

# 2021학년도 2학기 언어와 컴퓨터

## 제1강 강의 소개

박수지

서울대학교 인문대학 언어학과

2021년 9월 1일 수요일

## 오늘의 목표

- 〈언어와 컴퓨터〉에서 한 학기 동안 다룬 내용을 개략적으로 설명할 수 있다.
- Word2Vec 데모에 적절한 예시를 넣어 실행할 수 있다.
- Google Colab에서 파이썬 코드를 입력하고 실행해 본다.

# 규칙 기반 텍스트 처리

## 사실

(현대의) 컴퓨터는 0과 1로 이루어진 신호를 처리한다.

## 문제

- 컴퓨터는 ㄱ, ㄴ, ㄷ 등을 모른다.
  - ㄱ을 어떻게 0과 1의 연쇄로 대응시키는가?
  - 텍스트에서 한글만 남기고 싶다면 어떻게 하는가?
- 사람과 컴퓨터가 처리하는 어절의 정의도 다르다.
  - 문장 성분의 최소 단위 vs. 공백 문자로 분리되는 문자열
- 사람과 컴퓨터가 처리하는 문자의 범위도 다르다.
  - “ ” 안에 어떤 문자가 들어있는가 — 없음 vs. 공백 문자

# 규칙 기반 텍스트 처리

활용: ELIZA 챗봇

## 예시

Human Men are all alike.

ELIZA IN WHAT WAY

Human They are always bugging us about something or other.

ELIZA CAN YOU THINK OF A SPECIFIC EXAMPLE

Human Well, my boyfriend made me come here.

ELIZA YOUR BOYFRIEND MADE YOU COME HERE

Human He says I' m depressed much of the time.

ELIZA I AM SORRY TO HEAR YOU ARE DEPRESSED

# 규칙 기반 텍스트 처리

활용: ELIZA 챗봇

- Weizenbaum, Joseph. (1966). “ELIZA – A Computer Program For the Study of Natural Language Communication Between Man and Machine.” Communications of the ACM 9(1), p.p. 36-45.
- Eliza (elizabot.js) <https://www.masswerk.at/elizabot/>

# 통계적 자연언어처리

## 예시

- 이 단어 다음에 저 단어가 나올 **확률** (활용: 자동 문장 완성, 오타 수정)
- 주어진 문장이 표현하는 감정이 긍정적일 **확률** (활용: 감정 분석)

## 목표

데이터로부터 확률 추정하기 — 통계 모형, 기계학습

## 할 일

- 1 통계 개념을 이해하고 파이썬에서 구현한다.
- 2 텍스트로 된 데이터를 컴퓨터가 읽을 수 있는 형태로 가공한다.

# 통계적 자연언어처리

## ■ Sentiment Analysis - AllenNLP - Demo

<https://demo.allennlp.org/sentiment-analysis>

Input

hot pizza

Run >

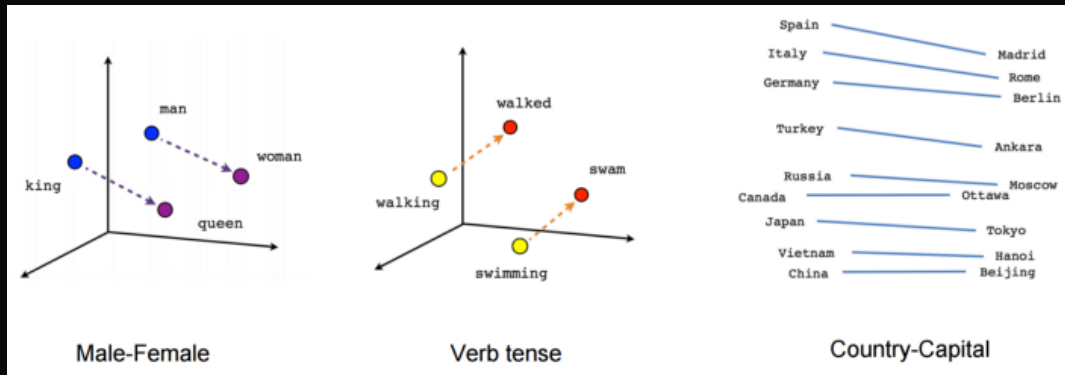
Answer

The model is quite sure the sentence is **Positive**. (96.8%)

# 단어의 의미

## Word2Vec

의미를 수치(벡터)로 표현하기 → 단어 사이의 의미 관계를 계산할 수 있다.



<https://towardsdatascience.com/>



# 단어의 의미

## Word2Vec

### 예시

단어쌍 사이의 공통된 의미 관계

- 아빠 : 엄마 = 할아버지 : 할머니 (= 남자 : 여자)
- [벡터공간에서] 아빠 - 엄마 = 할아버지 - 할머니

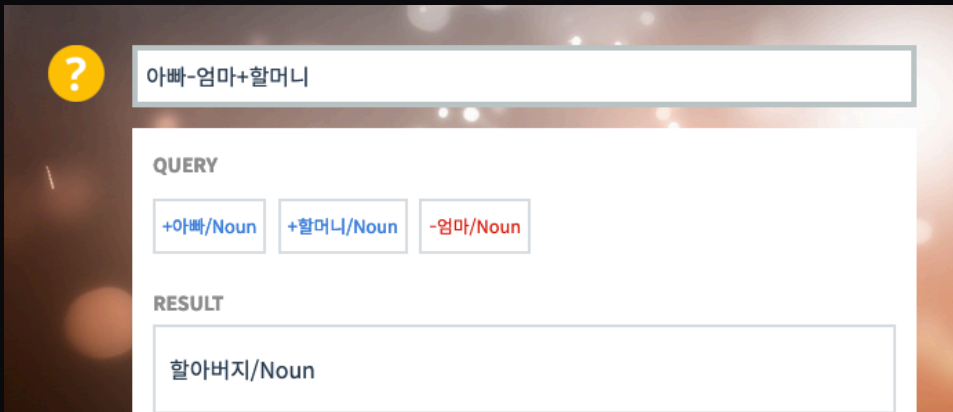
$$A : B = X : D \quad \Rightarrow \quad X = A - B + D$$

# 단어의 의미

## Word2Vec

아빠 : 엄마 = 할아버지 : 할머니 관계가 성립한다는 것을 어떻게 확인할 수 있는가?

- Korean Word2Vec <https://word2vec.kr>



The screenshot shows the Korean Word2Vec web application interface. At the top left, there is a yellow circle with a question mark. Below it, a search bar contains the text "아빠-엄마+할머니". Under the search bar, the word "QUERY" is displayed. Below the query, there are three buttons: "+아빠/Noun", "+할머니/Noun", and "-엄마/Noun". Below these buttons, the word "RESULT" is displayed. Below the result, there is a box containing the text "할아버지/Noun".

# 단어의 의미

Word2Vec

## 다양한 의미 관계

- 나라-국기
- 주군-책사
- 반의 관계
- 동사 활용

## 데이터의 한계

- 나라-대통령

# 질문

# Google Colaboratory

<https://colab.research.google.com>

- 파이썬과 인사하기
- 파이썬을 계산기처럼 사용하기