**Assignment 1: Distributional Similarity**

**Generating lists of the most 20 similar words according to word2vec**

**foccacia\_bread:**

[('roasted\_cherry\_tomatoes', 0.7796332836151123), ('brandy\_cream', 0.7609724402427673), ('grilled\_flatbread', 0.7598520517349243), ('Grilled\_shrimp', 0.759242832660675), ('toasted\_brioche', 0.7520718574523926), ('shredded\_chicken', 0.749045729637146), ('mozarella', 0.7487182021141052), ('wasabi\_aioli', 0.7476052641868591), ('creamy\_horseradish', 0.7466524839401245), ('blue\_cheese\_crumbles', 0.7451153993606567), ('creamy\_mash', 0.7447883486747742), ('marinated\_lamb', 0.7441205382347107), ('spinach\_ricotta', 0.7440380454063416), ('pomodoro\_sauce', 0.7439367175102234), ('white\_bean\_puree', 0.7438353300094604), ('crumbled\_goat\_cheese', 0.743340015411377), ('marscapone', 0.7433167695999146), ('foccacia', 0.742948055267334), ('sautéed\_zucchini', 0.7428866624832153), ('melted\_cheddar\_cheese', 0.7425709366798401)]

**year:**

[('month', 0.7653314471244812), ('week', 0.6598176956176758), ('months', 0.5790778398513794), ('decade', 0.577674388885498), ('years', 0.5731882452964783), ('summer', 0.5514534115791321), ('year.The', 0.5041267275810242), ('September', 0.5011922717094421), ('weeks', 0.4975568652153015), ('season', 0.4971977770328522), ('June', 0.4965071678161621), ('yaer', 0.4945131838321686), ('weekend', 0.49340492486953735), ('July', 0.49157896637916565), ('January', 0.49015676975250244), ('twoyears', 0.48917046189308167), ('August', 0.4877557158470154), ('threeyear', 0.4866842031478882), ('October', 0.48396244645118713), ('December', 0.4830012023448944)]

**citzen:**

[('citizen', 0.509380042552948), ('NO\_RIGHTS', 0.47435349225997925), ('AMERICAN\_CITIZEN', 0.4661392867565155), ('citizens', 0.4654591679573059), ('AMERICAN\_CITIZENS', 0.46374186873435974), ('citzens', 0.46252426505088806), ('law\_abiding\_taxpaying', 0.4615214467048645), ('MORALLY', 0.4562292695045471), ('ABSOLUTELY\_NOT', 0.4540644586086273), ('singaporean', 0.4532110095024109), ('taxpaying\_citizen', 0.4486428499221802), ('THANK\_YOU\_THANK\_YOU', 0.4462531507015228), ('UNIDENTIFIED\_WOMAN', 0.4456477165222168), ('CINDY\_SHEEHAN', 0.44440552592277527), ('ENOUGH\_IS\_ENOUGH', 0.4437023103237152), ('south\_africans', 0.4415375590324402), ('peop', 0.441243976354599), ('people.I', 0.43930524587631226), ('sincerly', 0.4391438066959381), ('OBAMA\_BIDEN', 0.4373474419116974)]

**first:**

[('second', 0.7971885800361633), ('third', 0.6932076215744019), ('fourth', 0.6732367277145386), ('fifth', 0.6571477651596069), ('sixth', 0.6237859129905701), ('seventh', 0.591548502445221), ('thefirst', 0.5863957405090332), ('last', 0.5842218995094299), ('eighth', 0.5558103919029236), ('final', 0.5550165772438049), ('fi\_rst', 0.5481294989585876), ('ninth', 0.5459919571876526), ('eleventh', 0.5380799770355225), ('##th', 0.5116368532180786), ('next', 0.4894919693470001), ('thirteenth', 0.4882149398326874), ('thesecond', 0.4813888669013977), ('##rd', 0.47677913308143616), ('only', 0.4762601852416992), ('fourteenth', 0.4741714298725128)]

**in:**

[('inthe', 0.5891957879066467), ('where', 0.5662435293197632), ('the', 0.5429296493530273), ('In', 0.5415117144584656), ('during', 0.5188906192779541), ('iin', 0.48737412691116333), ('at', 0.484235554933548), ('from', 0.48268404603004456), ('outside', 0.47092658281326294), ('for', 0.4566476047039032), ('since', 0.4543766379356384), ('near', 0.44844427704811096), ('last', 0.4461378753185272), ('####in', 0.432082861661911), ('throughout', 0.43173980712890625), ('after', 0.42844122648239136), ('іn', 0.42422354221343994), ('following', 0.4220481812953949), ('around', 0.41807663440704346), ('into', 0.41720473766326904)]

**Polysemous Words:**

1. Three polysemous words belong in the first group and their neighbors:

**Mouse meanings:** an animal, computer device.

**Top 10 neighbors:**

1. 'Logitech\_MX\_Revolution', 0.6175230741500854
2. 'Razer\_Mamba', 0.5994570851325989
3. 'mice', 0.5896884799003601
4. 'cordless\_laser', 0.5652030110359192
5. 'VX\_Nano', 0.5619357824325562
6. 'Logitech\_MX###', 0.5604779124259949
7. 'keyboard\_arrow\_keys', 0.5545550584793091
8. 'Logitech\_G9x', 0.5538491606712341
9. 'NOTE\_TO\_READERS\_Hovering', 0.5520266890525818
10. 'Razer\_Abyssus', 0.5489388108253479

**Head** meanings: can mean the upper part of the human body, the top or front of something, a person in charge, or the froth on top of a glass of beer.

**Top 10 neighbors:**

1. 'head', 0.596509575843811
2. 'Director', 0.5869857668876648
3. 'Managing\_Director', 0.5186715126037598
4. 'Vice\_President', 0.5178024172782898
5. 'General\_Manager', 0.5169602036476135
6. 'Associate\_Director', 0.496967077255249
7. 'Senior\_Vice', 0.4953590929508209
8. 'Manager', 0.4950293004512787
9. 'Shoulders\_shampoo\_Pringles', 0.4934467375278473
10. 'VP', 0.48192524909973145

**Bat** meanings: Flying Mammal, Sports Equipment, Swinging Motion.

**Top 10 neighbors:**

1. 'bats', 0.767751932144165
2. 'batting', 0.6346984505653381
3. 'Pinch\_hitter\_Brayan\_Pena', 0.6011940836906433
4. 'batsman', 0.5579798817634583
5. 'batted', 0.5542199611663818
6. 'Hawaiian\_hoary', 0.5447419881820679
7. 'Lelands.com\_auctioned', 0.5397745966911316
8. 'yelled\_Cheater', 0.538004457950592
9. 'wicketkeeper\_Andrew\_Hodd', 0.5371009111404419
10. 'lefthanded\_batter', 0.53566575050354
11. Three polysemous words belong in the second group:

**Bank** meanings: **financial institution**, a person conducting a gambling house or game, a set in a row (as in a "bank of lights"), one of the horizontal and usually secondary or lower divisions of a headline.

**Financial institutions are reflected in the top 10 neighbors.**

**Top 10 neighbors:**

1. 'banks', 0.7440759539604187),
2. 'banking', 0.690161406993866),
3. 'Bank', 0.6698698401451111)
4. 'lender', 0.6342284679412842
5. 'banker', 0.6092953085899353
6. 'depositors', 0.6031531691551208
7. 'mortgage\_lender', 0.5797975659370422
8. 'depositor', 0.5716427564620972
9. 'BofA', 0.5714625120162964
10. 'Citibank', 0.5589520335197449

**Gallery**: a roofed promenade, an outdoor balcony, a long and narrow passage, apartment, or corridor, a room or building devoted to the exhibition of works of art, a small ornamental barrier or railing.

**A room or building devoted to the exhibition of works of art reflected in the top 10 neighbors.**

**Top 10 neighbors of the meaning :**

1. 'galleries', 0.8196952939033508
2. 'Gallery', 0.6979522705078125
3. 'art\_gallery', 0.6867443919181824
4. 'Galleries', 0.6668725609779358
5. 'Art\_Gallery', 0.6286720633506775
6. 'Greene\_Naftali', 0.5900658965110779
7. 'Whitechapel\_Art\_Gallery', 0.5814566612243652
8. 'Sydney\_Roslyn\_Oxley9', 0.5809972286224365
9. 'David\_Kesting', 0.573269248008728
10. 'Lisson\_Gallery', 0.5683467388153076

**Book** meanings: to park and **a park**.

**A park reflected in the top 10 neighbors.**

**Top 10 neighbors:**

1. 'parks', 0.7697824835777283
2. 'Park', 0.613426685333252
3. 'superintendent\_Dave\_Uberuaga', 0.5954588055610657
4. 'skate\_park', 0.5911567211151123
5. 'parkland', 0.5799823999404907
6. 'parkland', 0.5719486474990845
7. 'campground', 0.5696098804473877
8. 'Taraji\_Henson\_knocked', 0.5658068060874939
9. 'Castaway\_Cove', 0.5577318668365479
10. 'skateboard\_park', 0.5547470450401306
11. The second group of neighbors reflect only one sense because the other meanings are rarer and less used. The neighbors are affected from the context of the word. I can assume the text the model has been trained on contains more context of those words meaning. May be the model ignored those meaning because they are rare or maybe they aren’t in the top 10 neighbors.

**Synonyms and Antonyms**

1. Triple:

word1 = "happy"

word2 = "joyful"

word3 = "sad"

model.similarity(word1, word2) = 0.42381963

model.similarity(word1, word3) = 0.5354614

1. Word2vec represents words as vectors, and similarity function determines the similarity between words based on their context. In the text the model was trained on, I assume that the context of happy is closer (more similar) to the context of sad than the context of joyful and happy.

**The Effect of Different Corpora**

5 words whose top 10 neighbors based on the news corpus are very similar to their top 10 neighbors based on the twitter corpus

|  |  |  |
| --- | --- | --- |
| word | glove-wiki-gigaword-200 | glove-twitter-200 |
| Weather | [('conditions', 0.65), ('inclement', 0.6471),  ('rainy', 0.6410),  ('temperatures', 0.6410),  ('rain', 0.6296),  ('storms', 0.6172),  ('winter', 0.5967), ('meteorologists', 0.5926),  ('winds', 0.5916),  ('cold', 0.5842)] | [('rain', 0.7422),  ('snow', 0.7304),  ('cold', 0.7230),  ('freezing', 0.7103),  ('winter', 0.7007),  ('rainy', 0.6836),  ('outside', 0.6656),  ('forecast', 0.6596),  ('storm', 0.6582),  ('conditions', 0.6533)] |
| Apple | [('samsung', 0.7578), ('microsoft', 0.7387),  ('iphone', 0.7183),  ('google', 0.6872),  ('blackberry', 0.6836),  ('galaxy', 0.6636),  ('nexus', 0.6629),  ('ipad', 0.6608),  ('nokia', 0.6555),  ('ios', 0.6468)] | [('iphone', 0.6271),  ('microsoft', 0.6061),  ('intel', 0.5992),  ('macintosh', 0.5989),  ('ipod', 0.5908),  ('ibm', 0.5888),  ('ipad', 0.5876),  ('software', 0.5711),  ('google', 0.5554),  ('itunes', 0.5478)] |
| nutrition | [('health', 0.7356),  ('nutritional', 0.6997), ('wellness', 0.6812),  ('fitness', 0.6731), ('supplements', 0.6500), ('foods', 0.6480),  ('protein', 0.6295),  ('healthy', 0.6166),  ('diet', 0.6121),  ('supplement', 0.6070)] | [('nutritional', 0.7267),  ('diet', 0.6209),  ('health', 0.6071),  ('food', 0.6017),  ('wellness', 0.5752),  ('hygiene', 0.5729),  ('dietetics', 0.5704),  ('care', 0.5685),  ('prevention', 0.5402),  ('foods', 0.5295)] |
| football | [('soccer', 0.8105),  ('basketball', 0.7896),  ('league', 0.7167),  ('baseball', 0.7023),  ('rugby', 0.7006),  ('hockey', 0.6886),  ('club', 0.6874),  ('team', 0.6729),  ('teams', 0.6543),  ('players', 0.6527)] | [('soccer', 0.8488),  ('basketball', 0.7915),  ('sports', 0.7718),  ('players', 0.7587),  ('baseball', 0.7571),  ('league', 0.7383),  ('rugby', 0.7382),  ('game', 0.7134),  ('player', 0.6989),  ('softball', 0.6956)] |
| car | [('cars', 0.8199),  ('vehicle', 0.8115),  ('driver', 0.7699),  ('truck', 0.7657),  ('driving', 0.7070),  ('vehicles', 0.6832), ('motorcycle', 0.6503),  ('parked', 0.6456),  ('drivers', 0.6364),  ('bus', 0.6304)] | [('truck', 0.7657),  ('cars', 0.7584),  ('driving', 0.7157),  ('drive', 0.6923),  ('front', 0.6751),  ('vehicle', 0.6674),  ('bus', 0.6626),  ('bike', 0.6559),  ('cause', 0.6432),  ('house', 0.6365)] |

5 words who’s top 10 neighbors based on the news corpus which are substantially different from the top 10 neighbors based on the twitter corpus:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Word | glove-wiki-gigaword-200 | Meaning | Glove-twitter-200 | Meaning |
| vector | [('vectors', 0.7915065884590149), ('formula\_3', 0.7009355425834656), ('formula\_1', 0.6953726410865784), ('formula\_2', 0.6908257007598877), ('formula\_5', 0.684027373790741), ('formula\_4', 0.6806007623672485), ('formula\_6', 0.6665896773338318), ('tangent', 0.6512946486473083), ('formula\_7', 0.6493140459060669), ('formula\_11', 0.6435732841491699)] | vector formulas | [('illustration', 0.6054693460464478), ('psd', 0.5514928698539734), ('template', 0.5432111024856567), ('backgrounds', 0.5422154664993286), ('graphics', 0.5403188467025757), ('graphic', 0.5324743390083313), ('templates', 0.5264845490455627), ('abstract', 0.5254871845245361), ('illustrator', 0.5142496824264526), ('illustrations', 0.5118792653083801)] | Math aspect |
| head | [('heads', 0.7668997645378113), ('headed', 0.6344295144081116), (**'chief'**, 0.6314131617546082), ('body', 0.6098024249076843), (**'assistant'**, 0.6064105033874512), (**'director'**, 0.6037707328796387), (**'deputy'**, 0.5836146473884583), ('hand', 0.5738338232040405), ('left', 0.5574275255203247), ('arm', 0.5565925240516663)] | the upper part of the human body and a person in charge | [('heads', 0.7717670202255249), ('up', 0.7592162489891052), ('right', 0.7493096590042114), ('down', 0.722529947757721), ('my', 0.7108836770057678), ('out', 0.7091900110244751), ('that', 0.7060001492500305), ('his', 0.6992073655128479), ('him', 0.6986932158470154), ('going', 0.6972156763076782)]  [ | Direction or the upper part of the human body |
| paper | [('newspaper', 0.671421229839325), ('papers', 0.6713257431983948), ('printed', 0.6686532497406006), ('sheet', 0.6124283671379089), ('printing', 0.6033082604408264), ('newspapers', 0.5930173397064209), ('print', 0.5892402529716492), ('piece', 0.5870198607444763), ('published', 0.581505298614502), ('book', 0.5597691535949707)] | A newspaper or an academic article | [('papers', 0.7432311177253723), ('essay', 0.7063133120536804), ('exam', 0.6881474256515503), ('assignment', 0.6701089143753052), ('homework', 0.6578236222267151), ('study', 0.6473684310913086), ('class', 0.6331662535667419), ('book', 0.6290728449821472), ('math', 0.6230848431587219), ('presentation', 0.6212742328643799)] | Studies |
| rose | [('fell', 0.8498217463493347), ('climbed', 0.7398364543914795), ('surged', 0.7233148813247681), ('jumped', 0.7221606969833374), ('shares', 0.7005136013031006), ('percent', 0.6983330249786377), ('dropped', 0.6880760788917542), ('slipped', 0.6827592253684998), ('soared', 0.6758185029029846), ('dipped', 0.6733461618423462)] | Adventure | [('derrick', 0.6237108707427979), ('flower', 0.5979903340339661), ('blue', 0.5857720375061035), ('green', 0.5798736810684204), ('roses', 0.562809407711029), ('diamond', 0.561917245388031), ('violet', 0.5611094236373901), ('white', 0.5601827502250671), ('pink', 0.5569117069244385), ('brown', 0.5545764565467834)] | Colors, flower, diamond |
| function | [('functions', 0.8763212561607361), ('i.e.', 0.6624270081520081), ('functional', 0.6439272165298462), ('defined', 0.6280918717384338), ('hence', 0.6072871088981628), ('parameters', 0.5982860922813416), ('integral', 0.5973845720291138), ('e.g.', 0.5824131965637207), ('normal', 0.5801949501037598), ('thus', 0.5795422792434692)] | Math | [('functions', 0.5973495244979858), ('study', 0.5633230209350586), ('properly', 0.5596754550933838), ('cannot', 0.5530847311019897), ('functioning', 0.5342937111854553), ('type', 0.5328474640846252), ('handle', 0.5282319784164429), ('form', 0.5273945927619934), ('system', 0.5240757465362549), ('work', 0.5239363312721252)] | Functioning |

My strategy was to find words with duplicate meaning where one meaning is used more in wiki and the other meaning is used more in twitter posts.

**A green and blue dots

Description automatically generatedPlotting words in 2D**

The green circles are words ending with “ing” and the blue circle are words ending with “ed”. There are 414 blue circles and 294 green circles (120 words difference). Words ending with "ing" are predominantly found on the negative x-axis and words ending with "ed" are on the positive x-axis, it suggests that there is a pattern or structure in the data that separates these 2 groups along the first principal component. Possible explanation could include:

1. Uses: The distribution of words in the dataset might result in a separation along the first principal component if words ending with "ing" are **more frequent in certain contexts** and words ending with "ed" in others, PCA might capture this distinction.
2. Semantic Patterns: There might be certain semantic patterns in words that end with “ing” and “ed” that are captured by the PCA. For example, words ending with “ing” may represent ongoing actions or processes, while words ending with “ed” may represent completed actions and that’s why the context of each word will be different, so the words split to 2 groups.

**Related words**

|  |  |  |
| --- | --- | --- |
|  | Chat GPT | Word2vec-google-news-300 |
| foccacia bread | Ciabatta  Baguette  Sourdough  Pita  Naan  Flatbread  Challah  Rye bread  Whole grain bread  Multigrain bread  Artisan bread  Olive bread  Garlic bread  Herb-infused bread  Crusty bread  Italian bread  French bread  Panini  Bruschetta  Rustic bread | roasted\_cherry\_tomatoes  brandy\_cream  grilled\_flatbread  Grilled\_shrimp  toasted\_brioche  shredded\_chicken  mozarella  wasabi\_aioli  creamy\_horseradish  blue\_cheese\_crumbles  creamy\_mash  marinated\_lamb  spinach\_ricotta  pomodoro\_sauce  white\_bean\_puree  crumbled\_goat\_cheese  marscapone  foccacia  sautéed\_zucchini  melted\_cheddar\_cheese |
| citizen | Resident  Inhabitant  Denizen  National  Subject  Voter  Civic  Member  Native  Countryman  Dweller  Fellow citizen  Nationality  Townsfolk  Peasant  Townsman  Burgher  Civilian  Countrywoman  Native-born | citizen  NO\_RIGHTS  AMERICAN\_CITIZEN  citizens  AMERICAN\_CITIZENS  citzens  law\_abiding\_taxpaying  MORALLY  ABSOLUTELY\_NOT  singaporean  taxpaying\_citizen  THANK\_YOU\_THANK\_YOU  UNIDENTIFIED\_WOMAN  CINDY\_SHEEHAN  ENOUGH\_IS\_ENOUGH  south\_africans  peop  people.I  sincerly  OBAMA\_BIDEN |
| year | Annual  Calendar year  Twelve months  Fiscal year  Timeframe  Period  Cycle  Season  Anniversary  Decade  Epoch  Span  Term  Duration  Era  Age  Time  Quarter  Half-year  Solar year | month  week  months  decade  years  summer  year.The  September  weeks  season  June  yaer  weekend  July  January  twoyears  August  threeyear  October |
| in | Inside  Within  Interior  Amidst  Among  Enclosed  Surrounded  During  Amid  Inside of  Within the bounds of  Throughout  On the inside  In the midst of  In the heart of  Contained by  Embedded  Located in  Imbedded  Situated within | inthe  where  the  In  during  iin  at  from  outside  for  since  near  last  ####in  throughout  after  іn  following  around  into |
| first | Initial  Primary  Inaugural  Premier  Debut  Foremost  Maiden  Opening  Introductory  Pioneering  Head  Leading  Primary  Commencing  Beginning  Prime  Genesis  Inceptive  Original  Starting | second  third  fourth  fifth  sixth  seventh  thefirst  last  eighth  final  fi\_rst  ninth  eleventh  ##th  next  thirteenth  thesecond  ##rd  only  fourteenth |

The output of Chat GPT is very similar to the given word, similar in meaning or in type. For example, for foccacia bread it gave different types of bread, for first/in it gave **synonym** words, and for year/citizen it gave **related words to the meaning** of them. Where google gave words which **related to context** **of uses** and not to the meaning. For example for citizen it gave no\_rights or American\_citizen which related to specific field that related to citizen context. For the word in, google gives words which can be in the same sentence with in - the, where, at and not words that can replace the word in.

Chat GPT gives diversity result and google give specific result relating to one field and almost in each word there are duplicate outputs. In Chat GPT most of the output is one word and in google some of the outputs are multi word phrases (NO\_RIGHTS, AMERICAN\_CITIZEN). I can’t make google model produce only words because I will need to change the given word for this.

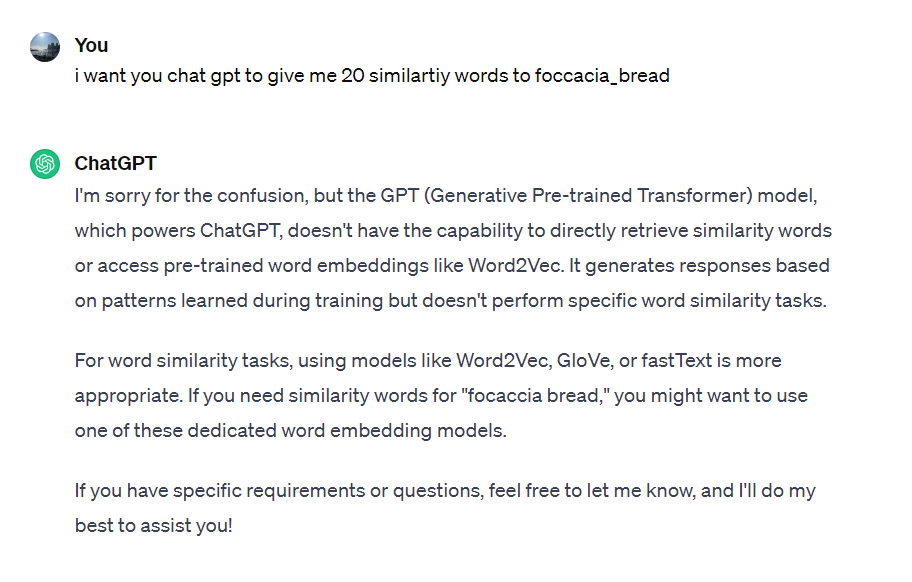
**Increasing the number of neighbors from 20 to 100, for both ChatGPT and word2vec**

|  |  |  |
| --- | --- | --- |
| Word | Chat GPT | google |
| foccacia bread | Bread  Italian bread  Artisan bread  Flatbread  Herb bread  Olive bread  Ciabatta  Baguette  Sourdough  Pita  Naan  Challah  Rye bread  Whole grain bread  Multigrain bread  Garlic bread  Rosemary bread  Thyme bread  Mediterranean bread  Crusty bread  Rustic bread  White bread  Wheat bread  Grain bread  Homemade bread  Focaccia-style bread  Italian flatbread  Artisanal bread  Breadstick  Olive oil bread  Herbed focaccia  Onion focaccia  Tomato focaccia  Pepper focaccia  Parmesan focaccia  Cheese bread  Sun-dried tomato bread  Basil bread  Flavored bread  Traditional bread  Crispy bread  Fresh bread  Baked bread  Golden bread  Mediterranean-style bread  Handmade bread  Bistro bread  Gourmet bread  Bakery bread  Roll  Sandwich bread  Appetizer bread  Focaccia roll  Rosemary-infused bread  Onion-infused bread  Savory bread  Fluffy bread  Soft bread  Square bread  Italian-style bread  Yeast bread  Doughy bread  Crispy crust bread  Airy bread  Flaky bread  Traditional flatbread  Seasoned bread  Italian herb-infused bread  Focaccia loaf  Mediterranean herb bread  Olive oil-infused bread  Garlic-infused bread  Focaccia square  Bakery-style bread  Homemade focaccia  Bread with toppings  Baked dough  Handcrafted bread  Herb and cheese bread  Freshly baked bread  Parmesan herb bread  Crispy crust flatbread  Gourmet flatbread  Mediterranean-style flatbread  Olive oil artisan bread  Italian-inspired bread  Seasoned flatbread  Herb-dusted bread  Artisan herb bread  Garlic and herb-infused bread  Parmesan-dusted bread  Freshly baked focaccia  Herb-flavored bread  Italian cuisine bread  Crunchy crust bread  Focaccia breadstick  Garlic rosemary bread  Artisan olive bread  Fluffy focaccia  Traditional herb bread | roasted\_cherry\_tomatoes  brandy\_cream  grilled\_flatbread  Grilled\_shrimp  toasted\_brioche  shredded\_chicken  mozarella  wasabi\_aioli  creamy\_horseradish  blue\_cheese\_crumbles  creamy\_mash  marinated\_lamb  spinach\_ricotta  pomodoro\_sauce  white\_bean\_puree  crumbled\_goat\_cheese  marscapone  foccacia  sautéed\_zucchini  melted\_cheddar\_cheese  brioche\_bread  broiled\_salmon  sliced\_baguette  grilled\_peppers  homemade\_croutons  creamy\_horseradish\_sauce  potato\_croquettes  truffle\_honey  apple\_cider\_vinaigrette  grilled\_flank\_steak  toasted\_garlic\_bread  applewood\_bacon  herbed\_goat\_cheese  pickle\_slices  toasted\_pita  fried\_wontons  portabella\_mushroom  spinach\_feta  red\_onion\_marmalade  toasted\_baguette  apple\_smoked\_bacon  shoestring\_fries  chicken\_liver\_pate  garlic\_aioli  eggplant\_caponata  lemon\_mayonnaise  marinara\_dipping\_sauce  roasted\_carrots  crispy\_fries  cheddar\_cheese\_sauce  onion\_marmalade  scallopini  crostinis  grilled\_zucchini  Spinach\_salad  demi\_glaze  garlic\_croutons  Marinara\_sauce  chopped\_pistachios  focaccia\_bread  baked\_eggplant  marinated\_vegetables  roasted\_mushrooms  marsala\_sauce  minced\_lamb  roast\_pork\_tenderloin  grilled\_artichokes  creamy\_potato\_salad  blueberry\_sauce  honey\_mustard\_dressing  warm\_pita\_bread  homemade\_coleslaw  grilled\_salmon\_fillet  oven\_dried\_tomatoes  thickly\_sliced  roasted\_veggie  sweet\_potato\_mash  caramelized\_shallots  melted\_provolone\_cheese  garlic\_toast  la\_plancha  cippolini\_onions  cheesy\_potatoes  caramelised\_onions  fruit\_compote  homemade\_tartar\_sauce  roasted\_veggies  al\_forno  frisee\_salad  herb\_butter  grilled\_fennel  ranchero\_sauce  melted\_mozzarella  brandy\_sauce  sourdough\_toast  shaved\_parmesan  chicken\_liver\_pâté  shredded\_Parmesan  hickory\_smoked\_bacon  boursin\_cheese |
| first | Initial  Primary  Inaugural  Premier  Debut  Foremost  Maiden  Opening  Introductory  Pioneering  Head  Leading  Commencing  Beginning  Prime  Genesis  Inceptive  Original  Starting  Initial  Initial  Primeval  Preliminary  Fundamental  Front  Primary  Antecedent  Earliest  Elemental  Inceptive  Preliminary  Basic  Essential  Inceptive  Precursor  Initiative  Opening  Preliminary  Headmost  Inceptive  Preliminary  Chief  Premature  Premier  Inaugural  Commencement  Leading  Genesis  Prime  Early  Foremost  Commencing  Originating  Pioneer  Principal  Essential  Preliminary  Starting  Initial  Inaugural  Early  First-rate  Primeval  Primitive  Primary  Basic  Earliest  Commencing  Preliminary  Elemental  Original  Fundamental  Inceptive  Premier  Opening  Maiden  Prime  Commencement  Antecedent  Pioneering  Premier  Initial  Head  Introductory  Inceptive  Foremost  Firstborn  Preliminary  Debut  Leading  Genesis  Primary  Prime  Opening  First-class  Starting  Commencing  Elementary  Essential  Preliminary | second  third  fourth  fifth  sixth  seventh  thefirst  last  eighth  final  fi\_rst  ninth  eleventh  ##th  next  thirteenth  thesecond  ##rd  only  fourteenth  twelfth  1st  fouth  ##nd  ##st  consecutive  tenth  ###th  2nd  after  before  th  prior  thefourth  4th  previous  when  fifteenth  youngest  seventeenth  lone  3rd  since  thrid  First  seond  another  7th  earlier  nine  midway\_through  beginning  ###rd  Second  six  6th  inaugural  5th  ###nd  fiftieth  ###st  three  later  nineteenth  during  twentieth  hisfirst  ####  four  biggest  shortest  one  Paul\_Konerko\_gloving  longest  f\_irst  firts  early  3\_rd  fisrt  drove\_compactor\_bulldozer  preceding  seven  Baby\_Bresnik  eighteenth  thethird  9th  latest  8th  initial  earliest  debut  eightth  opening  ago  half  late  Defending\_champion\_Andres\_Romero  highest  Sweden\_Lena\_Videkull  sixgame |

The first 20 words which Chat GPT gave for foccacia bread were words related to bread. The same type of words been given for top 100 words – words related to bread and google also gave the same type of words like the first 20 – words which related to bread topping like sauces and food to put inside the bread.

The same happened for the word first. Chat GPT continues to bring words similar in the meaning to first and google give more words related to the context of first.

Note: just the first prompt didn’t word well for me:



**Synonyms and Antonyms**

Yes, ChatGPT produced reliably synonyms. The prompot I used: can you give me synonyms for the word anger?

Yes, ChatGPT produced reliably antonyms. The prompot I used: can you give me antonyms for the word accept?

**Polysemy**

ChatGPT bring **synonym** or **related words to the meaning** of the given word and google bring words which **related to context** **of uses** and not to the meaning.

**Mean Average Precision (MAP) evaluation**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Word | ChatGPT | Judgment topically related to the target word. | semantic Judgment  context | Google | topical similarity Judgment | semantic Judgment |
| first | 1. Initial 2. Primary 3. Inaugural 4. Premier 5. Debut 6. Foremost 7. Maiden 8. Opening 9. Introductory 10. Pioneering 11. Head 12. Leading 13. Primary 14. Commencing 15. Beginning 16. Prime 17. Genesis 18. Inceptive 19. Original 20. Starting | +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  + | -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  - | 1. second 2. third 3. fourth 4. fifth 5. sixth 6. seventh 7. thefirst 8. last 9. eighth 10. final 11. fi\_rst 12. ninth 13. eleventh 14. ##th 15. next 16. thirteenth 17. thesecond 18. ##rd 19. only 20. fourteenth | -  -  -  -  -  -  +  -  -  -  +  -  -  -  -  -  -  -  -  - | +  + +  +  +  +  -  +  +  +  -  +  +  +  +  +  +  +  +  + |
| foccacia bread | 1. Ciabatta 2. Baguette 3. Sourdough 4. Pita 5. Naan 6. Flatbread 7. Challah 8. Rye bread 9. Whole grain bread 10. Multigrain bread 11. Artisan bread 12. Olive bread 13. Garlic bread 14. Herb-infused bread 15. Crusty bread 16. Italian bread 17. French bread 18. Panini 19. Bruschetta 20. Rustic bread | +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  + | -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  -  - | 1. roasted\_cherry\_tomatoes 2. brandy\_cream 3. grilled\_flatbread 4. Grilled\_shrimp 5. toasted\_brioche 6. shredded\_chicken 7. mozarella 8. wasabi\_aioli 9. creamy\_horseradish 10. blue\_cheese\_crumbles 11. creamy\_mash 12. marinated\_lamb 13. spinach\_ricotta 14. pomodoro\_sauce 15. white\_bean\_puree 16. crumbled\_goat\_cheese 17. marscapone 18. foccacia 19. sautéed\_zucchini 20. melted\_cheddar\_cheese | -  -  +  -  +  -  -  -  -  -  -  -  -  -  -  -  -  +  -  - | +  + -  +  -  +  +  +  +  +  +  +  +  +  +  +  +  -  +  + |

The AP for the word **“first”** in google is:

Topical similarity judgment:

AP=(6+7/8+8/9+9/10+10/12+11/13+12/14+13/15+14/16+15/17+16/18+17/19+18/20)/20=

0.82

Semantic judgment:

AP=(1/7+2/11)/20=0.016

In chat GPT model is :

Topical similarity judgment: AP=1

Semantic judgment: AP=0

The AP for the word in **“foccacia bread”** in google model is:

Topical similarity judgment:

AP=(1/3+2/5+3/18)/20 = 0.045

Semantic judgment:

AP=(2+3/4+4/6+5/7+6/8+7/9+8/10+9/11+10/12+11/13+12/14+13/15+14/16+15/17+16/19+17/20)/20 = 0.7

In chat GPT model is :

Topical similarity judgment: AP=1

Semantic judgment: AP=0

MAP using ChatGPT with topical similarity judgment: MAP=1

MAP using ChatGPT with semantic judgment: MAP=0

MAP using google with topical similarity judgment: MAP=(0.82+0.045)/2=0.4325

MAP using google with semantic judgment: MAP=(0.016+0.7)/2=0.358