכולות שהורדנו מהפרסר :

הורדנו את שמירת התיוגים. הורדנו את המידע הזה בגלל שהבנו שהוא לא מוסיף לנו מידע , אין חיפוש שלהם בשאילתות וגם רוב השמות של המשתמשים זה או לא מילים אמיתיות או חיבור של כמה מילים ובגלל שאנחנו לא מפרסרים אותם המידע הזה לא מוסיף לנו.

הוספות :

בחרנו לממש שיטה נוספת שהיא advanced parser ולכן בשבילה הוספנו כמה חוקים:

* coronaNames = ["covid", "cov", "corona","coronavirus","covid19","cov19","coronavir","coronovirus"]- הוספנו את מילון זה בשביל להפוך את כל הצורות שהם קורונה יכולה להיות רשומה , לצורה אחת , דבר שעזר מאוד גם במדד הMAP.
* self.curse\_word=["shit","fuck","fucking","bull","bullshit","asshole","dickhead","dick","wtf","bastard","nigger","bitch","cunt","cock","pussy","motherfuker","motherfuck","dickhole","crap","nigaa","twat","slut","whore","dork","thot","simp","scumbag","pimp","boomer","alm","master","slave"]- הוספנו מילון זה גם בשביל החלק היצירתי , על מנת להפוך את המנוע שלנו למנוע ידידותי למשתמש החלטנו להוריד את רוב המילים הלא נאותות בשפה ובנוסף, גם החלטנו שהמילים הללו לא מוסיפות לנו מידע רלוונטי אודות הנושא שעליו עוסק המנוע.
* הורדנו אימוג'ים וכל דבר שלא נמצא בטבלת ההאסקי הבסיסית (כמו דברים שהם בbold)
* הורדנו כל מה שהוא לא אנגלית
* מצאנו תאריכים בעזרת regex
* הפכנו כמה נקודות לנק' אחת – "...." יהפוך ל "."
* הפרדנו מילים אשר בינהן יש נקודה ולא רווח
* הורדנו תווי רווח שהם נראים בפורמט של nbsm

3. לפי הפרקט שקיבלנו בחלק ג' :

סהכ המילים שהיו בפרקט : 12040 מילים

לאחר advanced parser : 8864 מילים

לאחר הורדת מילים שמופיעות רק פעם אחת : 4318 מילים

הורדנו רק מילים שחוזרות פעם אחת כי ראינו שזה עוזר לתוצאות מבחינת ממד הMAP , אך כשהורדנו יותר ראינו כי הMAP יורד ולכן החלטנו להישאר עם הורדה רק של מילים שחוזרות פעם אחת.

4. **האינדקסר:**

כל המידע המפורסר מגיע ישירות לאינדקסר.   
במילון שלנו המפתחות הם המונחים.

בנוסף , שמרנו את מספר טוויטים שבהם המונח הופיע, מספר הפעמים שהמונח הופיע בכל הקורפוס ומילון שממפה כל מסמך לכמות הפעמים שמונח זה חזר בו.

כמות הפעמים שהמונח חזר במסמך מסויים מנורמל אצלינו לפי המילה השכיחה ביותר במסמך זה.

המילון נראה באופן הבא:

Term -> N\_doc , N\_in\_corpus , {twitt\_id : freq1, twitt\_id2 : freq2 }

החלטנו לעשות את התהליך הזה בלי סטמינג כי אנחנו מישמנו wordnet ולכן סטמינג יפגע בביצועי שיטה זו ובנוסף סטמינג מכפיל את זמני הריצה ולכן בחרנו לעשות את זה ללא סטמינג.

5.

הפעולות שנקטנו על מנת להגיע לזמן ריצה מיטבי :

* הפכנו את כל המילים שקורונה יכולה להירשם בהם לקוביד , דבר שגם מצמם לנו את המילון
* הורדנו stop words נוספות שראינו שחוזרות הרבה ולא מוסיפות לנו מידע self.secondStop\_word = ['rt', 'i', 'p', 'etc', 'oh', 'im', 'also','amp','https','co','pm','am','pm et','et','wt','ps','ppp']

וקללות , דבר שממצמם לנו את המילון

* הורדנו מילים שחוזרות פעם אחת בלבד , דבר שממצמם לנו את המילון
* הורדנו את כל התיוגים , דבר שממצמם לנו את המילון
* בנינו tf-idf רק על המילים שחוזרות בשאילתא , דבר שמקטין באופן משמעותי את הגודל של המטריצה ואת החישובים של הדימיון.

6. הפעולות שנקטנו על מנת לשפר את איכות התוצאות:

* הפכנו את כל המילים שקורונה יכולה להירשם בהם לקוביד , דבר אשר שיפר את איכות התוצאות
* הורדנו stop words נוספות שראינו שחוזרות הרבה ולא מוסיפות לנו מידע self.secondStop\_word = ['rt', 'i', 'p', 'etc', 'oh', 'im', 'also','amp','https','co','pm','am','pm et','et','wt','ps','ppp']

וקללות , דבר אשר שיפר את איכות התוצאות

* הורדנו את כל התיוגים , דבר שממצמם לנו את המילון, דבר אשר שיפר את איכות התוצאות
* הורדנו מילים שחוזרות פעם אחת בלבד , דבר אשר שיפר את איכות התוצאות
* בנינו tf-idf רק על המילים שחוזרות בשאילתא , דבר אשר שיפר את איכות התוצאות , בניגוד למודל הlsi אשר בנינו בחלק א'
* שילבנו כמה שיטות על מנת לשפר את התוצאות השילוב שהכי תרם לשיפור הוא : wordnet יחד עם הspell correction והadvanced parser

7. בהתחלה , ניסינו את כל השיטות שבנינו באופן אינדיבידואלי. ממה שראינו היה :

-lsi – מוריד תוצאות

-wordnet- העלה תוצאות

-advanced parser – העלה תוצאות

-thesarus – העלה תוצאות אבל ברמה יותר פחותה מהwordnet

- spell correction – לא שינה את התוצאות

לפי מה שראינו מההרצה באופן אינדיבידואלי ניסינו לחבר את הlsi עם השיטות השונות , דבר אשר לא שיפר את תוצאותיו , בנוסף גם לקבל מסמכים אשר דורגו גבוהה לאחר tf-idf ולתת אותם למודל הlsi ועל זה להראות תוצאות סופיות גם לא שיפר את תוצאות המודל.

ולכן בסופו של דבר החלטנו לחבר יחד את 2 השיטות אשר הניבו את התוצאות הגבוהות ביותר :

Wordnet וadvanced parser

8. בעבודה נעזרנו רבות בספריית genism ובקוד שניר סיפק לנו.   
genism מאפשרת לנו לבנות מודל, לשמור אותו, להפוך את השאילתא לוקטור, לבנות מטריצת similarity.   
השתמשנו במילון הקיים אצלם.   
בנוסף , הקוד של ניר עזר לנו להבין איך להשתמש נכון בכל אחד מהדברים הללו ואיך להזרים קורפוס למודל.

<https://radimrehurek.com/gensim/models/lsimodel.html>

<https://radimrehurek.com/gensim/corpora/dictionary.html>

<https://radimrehurek.com/gensim/auto_examples/core/run_similarity_queries.html>

<https://radimrehurek.com/gensim/auto_examples/core/run_topics_and_transformations.html>

בנוסף השתמשנו בnltk בשביל הwordnet והthesaurus ובשביל הstop words.

<https://www.nltk.org/api/nltk.corpus.html>

<https://www.nltk.org/_modules/nltk/corpus/reader/lin.html>

בנוסף התמשנו בקוד פתוח של גיטהאב מהמימושים האפשריים לשיטות אשר סופקו לנו מסגל הקורס

<https://github.com/barrust/pyspellchecker>

9. מימשנו בסה"כ 5 שיטות ובבדיקות האוטומטיות נבדקות רק 2 לכן חשוב להוסיף לבדיקות את 3 השיטות הנוספות שמומשו.

יצירתיות –

ברוכים הבאים למנוע הנחמד!   
המנוע שלנו מוריד קללות   
בנוסף , בשביל הtf-idf השתמשנו בנוחסא משופרת (פירוט עליה ב3 סיבות למה המנוע שלנו יותר טוב משל אחרים)

ובנוסף , אנחנו מדפיסים fun facts להעשרה בדברים הקשורים למנוע שלנו.

הם נמצאים בsearch\_engine\_best בפוקנציה ששמה funFacts , לא הוספנו הדפסות על מנת שלא לפגוע בטסטים האוטומטיים, יכולים להציץ שם ולהפעיל אותה.

10.

Wordnet

שאילתא 1 :

Dr. Anthony Fauci wrote in a 2005 paper published in Virology Journal that hydroxychloroquine was effective in treating SARS.

1.

tweet id: 1291683169557721088

Dr Fauci wrote medical paper on Hydroxychloroquine was only medicine to prevent and treat from SARS and Corona Virus ask him why he down plays it Now

הסבר: טוויט זה מתחייס לנאמר בשאילתא שד"ר אנטוני כתב מאמר על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

2.

tweet id: 1287163289084936192

Fauci knew this in 2005 when he wrote a paper saying Chloroquine (even milder than Hydroxychloroquine) is practically a vaccine against ALL Coronaviruses &amp; SARS.

Relatives of those who died from CORONA should be able to SUE HIM! He killed them! <https://t.co/WanX8YP9Vq>

הסבר: טוויט זה מתחייס לנאמר בשאילתא פוצי כתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

3.

tweet id: 1285169749333356544

Published online 2005 Aug 22. #Virology journal publishes study showcasing the positive effects of #chloroquine aganist #SARS type #coronavirus #FauciTheFraud knew #Hydroxychloroquine works #Scamdemic #COVID19 #depopulation #Corruption

הסבר: טוויט זה מתחייס לנאמר בשאילתא שנככתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס בנוסף יש ב# את שמו של פוצי ואת המילה Hydroxychloroquine ולכן זו תוצאה בדירוג גבוה.

4.

tweet id: 1291133082976899072

RT @USAMomUtah: @CindyProUSA Yes. Dr Fauci published a paper in 2005 praising the use of hydrochloroquine in the originals SARS COVID outbr…

הסבר: טוויט זה מתחייס לנאמר בשאילתא שד"ר פוצי כתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

5.

tweet id: 1285591112283897856

@Nationals .@NCBI published a paper in 2005 confirming the effectiveness of chloroquine against SARS #coronavirus, but Fauci The Fraud has been gaslighting us by claiming HCQ is not effective against #COVID19! Screw him AND you! #BoycottMLB <https://t.co/myZoDbx4Xe>

הסבר: טוויט זה מתחייס לנאמר בשאילתא שנככתב מאמר ב2005 על איך chloroquine הועיל בטיפול בסארס ושפוצי אמר דברים נוספים ולכן זו תוצאה בדירוג גבוה.

שאילתא 2:

The seasonal flu kills more people every year in the U.S. than COVID-19 has to date

1.

tweet id: 1282348840264388609

@brycetache Every year influenza kills people as well. I think those just focusing on Coronavirus are morons, sorry.

הסבר: הטוויט מדבר על כך ששפעת הורגת אנשים גם ואין צורך להתייחס יותר מידי לקורונה כפי שהשאילתא אומרת ולכן הוא מדורג גבוה. ניתן לראות גם כי הwordnet החזיר מילה נרדפת לשפעת.

2.

tweet id: 1284128251309260801

@mamalionpack Ah the flu, you know that one virus that does actually kill a percentage of children every year. I bet any flu deaths will be called covid deaths and we will get a big media chorus of “see covid kills kids too” and shutter the schools again.

הסבר: הטוויט מדבר על כך ששפעת הורגת אנשים גם ושיגידו על מוות משפעת שהוא נגרם מקורונה , מה שאומר ששפעת יותר חמורה מהקורונה כפי שהשאילתא אומרת ולכן הוא מדורג גבוה.

3.

tweet id: 1285703407353892865

Why if the death rate from Covid-19 is less than Influenza, do we need a mandatory vaccine if we don't require one for the flu?

הסבר: הטוויט מדבר על כך ששפעת הורגת יותר אנשים מהקורונה כפי שהשאילתא ולכן הוא מדורג גבוה.

4.

tweet id: 1286664336140795905

@dtb479353 @Junedocherty2 @GeeMinx @STVNews Covid doesn’t kill less than the flu. That’s just a straight up lie.

הסבר: הטוויט מדבר על כך שקורונה הורגת יותר אנשים משפעת ,השאילתא אומרת משהו הפוך אך מילות המפתח נורא דומות ולכן הוא מדורג גבוה.

5.

tweet id: 1291802650980343809

RT @snopes: ❌ No, the seasonal flu does not kill more people every year in the U.S. than COVID-19 has to date. <https://t.co/eiyBjfZsTM>

הסבר: הטוויט מדבר על כך שקורונה הורגת יותר אנשים משפעת ,השאילתא אומרת משהו הפוך אך מילות המפתח נורא דומות ולכן הוא מדורג גבוה.

שאילתא 4:

The coronavirus pandemic is a cover for a plan to implant trackable microchips and that the Microsoft co-founder Bill Gates is behind it

1.

tweet id: 1287384513950162944

@SHurrell1 It's on Natural News but Twitter wont allow the link to article: 'Enzyme that will make Bill Gates’ vaccine microchip implant work known as luciferase.'

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

2.

tweet id: 1286580387062276097

Bill Gates debunks 'Coronavirus vaccine is my 5G mind control microchip implant' conspiracy theory <https://t.co/7y3trI75Tj>

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

3.

tweet id: 1286590278791569408

RT @TheRegister: Bill Gates debunks 'Coronavirus vaccine is my 5G mind control microchip implant' conspiracy theory <https://t.co/KuOwWfVS3t>

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

4.

tweet id: 1286391004946472971

RT @Lioness\_1970: Bill Gates shot down a conspiracy theory that he wants a global coronavirus vaccine rollout so he can implant microchips…

הסבר: ניתן לראות כי הטוויט מדבר על כך שביל גייטס סותר את הקונספירציה ועל כך שמתכננים לשים ציפים בחיסון הפוך ממה שמדובר בשאילתא אבל המילות מפתח הן אותו הדבר ולכן זה חזר בדירוג גבוה

5.

tweet id: 1286063330126430209

RT @CBSNews: .@NorahODonnell: “To be clear, do you want a vaccine so that you can implant microchips into people?”

Bill Gates: “No…I don't…

הסבר: ניתן לראות כי הטוויט מדבר על כך שביל גייטס סותר את הקונספירציה ועל כך שמתכננים לשים ציפים בחיסון הפוך ממה שמדובר בשאילתא אבל המילות מפתח הן אותו הדבר ולכן זה חזר בדירוג גבוה

שאילתא 7:

Herd immunity has been reached.

1.

tweet id: 1284101206382649345

Oxford epidemiologists declare herd immunity has been reached in the UK.

Only 20% of the population was ever susceptible to Covid.

#FridayFeeling <https://t.co/ALKcFiJb6x>

הסבר: ניתן לראות כי השאילתא חזרה בשלמותה בטוויט ולכן הוא מדורג גבוה.

2.

tweet id: 1291360894123945984

@NYGovCuomo @WRicciardi Pandemic is over in NY because you have reached herd immunity.

הסבר: ניתן לראות כי הטוויט מדבר על זה שהשגנו חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

3.

tweet id: 1284044533009993731

Oxford Study say's we've reached herd immunity in the UK <https://t.co/fePhBFOisC>

הסבר: ניתן לראות כי הטוויט מדבר על זה שהשגנו חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

4.

tweet id: 1291042926793908224

@yashar They probably reached herd immunity ....

הסבר: ניתן לראות כי הטוויט מדבר על זה שהשגנו חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

5.

tweet id: 1287373820186304512

@athome\_waiting They would have reached herd immunity quicker... Its the only way to get thru this

הסבר: ניתן לראות כי הטוויט מדבר על זה שיכלו להשיג חיסון העדר מהר יותר והמילות מפתח מופיעות בשאילתא ולכן זה חזר בדירוג גבוה.

שאילתא 8:

Children are “almost immune from this disease.”

1.

tweet id: 1291094104244457474

Hello @realDonaldTrump 👋🏻

I sent a kid to the ER last week for suspected multi system inflammatory syndrome post COVID-19 infection.

Please stop telling the country that “children are almost immune” because that is misinformation

הסבר : הטוויט מדבר על כך שילדים לא חסינים מפני המחלה אך מופיעות הרבה מילות מפתח מהשאילתא ולכן מדורג גבוהה

2.

tweet id: 1291153948980764672

Facebook has removed a post on Donald Trump's personal page after he claimed that children are "almost immune" from coronavirus <https://t.co/FGAuMJ6MHf>

הסבר : הטוויט מדבר על כך שטראמפ פרסם שילדים חסינים מפני המחלה וזה בידיוק מה שהשאילתא מדברת עליו ולכן מדורג גבוהה

3.

tweet id: 1291371404726882304

Facebook and Twitter remove video of Trump falsely claiming children are "almost immune" to the coronavirus https://t.co/OVOvi4uwDb via @CBSNews

הסבר : הטוויט מדבר על כך שטראמפ פרסם שילדים חסינים מפני המחלה וזה בידיוק מה שהשאילתא מדברת עליו ולכן מדורג גבוהה

4.

tweet id: 1291660938404343808

Children are not 'almost immune' to the Coronavirus: Everything we know on how the virus infects them <https://t.co/ZHO38JpCza>

הסבר : הטוויט מדבר על כך שילדים לא חסינים מפני המחלה אך מופיעות הרבה מילות מפתח מהשאילתא ולכן מדורג גבוהה

5.

tweet id: 1291280042480283648

It is a lie to say children are immune to CoronaVirus. Last week, a 13 year old was rushed to the hospital with the disease and died within one hour

הסבר : הטוויט מדבר על כך שילדים לא חסינים מפני המחלה אך מופיעות הרבה מילות מפתח מהשאילתא ולכן מדורג גבוהה

LSI

שאילתא 1 :

Dr. Anthony Fauci wrote in a 2005 paper published in Virology Journal that hydroxychloroquine was effective in treating SARS.

1.

tweet id: 1285849606711123968

Iceland's tiny museums – and the joy of niche collections #iceland #icelandicart #art #artists #galleries #artgallery #nichegalleries #Exhibitions #WashYourHands #WearAMask #WearAMaskSaveALife #WearAMaskPlease #WearADamnMask #Masks4All <https://t.co/Ndy6ZigQzf>

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

2.

tweet id: 1282550461682393088

@GregAbbott\_TX 👇🏻👇🏻👇🏻👇🏻Why is it a problem if people get a virus that they recover from? Happens with the flu every year!! Stop this madness!

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

3.

tweet id: 1288840520005873665

RT @CueGirl2020: Article from the Virology Journal (August 22, 2005). Dr Fauci makes bold statement by saying Hydroxychloroquine functions…

הסבר : הטוויט מדבר על מאמר מ2005 של פוצי על הפוקנציונאליות של Hydroxychloroquine כפי שהשאילתא מדבר ולכן זה מדורג גבוה.

4.

tweet id: 1285926549967876096

RT @ilanasorokin: 😱 Scandalous disinformation campaign: In 2005 the NIH under the direction of Dr Fauci 🤥 published “Chloroquine is a poten…

הסבר : הטוויט מדבר על מאמר מ2005 של פוצי על הפוקנציונאליות של Hydroxychloroquine כפי שהשאילתא מדבר ולכן זה מדורג גבוה.

5.

tweet id: 1281212739482988544

RT @RichHiggins\_DC: Fauci wasn’t “wrong.”

Fauci was “lying.”

Fauci knew in 2005 that HCQ worked on Coronaviruses.

Fauci will now be…

הסבר : הטוויט מדבר על פוצי ועל שנת 2005 כמו בשאילתא ולכן זה מדורג גבוה.

שאילתא 2:

The seasonal flu kills more people every year in the U.S. than COVID-19 has to date

1.

tweet id: 1286690976413581312

@charliekirk11 Social justice warriors reached herd immunity months ago

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

2.

tweet id: 1281703043977928704

Fucking wear your masks, people!

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

3.

tweet id: 1282099289922187264

Wear your fucking masks now people

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

4.

tweet id: 1284030877454888961

Wear your fucking masks people!

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

5.

tweet id: 1280970034987270144

"People should NOT wear masks when exercising as masks may reduce the ability to breathe comfortably."

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

שאילתא 4:

The coronavirus pandemic is a cover for a plan to implant trackable microchips and that the Microsoft co-founder Bill Gates is behind it

1.

tweet id: 1283907143779774464

@kgopinion Yes, it helps and is the best way for us to contain it. If we all wear masks we'll get a handle on the virus. It's been proven.

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

2.

tweet id: 1285952321420230659

There's a coin shortage. Restaurants, grocery stores, etc., have posted signs requesting people pay with credit cards/exact change. The U.S. Mint decreased staffing b/c of the pandemic. As the nation's businesses have reopened, demand for coins has exceeded the available supply. <https://t.co/l31nQGGFtT>

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

3.

tweet id: 1281198624576360450

According to the latest COVID Symptom Study data, rates of new COVID cases have stopped declining with over 23,000 suspected cases in the UK.

Ask yourself, do I REALLY need that haircut or pint? 🤔

<https://t.co/ctaPFLNmjt>

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

4.

tweet id: 1284518485440290820

Here's one that's probably true: Coronavirus did not originate from a bat to a pig to a human. It was created in a lab as part of "gain of function" research and either escaped accidentally or intentionally.

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

5.

tweet id: 1286314377273016320

RT @JonGar00990153: @lucyallan Research - Bill Gates ID2020 Vaccine, Bill Gates LUCIFERASE, Bill Gates Microsoft Patent 060606 digital curr…

הסבר : ניתן לראות כי הטטויט מדבר על ביל גייטס ועל החיסון כפי שהשאילתא מדברת ולכן חזר בדירוג גבוה

שאילתא 7:

Herd immunity has been reached.

1.

tweet id: 1286584569324539904

Did you wear a mask before leaving home today?

Wearing face masks appropriately can help prevent the spread of #COVID19, so the more you wear your mask, the better

Remember, face shields cannot protect you from… <https://t.co/ZOMj4HNnpy>

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

2.

tweet id: 1281605225489379329

@Monideepa62 We have reached herd immunity to injustice.

הסבר: ניתן לראות כי הטוויט מדבר על זה שהשגנו חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

3.

tweet id: 1282424367444455424

@NateSilver538 Herd immunity - achieved!

הסבר: ניתן לראות כי הטוויט מדבר על זה שהשגנו חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

4.

tweet id: 1287133767560298496

@dallasnews Approaching herd immunity.

הסבר: ניתן לראות כי הטוויט מדבר על זה שאנחנו סדרך לחיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה

5.

tweet id: 1290499319632547847

@GodFamilyJesus Nope, herd immunity

הסבר: ניתן לראות כי הטוויט מדבר על חיסון העדר וגם השאילתא מדברת על זה ולכן זה חזר בדירוג גבוה.

שאילתא 8:

Children are “almost immune from this disease.”

1.

tweet id: 1287157292908122112

@tedcruz But it’s ok cause many are wearing masks (albeit like chinstraps)...

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

2.

tweet id: 1285776250880000001

5G &amp; COVID-19: Myths and Facts <https://t.co/iAQJxeRYI2>

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

3.

tweet id: 1285676020964700167

@RachelNotley Looks like the spell has broken !

Healthy people don't wear masks

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

4.

tweet id: 1286672417050501120

@NunyaBiznez101 @hodlonaut You breath out CO2. The masks are protecting the climate as the CO2 stays in your body if you wear a mask. It’s a climate hoax!

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

5.

tweet id: 1285556567262273536

Every thread by @ASlavitt is worth your time

הסבר : אין לנו הסבר למה דווקא התוצאה הזו חזרה. בגלל שאנחנו משתמשים במודל הLSI יכול להיות שבמטריצה הקונספטים שלו מאימון המודל הוא החליט שהשאילתא וטוויט זה נמצאים תחת אותו קונספט

Best = wordnet + spell correction + advanced parser

שאילתא 1 :

Dr. Anthony Fauci wrote in a 2005 paper published in Virology Journal that hydroxychloroquine was effective in treating SARS.

1.

tweet id: 1291683169557721088

Dr Fauci wrote medical paper on Hydroxychloroquine was only medicine to prevent and treat from SARS and Corona Virus ask him why he down plays it Now

הסבר: טוויט זה מתחייס לנאמר בשאילתא שד"ר אנטוני כתב מאמר על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

2.

tweet id: 1287163289084936192

Fauci knew this in 2005 when he wrote a paper saying Chloroquine (even milder than Hydroxychloroquine) is practically a vaccine against ALL Coronaviruses &amp; SARS.

Relatives of those who died from CORONA should be able to SUE HIM! He killed them! <https://t.co/WanX8YP9Vq>

הסבר: טוויט זה מתחייס לנאמר בשאילתא פוצי כתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

3.

tweet id: 1285169749333356544

Published online 2005 Aug 22. #Virology journal publishes study showcasing the positive effects of #chloroquine aganist #SARS type #coronavirus #FauciTheFraud knew #Hydroxychloroquine works #Scamdemic #COVID19 #depopulation #Corruption

הסבר: טוויט זה מתחייס לנאמר בשאילתא שנככתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס בנוסף יש ב# את שמו של פוצי ואת המילה Hydroxychloroquine ולכן זו תוצאה בדירוג גבוה.

4.

tweet id: 1285591112283897856

@Nationals .@NCBI published a paper in 2005 confirming the effectiveness of chloroquine against SARS #coronavirus, but Fauci The Fraud has been gaslighting us by claiming HCQ is not effective against #COVID19! Screw him AND you! #BoycottMLB <https://t.co/myZoDbx4Xe>

הסבר: טוויט זה מתחייס לנאמר בשאילתא שנככתב מאמר ב2005 על איך chloroquine הועיל בטיפול בסארס ושפוצי אמר דברים נוספים ולכן זו תוצאה בדירוג גבוה.

5.

tweet id: 1291133082976899072

RT @USAMomUtah: @CindyProUSA Yes. Dr Fauci published a paper in 2005 praising the use of hydrochloroquine in the originals SARS COVID outbr…

הסבר: טוויט זה מתחייס לנאמר בשאילתא שד"ר פוצי כתב מאמר ב2005 על איך Hydroxychloroquine הועיל בטיפול בסארס ולכן זו תוצאה בדירוג גבוה.

שאילתא 2:

The seasonal flu kills more people every year in the U.S. than COVID-19 has to date

1.

tweet id: 1282348840264388609

@brycetache Every year influenza kills people as well. I think those just focusing on Coronavirus are morons, sorry.

הסבר: הטוויט מדבר על כך ששפעת הורגת אנשים גם ואין צורך להתייחס יותר מידי לקורונה כפי שהשאילתא אומרת ולכן הוא מדורג גבוה. ניתן לראות גם כי הwordnet החזיר מילה נרדפת לשפעת.

2.

tweet id: 1284128251309260801

@mamalionpack Ah the flu, you know that one virus that does actually kill a percentage of children every year. I bet any flu deaths will be called covid deaths and we will get a big media chorus of “see covid kills kids too” and shutter the schools again.

הסבר: הטוויט מדבר על כך ששפעת הורגת אנשים גם ושיגידו על מוות משפעת שהוא נגרם מקורונה , מה שאומר ששפעת יותר חמורה מהקורונה כפי שהשאילתא אומרת ולכן הוא מדורג גבוה.

3.

tweet id: 1285703407353892865

Why if the death rate from Covid-19 is less than Influenza, do we need a mandatory vaccine if we don't require one for the flu?

הסבר: הטוויט מדבר על כך ששפעת הורגת יותר אנשים מהקורונה כפי שהשאילתא ולכן הוא מדורג גבוה.

4.

tweet id: 1282097456788975616

@nolagoblue @mark\_ca @RealSoulbuilder @dreams\_in\_green @NBCNews Yet the science fact is that the coronavirus is considerably much more contagious (and potentially deadlier for some) than the various influenza viruses. Approximately 8,200 to 20,000 people die from the flu each year . . .

הסבר: הטוויט מדבר על כך שקורונה הורגת יותר אנשים משפעת ויותר מדביקה ,השאילתא מדברת על זה גם ולכן הוא מדורג גבוה.

5.

tweet id: 1291786635097657344

RT @snopes: ❌ No, the seasonal flu does not kill more people every year in the U.S. than COVID-19 has to date. <https://t.co/eiyBjfZsTM>

הסבר: הטוויט מדבר על כך שקורונה הורגת יותר אנשים משפעת ,השאילתא אומרת משהו הפוך אך מילות המפתח נורא דומות ולכן הוא מדורג גבוה.

שאילתא 4:

The coronavirus pandemic is a cover for a plan to implant trackable microchips and that the Microsoft co-founder Bill Gates is behind it

1.

tweet id: 1286590278791569408

RT @TheRegister: Bill Gates debunks 'Coronavirus vaccine is my 5G mind control microchip implant' conspiracy theory <https://t.co/KuOwWfVS3t>

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

2.

tweet id: 1287384513950162944

@SHurrell1 It's on Natural News but Twitter wont allow the link to article: 'Enzyme that will make Bill Gates’ vaccine microchip implant work known as luciferase.'

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

3.

tweet id: 1286580387062276097

Bill Gates debunks 'Coronavirus vaccine is my 5G mind control microchip implant' conspiracy theory <https://t.co/7y3trI75Tj>

הסבר: ניתן לראות כי הטוויט מדבר על ביל גייטס ועל כך שמתכננים לשים ציפים בחיסון כפי שמדובר בשאילתא ולכן זה חזר בדירוג גבוה

4.

tweet id: 1286022954204237829

RT @CBSNews: .@NorahODonnell: “To be clear, do you want a vaccine so that you can implant microchips into people?”

Bill Gates: “No…I don't…

הסבר: ניתן לראות כי הטוויט מדבר על כך שביל גייטס סותר את הקונספירציה ועל כך שמתכננים לשים ציפים בחיסון הפוך ממה שמדובר בשאילתא אבל המילות מפתח הן אותו הדבר ולכן זה חזר בדירוג גבוה

5.

tweet id: 1286034337947742210

RT @CBSNews: .@NorahODonnell: “To be clear, do you want a vaccine so that you can implant microchips into people?”

Bill Gates: “No…I don't…

הסבר: ניתן לראות כי הטוויט מדבר על כך שביל גייטס סותר את הקונספירציה ועל כך שמתכננים לשים ציפים בחיסון הפוך ממה שמדובר בשאילתא אבל המילות מפתח הן אותו הדבר ולכן זה חזר בדירוג גבוה

שאילתא 7:

Herd immunity has been reached.

1.

tweet id: 1291042926793908224

@yashar They probably reached herd immunity ....

הסבר : הטוויט מדבר על זה שהגיעו לחיסון המונים וגם השאילתא ולכן זה מדורג גבוה.

2.

tweet id: 1291231983234945024

@PnPCBC @CPHO\_Canada I doubt it. It looks to me we have already reached herd immunity in much of Canada.

הסבר : הטוויט מדבר על זה שהגיעו לחיסון המונים וגם השאילתא ולכן זה מדורג גבוה.

3.

tweet id: 1285647976086605824

RT @boriquagato: so, daily deaths and ICU have dropped to basically zero in sweden.

it's over. herd immunity was reached and worked. how…

הסבר : הטוויט מדבר על זה שהגיעו לחיסון המונים וגם השאילתא ולכן זה מדורג גבוה.

4.

tweet id: 1283782121471893510

@FrankFshagan @ty\_fried Where do you get herd immunity reached at 20%?

הסבר : הטוויט שואל על מה נחשב להגעה לחיסון המונים ובשאילתא מופיעות המלות מפתח הללו ולכן זה מדורג גבוה.

5.

tweet id: 1284024029209407488

Britain may have reached herd immunity, according to Oxford University scientists- @Telegraph https://t.co/IkjCzP7cRh

הסבר : הטוויט מדבר על זה שהגיעו לחיסון המונים וגם השאילתא ולכן זה מדורג גבוה.

שאילתא 8:

Children are “almost immune from this disease.”

1.

tweet id: 1291094104244457474

Hello @realDonaldTrump 👋🏻

I sent a kid to the ER last week for suspected multi system inflammatory syndrome post COVID-19 infection.

Please stop telling the country that “children are almost immune” because that is misinformation

הסבר : הטוויט מדבר על כך שילדים לא חסינים מפני המחלה אך מופיעות הרבה מילות מפתח מהשאילתא ולכן מדורג גבוהה

2.

tweet id: 1291358157344825345

RT @DrEricDing: WOW. Trump tweeted a video clip that children are “almost immune” from covid-19, violating Twitter’s rules against coronavi…

הסבר : הטוויט מדבר על כך שנאמר שילדים חסינים מפני המחלה וגם השאילתא מדברת על זה ולכן מדורג גבוהה

3.

tweet id: 1291144910482153472

RT @CBSEveningNews: Pres. Trump said children are “almost immune” from COVID-19, in unsubstantiated claims about how the virus impacts kids…

הסבר : הטוויט מדבר על כך שנאמר שילדים חסינים מפני המחלה וגם השאילתא מדברת על זה ולכן מדורג גבוהה

4.

tweet id: 1291185086948429831

RT @DrEricDing: WOW. Trump tweeted a video clip that children are “almost immune” from covid-19, violating Twitter’s rules against coronavi…

הסבר : הטוויט מדבר על כך שנאמר שילדים חסינים מפני המחלה וגם השאילתא מדברת על זה ולכן מדורג גבוהה

5.

tweet id: 1291214264342999040

Facebook and Twitter punish Trump over video where he says children 'almost immune' to Covid-19 <https://t.co/cq1s4DwdqM>

הסבר : הטוויט מדבר על כך שנאמר שילדים חסינים מפני המחלה וגם השאילתא מדברת על זה ולכן מדורג גבוהה

11. עבור כל השאילתות בכל שיטה :

Wordnet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [precision@5](mailto:precision@5) | [precision@10](mailto:precision@10) | [precision@50](mailto:precision@50) | precision | recall |
| 1 | 1 | 1 | 0.86 | 0.717277 | 0.717277 |
| 2 | 0.6 | 0.8 | 0.82 | 0.816425 | 0.816425 |
| 3 | 0.8 | 0.8 | 0.94 | 0.871698 | 0.987179 |
| 4 | 1 | 1 | 0.94 | 0.738462 | 0.86747 |
| 5 | 0.2 | 0.1 | 0.42 | 0.367742 | 0.6 |
| 6 | 0.2 | 0.4 | 0.48 | 0.442529 | 0.631148 |
| 7 | 0.8 | 0.8 | 0.94 | 0.845018 | 0.991342 |
| 8 | 1 | 1 | 1 | 0.985401 | 1 |
| 9 | 0.6 | 0.5 | 0.58 | 0.799087 | 0.817757 |
| 10 | 0.8 | 0.9 | 0.84 | 0.883534 | 0.952381 |
| 11 | 1 | 1 | 1 | 0.795367 | 0.980952 |
| 12 | 1 | 0.8 | 0.2 | 0.095238 | 0.96 |
| 13 | 0.8 | 0.9 | 0.56 | 0.32967 | 0.714286 |
| 14 | 0.6 | 0.7 | 0.88 | 0.907895 | 0.8625 |
| 15 | 0.8 | 0.7 | 0.72 | 0.798319 | 0.931373 |
| 16 | 1 | 1 | 0.92 | 0.8 | 0.946237 |
| 17 | 0.8 | 0.9 | 0.92 | 0.675393 | 0.914894 |
| 18 | 0.8 | 0.8 | 0.72 | 0.746094 | 0.979487 |
| 19 | 1 | 1 | 0.8 | 0.561905 | 0.855072 |
| 20 | 0.8 | 0.7 | 0.48 | 0.230769 | 0.818182 |
| 21 | 1 | 1 | 0.84 | 0.43346 | 1 |
| 22 | 0.6 | 0.7 | 0.32 | 0.276471 | 0.52809 |
| 23 | 0.8 | 0.8 | 0.82 | 0.709877 | 0.737179 |
| 24 | 0.6 | 0.7 | 0.64 | 0.587209 | 0.90991 |
| 25 | 1 | 0.7 | 0.38 | 0.222222 | 1 |
| 26 | 0.8 | 0.9 | 0.9 | 0.792553 | 0.886905 |
| 27 | 1 | 1 | 0.98 | 0.730469 | 0.994681 |
| 28 | 1 | 1 | 0.8 | 0.633721 | 0.813433 |
| 29 | 0.8 | 0.9 | 0.9 | 0.57971 | 0.779221 |
| 30 | 0.8 | 0.9 | 0.82 | 0.521277 | 0.942308 |
| 31 | 0.8 | 0.7 | 0.8 | 0.754386 | 0.886598 |
| 32 | 0.4 | 0.4 | 0.3 | 0.317568 | 0.723077 |
| 33 | 1 | 1 | 0.98 | 0.955645 | 0.955645 |
| 34 | 1 | 1 | 0.78 | 0.836449 | 0.94709 |
| 35 | 0.2 | 0.4 | 0.5 | 0.489712 | 0.88806 |

LSI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [precision@5](mailto:precision@5) | [precision@10](mailto:precision@10) | [precision@50](mailto:precision@50) | precision | recall |
| 1 | 0.8 | 0.6 | 0.74 | 0.691176 | 0.717277 |
| 2 | 1 | 0.9 | 0.76 | 0.758242 | 0.816425 |
| 3 | 0.6 | 0.7 | 0.86 | 0.852941 | 0.987179 |
| 4 | 0.8 | 0.6 | 0.78 | 0.611111 | 0.86747 |
| 5 | 0.2 | 0.4 | 0.44 | 0.346715 | 0.6 |
| 6 | 0.6 | 0.6 | 0.38 | 0.44 | 0.631148 |
| 7 | 1 | 1 | 0.9 | 0.844444 | 0.991342 |
| 8 | 1 | 0.9 | 0.96 | 0.985401 | 1 |
| 9 | 0.8 | 0.9 | 0.82 | 0.775362 | 0.817757 |
| 10 | 0.6 | 0.7 | 0.86 | 0.86 | 0.952381 |
| 11 | 0.8 | 0.7 | 0.94 | 0.778243 | 0.980952 |
| 12 | 0 | 0.2 | 0.12 | 0.09375 | 0.96 |
| 13 | 1 | 0.7 | 0.52 | 0.165354 | 0.714286 |
| 14 | 1 | 1 | 1 | 0.910638 | 0.8625 |
| 15 | 0 | 0.2 | 0.64 | 0.396078 | 0.931373 |
| 16 | 0.8 | 0.7 | 0.5 | 0.726141 | 0.946237 |
| 17 | 0.4 | 0.3 | 0.54 | 0.54023 | 0.914894 |
| 18 | 1 | 1 | 0.9 | 0.746154 | 0.979487 |
| 19 | 0.2 | 0.2 | 0.24 | 0.485981 | 0.855072 |
| 20 | 0.2 | 0.1 | 0.16 | 0.270936 | 0.818182 |
| 21 | 0.4 | 0.4 | 0.6 | 0.477707 | 1 |
| 22 | 0.2 | 0.2 | 0.08 | 0.334586 | 0.52809 |
| 23 | 0 | 0.4 | 0.68 | 0.606557 | 0.737179 |
| 24 | 0.8 | 0.8 | 0.68 | 0.579787 | 0.90991 |
| 25 | 0 | 0.2 | 0.36 | 0.301075 | 1 |
| 26 | 0.8 | 0.6 | 0.66 | 0.634538 | 0.886905 |
| 27 | 0.8 | 0.6 | 0.74 | 0.694836 | 0.994681 |
| 28 | 0.4 | 0.3 | 0.34 | 0.421348 | 0.813433 |
| 29 | 0.6 | 0.5 | 0.64 | 0.568528 | 0.779221 |
| 30 | 0.8 | 0.9 | 0.74 | 0.395349 | 0.942308 |
| 31 | 1 | 0.7 | 0.78 | 0.477833 | 0.886598 |
| 32 | 0.6 | 0.4 | 0.28 | 0.248092 | 0.723077 |
| 33 | 1 | 1 | 0.84 | 0.946565 | 0.955645 |
| 34 | 1 | 0.9 | 0.62 | 0.737052 | 0.94709 |
| 35 | 1 | 0.9 | 0.56 | 0.509653 | 0.88806 |

Spelling correction

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | precision@5 | precision@10 | precision@50 | precision | recall |
| 1 | 0.8 | 0.7 | 0.84 | 0.822222 | 0.581152 |
| 2 | 0.6 | 0.8 | 0.86 | 0.81407 | 0.782609 |
| 3 | 0.8 | 0.9 | 0.96 | 0.877863 | 0.982906 |
| 4 | 0.8 | 0.9 | 0.94 | 0.744898 | 0.879518 |
| 5 | 0.2 | 0.1 | 0.3 | 0.375758 | 0.652632 |
| 6 | 0.2 | 0.3 | 0.48 | 0.44186 | 0.622951 |
| 7 | 1 | 1 | 0.96 | 0.845018 | 0.991342 |
| 8 | 1 | 1 | 0.94 | 0.988462 | 0.951852 |
| 9 | 0.6 | 0.6 | 0.66 | 0.78972 | 0.78972 |
| 10 | 1 | 0.8 | 0.891304 | 0.891304 | 0.177489 |
| 11 | 1 | 1 | 0.98 | 0.796154 | 0.985714 |
| 12 | 1 | 0.8 | 0.24 | 0.1 | 0.88 |
| 13 | 0.8 | 0.9 | 0.604167 | 0.604167 | 0.690476 |
| 14 | 0.8 | 0.8 | 0.92 | 0.926829 | 0.316667 |
| 15 | 0.8 | 0.8 | 0.72 | 0.798319 | 0.931373 |
| 16 | 1 | 1 | 0.92 | 0.8 | 0.946237 |
| 17 | 1 | 1 | 0.92 | 0.678947 | 0.914894 |
| 18 | 0.8 | 0.9 | 0.7 | 0.746094 | 0.979487 |
| 19 | 1 | 0.7 | 0.8 | 0.561905 | 0.855072 |
| 20 | 0.8 | 0.8 | 0.46 | 0.252174 | 0.878788 |
| 21 | 1 | 1 | 0.7 | 0.479769 | 0.72807 |
| 22 | 0.6 | 0.6 | 0.32 | 0.274286 | 0.539326 |
| 23 | 1 | 0.8 | 0.8 | 0.682081 | 0.75641 |
| 24 | 1 | 0.7 | 0.54 | 0.627586 | 0.81982 |
| 25 | 0.8 | 0.6 | 0.4 | 0.222222 | 1 |
| 26 | 1 | 1 | 0.88 | 0.798883 | 0.85119 |
| 27 | 1 | 1 | 0.96 | 0.729412 | 0.989362 |
| 28 | 1 | 1 | 0.78 | 0.633721 | 0.813433 |
| 29 | 0.6 | 0.8 | 0.88 | 0.788136 | 0.603896 |
| 30 | 1 | 1 | 0.98 | 0.507692 | 0.951923 |
| 31 | 0.8 | 0.8 | 0.8 | 0.754386 | 0.886598 |
| 32 | 0.4 | 0.3 | 0.3 | 0.317568 | 0.723077 |
| 33 | 1 | 1 | 0.98 | 0.955645 | 0.955645 |
| 34 | 0.8 | 0.9 | 0.74 | 0.840206 | 0.862434 |
| 35 | 0.4 | 0.5 | 0.42 | 0.489712 | 0.88806 |

Thesaurus

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | precision@5 | precision@10 | precision@50 | precision | recall |
| 1 | 0.8 | 0.8 | 0.8 | 0.717391 | 0.691099 |
| 2 | 0.8 | 0.9 | 0.84 | 0.827225 | 0.763285 |
| 3 | 1 | 0.9 | 0.94 | 0.885375 | 0.957265 |
| 4 | 1 | 0.9 | 0.94 | 0.73545 | 0.837349 |
| 5 | 0 | 0.2 | 0.24 | 0.371257 | 0.652632 |
| 6 | 0.8 | 0.6 | 0.52 | 0.442529 | 0.631148 |
| 7 | 1 | 0.9 | 0.96 | 0.845018 | 0.991342 |
| 8 | 1 | 1 | 1 | 0.985401 | 1 |
| 9 | 0.6 | 0.8 | 0.72 | 0.80543 | 0.831776 |
| 10 | 1 | 1 | 0.9 | 0.884 | 0.95671 |
| 11 | 1 | 1 | 0.96 | 0.833333 | 0.97619 |
| 12 | 1 | 0.8 | 0.24 | 0.108247 | 0.84 |
| 13 | 0.8 | 0.9 | 0.630435 | 0.630435 | 0.690476 |
| 14 | 0.6 | 0.7 | 0.88 | 0.916667 | 0.870833 |
| 15 | 0.8 | 0.7 | 0.68 | 0.798319 | 0.931373 |
| 16 | 0.8 | 0.9 | 0.94 | 0.809302 | 0.935484 |
| 17 | 0.6 | 0.7 | 0.9 | 0.75 | 0.851064 |
| 18 | 0.8 | 0.8 | 0.72 | 0.746094 | 0.979487 |
| 19 | 0.6 | 0.7 | 0.76 | 0.390411 | 0.826087 |
| 20 | 0.4 | 0.5 | 0.52 | 0.46988 | 0.590909 |
| 21 | 0.4 | 0.7 | 0.8 | 0.428571 | 0.973684 |
| 22 | 0.6 | 0.3 | 0.3 | 0.274286 | 0.539326 |
| 23 | 0.6 | 0.7 | 0.82 | 0.645161 | 0.769231 |
| 24 | 0.6 | 0.7 | 0.64 | 0.615385 | 0.936937 |
| 25 | 0.8 | 0.7 | 0.42 | 0.233696 | 0.767857 |
| 26 | 1 | 1 | 0.9 | 0.829932 | 0.72619 |
| 27 | 0.8 | 0.9 | 0.92 | 0.730769 | 0.909574 |
| 28 | 1 | 1 | 0.8 | 0.637427 | 0.813433 |
| 29 | 0.8 | 0.9 | 0.88 | 0.57767 | 0.772727 |
| 30 | 1 | 0.9 | 0.98 | 0.505102 | 0.951923 |
| 31 | 0.8 | 0.8 | 0.8 | 0.754386 | 0.886598 |
| 32 | 0.6 | 0.4 | 0.36 | 0.315068 | 0.707692 |
| 33 | 1 | 1 | 0.98 | 0.951613 | 0.951613 |
| 34 | 0.8 | 0.8 | 0.72 | 0.835052 | 0.857143 |
| 35 | 1 | 0.7 | 0.54 | 0.478992 | 0.850746 |

Advanced parser

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [precision@5](mailto:precision@5) | [precision@10](mailto:precision@10) | [precision@50](mailto:precision@50) | precision | recall |
| 1 | 0.8 | 0.9 | 0.78 | 0.72043 | 0.701571 |
| 2 | 0.6 | 0.8 | 0.88 | 0.811881 | 0.792271 |
| 3 | 1 | 1 | 0.86 | 0.878327 | 0.987179 |
| 4 | 1 | 0.9 | 0.94 | 0.703518 | 0.843373 |
| 5 | 0.2 | 0.1 | 0.34 | 0.375758 | 0.652632 |
| 6 | 0.4 | 0.6 | 0.34 | 0.435897 | 0.696721 |
| 7 | 0.8 | 0.9 | 0.96 | 0.845018 | 0.991342 |
| 8 | 1 | 1 | 1 | 0.985401 | 1 |
| 9 | 0.6 | 0.8 | 0.66 | 0.778802 | 0.78972 |
| 10 | 1 | 1 | 0.96 | 0.882114 | 0.939394 |
| 11 | 1 | 1 | 0.98 | 0.796154 | 0.985714 |
| 12 | 1 | 0.8 | 0.32 | 0.127072 | 0.92 |
| 13 | 0.8 | 0.9 | 0.659091 | 0.659091 | 0.690476 |
| 14 | 0.8 | 0.8 | 0.84 | 0.913934 | 0.929167 |
| 15 | 0.8 | 0.9 | 0.7 | 0.798319 | 0.931373 |
| 16 | 1 | 1 | 0.96 | 0.79638 | 0.946237 |
| 17 | 0.8 | 0.9 | 0.92 | 0.678947 | 0.914894 |
| 18 | 0.8 | 0.9 | 0.76 | 0.746094 | 0.979487 |
| 19 | 1 | 0.8 | 0.82 | 0.567308 | 0.855072 |
| 20 | 0.8 | 0.8 | 0.48 | 0.219828 | 0.772727 |
| 21 | 0.4 | 0.5 | 0.82 | 0.431298 | 0.991228 |
| 22 | 0.8 | 0.4 | 0.28 | 0.294872 | 0.516854 |
| 23 | 0.8 | 0.7 | 0.74 | 0.6875 | 0.775641 |
| 24 | 0.8 | 0.8 | 0.64 | 0.575 | 0.828829 |
| 25 | 1 | 0.7 | 0.5 | 0.222222 | 1 |
| 26 | 1 | 1 | 0.88 | 0.805714 | 0.839286 |
| 27 | 1 | 1 | 0.94 | 0.729412 | 0.989362 |
| 28 | 1 | 1 | 0.98 | 0.632184 | 0.820896 |
| 29 | 1 | 1 | 0.92 | 0.57971 | 0.779221 |
| 30 | 1 | 0.9 | 0.98 | 0.507692 | 0.951923 |
| 31 | 0.8 | 0.8 | 0.84 | 0.756522 | 0.896907 |
| 32 | 0.4 | 0.2 | 0.32 | 0.316901 | 0.692308 |
| 33 | 1 | 1 | 0.98 | 0.950673 | 0.854839 |
| 34 | 1 | 0.8 | 0.68 | 0.850515 | 0.873016 |
| 35 | 0.4 | 0.4 | 0.36 | 0.487603 | 0.880597 |

Best = advanced parser + spelling correction + wordnet

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | [precision@5](mailto:precision@5) | [precision@10](mailto:precision@10) | [precision@50](mailto:precision@50) | precision | recall |
| 1 | 1 | 1 | 0.86 | 0.705882 | 0.691099 |
| 2 | 0.8 | 0.9 | 0.84 | 0.815166 | 0.830918 |
| 3 | 1 | 1 | 0.86 | 0.878327 | 0.987179 |
| 4 | 1 | 1 | 0.94 | 0.716418 | 0.86747 |
| 5 | 0.2 | 0.1 | 0.34 | 0.380645 | 0.621053 |
| 6 | 0.4 | 0.4 | 0.32 | 0.436548 | 0.704918 |
| 7 | 0.8 | 0.9 | 0.94 | 0.845018 | 0.991342 |
| 8 | 1 | 1 | 1 | 0.985401 | 1 |
| 9 | 0.4 | 0.6 | 0.6 | 0.8 | 0.82243 |
| 10 | 0.8 | 0.9 | 0.96 | 0.885714 | 0.939394 |
| 11 | 1 | 1 | 0.96 | 0.796154 | 0.985714 |
| 12 | 1 | 0.8 | 0.26 | 0.097276 | 1 |
| 13 | 0.8 | 0.9 | 0.58 | 0.32967 | 0.714286 |
| 14 | 0.8 | 0.8 | 0.88 | 0.913223 | 0.920833 |
| 15 | 0.8 | 0.7 | 0.74 | 0.798319 | 0.931373 |
| 16 | 1 | 1 | 0.96 | 0.79638 | 0.946237 |
| 17 | 0.8 | 0.7 | 0.92 | 0.675393 | 0.914894 |
| 18 | 0.8 | 0.9 | 0.84 | 0.746094 | 0.979487 |
| 19 | 1 | 1 | 0.8 | 0.567308 | 0.855072 |
| 20 | 0.8 | 0.8 | 0.46 | 0.243478 | 0.848485 |
| 21 | 0.4 | 0.6 | 0.82 | 0.431298 | 0.991228 |
| 22 | 1 | 0.6 | 0.38 | 0.317568 | 0.52809 |
| 23 | 0.8 | 0.8 | 0.76 | 0.685393 | 0.782051 |
| 24 | 0.6 | 0.6 | 0.64 | 0.575 | 0.828829 |
| 25 | 1 | 0.7 | 0.5 | 0.222222 | 1 |
| 26 | 0.8 | 0.9 | 0.92 | 0.792553 | 0.886905 |
| 27 | 1 | 1 | 0.98 | 0.730469 | 0.994681 |
| 28 | 1 | 1 | 0.98 | 0.632184 | 0.820896 |
| 29 | 1 | 1 | 0.92 | 0.581731 | 0.785714 |
| 30 | 0.8 | 0.9 | 0.82 | 0.521277 | 0.942308 |
| 31 | 1 | 0.9 | 0.82 | 0.763158 | 0.896907 |
| 32 | 0.2 | 0.3 | 0.42 | 0.319444 | 0.707692 |
| 33 | 1 | 1 | 1 | 0.950673 | 0.854839 |
| 34 | 1 | 1 | 0.72 | 0.836449 | 0.94709 |
| 35 | 0.8 | 0.6 | 0.56 | 0.489712 | 0.88806 |

עבור כל השאילתות בכל שיטה:

Wordnet

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.637 | 0.711 | 0.095 | 0.985 |
| [percision@5](mailto:percision@5) | 0.834 | 0.800 | 0.200 | 1.000 |
| [percision@10](mailto:percision@10) | 0.811 | 0.900 | 0.100 | 1.000 |
| [percision@50](mailto:percision@50) | 0.743 | 0.820 | 0.200 | 1.000 |
| recall | 0.868 | 0.897 | 0.539 | 1.000 |
| MAP | 0.755 |  |  |  |

Lsi

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.577 | 0.580 | 0.094 | 0.985 |
| [percision@5](mailto:percision@5) | 0.634 | 0.800 | 0.000 | 1.000 |
| [percision@10](mailto:percision@10) | 0.606 | 0.600 | 0.100 | 1.000 |
| [percision@50](mailto:percision@50) | 0.619 | 0.660 | 0.080 | 1.000 |
| recall | 0.898 | 0.984 | 0.186 | 1.000 |
| MAP | 0.622 |  |  |  |

Spelling correction

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.653 | 0.732 | 0.097 | 0.988 |
| [percision@5](mailto:percision@5) | 0.834 | 0.800 | 0.200 | 1.000 |
| [percision@10](mailto:percision@10) | 0.826 | 0.900 | 0.100 | 1.000 |
| [percision@50](mailto:percision@50) | 0.737 | 0.800 | 0.300 | 0.980 |
| recall | 0.801 | 0.840 | 0.177 | 1.000 |
| MAP | 0.755 |  |  |  |

Thesaurus

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.647 | 0.721 | 0.099 | 0.985 |
| [percision@5](mailto:percision@5) | 0.777 | 0.800 | 0.000 | 1.000 |
| [percision@10](mailto:percision@10) | 0.786 | 0.800 | 0.200 | 1.000 |
| [percision@50](mailto:percision@50) | 0.734 | 0.800 | 0.240 | 1.000 |
| recall | 0.833 | 0.851 | 0.539 | 1.000 |
| MAP | 0.751 |  |  |  |

Advanced parser

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.644 | 0.708 | 0.101 | 0.985 |
| [percision@5](mailto:percision@5) | 0.840 | 1.000 | 0.200 | 1.000 |
| [percision@10](mailto:percision@10) | 0.823 | 0.900 | 0.100 | 1.000 |
| [percision@50](mailto:percision@50) | 0.741 | 0.820 | 0.260 | 1.000 |
| recall | 0.856 | 0.855 | 0.528 | 1.000 |
| MAP | 0.759 |  |  |  |

Best = advanced parser + spelling correction + wordnet

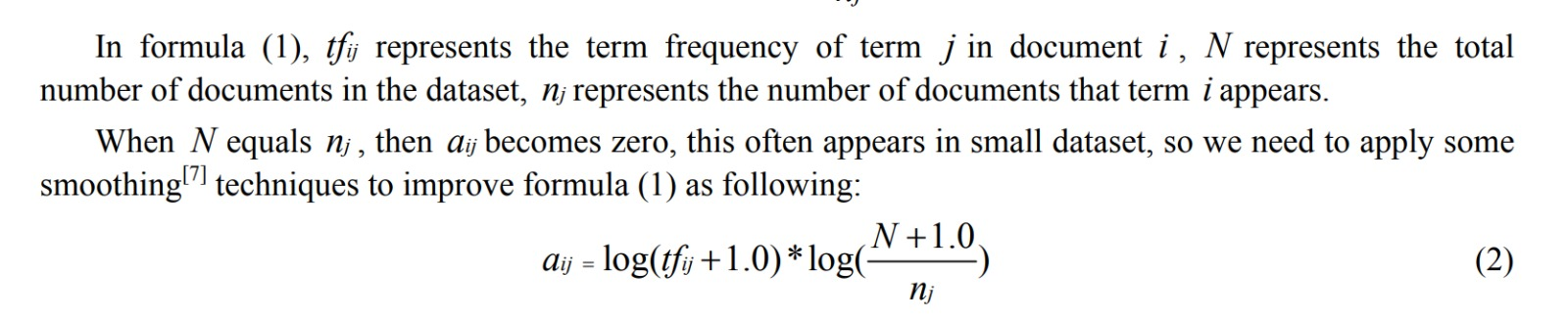
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | average | median | min | max |
| precision | 0.635 | 0.706 | 0.097 | 0.985 |
| [percision@5](mailto:percision@5) | 0.823 | 1.000 | 0.000 | 1.000 |
| [percision@10](mailto:percision@10) | 0.834 | 0.900 | 0.100 | 1.000 |
| [percision@50](mailto:percision@50) | 0.751 | 0.820 | 0.280 | 1.000 |
| recall | 0.864 | 0.887 | 0.483 | 1.000 |
| MAP | 0.764 |  |  |  |

השוואות כל השיטות לפי המדדים השונים :

12.

המנוע שלנו טוב יותר ב:   
1. הוא משתמש ב3 שיטות יחד לא כמו רוב המנועים האחרים , מה שגורם להחזרת תוצאות יותר מדוייקות.

1. השתמשנו בנוסחאת tf-idf משופרת שמצאנו באינטרנט



3. יש לנו pre processing של כל היישויות והאותיות גדולות קטנות , דבר שמקל עליינו מלעשות את זה באינדקסר.

המנוע שלנו פחות טוב משל אחרים ב:

1. כי לא השתמשנו בstemming דבר אחר יחזיר תוצאות פחות טובות
2. בגלל שאנחנו מתמשים בwordnet ובspell correction שתי שיטות שמרחיבות את השאילתא בזמן ריצה , זה יכול לפגוע מעט בזמני הריצה
3. במנוע שלנו אין חשיבות למיקום של מילה