

# 위키백과와 위키데이터를 이 용한 다국어 검색

2017. 12. 20

이만재 ([manjai48@gmail.com](mailto:manjai48@gmail.com))

한국위키미디어 협회 이사장/ 인공지능연구원 전문위원

# 들어가면서

- 자연어 처리 연구에 있어 데이터의 중요성
  - 인터넷을 통한 다량의 텍스트 정보 확보 가능
  - 분량보다 중요한 것은 데이터의 품질
  - 공개적으로 사용가능한 유일한 데이터

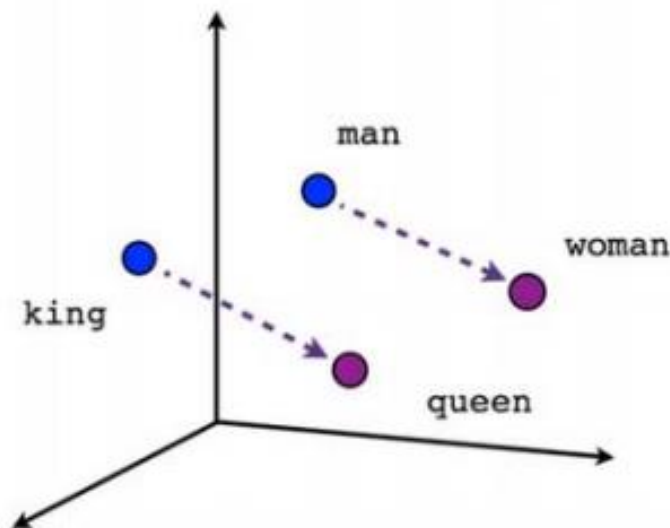
## 위키백과

# Word2vec

- Tomas Mikolov (2013)
  - 대량의 텍스트 데이터를 벡터화하는 방식
  - 모든 단어는 주변에 함께하는 단어로부터 의미를 알 수 있다
- 다음 단어로 알 수 있는 단어?
  - 대한민국, 교육인, 과학자, 인터넷, 서울대학교
  - 대한민국, 교육인, 과학자, 인터넷, 서울대학교, 전기공학, 스탠포드, 텍사스, ETRI, 삼보컴퓨터, 책임연구원, 디지털, 콘텐츠
- 모든 언어에 적용가능함 - 당연히 한글도 포함됨

# Word2vec / doc2vec

- 단어의 벡터 값을 안다면 관련어를 얻을 수 있음
  - king - man + woman => queen
  - 장비: 삼국지의 장비(?) 계축기의 장비(?)
- Doc2Vec : 특정 문서의 word2vec 의 합
- 문제는 어떻게 벡터를 구하는 가?



# Word2Vec을 얻는 방법

- [GloVe: Global Vectors for Word Representation](#)
  - Jeffrey Pennington(2014)
- Wikipedia 2014 + Gigaword 데이터 이용 Word vector 발표
  - 60억개의 token, 40만개 단어
  - 50, 100, 300. 300 차원의 벡터
  - 822MB 용량
- 문제는 영어단어만 제공된다는 것
  - 한글 처리를 위한 공개된 Word2vec은 아직 없음

# Word2vec / Doc2vec

Input:  
text

Lorem ipsum dolor  
sit amet, consete-  
tur sadipscing elitr,  
sed diam nonumy  
elitmod tempor  
invidunt ut labore  
et dolore magna  
aliquyam erat, sed  
diam voluptua. At  
vero eos et

  
train for  
each word  
a word vector

## word2vec

Model:



vector space:  
consists of **word vectors**  
for each word

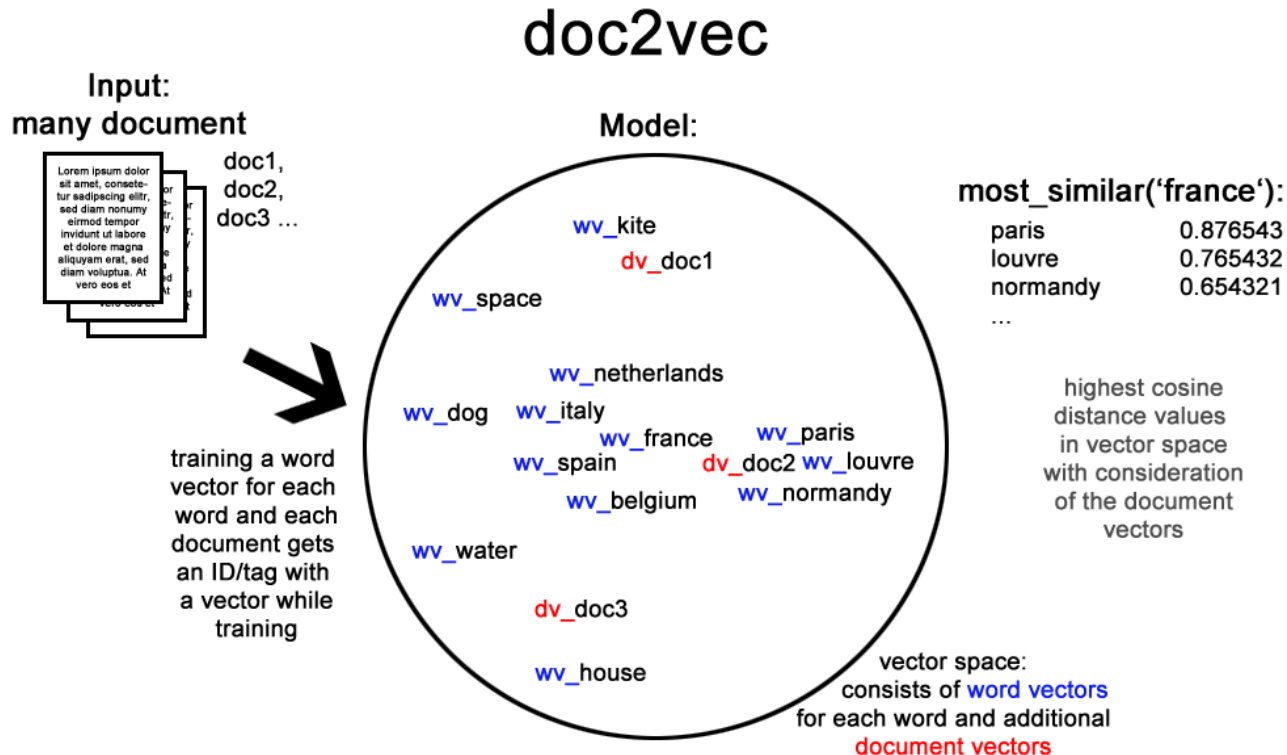
most\_similar('france'):

spain	0.678515
belgium	0.665923
netherlands	0.652428
italy	0.633130

highest cosine  
distance values  
in vector space  
of the nearest  
words

# Doc2Vec

- 단어 대신 문서/문장의 벡터를 계산
  - doc2vec 벡터와 word2vec 벡터는 같은 공간에 위치함



# 위키백과 텍스트 데이터 수집

- 가치있는 텍스트 말뭉치(Corpus)
  - 위키백과의 텍스트가 유일함( 한글 40만 문서)
  - 매월 모든 언어 버전의 주기적인 덤프 제공
- Wikimedia Downloads
  - <https://dumps.wikimedia.org/backup-index.html>
  - [kowiki-20171001-pages-meta-current.xml.bz2](https://dumps.wikimedia.org/ko/2017/20171001-pages-meta-current.xml.bz2) (613.9 MB)
  - 영어의 경우에는 ko 대신 en 사용
- WikiExtractor
  - <https://github.com/attardi/wikiextractor>
  - Xml 파일에서 본문 텍스트만 추출



# 위키백과 덤프 파일 구조

- page-meta-current.xml

- 표준 xml 포맷으로 위키백과 전체 내용 포함
- 한국어 : 2.9 GB, 영어 127 GB의 대용량 파일 ( 압축 해제시)

```
<mediawiki xmlns="http://www.mediawiki.org/xml/export-0.10/" ...  
<siteinfo>  
  <sitename>위키백과</sitename>  
  <dbname>kowiki</dbname>  
  <base>https://ko.wikipedia.org/wiki/% ...  
  <generator>MediaWiki 1.28.0-wmf.12</generator>  
  <case>first-letter</case>  
  <namespaces>  
    <namespace key="-2" case="first-letter">미디어</namespace>  
    <namespace key="-1" case="first-letter">특수</namespace>  
    <namespace key="0" case="first-letter" />  
    <namespace key="1" case="first-letter">토론</namespace>  
    <namespace key="2" case="first-letter">사용자</namespace>  
  ...
```

# Wiki Extractor 처리 후

<doc id="5" url="https://ko.wikipedia.org/wiki?curid=5" title="지미 카터">

지미 카터

제임스 얼 "지미" 카터 주니어(, 1924년 10월 1일 ~ )는 민주당 출신 미국 39번째 대통령 (1977년 ~ 1981년)이다. 지미 카터는 조지아 주 섬터 카운티 플레인스 마을에서 태어났다. 조지아 공과대학교를 졸업하였다. 그 후 해군에 들어가 전함.원자력.잠수함의 승무원으로 일하였다. 1953년 미국 해군 대위로 예편하였고 이후 땅콩·면화 등을 가꿔 많은 돈을 벌었다. 그의 별명이 "땅콩 농부" (Peanut Farmer)로 알려졌다.

...

</doc>

<doc id="9" url="https://ko.wikipedia.org/wiki?curid=9" title="수학">

수학

수학(數學)은 양, 구조, 공간, 변화 등 개념을 다루는 학문이다. 현대 수학은 형식 논리를 이용해서 공리로 구성된 추상적 구조를 연구하는 학문으로 여겨지기도 한다. 수학은 그 구조와 발전 과정에서는 자연과학에 속하는 물리학을 비롯한 다른 학문들과 깊은 연관을 맺고 있으나, 어느 과학의 분야들과는 달리, 자연계에서 관측되지 않는 개념들에 대해서까지 이론을 일반화 및 추상화시킬 수 있다는 차이가 있다. 수학자들은 그러한 개념들에 대해서 추측을 하고, 적절하게 선택된 정의와 공리로부터의 엄밀한 연역을 통해서 추측들의 진위를 파악한다.

...

</doc>

# Text Cleaning

- 문서 제목 정리 , 제목과 본문 분리
- 괄호 등 불필요한 내용 삭제
- period, comma, quotation mark는 delimiter로 사용
- 본문 내부의 영문은 남겨둠

**지미 카터** 제임스 얼 "지미" 카터 주니어는 민주당 출신 미국 39번째 대통령 이다. 지미 카터는 조지아 주 섬터 카운티 플레인스 마을에서 태어났다. 조지아 공과대학교를 졸업하였다. 그 후 해군에 들어가 전함·원자력·잠수함의 승무원으로 일하였다. 1953년 미국 해군 대위로 예편하였고 이후 땅콩·면화 등을 가꿔 많은 돈을 벌었다. 그의 별명이 "땅콩 농부" 로 알려졌다.

...

**수학** 수학은 양, 구조, 공간, 변화 등 개념을 다루는 학문이다. 현대 수학은 형식 논리를 이용해서 공리로 구성된 추상적 구조를 연구하는 학문으로 여겨지기도 한다. 수학은 그 구조와 발전 과정에서는 자연과학에 속하는 물리학을 비롯한 다른 학문들과 깊은 연관을 맺고 있으나, 어느 과학의 분야들과는 달리, 자연계에서 관측되지 않는 개념들에 대해서까지 이론을 일반화 및 추상화시킬 수 있다는 차이가 있다.

...

# 한국어 명사 추출

- 한국어 형태소 분석기
  - [konlpy](#) - 꼬꼬마 모듈 사용 (kkma)

```
from konlpy.tag import Twitter
kkma = Twitter()

...
for sentence in sentences:
    buf = " ".join(kkma.nouns(sentence))
    obuf = obuf + " " + buf
```

**지미 카터** | 지미 카터 제임스 얼 지미 카터 주니어 민주당 출신 미국 대통령 지미 카터 조지아 주 섬터 카운티 플레인스 마을 조지아 공과 대학교 그 후 해군 전함 원자력 잠수함 승무원 년 미국 해군 대위 예 이후 땅콩 면화 등 돈 그 별명 땅콩 농부

**수학** | 수학 수학 양 구조 공간 변화 등 개념 학문 현대 수학 형식 논리 공리 추상 구조 학문 기도 수학 그 구조 발전 과정 자연과학 물리학 다른 학문 연관 여느 과학 분야 달리 자연계 관측 개념 이론 일반 및 추상화 수 차이

# 영문 명사 추출

- nlTK 라이브러리 이용
  - 문장을 delimiter 로 분리
  - nltk.tokenize 함수

```
import nltk
...
sentences = texts.split('.')
text= ' '.join(sentences).lower()
tokens = nltk.word_tokenize(text)
tags = nltk.pos_tag(tokens)

# nouns 에 분석 결과 저장
# pos : part of speech = 단어의 품사, 여기서는 모든 명사 선택
nouns = [word for word, pos in tags if (pos == 'NN' or pos == 'NNP'
    or pos == 'NNS' or pos == 'NNPS')]
```

# Word2vec 만들기

- 사용목적에 따라 명사만 추출
  - 언어의 특성에 따라 국가마다 다른 형태소 분석
- Doc2vec을 위해 위키백과의 문서명을 기준으로 분류
- Gensim 텍스트 분석 라이브러리 사용
  - GPU 대신 multi CPU 기능을 활용
  - 단어사이의 거리와 빈도를 파라미터로 사용

# size: 벡터 차원 , window : 단어 사이의 거리, workers : thread 수

```
model1 = doc2vec.Doc2Vec(docs, size = 100, window = 6, min_count = 3, workers = 4)
```

# Word2vec 테스트

> 런던 - 영국 + 프랑스

0.6759	파리
0.5170	브뤼셀
0.4466	상젤리제
0.4382	제네바
0.4370	리옹
0.4336	베를린
0.4322	마드리드
0.4235	밀라노
0.4230	스톡홀름
0.4203	살롱

> 왕자 - 남자 + 여자

0.6482	공주
0.5717	왕비
0.5569	왕인
0.5565	왕녀
0.5520	왕세자
0.5500	왕태자
0.5406	부군
0.5235	왕세자빈
0.5191	대공비
0.5105	여왕

# Word2vec 테스트

## > 장비

0.6871	설비
0.6844	기기
0.6595	무전기
0.6488	탄약
0.6461	화기
0.6205	부품
0.6183	휴대
0.6130	중장비
0.5965	박격포
0.5952	이동식

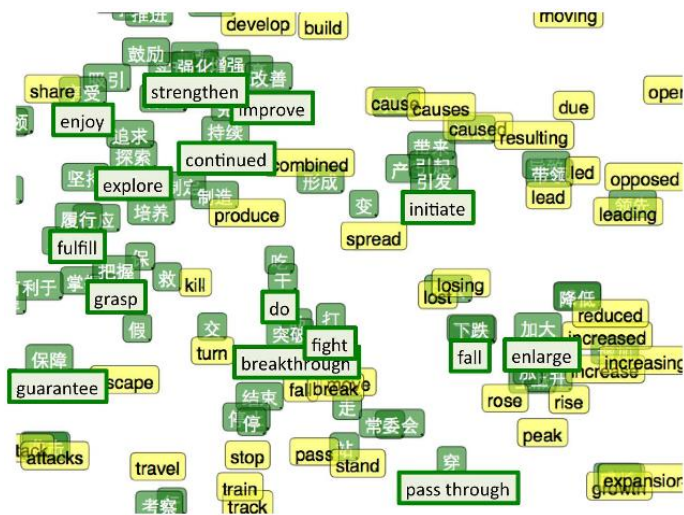
## > 장비 + 유비

0.7863	조운
0.7082	관우
0.7028	조조
0.6834	유비_관우
0.6821	손견
0.6815	조선_수군
0.6772	제갈량
0.6734	손권
0.6734	주유
0.6671	여포

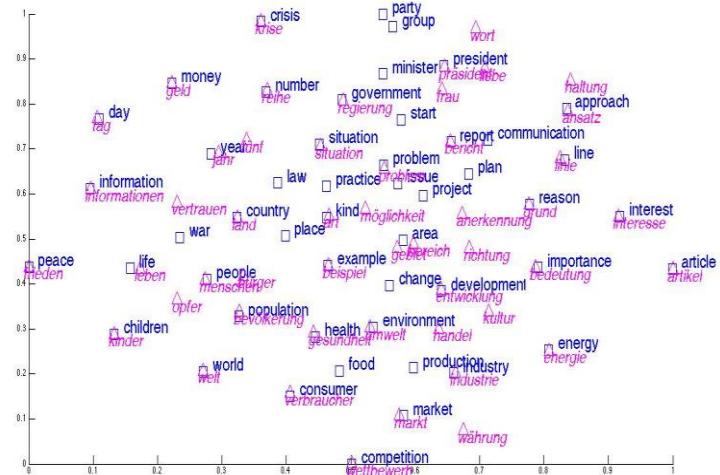


# Cross-lingual embedding model

- 서로 다른 언어의 embedding 정보 활용
- [Survey of cross-lingual embedding models](#)
- 다른 언어를 연결해 주는 데이터세트가 있어야 함



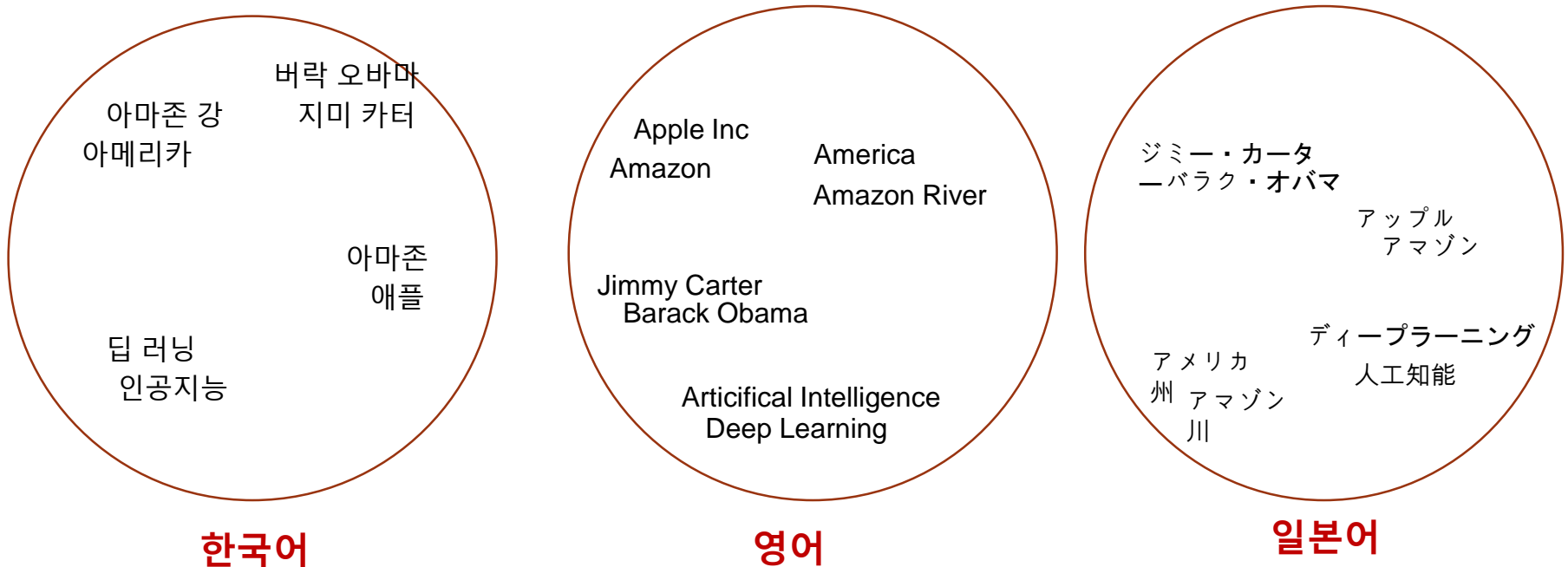
영어와 중국어



영어와 독일어

# 언어별 단어의 공간 배치

- 위키백과로부터 Word2Vec/Doc2Vec 추출
  - 비슷한 단어끼리 가까이 배치되나 언어마다 다른 위치를 차지함

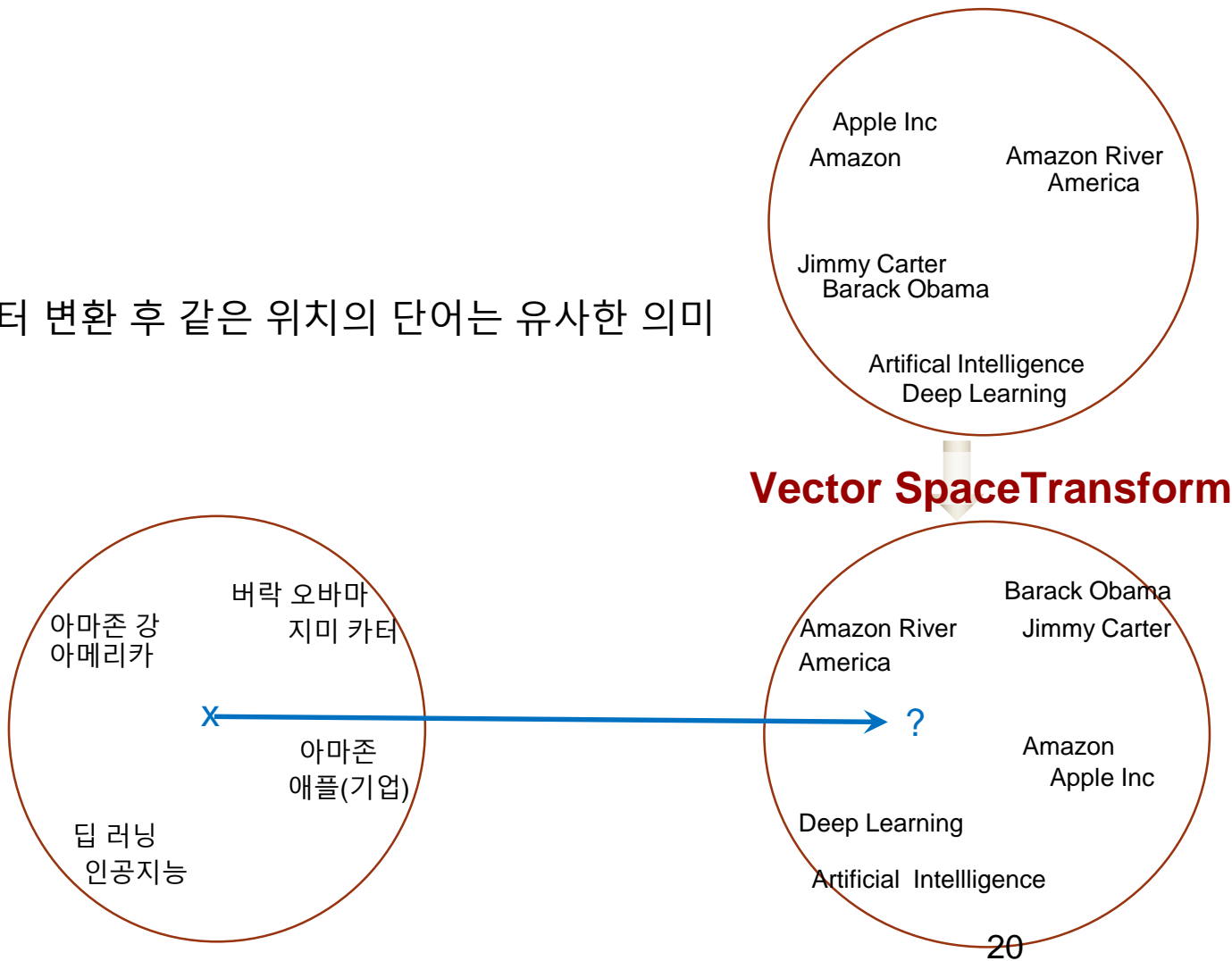


# 위키데이터의 Q번호

Q 번호	한국어	영어	일본어
Q312	애플(기업)	Apple Inc.	アップル
Q3783	아마존 강	Amazon River	アマゾン川
Q456120	아마존	Amazon	アマゾン
Q828	아메리카	Americas	アメリカ州
Q76	버락 오바마	Barack Obama	バラク・オバマ
Q23685	지미 카터	Jimmy Carter	ジミー・カーター
Q11660	인공지능	Artificial intelligence	人工知能
Q197536	딥 러닝	Deep learning	ディープラーニング

# WordCross (Word2Vec 공간) 변환

벡터 변환 후 같은 위치의 단어는 유사한 의미



# 위키백과와 위키데이터

위키백과	위키데이터
	
<ul style="list-style-type: none"> <li>● 사람이 읽기 쉽게 작성된 텍스트를 기반으로 함</li> </ul>	<ul style="list-style-type: none"> <li>● 기계를 위한 데이터 (궁극적으로는 사람을 위함)</li> </ul>
<ul style="list-style-type: none"> <li>● 언어별로 따로 작성 ko, en, ja, zh 등</li> </ul>	<ul style="list-style-type: none"> <li>● 언어 중립적이며 하나로 통일</li> </ul>
<ul style="list-style-type: none"> <li>● 언어별로 별도의 id 부여</li> </ul>	<ul style="list-style-type: none"> <li>● 모든 항목에 Q 번호라는 식별자 부여</li> </ul>
<ul style="list-style-type: none"> <li>● 위키백과의 언어별 설명은 같은 Q 번호의 경우 같은 내용을 다른 언어로 설명하는 것임</li> <li>● Cross-lingual embedding 언어 분석에 가장 좋은 데이터 소스</li> </ul>	

# 위키데이터의 언어별 표시

**label** — Douglas Adams (Q42) — item identifier

**description** — English writer and humorist  
Douglas Noël Adams | Douglas Noel Adams — aliases  
▶ In more languages

**property** — **educated at** — value

**rank** — **St John's College** — qualifiers

end time	1974
academic major	English literature
academic degree	Bachelor of Arts
start time	1971

**statement group** — **2 references** — opened references

stated in	Encyclopædia Britannica Online
reference URL	http://www.nndb.com/people/731/000023662/
original language of work	English
retrieved	7 December 2013
publisher	NNDB
title	Douglas Adams (English)

+ add reference

**Brentwood School**

end time	1970
start time	1959

**▶ 0 references** — collapsed reference

+ add (statement)

**레이블** — 더글러스 애덤스 (Q42) — 고유 식별자

**설명** — 영국의 작가  
더글러스 애덤스 | 더글러스 노엘 애덤스 — 다른 이름  
▶ 다른 언어

**속성** — **다녔던 학교** — 값

**등급** — **세인트존스 칼리지** — 한정자

종료된 날	1974
전공	영문학
학위	문학사
시작된 날	1971

**서술 목록** — **2 각주** — 출처 목록 (열린 상태)

다음에 서술되어 있음	브리태니커 백과사전 온라인
읽고한 URL	http://www.nndb.com/people/731/000023662/
저작물의 원어	영어
확인한 날짜	7 12월 2013
발행사	NNDB
제목	Douglas Adams (영어)

+ 각주 추가

**브렌트우드 스쿨**

종료된 날	1970
시작된 날	1959

**▶ 0 각주** — 출처 목록 (닫힌 상태)

+ 추가

# 위키데이터의 다국어 항목 추출

- [Wikidata Tutorial](#)
- 위키데이터에서 지정된 언어에 문서가 있는 항목만 추출

Q15|아프리카|Africa|Afrika|África|Afrique|アフリカ  
Q17|일본|Japan|Japan|Japón|Japon|日本  
Q18|남아메리카|South America|Südamerika|América del Sur|Amérique du Sud|南アメリカ  
Q20|노르웨이|Norway|Norwegen|Noruega|Norvège|ノルウェー  
Q21|잉글랜드|England|England|Inglaterra|Angleterre|イングランド  
Q22|스코틀랜드|Scotland|Schottland|Escocia|Écosse|スコットランド  
Q25|웨일스|Wales|Wales|Gales|Pays de Galles|ウェールズ

# 동일한 단어의 한영 데이터 명사 추출

- 같은 Q 번호를 사용하는 한영 문서의 제목 및 명사 모음
- 제목의 경우 한 단어 이상의 복합명사 가능

## Q23685|지미 카터|Jimmy Carter

|제임스 얼 지미 카터 주니어 민주당 출신 미국 대통령 지미 카터 조지아 주 섬터 카운티 플레인스 마을 조지아 공과 대학교 그 후 해군 전함 원자력 잠수함 승무원 년 미국 해군 대위 예 이후 땅콩 면화 등 돈 그 별명 땅콩 농부 로 년 조지아 주 상원 의원 선거 ...

|james carter jr politician president states member party governor prior election president carter life post-presidency peace prize work carter center carter states school submarines navy peanut farmer carter georgia ...

## Q395|수학|Mathematics

|수학 양 구조 공간 변화 등 개념 학문 현대 수학 형식 논리 공리 추상 구조 학문 기도 수학 그 구조 발전 과정 자연 과학 물리학 다른 학문 연관 여느 과학 분야 달리 자연계 관측 개념 이론 일반 및 추상화 수 차이 수학자 개념 추측 선택 정의 공리 로부터 엄밀 연역 ...

|mathematics study topics quantity structure space change range views mathematicians philosophers scope definition mathematics mathematicians patterns conjectures mathematicians truth falsity conjectures proof structures ...



# 복합명사 처리 결과

## Q23685|지미 카터|Jimmy Carter

|제임스 얼 지미\_카터 주니어 민주당 출신 미국 대통령 지미\_카터 조지아 주 섬터 카운티 플레인스 마을 조지아 공과\_대학교 그 후 해군 전함 원자력\_잠수함 승무원 년 미국 해군 대위 에 이후 땅콩 면화 등 돈 그 별명 땅콩 농 부 로 년 조지아 주 상원\_의원 선거

|james carter jr politician president states member party governor prior election president carter life post-presidency peace prize work carter center carter states school submarines navy peanut farmer carter georgia state senate carter election governor georgia governor carl sanders primary governor outside georgia start campaign carter nomination election carter president gerald\_ford election day office carter evaders

## Q395|수학|Mathematics

|자연과학 물리학 다른 학문 연관 여느 과학 분야 달리 자연계 관측 개념 이론 일반 및 추상화 수 차이 수학자 개념 추측 선택 정의 공리 로부터 엄밀 연역 추측 진위 수학 숫자 세기 계산 측정 및 물리 대상 모양 움직임 추상화 이 논리 추론 기본 개념 고대 이집트 메소포타미아 고대 인도 고대 중국 및 고대 그리스 수학 책 수 유클리드 원론 엄밀 논증 발전 그 뒤 도 계속 세기 르네상스 수학 발전 과학 방법 상호\_작용 혁명 연구 |mathematics study topics quantity structure space change range views mathematicians philosophers scope definition mathematics mathematicians patterns conjectures mathematicians truth\_falsity conjectures proof structures models phenomena reasoning insight predictions nature use abstraction logic mathematics counting calculation measurement study shapes motions objects mathematics activity records research problems years centuries inquiry arguments mathematics euclid\_elements work peano david\_hilbert

# Vector Transform Model

- 단순 regression model
  - loss: cosine\_proximity
  - metrics : mae

```
model = models.Sequential()  
model.add(layers.Dense(5000, activation='relu', input_shape=(DIM,)))  
model.add(layers.Dense(5000, activation='relu'))  
model.add(layers.Dense(5000, activation='relu'))  
model.add(layers.Dense(DIM))
```

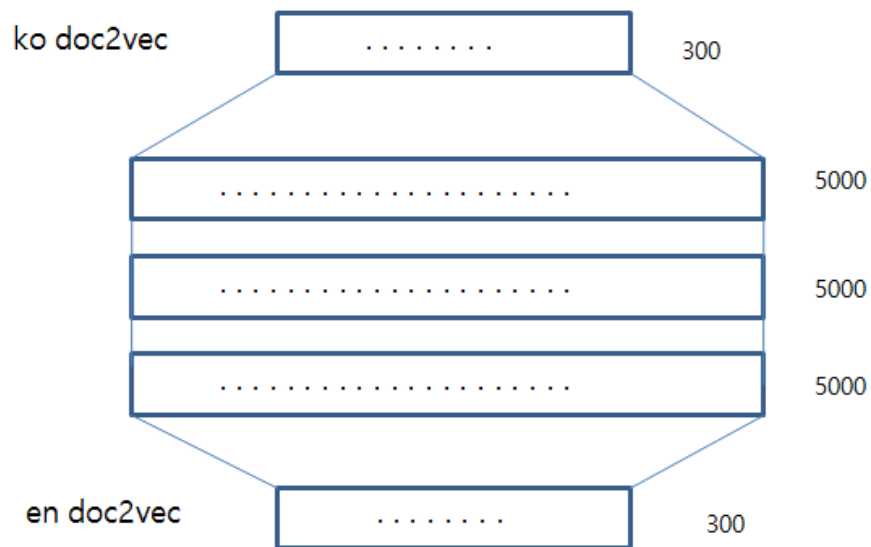
```
model.summary()  
model.compile(optimizer='rmsprop', loss='cosine_proximity', metrics=['mae'])
```

# Simple Neural Network Model

- ko - en, en - ko 변환하는 신경망 구성
- Multiple Regression

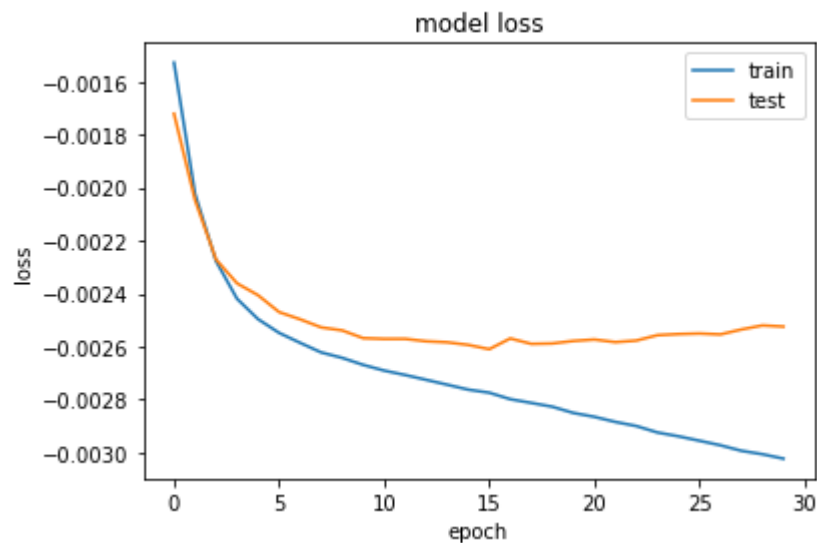
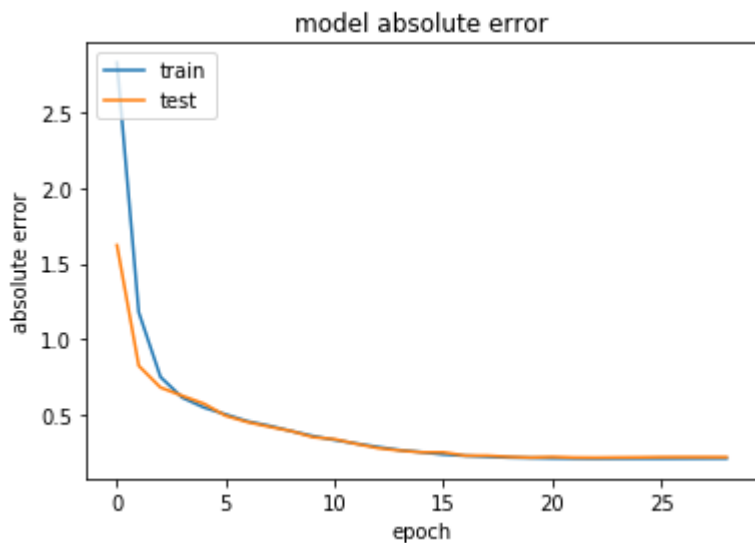
Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 5000)	1505000
dense_2 (Dense)	(None, 5000)	25005000
dense_3 (Dense)	(None, 5000)	25005000
dense_4 (Dense)	(None, 300)	1500300

=====  
Total params: 53,015,300  
Trainable params: 53,015,300  
Non-trainable params: 0



# Deep Learning Model validation

- 더 많은 층을 쌓아도 개선 없었음
- 한영변환에 ko2en, en2ko의 두 모델 필요



# word2vec 과 doc2vec 초기 결과

## ㉔ 임진왜란

0.8313	왜구
0.8170	칠천량 해전
0.8087	고니시 유키나가
0.8015	나라 시대
0.7996	도요토미 히데요시
0.7721	원나라의 일본 원정
0.7696	명나라
0.7695	이순신
0.7658	고려
0.7609	조선

## ㉕ 임진왜란

0.8289	정유재란
0.7997	병자호란
0.7713	중일전쟁
0.7618	청일전쟁
0.7395	중일 전쟁
0.7293	오닌 난
0.7188	러일전쟁
0.7140	청일 전쟁
0.7123	러일 전쟁

WordCross Demo

keyword  
Barack Obama Find

Korean	English	Japanese	Chinese
1.0000  버락 오바마	1.0000  Barack Obama	1.0000  バラク・オバマ	1.0000 贝拉克·奥巴马
0.7945  힐러리 클린턴	0.7563  George W. Bush		0.8243 希拉里·克林顿
0.7553  밋 롬니	0.7535  Joe Biden		0.7948 比尔·克林顿
0.7207  1960년 미국 대통령 선거	0.7358  Nancy Pelosi		0.7894 骆家辉
0.7069  워런 G. 하딩	0.7085  Newt Gingrich		0.7856 约翰·肯尼迪
0.7054  제럴드 포드	0.7082  Bill Clinton		0.7850 唐纳德·拉姆斯菲尔德
	0.7035  Tim Kaine		0.7814 泰德·肯尼迪
			0.7802 1980년美国总统选举

WIKIPEDIA

## 버락 오바마

미합중국 제44대 대통령.

오바마는 여기로 연결됩니다. 다른 뜻에 대해서는 오바마 (동음이의) 문서를 참조하십시오.

이 문서는 버락 오바마 시니어의 아들인 미국의 대통령에 관한 것입니다. 미국 대통령 버락 오바마의 아버지인 케냐의 경제학자에 대해서는 버락 오바마 시니어 문서를 참조하십시오.


**버락 후세인 오바마 주니어**(영어: Barack Hussein Obama, Jr., 문화어: 바라크 후세인 오바마, 1961년 8월 4일 ~ , 미국 하와이 주 호놀룰루 출생)는 미국의 정치인으로 제44대 대통령이다. 케냐 출신의 아버지와 유령계 백인 어머니 사이에서 태어난 몰라도로, 2008년 미국 대통령 선거에 민주당 소속으로 출마, 미국 최초로 아프리카계 미국인이자 하와이 출신으로 대통령에 당선되었으며, 2009년 1월 20일부터 임기를 시작하였다. 2005년 1월 알리노

WIKIPEDIA

## Barack Obama

"Barack" and "Obama" redirect here. For other uses, see [Barack \(disambiguation\)](#) and [Obama \(disambiguation\)](#).

**Barack Hussein Obama II** (/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/ <sup>ⓘ</sup> <sup>ⓘ</sup> listen);<sup>[1]</sup> born August 4, 1961) served as the **44th President of the United States** from 2009 to 2017. The first **African American** to assume the presidency, he previously served as a **United States Senator**, representing **Illinois** from 2005 to 2008. He also served in the **Illinois State Senate** from 1997 to 2004.




ウィキペディア

## バラク・オバマ

ページの問題点  
アメリカ合衆国の政治家、第44代大統領

**バラク・フセイン・オバマ2世** (**Barack Hussein Obama II**<sup>[2]</sup> /bəˈrɑːk huːˈseɪn oʊˈbɑːmə/ (<sup>ⓘ</sup> 音声ファイル), 1961年8月4日 - ) は、アメリカ合衆国の政治家である。民主党所属。上院議員 (1期)、イリノイ州上院議員 (3期)、第44代アメリカ合衆国大統領を歴任した。

バラク・オバマ  
**Barack Obama**




WIKIPEDIA


## 贝拉克·奥巴马

第44任美国总统

奥巴马重定向至此。此条目介绍的是第44任美国总统。关于奥巴马的其他资讯，请参阅奥巴马 (消歧义)。

**巴拉克·奥巴马**（英語：Barack Obama，1961年8月4日－），全名**巴拉克·侯赛因·奥巴马二世**（英語：Barack Hussein Obama II，美國/bəˈrɑːk huːˈseɪn oʊˈbɑːmə/），是美国政治人物，从2009年至2017年任第44任美国总统。他是第一位美国非裔总统，也是第一位出生于美国本土之外的美国总统。在这之前，他从2005年至2008年代表伊利诺伊州担任美国联邦参议员，从1997年至2004年担任伊利诺伊州参议员。

 **巴拉克·奥巴马**  
Barack Obama



keyword 이순신 Find

Korean	English	Japanese	Chinese
1.0000 이순신	1.0000 Yi Sun-sin	1.0000 李舜臣	1.0000 李舜臣 (朝鮮)
0.8322 송희립	0.8218 Won Gyun	0.7224 順天城の戦い	0.8604 楊鎬
0.8280 칠천량 해전	0.7673 The Admiral: Roaring Currents	0.7216 閑山島海戦	0.8599 不滅的李舜臣
0.8219 원균	0.7617 Battle of Noryang	0.7073 劉テイ (明)	0.8501 天命 (后金)
0.7886 안위 (1563년)	0.7480 Battle of Chilcheollyang	0.7054 鳴梁海戦	0.8490 后梁
0.7867 이순신 (1553년)		0.7016 蔚山城の戦い	0.8432 宋庆

WIKIPEDIA 이순신

조선의 무관으로, 시호는 충무공이다.

다른 뜻에 대해서는 이순신 (동음이의) 문서를 참조하십시오.

이순신(李舜臣, 1545년 4월 28일<sup>[2]</sup> ~ 1598년 12월 16일)은 조선 중기의 무신이다. 본관은 덕수(德水), 자는 여해(汝諧), 시호는 충무(忠武)이며, 한성 출신이다. 문반 가문 출신으로 1576년(선조 9년) 무과(武科)에 급제<sup>[3]</sup>하여 그 관직이 동구비보 권관, 훈련원 봉사, 발포진 수군만호, 조산보 만호, 전라좌도 수군절도사를 거쳐 정헌대부 삼도수군통제사에 이르렀다.

이순신 李舜臣

WIKIPEDIA Yi Sun-sin

Page issues

This is a Korean name; the family name is Yi.

**Yi Sun-sin** (**Hunminjeongeum**: 이순신; **Hanja**: 李舜臣; April 28, 1545 – December 16, 1598) was a Korean naval commander famed for his victories against the Japanese navy during the *Imjin war* in the *Joseon Dynasty*, who became an exemplar of conduct to both the Koreans and Japanese.<sup>[1]</sup> Despite the fact that he had no prior naval training, Admiral Yi was never defeated at sea, and military historians have placed him on par with **Admiral Horatio Nelson** as one of the greatest naval commanders in history.<sup>[2][3][4]</sup> His title of *Samdo Sugun Tongjesa* (삼도 수군 통제사; 三道水軍統制使), literally meaning "Naval Commander of the

ウィキペディア 李舜臣

この項目では、李氏朝鮮の武将について記述しています。その他の同名の人物については「李舜臣 (曖昧さ回避)」をご覧ください。

李舜臣 (日本語読み: りしゅんしん、朝鮮語読み: イ・スンシン、朝: 이순신, 1545年4月18日(明暦: 嘉靖24年3月8日) - 1598年12月16日(明暦: 万暦26年11月19日))は、李氏朝鮮の將軍。字は汝諧(ヨハ、여해)。文祿・慶長の役において朝鮮水軍を率いて日本軍と戦い活躍したとされるが、党争<sup>[1]</sup>の影響でしばしば讒言を受け、また、抗命したことで兵卒に落とされするなど浮き沈みが激しく、元均の戦死で復権したが、露梁海戦で戦死した。死後に贈られた諡は忠武公(충무공)。

WIKIPEDIA 李舜臣 (朝鮮)

李氏朝鮮時將領

本文介紹的是朝鮮的將領，關於其他同名人物，請見「李舜臣」。

李舜臣 (韓語: 리순신/이순신, 1545年4月28日 - 1598年12月16日), 字汝諧, 本貫德水, 朝鮮京畿開丰(今开城)人, 李氏朝鮮時期名將, 諡号忠武, 死後追贈領議政, 1604年朝鮮宣祖(李昖)追封其為孝忠仗義迪毅協力宣武功臣<sup>[1][2]</sup>, 同列者有權慄和元均, 加贈德豐府院君。

李舜臣



keyword

조선

Find

Korean

English

Japanese

Chinese

- 1.0000 조선
- 0.8519 유비
- 0.8078 손권
- 0.7977 원소 (후한)
- 0.7927 관우
- 0.7901 위 (삼국)
- 0.7884 위숙

- 1.0000 Cao Cao
- 0.8463 Xiahou Dun
- 0.8281 Ma Chao
- 0.8229 Liu Bei
- 0.8207 Zhao Yun
- 0.8179 Deng Ai
- 0.8166 Sima Yi

- 1.0000 曹操
- 0.8280 劉備
- 0.8119 諸葛亮
- 0.8044 荀イク
- 0.7995 魏 (三国)
- 0.7804 張飛
- 0.7775 蜀漢

- 1.0000 曹操
- 0.9491 周瑜
- 0.9345 祢衡
- 0.9301 荀彧
- 0.9289 刘禅
- 0.9289 孔融
- 0.9282 郭嘉
- 0.9249 曹丕

WIKIPEDIA

### 조선

문서 미수

자는 맹덕(孟德). 중국 후한 말기의 정치가이자, 군인, 시인이다.

다른 뜻에 대해서는 조선 (동음이의) 문서를 참조하십시오.

조위 태조 무황제 조조(曹魏 太祖 武皇帝 曹操, 155년 음력 6월 3일 ~ 220년 음력 1월 23일)는 중국 후한 말기의 정치가이자, 군인이며 시인이다. 자는 맹덕(孟德)이며, 훗날 위가 건국된 이후 추증된 묘호는 태조(太祖), 시호는 무황제(武皇帝)이다. 위나라의 초대 황제는 조조의 아들 조비이지만 실질적으로 위 건국의 기틀을 마련한 것은 조비의 아버지 조조이다. 후한이 그 힘을 잃어가던 시기에 비상하고 탁월한 재능으로 두각을 드러내, 여러 제후들을 연달아 격파하고 중국대륙의 대부분을 통일하여, 위나라가 세워질 수 있는 기틀을 닦았다. 조조는 삼국지인 영웅들 가운데 패왕(覇王)으로 불린다.

WIKIPEDIA

### Cao Cao

For other uses, see Cao Cao (disambiguation).

This is a Chinese name; the family name is Cao.

**Cao Cao** ([tʃhɑu tʃhɑu]; Chinese: 曹操; 155 – 15 March 220),<sup>[1]</sup> courtesy name **Mengde**, was a Chinese warlord and the penultimate **Chancellor of the Eastern Han dynasty** who rose to great power in the **final years of the dynasty**. As one of the central figures of the **Three Kingdoms** period, he laid the foundations for what was to become the state of **Cao Wei** and was posthumously honoured as "Emperor Wu of Wei". He is often portrayed as a cruel and merciless tyrant in subsequent literature; he has also been praised as a brilliant ruler and military genius who treated his subordinates like his family. During the fall of the Eastern Han dynasty, Cao Cao was able to secure the most populated and prosperous cities of the


ウィキペディア

### 曹操

後漢の魏王

曹操 (そう そう、拼音: Cáo Cǎo、永寿元年(155年) - 建安25年1月23日(220年3月15日)) は、後漢末期の武将、政治家。詩人、兵法家としても業績を残した。字は孟徳(もうとく)、幼名は阿瞞、また吉利。沛国譙県(現在の安徽省亳州市、また河南省永城市という説もある)の出身。

魏武帝



WIKIPEDIA

### 曹操


中国东汉末年的军事家、政治家和诗人

曹操 (155年–220年3月15日), 字孟德, 小名吉利, 小字阿瞞, 沛国譙 (今安徽省亳州市) 人。東漢末年著名軍事家、政治家和詩人, 三國時代魏國奠基者和主要締造者。曹操在世时官至漢丞相, 爵至魏王, 谥号武王。<sup>[2]</sup>其子曹丕稱帝後, 追尊其為武皇帝, 廟號太祖。

魏武帝

漢丞相魏王, 領冀州牧

前任: 無  
繼任: 曹丕





# 맺음말

- 위키백과/위키데이터를 이용한 언어간 변환
  - 모든 언어쌍에 적용가능함
  - 한국어, 영어, 일어, 중국어에 적용
  - 개별언어 형태소 분석기 활용
- 위키백과/위키데이터의 중요성
  - Sum of all human knowledge
  - 지식 제공 이상의 기여
- Next Personal Project : [Wikiplorer](#) 기능개선