searh_engine report

183014 이해건

1. Introduction

   A. Project Purpose and Background

      This project is about putting Python into practice, which we learned for seven weeks.

   B. Goal

      Learn how to develop a basic search engine that searches files for sentences that are highly similar to the user's sentences

2. 요구사항 Requirements

   A. User requirements

      Search engine that searches for similar sentences when a user enters a sentence

   B. Functional Requirements:

      i. sentences in the search target to the list after preprocessing

      ii. Receive input from the user, store it in the list, and remove duplicate values.

      iii. Calculate similarities based on a same token set

      iv. Sort the similarity list

      v. Print the result

3. Design and Implementation

   A. Implementation Details

      i. Save sentences in the search target to the list after preprocessing

```
# 1. Indexing
#파일을 가져와서 문장들을 가져와서 file_tokens_pairs 리스트에 한줄씩 추가한다
file_name = "jhe-koen-dev.en"
file_tokens_pairs = indexing(file_name)
```

```
def indexing(file_name):
    file_tokens_pairs = []
    lines = open(file_name, "r", encoding="utf8").readlines()
    for line in lines:
        #개행문자 떼고 ""를 기준으로 자른 것들을 토큰 변수에 저장하고 그걸 하나하나 리스트에 추가
        tokens =preprocess(line)
        file_tokens_pairs.append(tokens)
    return file_tokens_pairs
```

1. Input

   Jhe-koen-dev.en file

2. Result

   Converted to a list to which all lines in the file are members

3. description

   Take out the sentences in each file as a variable line and save them to the list

ii. Receive input from the user, store it in the list, and remove duplicate values.

```
# 2. Input the query
#유저에게 입력을 받고
query = input("영어 쿼리를 입력하세요.")
#입력받은 값도 또한 개행문자 때고 ""를 기준으로 잘라서 리스트에 넣는다.
preprocessed_query = preprocess(query.lower())
#리스트에 넣은걸 set으로 해서 중복값 날림
query_token_set = set(preprocessed_query)
```

```
def preprocess(sentence):
    preprocessed_sentence = sentence.strip().split(" ")
    return preprocessed_sentence
```

1. Input

   User input = query

2. Result

   User input .strip and .split value

3. description

   The values input by the user are listed and aggregated.

iii. similarities based on a same token set

```
# 3. Calculate similarities based on a same token set
score_dict = calc_similarity(query_token_set, file_tokens_pairs)
```

```
def calc_similarity(preprocessed_query, preprocessed_sentences):
    score_dict = {}
    #파일 속의 문장 리스트를 반복문으로 하나하나 뽑아서
    for i in range(len(file_tokens_pairs)):
        #집합에 넣어버리고
        #파일 중 한개씩 리스트에서 가져온걸 a에 저장
        a = file_tokens_pairs[i]
        #b는 a 리스트의 값들을 " "를 넣고 구분하여 문자열로 변환
        b = (" ".join(a))
        #문자열 b를 소문자화하여 lower로 저장
        lower = b.lower()
        #그것을 개행문자 때고 ""를 기준으로 때서 리스트로 저장하는 함수로 보낸걸 집합으로
        lower_pairs= set(preprocess(lower))

        #유저 입력값과 파일 i 줄 리스트의 집합의 합집합이 all takens
        all_tokens = query_token_set | lower_pairs
        #유저 입력값과 파일 i 줄의 교집합을 same tokens
        same_tokens = query_token_set & lower_pairs
        #교집합 길이를 합집합으로 나눈게 유사도
        similarity = len(same_tokens) / len(all_tokens)
        #score_dict[i] 는 유사도로 해서 저장
        score_dict[i] = similarity
    return score_dict
```

1. Input

   query_token_set, file_tokens_pairs

2. Result

   Score_dict

3. description

   Extract the sentences from the list one by one as a repeating statement. The user input value, the set of sentences in the file list, and the intersection are obtained, and the similarity is obtained by dividing it.

iv. Sort the similarity list.

```
# 4. Sort the similarity list
#내림차순으로 정렬한 리스트를 sorted score list 리스트로 저장
sorted_score_list = sorted(score_dict.items(), key = operator.itemgetter(1), reverse=True
```

1. Input x

2. Result x

3. description

   The values calculated for similarity are sorted in descending order in the order of high similarity and stored as sorted_score_list.

v. Print the result

```
# 5. Print the result
#0.1의 값이 0.00이면 유사한 문장이 없다고 출력
if sorted_score_list[0][1] == 0.0:
    print("There is no similar sentence.")
#아니면 랭크 인덱스 점수 문장, 을 탭으로 출력
else:
    print("rank", "Index", "score", "sentence", sep = "₩t")
    #랭크를 1부터 지정
    rank = 1
    #반복문으로 랭크 10까지 출력
    for i, score  in sorted_score_list:
        print(rank, i, score, ' '.join(file_tokens_pairs[i]), sep = "₩t")
        if rank == 10:
            break
        rank = rank + 1
```

1. Input x

2. Result

If the highest similarity is 0.0, print “There is no similar sentence.”

Or print “rank", "Index", "score", "sentence”
Repeated statements output up to similarity order rank == 10

3. Description

If the highest similarity is 0.0, print “There is no similar sentence.” Alternatively, all the contents from the ranking to the 10th place with high similarity are output through the repeated statement.

4. Testing

A. Test Results for Each Functionality

i. Save sentences in the search target to the list after preprocessing

```
[["You'll", 'be', 'picking', 'fruit', 'and', 'generally', 'helping', 'us', 'do', 'all', 'the', 'usual', 'farm', 'work.'], ['In', 'the',
'Middle', 'Ages,', 'cities', 'were', 'not', 'very', 'clean,', 'and', 'the', 'streets', 'were', 'filled', 'with', 'garbage.'], ['For',
'the', 'moment', 'they', 'may', 'yet', 'be', 'hiding', 'behind', 'their', 'apron', 'strings,', 'but', 'sooner', 'or', 'later', 'their',
'society', 'will', 'catch', 'up', 'with', 'the', 'progressive', 'world.'], ['Do', 'you', 'know', 'what', 'the', 'cow', 'answered?"', 's
aid', 'the', 'minister.'], ['Poland', 'and', 'Italy', 'may', 'seem', 'like', 'very', 'different', 'countries.'], ['Mr.', 'Smith', 'an
d', 'I', 'stayed', 'the', 'whole', 'day', 'in', 'Oxford.'], ['The', 'sight', 'of', 'a', 'red', 'traffic', 'signal', 'gave', 'him', 'a
n', 'idea.'], ['So', 'they', 'used', 'pumpkins', 'instead.'], ['2.', 'a', 'particular', 'occasion', 'of', 'state', 'of', 'affairs:', 'T
hey', 'might', 'not', 'offer', 'me', 'much', 'money.'], ["I'm", 'especially', 'interested', 'in', 'learning,', 'horse-riding', 'skill
s,', 'so', 'I', 'hope', "you'll", 'include', 'information', 'about', 'this.'], ['Instead,', 'the', 'devil', 'gave', 'him', 'a', 'singl
e', 'candle', 'to', 'light', 'his', 'way', 'through', 'the', 'darkness.'], ['It', 'shines', 'over', 'the', 'sea.'], ['He,', 'too,', 'wa
s', 'arrested,', 'and', 'a', 'bomb', 'was', 'thrown', 'at', 'his', 'house.'], ['It', 'seems', 'that', 'the', 'high', 'temperature', 'an
d', 'pressure', 'on', 'the', 'star', 'made', 'its', 'carbon', 'surface', 'turn', 'to', 'diamond.'], ['"The', 'pig', 'was', 'unpopular',
'while', 'the', 'cow', 'was', 'loved', 'by', 'everyone.'], ['Books', 'give', 'a', 'lot', 'of', 'things', 'to', 'us.'], ['Jimmy', 'and',
'Timmy', 'were', 'identical', 'twins.'], ['It', 'is', 'a', 'chemical', 'that', 'cause', 'cancer.'], ['Ziege', 'from', 'Germany', 'and',
'Brazilian', 'superstars', 'Ronaldo', 'and', 'Roberto', 'Carlos', 'belonged', 'to', 'the', 'bald', 'club.'], ['Now', 'the', 'Taliban',
'are', 'gone', 'and', 'things', 'have', 'begun', 'to', 'change.'], ['Is', 'your', 'skin', 'clear,', 'smooth,', 'and', 'shining', 'wit
h', 'health?'], ['As', 'a', 'result,', 'there', 'is', 'a', 'great', 'deal', 'of', 'traffic', 'and', 'usually', 'not', 'enough', 'road
```

ii. Receive input from the user, store it in the list, and remove duplicate values.

```
영어 쿼리를 입력하세요.i am so samrt
{'so', 'am', 'i', 'samrt'}
```

iii. Calculate similarities based on a same token set

{0: 0.0, 1: 0.0, 2: 0.0, 3: 0.0, 4: 0.0, 5: 0.07692307692307693, 6: 0.0, 7: 0.125, 8: 0.0, 9: 0.11764705882352941, 10: 0.0, 11: 0.0, 12: 0.0, 13: 0.0, 14: 0.0, 15: 0.0, 16: 0.0, 17: 0.0, 18: 0.0, 19: 0.0, 20: 0.0, 21: 0.0, 22: 0.0, 23: 0.0625, 24: 0.0, 25: 0.0, 26: 0.0, 27: 0.0, 28: 0.0, 29: 0.0, 30: 0.0, 31: 0.0, 32: 0.047619047619047616, 33: 0.0, 34: 0.0, 35: 0.0, 36: 0.0, 37: 0.0, 38: 0.0, 39: 0.0, 40: 0.043478260869565216, 41: 0.05555555555555555, 42: 0.06666666666666667, 43: 0.0, 44: 0.0, 45: 0.0, 46: 0.0, 47: 0.0, 48: 0.0, 49: 0.0, 50: 0.0, 51: 0.0, 52: 0.0, 53: 0.0, 54: 0.0, 55: 0.0, 56: 0.0, 57: 0.0, 58: 0.0, 59: 0.0, 60: 0.0, 61: 0.0, 62: 0.1, 63: 0.0, 64: 0.0, 65: 0.0, 66: 0.0, 67: 0.0, 68: 0.0, 69: 0.0, 70: 0.0, 71: 0.0, 72: 0.0, 73: 0.0, 74: 0.0, 75: 0.0, 76: 0.0, 77: 0.0, 78: 0.0, 79: 0.0, 80: 0.0, 81: 0.0625, 82: 0.0, 83: 0.0, 84: 0.0, 85: 0.0, 86: 0.0, 87: 0.0, 88: 0.0, 89: 0.0, 90: 0.0, 91: 0.0, 92: 0.0, 93: 0.0, 94: 0.0, 95: 0.0, 96: 0.0, 97: 0.0, 98: 0.0, 99: 0.0, 100: 0.0, 101: 0.0, 102: 0.0, 103: 0.0, 104: 0.0625, 105: 0.0, 106: 0.05263157894736842, 107: 0.0, 108: 0.0, 109: 0.0, 110: 0.0, 111: 0.0, 112: 0.07692307692307693, 113: 0.0, 114: 0.0, 115: 0.0, 116: 0.0, 117: 0.0, 118: 0.0, 119: 0.07142857142857142, 120: 0.0, 121: 0.0, 122: 0.0, 123: 0.0, 124: 0.0, 125: 0.0, 126: 0.0, 127: 0.0, 128: 0.04, 129: 0.0, 130: 0.0, 131: 0.0, 132: 0.0, 133: 0.0, 134: 0.0, 135: 0.0, 136: 0.0, 137: 0.0, 138: 0.0, 139: 0.0, 140: 0.0, 141: 0.0, 142: 0.0, 143: 0.0, 144: 0.0, 145: 0.0, 146: 0.0, 147: 0.07692307692307693, 148: 0.0, 149: 0.0, 150: 0.0, 151: 0.0, 152: 0.0, 153: 0.0, 154: 0.0, 155: 0.0, 156: 0.0, 157: 0.0, 158: 0.0, 159: 0.0, 160: 0.15384615384615385, 161: 0.0, 162: 0.0, 163: 0.05555555555555555, 164: 0.0, 165: 0.0, 166: 0.0, 167: 0.0, 168: 0.0666666666666667, 169: 0.0, 170: 0.045454545454545456, 171: 0.0, 172: 0.0, 173: 0.1, 174: 0.0, 175: 0.0, 176: 0.0, 177: 0.0, 178: 0.0, 179: 0.0, 180: 0.0, 181: 0.0, 182: 0.0, 183: 0.0, 184: 0.0, 185: 0.14285714285714285, 186: 0.0, 187: 0.0, 188: 0.0, 189: 0.0, 190: 0.0, 191: 0.0, 192: 0.0, 193: 0.0, 194: 0.0, 195: 0.0, 196: 0.0, 197: 0.0, 198: 0.0, 199: 0.0, 200: 0.0, 201: 0.0, 202: 0.0625, 203: 0.0, 204: 0.0, 205: 0.0, 206: 0.0, 207: 0.0, 208: 0.0, 209: 0.0, 210: 0.0, 211: 0.0, 212: 0.0, 213: 0.0, 214: 0.0, 215: 0.0, 216: 0.0, 217: 0.0, 218: 0.0, 219: 0.0, 220: 0.0, 221: 0.0, 222: 0.0, 223: 0.0, 224: 0.0, 225: 0.0, 226: 0.0, 227: 0.0, 228: 0.0, 229: 0.0, 230: 0.0, 231: 0.0, 232: 0.0, 233: 0.0, 234: 0.0, 235: 0.0, 236: 0.043478260869565216, 237: 0.0, 238: 0.0, 239: 0.0, 240: 0.0, 241: 0.0, 242: 0.0, 243: 0.0, 244: 0.0, 245: 0.0, 246: 0.0, 247: 0.0, 248: 0.0, 249: 0.0, 250: 0.0, 251: 0.2857142857142857, 252: 0.0, 253: 0.0, 254: 0.0, 255: 0.0, 256: 0.0, 257: 0.07142857142857142, 258: 0.0, 259: 0.0, 260: 0.0, 261: 0.0, 262: 0.0, 263: 0.0, 264: 0.0, 265: 0.0, 266: 0.09090909090909091, 267: 0.0, 268: 0.0, 269: 0.0, 270: 0.0, 271: 0.07692307692307693, 272: 0.0, 273: 0.1111111111111111, 274: 0.0, 275: 0.0, 276: 0.0, 277: 0.0, 278: 0.05, 279: 0.0, 280: 0.0, 281: 0.0, 282: 0.0, 283: 0.0, 284: 0.0, 285: 0.0, 286: 0.0, 287: 0.07142857142857142, 288: 0.0, 289: 0.0, 290: 0.0, 291: 0.03571428571428571, 292: 0.0, 293: 0.0, 294: 0.0, 295: 0.0, 296: 0.0, 297: 0.0, 298: 0.0, 299: 0.0, 300: 0.0, 301: 0.0, 302: 0.0, 303: 0.0, 304: 0.0, 305: 0.0, 306: 0.0, 307: 0.0, 308: 0.0, 309: 0.0, 310: 0.0, 311: 0.0, 312: 0.0, 313: 0.0, 314: 0.0, 315: 0.0, 316: 0.0, 317: 0.0, 318: 0.0, 319: 0.07692307692307693, 320: 0.0, 321: 0.0, 322: 0.0, 323: 0.0, 324: 0.0, 325: 0.0, 326: 0.0, 327: 0.0, 328: 0.0, 329: 0.0, 330: 0.0, 331: 0.0, 332: 0.0, 333: 0.0, 334: 0.07142857142857142, 335: 0.0, 336: 0.0, 337: 0.0, 338: 0.0, 339: 0.0, 340: 0.0, 341: 0.0, 342: 0.0, 343: 0.0, 344: 0.0, 345: 0.0, 346: 0.0, 347: 0.0, 348: 0.0, 349: 0.0, 350: 0.0, 351: 0.0, 352: 0.0, 353: 0.0, 354: 0.0, 355: 0.0, 356: 0.0, 357: 0.0625, 358: 0.0, 359: 0.0, 360: 0.0, 361: 0.0, 362: 0.0, 363: 0.0, 364: 0.0, 365: 0.0, 366: 0.0, 367: 0.0, 368: 0.0, 369: 0.0, 370: 0.0, 371: 0.0, 372: 0.0, 373: 0.0, 374: 0.0, 375: 0.0, 376: 0.0, 377: 0.0, 378: 0.0, 379: 0.0, 380: 0.0, 381: 0.0, 382: 0.0, 383: 0.1, 384: 0.0, 385: 0.0, 386: 0.0, 387: 0.0, 388: 0.1111111111111111, 389: 0.0, 390: 0.05263157894736842, 391: 0.0, 392: 0.0, 393: 0.041666666666666664, 394: 0.0, 395: 0.0, 396: 0.0, 397: 0.0, 398: 0.0, 399: 0.0, 400: 0.0, 401: 0.0, 402: 0.0, 403: 0.0, 404: 0.0, 405: 0.125, 406: 0.0, 407: 0.0, 408: 0.0, 409: 0.0, 410: 0.0, 411: 0.0, 412: 0.0, 413: 0.0, 414: 0.0, 415: 0.0625, 416: 0.0, 417: 0.0, 418: 0.0, 419: 0.0, 420: 0.0, 421: 0.0, 422: 0.0, 423: 0.0, 424: 0.0, 425: 0.0, 426: 0.058823529411764705, 427: 0.0, 428: 0.0, 429: 0.0, 430: 0.0, 431: 0.0, 432: 0.07692307692307693, 433: 0.0, 434: 0.0, 435: 0.0, 436: 0.09090909090909091, 437: 0.0, 438: 0.0, 439: 0.09090909090909091, 440: 0.0, 441: 0.18181818181818182, 442: 0.0, 443: 0.0, 444: 0.0, 445: 0.0, 446: 0.08333333333333333, 447: 0.0, 448: 0.0, 449: 0.0, 450: 0.0, 451: 0.0, 452: 0.0, 453: 0.0, 454: 0.0, 455: 0.0, 456: 0.0, 457: 0.0, 458: 0.0, 459: 0.0, 460: 0.0, 461: 0.0, 462: 0.0, 463: 0.0, 464: 0.0, 465: 0.0, 466: 0.0, 467: 0.0, 468: 0.0, 469: 0.0, 470: 0.0, 471: 0.0, 472: 0.1, 473: 0.0, 474: 0.0, 475: 0.1, 476: 0.0, 477: 0.0, 478: 0.0, 479: 0.0, 480: 0.0, 481: 0.0, 482: 0.0, 483: 0.0, 484: 0.0, 485: 0.0, 486: 0.05555555555555555, 487: 0.0, 488: 0.0, 489: 0.0, 490: 0.0, 491: 0.0, 492: 0.0, 493: 0.0, 494: 0.0, 495: 0.0, 496: 0.0, 497: 0.0, 498: 0.058823529411764705, 499: 0.0, 500: 0.0, 501: 0.0, 502: 0.0, 503: 0.0, 504: 0.0, 505: 0.0, 506: 0.058823529411764705, 507: 0.0, 508: 0.06666666666666667, 509: 0.0, 510: 0.04, 511: 0.058823529411764705, 512: 0.0, 513: 0.0, 514: 0.0, 515: 0.0, 516: 0.125, 517: 0.0, 518: 0.0, 519: 0.0, 520: 0.0, 521: 0.041666666666666664, 522: 0.0, 523: 0.0, 524: 0.0, 525: 0.0, 526: 0.0, 527: 0.0, 528: 0.0, 529: 0.0, 530: 0.0, 531: 0.0, 532: 0.0, 533: 0.0625, 534: 0.1, 535: 0.0, 536: 0.0, 537: 0.09090909090909091, 538: 0.0, 539: 0.0, 540: 0.0, 541: 0.0, 542: 0.0, 543: 0.05555555555555555, 544: 0.0, 545: 0.0, 546: 0.0, 547: 0.0, 548: 0.0, 549: 0.0625, 550: 0.0, 551: 0.0, 552: 0.0, 553: 0.0, 554: 0.0, 555: 0.0, 556: 0.1111111111111111, 557: 0.0, 558: 0.0, 559: 0.0, 560: 0.0, 561: 0.0, 562: 0.0, 563: 0.0, 564: 0.1111111111111111, 565: 0.0, 566: 0.0, 567: 0.0, 568: 0.0, 569: 0.0, 570: 0.0, 571: 0.0625, 572: 0.0, 573: 0.0, 574: 0.0, 575: 0.0625, 576: 0.0, 577: 0.0, 578: 0.0, 579: 0.058823529411764705, 580: 0.0, 581: 0.09090909090909091, 582: 0.0, 583: 0.0, 584: 0.0, 585: 0.0, 586: 0.07142857142857142, 587: 0.0, 588: 0.0, 589: 0.0, 590: 0.047619047619047616, 591: 0.0, 592: 0.0, 593: 0.0, 594: 0.0, 595: 0.0, 596: 0.0, 597: 0.0, 598: 0.0, 599: 0.0, 600: 0.0, 601: 0.07142857142857142, 602: 0.0, 603: 0.0, 604: 0.0, 605: 0.0, 606: 0.0, 607: 0.0, 608: 0.0, 609: 0.0, 610: 0.07692307692307693, 611: 0.0, 612: 0.0, 613: 0.0, 614: 0.0, 615: 0.0, 616: 0.0, 617: 0.0, 618: 0.0, 619: 0.0, 620: 0.1, 621: 0.07692307692307693, 622: 0.0, 623: 0.16666666666666666, 624: 0.0, 625: 0.0, 626: 0.0, 627: 0.0, 628: 0.0, 629: 0.0, 630: 0.0, 631: 0.0, 632: 0.0, 633: 0.0, 634: 0.0, 635: 0.0, 636: 0.0, 637: 0.0, 638: 0.05555555555555555, 639: 0.0, 640: 0.0, 641: 0.07692307692307693, 642: 0.0, 643: 0.0, 644: 0.0, 645: 0.0, 646: 0.0, 647: 0.0, 648: 0.0, 649: 0.08333333333333333, 650: 0.0, 651: 0.0, 652: 0.1, 653: 0.0, 654: 0.0, 655: 0.0, 656: 0.0, 657: 0.0, 658: 0.0, 659: 0.0, 660: 0.0, 661: 0.0, 662: 0.0, 663: 0.0, 664: 0.0, 665: 0.0, 666: 0.05, 667: 0.0, 668: 0.0, 669: 0.0, 670: 0.0, 671: 0.0, 672: 0.0, 673: 0.0, 674: 0.0, 675: 0.0, 676: 0.0, 677: 0.0, 678: 0.0, 679: 0.0, 680: 0.06666666666666667, 681: 0.0, 682: 0.0, 683: 0.0, 684: 0.0, 685: 0.0, 686: 0.0, 687: 0.0, 688: 0.0, 689: 0.0, 690: 0.0, 691: 0.05555555555555555, 692: 0.0, 693: 0.0, 694: 0.0, 695: 0.0, 696: 0.0, 697: 0.0, 698: 0.0, 699: 0.0, 700: 0.0, 701: 0.0, 702: 0.0, 703: 0.0, 704: 0.0, 705: 0.0, 706: 0.0, 707: 0.0, 708: 0.0, 709: 0.0, 710: 0.0, 711: 0.0, 712: 0.05, 713: 0.0, 714: 0.0, 715: 0.0, 716: 0.0, 717: 0.0, 718: 0.1111111111111111, 719: 0.0}

iv. Sort the similarity list

[(251, 0.2857142857142857), (441, 0.18181818181818182), (623, 0.16666666666666666), (160, 0.15384615384615385), (185, 0.14285714285714285), (7, 0.125), (405, 0.125), (516, 0.125), (9, 0.11764705882352941), (273, 0.1111111111111111), (388, 0.1111111111111111), (556, 0.1111111111111111), (564, 0.1111111111111111), (718, 0.1111111111111111), (62, 0.1), (173, 0.1), (383, 0.1), (472, 0.1), (475, 0.1), (534, 0.1), (620, 0.1), (652, 0.1), (266, 0.09090909090909091), (436, 0.09090909090909091), (439, 0.09090909090909091), (537, 0.09090909090909091), (581, 0.09090909090909091), (446, 0.08333333333333333), (649, 0.08333333333333333), (5, 0.07692307692307693), (112, 0.07692307692307693), (147, 0.07692307692307693), (271, 0.07692307692307693), (319, 0.07692307692307693), (432, 0.07692307692307693), (610, 0.07692307692307693), (621, 0.07692307692307693), (641, 0.07692307692307693), (119, 0.07142857142857142), (257, 0.07142857142857142), (287, 0.07142857142857142), (334, 0.07142857142857142), (586, 0.07142857142857142), (601, 0.07142857142857142), (42, 0.06666666666666667), (168, 0.06666666666666667), (508, 0.06666666666666667), (680, 0.06666666666666667), (23, 0.0625), (81, 0.0625), (104, 0.0625), (202, 0.0625), (357, 0.0625), (415, 0.0625), (533, 0.0625), (549, 0.0625), (571, 0.0625), (575, 0.0625), (426, 0.058823529411764705), (498, 0.058823529411764705), (506, 0.058823529411764705), (511, 0.058823529411764705), (579, 0.058823529411764705), (41, 0.05555555555555555), (163, 0.05555555555555555), (486, 0.05555555555555555), (543, 0.05555555555555555), (638, 0.05555555555555555), (691, 0.05555555555555555), (106, 0.05263157894736842), (390, 0.05263157894736842), (278, 0.05), (666, 0.05), (712, 0.05), (32, 0.047619047619047616), (590, 0.047619047619047616), (170, 0.045454545454545456), (40, 0.043478260869565216), (236, 0.043478260869565216), (393, 0.041666666666666664), (521, 0.041666666666666664), (128, 0.04), (510, 0.04), (291, 0.03571428571428571), (0, 0.0), (1, 0.0), (2, 0.0), (3, 0.0), (4, 0.0), (6, 0.0), (8, 0.0), (10, 0.0), (11, 0.0), (12, 0.0), (13, 0.0), (14, 0.0), (15, 0.0), (16, 0.0), (17, 0.0), (18, 0.0), (19, 0.0), (20, 0.0), (21, 0.0), (22, 0.0), (24, 0.0), (25, 0.0), (26, 0.0), (27, 0.0), (28, 0.0), (29, 0.0), (30, 0.0), (31, 0.0), (33, 0.0), (34, 0.0), (35, 0.0), (36, 0.0), (37, 0.0), (38, 0.0), (39, 0.0), (43, 0.0), (44, 0.0), (45, 0.0), (46, 0.0), (47, 0.0), (48, 0.0), (49, 0.0), (50, 0.0), (51, 0.0), (52, 0.0), (53, 0.0), (54, 0.0), (55, 0.0), (56, 0.0), (57, 0.0), (58, 0.0), (59, 0.0), (60, 0.0), (61, 0.0), (63, 0.0), (64, 0.0), (65, 0.0), (66, 0.0), (67, 0.0), (68, 0.0), (69, 0.0), (70, 0.0), (71, 0.0), (72, 0.0), (73, 0.0), (74, 0.0), (75, 0.0), (76, 0.0), (77, 0.0), (78, 0.0), (79, 0.0), (80, 0.0), (82, 0.0), (83, 0.0), (84, 0.0), (85, 0.0), (86, 0.0), (87, 0.0), (88, 0.0), (89, 0.0), (90, 0.0), (91, 0.0), (92, 0.0), (93, 0.0), (94, 0.0), (95, 0.0), (96, 0.0), (97, 0.0), (98, 0.0), (99, 0.0), (100, 0.0), (101, 0.0), (102, 0.0), (103, 0.0), (105, 0.0), (107, 0.0), (108, 0.0), (109, 0.0), (110, 0.0), (111, 0.0), (113, 0.0), (114, 0.0), (115, 0.0), (116, 0.0), (117, 0.0), (118, 0.0), (120, 0.0), (121, 0.0), (122, 0.0), (123, 0.0), (124, 0.0), (125, 0.0), (126, 0.0), (127, 0.0), (129, 0.0), (130, 0.0), (131, 0.0), (132, 0.0), (133, 0.0), (134, 0.0), (135, 0.0), (136, 0.0), (137, 0.0), (138, 0.0), (139, 0.0), (140, 0.0), (141, 0.0), (142, 0.0), (143, 0.0), (144, 0.0), (145, 0.0), (146, 0.0), (148, 0.0), (149, 0.0), (150, 0.0), (151, 0.0), (152, 0.0), (153, 0.0), (154, 0.0), (155, 0.0), (156, 0.0), (157, 0.0), (158, 0.0), (159, 0.0), (161, 0.0), (162, 0.0), (164, 0.0), (165, 0.0), (166, 0.0), (167, 0.0), (169, 0.0), (171, 0.0), (172, 0.0), (174, 0.0), (175, 0.0), (176, 0.0), (177, 0.0), (178, 0.0), (179, 0.0), (180, 0.0), (181, 0.0), (182, 0.0), (183, 0.0), (184, 0.0), (186, 0.0), (187, 0.0), (188, 0.0), (189, 0.0), (190, 0.0), (191, 0.0), (192, 0.0), (193, 0.0), (194, 0.0), (195, 0.0), (196, 0.0), (197, 0.0), (198, 0.0), (199, 0.0), (200, 0.0), (201, 0.0), (203, 0.0), (204, 0.0), (205, 0.0), (206, 0.0), (207, 0.0), (208, 0.0), (209, 0.0), (210, 0.0), (211, 0.0), (212, 0.0), (213, 0.0), (214, 0.0), (215, 0.0), (216, 0.0), (217, 0.0), (218, 0.0), (219, 0.0), (220, 0.0), (221, 0.0), (222, 0.0), (223, 0.0), (224, 0.0), (225, 0.0), (226, 0.0), (227, 0.0), (228, 0.0), (229, 0.0), (230, 0.0), (231, 0.0), (232, 0.0), (233, 0.0), (234, 0.0), (235, 0.0), (237, 0.0), (238, 0.0), (239, 0.0), (240, 0.0), (241, 0.0), (242, 0.0), (243, 0.0), (244, 0.0), (245, 0.0), (246, 0.0), (247, 0.0), (248, 0.0), (249, 0.0), (250, 0.0), (252, 0.0), (253, 0.0), (254, 0.0), (255, 0.0), (256, 0.0), (258, 0.0), (259, 0.0), (260, 0.0), (261, 0.0), (262, 0.0), (263, 0.0), (264, 0.0), (265, 0.0), (267, 0.0), (268, 0.0), (269, 0.0), (270, 0.0), (272, 0.0), (274, 0.0), (275, 0.0), (276, 0.0), (277, 0.0), (279, 0.0), (280, 0.0), (281, 0.0), (282, 0.0), (283, 0.0), (284, 0.0), (285, 0.0), (286, 0.0), (288, 0.0), (289, 0.0), (290, 0.0), (292, 0.0), (293, 0.0), (294, 0.0), (295, 0.0), (296, 0.0), (297, 0.0), (298, 0.0), (299, 0.0), (300, 0.0), (301, 0.0), (302, 0.0), (303, 0.0), (304, 0.0), (305, 0.0), (306, 0.0), (307, 0.0), (308, 0.0), (309, 0.0), (310, 0.0), (311, 0.0), (312, 0.0), (313, 0.0), (314, 0.0), (315, 0.0), (316, 0.0), (317, 0.0), (318, 0.0), (320, 0.0), (321, 0.0), (322, 0.0), (323, 0.0), (324, 0.0), (325, 0.0), (326, 0.0), (327, 0.0), (328, 0.0), (329, 0.0), (330, 0.0), (331, 0.0), (332, 0.0), (333, 0.0), (335, 0.0), (336, 0.0), (337, 0.0), (338, 0.0), (339, 0.0), (340, 0.0), (341, 0.0), (342, 0.0), (343, 0.0), (344, 0.0), (345, 0.0), (346, 0.0), (347, 0.0), (348, 0.0), (349, 0.0), (350, 0.0), (351, 0.0), (352, 0.0), (353, 0.0), (354, 0.0), (355, 0.0), (356, 0.0), (358, 0.0), (359, 0.0), (360, 0.0), (361, 0.0), (362, 0.0), (363, 0.0), (364, 0.0), (365, 0.0), (366, 0.0), (367, 0.0), (368, 0.0), (369, 0.0), (370, 0.0), (371, 0.0), (372, 0.0), (373, 0.0), (374, 0.0), (375, 0.0), (376, 0.0), (377, 0.0), (378, 0.0), (379, 0.0), (380, 0.0), (381, 0.0), (382, 0.0), (384, 0.0), (385, 0.0), (386, 0.0), (387, 0.0), (389, 0.0), (391, 0.0), (392, 0.0), (394, 0.0), (395, 0.0), (396, 0.0), (397, 0.0), (398, 0.0), (399, 0.0), (400, 0.0), (401, 0.0), (402, 0.0), (403, 0.0), (404, 0.0), (406, 0.0), (407, 0.0), (408, 0.0), (409, 0.0), (410, 0.0), (411, 0.0), (412, 0.0), (413, 0.0), (414, 0.0), (416, 0.0), (417, 0.0), (418, 0.0), (419, 0.0), (420, 0.0), (421, 0.0), (422, 0.0), (423, 0.0), (424, 0.0), (425, 0.0), (427, 0.0), (428, 0.0), (429, 0.0), (430, 0.0), (431, 0.0), (433, 0.0), (434, 0.0), (435, 0.0), (437, 0.0), (438, 0.0), (440, 0.0), (442, 0.0), (443, 0.0), (444, 0.0), (445, 0.0), (447, 0.0), (448, 0.0), (449, 0.0), (450, 0.0), (451, 0.0), (452, 0.0), (453, 0.0), (454, 0.0), (455, 0.0), (456, 0.0), (457, 0.0), (458, 0.0), (459, 0.0), (460, 0.0), (461, 0.0), (462, 0.0), (463, 0.0), (464, 0.0), (465, 0.0), (466, 0.0), (467, 0.0), (468, 0.0), (469, 0.0), (470, 0.0), (471, 0.0), (473, 0.0), (474, 0.0), (476, 0.0), (477, 0.0), (478, 0.0), (479, 0.0), (480, 0.0), (481, 0.0), (482, 0.0), (483, 0.0), (484, 0.0), (485, 0.0), (487, 0.0), (488, 0.0), (489, 0.0), (490, 0.0), (491, 0.0), (492, 0.0), (493, 0.0), (494, 0.0), (495, 0.0), (496, 0.0), (497, 0.0), (499, 0.0), (500, 0.0), (501, 0.0), (502, 0.0), (503, 0.0), (504, 0.0), (505, 0.0), (507, 0.0), (509, 0.0), (512, 0.0), (513, 0.0), (514, 0.0), (515, 0.0), (517, 0.0), (518, 0.0), (519, 0.0), (520, 0.0), (522, 0.0), (523, 0.0), (524, 0.0), (525, 0.0), (526, 0.0), (527, 0.0), (528, 0.0), (529, 0.0), (530, 0.0), (531, 0.0), (532, 0.0), (535, 0.0), (536, 0.0), (538, 0.0), (539, 0.0), (540, 0.0), (541, 0.0), (542, 0.0), (544, 0.0), (545, 0.0), (546, 0.0), (547, 0.0), (548, 0.0), (550, 0.0), (551, 0.0), (552, 0.0), (553, 0.0), (554, 0.0), (555, 0.0), (557, 0.0), (558, 0.0), (559, 0.0), (560, 0.0), (561, 0.0), (562, 0.0), (563, 0.0), (565, 0.0), (566, 0.0), (567, 0.0), (568, 0.0), (569, 0.0), (570, 0.0), (572, 0.0), (573, 0.0), (574, 0.0), (576, 0.0), (577, 0.0), (578, 0.0), (580, 0.0), (582, 0.0), (583, 0.0), (584, 0.0), (585, 0.0), (587, 0.0), (588, 0.0), (589, 0.0), (591, 0.0), (592, 0.0), (593, 0.0), (594, 0.0), (595, 0.0), (596, 0.0), (597, 0.0), (598, 0.0), (599, 0.0), (600, 0.0), (602, 0.0), (603, 0.0), (604, 0.0), (605, 0.0), (606, 0.0), (607, 0.0), (608, 0.0), (609, 0.0), (611, 0.0), (612, 0.0), (613, 0.0), (614, 0.0), (615, 0.0), (616, 0.0), (617, 0.0), (618, 0.0), (619, 0.0), (622, 0.0), (624, 0.0), (625, 0.0), (626, 0.0), (627, 0.0), (628, 0.0), (629, 0.0), (630, 0.0), (631, 0.0), (632, 0.0), (633, 0.0), (634, 0.0), (635, 0.0), (636, 0.0), (637, 0.0), (639, 0.0), (640, 0.0), (642, 0.0), (643, 0.0), (644, 0.0), (645, 0.0), (646, 0.0), (647, 0.0), (648, 0.0), (650, 0.0), (651, 0.0), (653, 0.0), (654, 0.0), (655, 0.0), (656, 0.0), (657, 0.0), (658, 0.0), (659, 0.0), (660, 0.0), (661, 0.0), (662, 0.0), (663, 0.0), (664, 0.0), (665, 0.0), (667, 0.0), (668, 0.0), (669, 0.0), (670,

v. Print the result

```
영어 쿼리를 입력하세요.bob is my brother.
rank    Index   score   sentence
1       526     1.0     Bob is my brother.
2       538     0.3333333333333333      My hobby is traveling.
3       679     0.3333333333333333      My name is Mike.
4       453     0.2857142857142857      My mother is sketching them.
5       241     0.25    My father is running with So-ra.
6       336     0.25    My family is at the park.
7       212     0.2222222222222222      My sister Betty is waiting for me.
8       505     0.2     My little sister Annie is five years old.
9       190     0.16666666666666666     It is Sunday.
10      314     0.16666666666666666     This is Washington.
```

vi. Final Test Screenshot

```
영어 쿼리를 입력하세요.i am so smart
rank    Index   score   sentence
1       251     0.2857142857142857      I am nine years old.
2       441     0.18181818181818182     So I opened the cage, and it flew away.
3       623     0.16666666666666666     I am sorry for my late reply to your letter.
4       160     0.15384615384615385     I just wanted you to know how proud I am of Bobby.
5       185     0.14285714285714285     So I didn't like the idea of my son becoming a cook.
6       7       0.125   So they used pumpkins instead.
7       405     0.125   I felt sorry for it.
8       516     0.125   I just got your flower.
9       9       0.11764705882352941     I'm especially interested in learning, horse-riding skills, so I hope you'll include information a
bout this.
10      273     0.1111111111111111      But I knew my parents would not approve, so I had to introduce them slowly to the idea.
```

5. Results and Conclusion

A. Result

Through the similarity of words inside the sentence, a program that searches for similar sentences could be established and supplemented.

B. Conclusion

At first, it felt quite difficult, but when I opened each line, it was amazing to see that it was built within what we know, and if I improve my skills later, I want to design something like this.