[337\_TFIDF.json]

x, idx2vocab = sent\_to\_word\_contexts\_matrix(  
 corpus,  
 windows=10, *#고려할 앞뒤 단어의 개수* min\_tf=0, *#10회 이하의 단어 무시* tokenizer=**lambda** x:x.split(), *# (default) lambda x:x.split(),* dynamic\_weight=**True**,  
 verbose=**True**)  
  
print(x.shape)  
  
glove = Glove(no\_components=100, learning\_rate=0.05, max\_count=30)  
model=glove.fit(x.tocoo(), epochs=10, no\_threads=4, verbose=**True**)

words = **'목록 축구관련 백만건 유료버전 푸시경고 최신스코어 시합통계 교체선수 리그뉴스 클럽뉴스 최고득점자 선호팀 축구리그들'**.split() *#이곳은 각 애플리케이션 문서의 키워드가 들어있음(키워드는 tf-idf로 추출함)*

**tf-idf**

**"목록"**: **"0.2518258895520309"**,  
**"축구관련"**: **"0.21983860633370508"**,  
**"백만건"**: **"0.21983860633370508"**,  
**"유료버전"**: **"0.21983860633370508"**,  
**"푸시경고"**: **"0.21983860633370508"**,  
**"최신스코어"**: **"0.21983860633370508"**,  
**"시합통계"**: **"0.21983860633370508"**,  
**"교체선수"**: **"0.21983860633370508"**,  
**"리그뉴스"**: **"0.21983860633370508"**,  
**"클럽뉴스"**: **"0.21983860633370508"**,  
**"최고득점자"**: **"0.21983860633370508"**,  
**"선호팀"**: **"0.21983860633370508"**,  
**"축구리그들"**: **"0.21983860633370508"**,

glove sim

스포츠 , 목록 = 0.10013973949280508

스포츠 , 축구관련 = -0.6086763133601077

스포츠 , 백만건 = -0.5454897650727982

스포츠 , 유료버전 = -0.41105925536527305

스포츠 , 푸시경고 = -0.1201977139444012

스포츠 , 최신스코어 = -0.3753809660695531

스포츠 , 시합통계 = -0.7923196996559604

스포츠 , 교체선수 = -0.299965048505893

스포츠 , 리그뉴스 = 0.30648268532731754

스포츠 , 클럽뉴스 = 0.21435958652696094

스포츠 , 최고득점자 = 0.20649817476082374

스포츠 , 선호팀 = 0.22743617360123008

스포츠 , 축구리그들 = 0.48041357253337497

glove sim \* tfidf 한 최종 weight

목록, 0.025976509941300473

**축구관련, -0.12489326706197329**

백만건, -0.11647800851473801

유료버전, -0.09093915439095396

푸시경고, -0.03258803396429703

최신스코어, -0.07807189468244866

**시합통계, -0.16353286570519363**

교체선수, -0.0631829919197551

리그뉴스, 0.04803980685230409

클럽뉴스, 0.034522467737328824

최고득점자, 0.033877541325774384

선호팀, 0.0344539999283516

축구리그들, 0.09627259055689807

|  |
| --- |
| word2vec  목록  0.068462305  축구  0.5668544  관련  0.16425689  백만  -0.044684812  건  0.006037943  유료  0.27997556  버전  0.12873086  푸시  0.21954091  경고  -0.059531238  최신  0.16937573  스코어  0.35617134  시합  0.2681474  통계  0.22526088  교체  0.055228114  선수  0.4029561  리그  0.38593853  뉴스  0.40567374  클럽  0.4043956  뉴스  0.40567374  최고  0.20783305  득점자  0.20107225  선호  0.12287942  팀  0.361611  축구  0.5668544  리그  0.38593853  들  0.09205809 |
| Final weight |

|  |
| --- |
| Category : 교육 /test file 130.TFIDF.json |
| Tfidf  **"별지"**: **"0.2251157417857172"**, **"행성"**: **"0.21823267659540663"**, **"별자리"**: **"0.2111026593564412"**, **"모드"**: **"0.19850645530637243"**, **"하늘"**: **"0.1862575410464857"**, **"시계"**: **"0.16061552286540492"**, **"88"**: **"0.15284943708341342"**, **"태양"**: **"0.15222960577372682"**, **"천문"**: **"0.14902684096414184"**, **"위치"**: **"0.1368438673492974"**, **"천체"**: **"0.12852703967049162"**, **"천구"**: **"0.12752846625208059"**, **"라이브"**: **"0.12702368305638195"**, **"중심"**: **"0.12063009106082355"**, **"디스플레이"**: **"0.1070960989932103"**, |
| words = **'별지 행성 별자리 모드 하늘 시계 88 태양 천문 위치 천체 천구 라이브 중심 디스플레이'**.split() *#이곳은 각 애플리케이션 문서의 키워드가 들어있음(키워드는 tf-idf로 추출함)* |
| x, idx2vocab = sent\_to\_word\_contexts\_matrix(  corpus,  windows=10, *#고려할 앞뒤 단어의 개수* min\_tf=0, *#10회 이하의 단어 무시* tokenizer=**lambda** x:x.split(), *# (default) lambda x:x.split(),* dynamic\_weight=**True**,  verbose=**True**) |
| glove = Glove(no\_components=100, learning\_rate=0.05, max\_count=30) glove.fit(x.tocoo(), epochs=10, no\_threads=4, verbose=**True**) |
| Glove sim weight  교육 , 별지 = -0.02058694480242076  교육 , 행성 = -0.09959176038179202  교육 , 별자리 = -0.0622146427693075  교육 , 모드 = 0.19419130804574383  교육 , 하늘 = -0.07256012241042203  교육 , 시계 = -0.1572574367173591  교육 , 88 = -0.15449500628184193  교육 , 태양 = -0.3370592458597273  교육 , 천문 = -0.22141501430126  교육 , 위치 = -0.06767654822523053  교육 , 천체 = -0.13263208250639957  교육 , 천구 = -0.47039230661842524  교육 , 라이브 = -0.021791404366475225  교육 , 중심 = -0.1506717523696949  교육 , 디스플레이 = -0.21917810334876767 |
| Final weight  별지, -0.0046344453502985645  행성, -0.021734176434966848  별자리, -0.013133676539511798  모드, 0.03854822821146845  하늘, -0.01351486997819721  시계, -0.025257985422831957  88, -0.023614474742377958  태양, -0.05131039611961595  천문, -0.032996780123347065  위치, -0.009261120587991776  천체, -0.017046808929879934  천구, -0.059988409399826186  라이브, -0.0027680244416006068  중심, -0.018175547208650153  디스플레이, -0.0234731198533837 |