

파이토치로 배우는  
자연어 처리

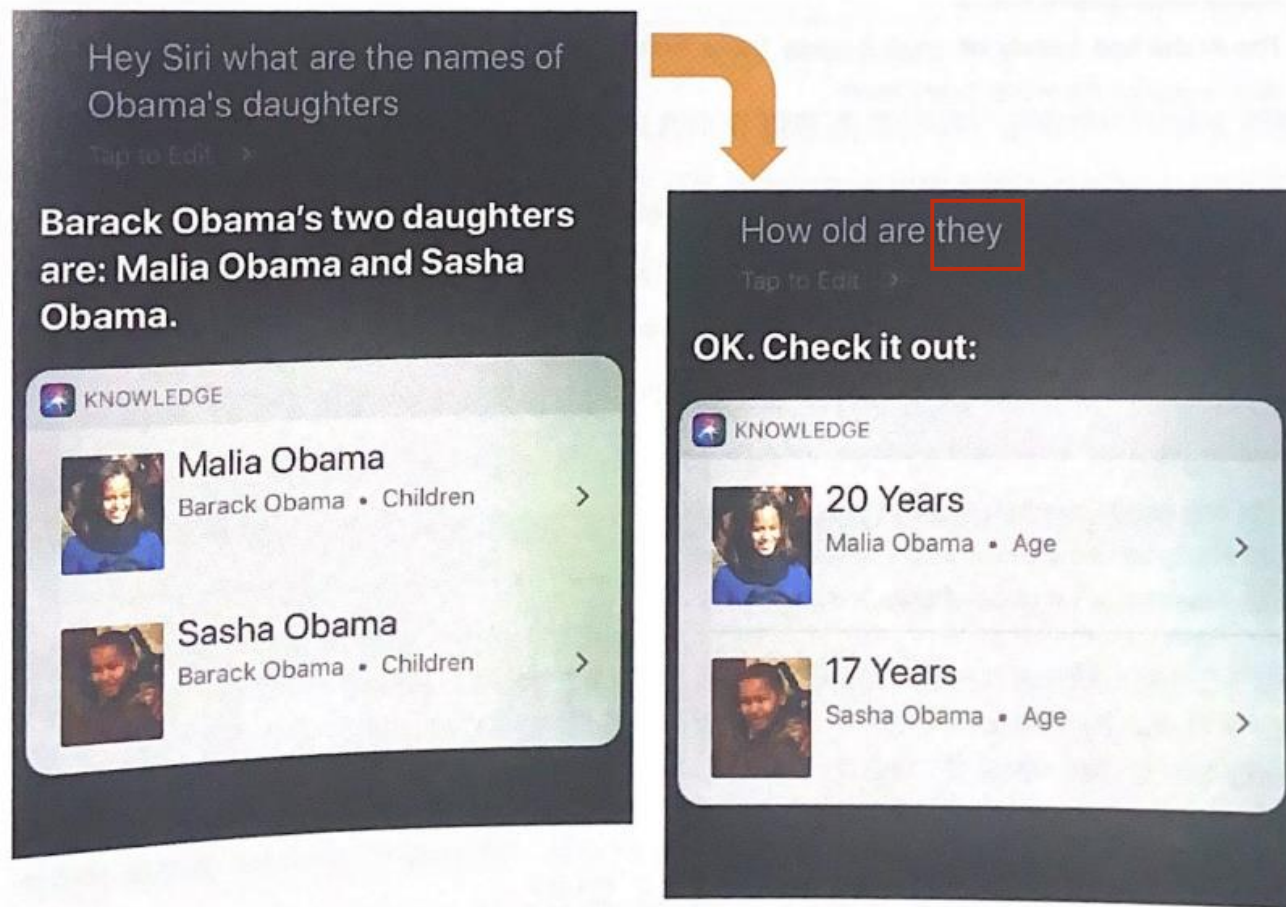
# 9장. 고전 모델, 최신 모델, 더 배울 것들

송선영

# CONTENTS

1. 전통적인 NLP 주제
2. 최근 NLP 모델
3. NLP 시스템의 설계 고려 사항

## 1) 대화식(상호작용) 시스템



## 2) 담화 분석

- 텍스트 문서의 부분-전체 특징을 이해하는 것

예시	담화 관계
GM officials want to get their strategy to reduce capacity and the workforce in place <u>before</u> those talks begin.	Temporal.Asynchronous. Precedence
But that ghost wouldn't settle for words, he wanted money and people—lots. <u>So</u> Mr. Carter formed three new Army divisions and gave them to a new bureaucracy in Tampa called the Rapid Deployment Force.	Contingency.Cause.Result
The Arabs had merely oil. <u>Implicit=while</u> These farmers may have a grip on the world's very heart	Comparison.Contrast

- (a) The dog chewed the bone. It was delicious.  
(b) The dog chewed the bone. It was a hot day.  
(c) Nia drank a tall glass of beer. It was chipped.  
(d) Nia drank a tall glass of beer. It was bubbly.

그림 9-2 대용어 복원에서 일어날 수 있는 몇 가지 문제. (a)에서 'it'은 dog 또는 bone을 의미하나요? (b)에서 'it'의 의미는 둘 다 아닙니다. (c)와 (d)에서 'it'은 각각 glass와 beer를 의미합니다. 유리잔보다 맥주에 거품이 있을 가능성이 높음을 아는 것은 참조 대상을 찾는 데 중요한 역할을 합니다(선택적 선호도(selectional preference)).

- Beijing imposed trade tariffs in response to tariffs on Chinese

-> 담화 파싱

: 두 문장의 문맥적인 연관성을 이해하는 것

-> 담화 이해

: 대용어 복원과 환유 감지 같은 문제를 해결하는 것

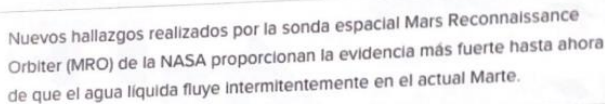
## 3) 정보 추출과 텍스트 마이닝

## 4) 문서 분석과 문서 추출

- 전통적인 NLP 연구에서 딥러닝 패러다임으로 전환
- 모델의 조합
  - 주어진 문제를 해결하는 데 여러 가지 모델을 조합
  - 작업에 따라 여러 구조를 조합할 수 있다는 점은 딥러닝의 가장 강력한 기능이며 작업을 성공적으로 수행하는 열쇠
- 시퀀스를 위한 합성곱
  - 합성곱 연산으로만 시퀀스를 모델링
  - 디코딩 단계에서는 전치 합성곱 연산 사용
- 필요한 것은 어텐션이 전부이다
  - 일반적으로 RNN과 CNN을 사용해 모델을 만듦
- 전이 학습
  - 한 작업에서 학습한 표현을 사용해 다른 작업의 학습을 향상시키는 방법
  - 최근 사전 훈련된 단어 벡터를 사용한 전이 학습 기법이 많이 사용됨

- ## 1) 온라인 시스템 vs 오프라인 시스템

- ## 2) 상호작용 시스템 vs 비상호작용 시스템



- ### 3) 유니모달 시스템 vs 멀티모달 시스템

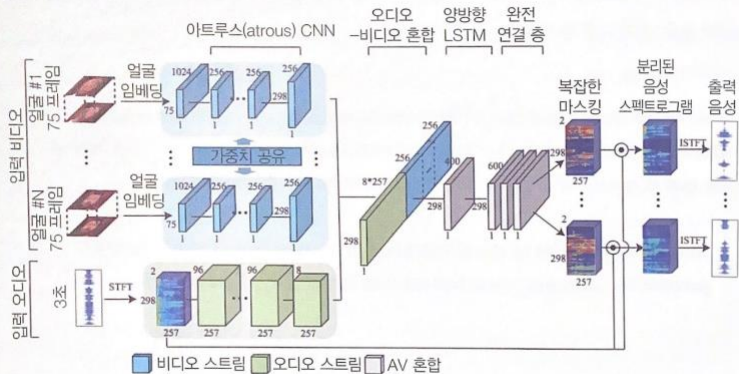
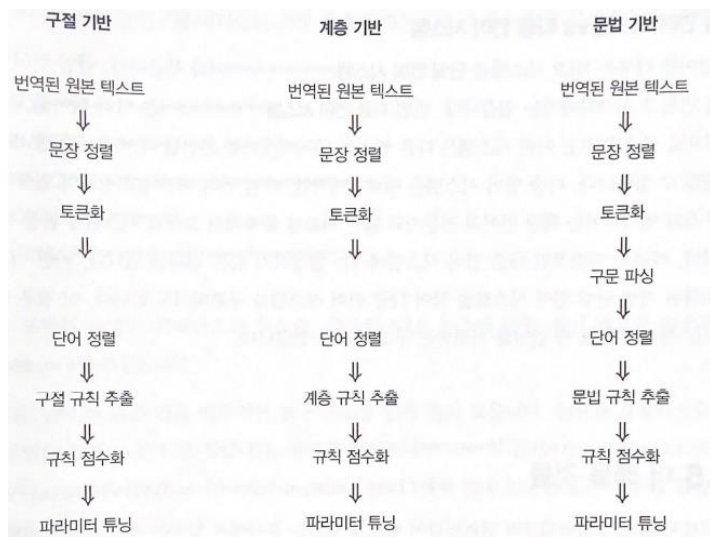


그림 9-5 카테일파티 문제와 같은 어려운 문제를 푸는 데 음성과 비디오 특성을 함께 사용하는 멀티모달 시스템(출처: 애플라트 등의 2018년 논문).

## 4) 엔드 투 엔드 시스템 vs 분할 시스템



## 5) 폐쇄형 도메인 시스템 vs 개방형 도메인 시스템

## 6) 단일 언어 시스템 vs 다중 언어 시스템

끝