

# 토키토키 기획안

## 1. 토키토키(Talkie Talkit)란?

토키토키는 한국어 회화를 배우고자 하는 외국인에게 도움을 주는 서비스입니다. 한국어는 하나의 어근(혹은 어간)에 하나 이상의 접사가 붙어 단어의 의미를 구성하는 특성을 가지고 있습니다. 예를 들어, “잡히셨겠더라”라는 어절은 어근 “잡-”, 파생 접사 “-히-”와 굴절 접사 “-(으)시-” + “-었-” + “-겠-” + “-더라”로 구성되어있습니다. 이를 형태소별로 끊어서 피동, 주체 높임, 과거, 추측, 전달의 의미를 판단하는 것은 매우 어려운 일입니다. 그렇기 때문에 한국어를 자연스럽게 습득할 수 있는 환경을 제공하여, 별도의 문법 공부 없이 유창한 회화 실력을 키워줄 수 있게 하는 것이 본 서비스의 목표입니다.

Talkie-Talkit는 Talkative(말이 많은)에서 착안한 단어입니다. 한국어 회화에 유창하지 않은 사람이라도, 본 서비스를 통해서 Talkie(Talkative)가 될 수 있다는 것을 의미합니다. 또한 Talk it(talkit)이라는 단어도 의미하고 있습니다. 질문이 주어지고 그에 맞는 의미를 지닌 문장을 말하면서 한국어를 습득하는 웹서비스 특성상, 주어진 질문에 맞는 한국어를 이야기 하라(talk it)는 의미를 내포하고 있습니다.

토키토키는 복잡한 한국어 문법체계를 학습하여 유창한 한국어를 구사하게 도와주는 시스템이 아닌, 사용자가 직접 말하면서 문맥의 의미를 파악하고, 보다 더 자연스러운 어순을 스스로 습득할 수 있게 도와주는 한국어 도우미입니다. 서비스는 웹을 통해 제공되며, 네이버 API를 통해 말하기와 듣기 기능을 제공합니다. 또한 여행 용어(음식점, 호텔예약 등), 필수 한국어 등 테마별로 회화들을 제공하여 보기 더욱 쉽게 하였습니다.

## 2. Q2

### 가. 목적

미국의 언어학자 스티븐 크라센(Stephen Krashen)은 Second Language Acquisition and Second Language Learning에서 습득-학습 구분 가설 (The acquisition-learning distinction)을 주장하였습니다. 어린아이가 모국어를 배우기 위한 과정은 습득(acquisition)의 대표적인 예시이고, 제 2 외국어를 배우는 학습자들의 과정은 학습(learning)의 예시입니다. 크라센은 체계적인 문법학습을 통하여 배운 외국어를 순간적인 발화에 사용하기 힘들다고 주장합니다. 순간적인 발화에서 자신이 표현하고자 하는 바를 효과적으로 전달하기 위해서는 습득의 과

정을 거쳐 익힌 언어가 더 도움이 됩니다. 예상하지 못 한 질문을 받게 되더라도 순발력 있게, 융통성 있게 질문에 대한 답변을 할 수 있게 됩니다. 위 가설들을 공부하다 보니 언어를 체계적으로 전달하려는 발화보다 의미를 전달하려는 발화가 모국어를 습득하는 어린아이의 과정에 더 수렴함을 알아냈습니다. 보다 더 모국어 사용자와 같은 방법으로 언어를 사용할 수 있게 언어를 교육하기 위해서는 위 교육방법이 도움이 될 것으로 예상합니다. 그래서 저희는 실생활에 적용 가능한 한국어 발화를 습득시키기 위하여 위 웹서비스를 제작하게 되었습니다.

## 나. 토크토크의 서비스

토크토크(Talkie-Talkie)는 3가지의 언어습득 방법을 제공합니다.

- 1) **Speak to go**: 발화를 CSR을 사용하여 텍스트로 바꿔준 후, 텍스트를 분석하여 결과값을 출력 해 줍니다. 사용자가 직접 말하여 보면서 언어를 습득할 수 있는 방식입니다.
- 2) **Write to go**: 한국어 발음이 능숙하지 못하여 CSR에서 발화가 잘 인식되지 않는 사용자들을 위하여 제공하는 서비스입니다. 발화가 아닌 한국어를 타이핑하여 회화를 습득하는 방식입니다.
- 3) **Dialouge to go**: 음식점 주문 상황, 호텔 예약 상황 등을 대화 형식으로 제공하여, 말하기 듣기로 회화를 익힐 수 있도록 하는 방식입니다.

## 3. 기술적 배경

Talkie-Talkit는 FastText, 네이버 CSR, 네이버 CSS, 네이버 NMT와 Konlpy를 사용합니다. 토크토크는 외국인이 한국어 회화능력을 키우기 위해 사용하는 서비스입니다. 단순한 학습, 암기가 아닌 회화의 듣고 말하는 습득 과정을 달성하기 위해서는 TTS, STT서비스가 제공되어야 합니다. 그렇기 때문에 네이버 클라우드 플랫폼의 CSR, CSS API를 사용하였습니다. 또한 습득에는 교정의 과정이 필요합니다. 사용자(외국인)의 발화가 한국어 문법에 부합한지, 의미가 제대로 전달되었는지를 판별하기 위해 Konlpy와 FastText를 사용하였습니다.

## 4. 구현 방식

사용자의 발화가 적절한가를 판별하는 방법으로 의미유사도, 문법정확도를 척도로 내세웠습니다. 2가지 요소를 종합적으로 고려하여 테스트의 점수를 매겨, Pass와 Fail을 판별합니다.

### 가. 의미유사도

의미유사도는 사용자의 발화가 원문과 얼마나 부합하는가를 판별하는 방법입니다. 예를 들어, 원문 “How can I get to the Gyeongbokgung-Palace?”가 주어

졌을 때, 사용자의 발화는 “경복궁에는 어떻게 가나요?”가 되어야 적절합니다. 만약 “경복궁 어떻게 가?”, “어떻게 가나요 경복궁?”가 주어진더라도 의미유사도는 높은 점수를 받을 수 있도록 하였습니다.

의미유사도는 FastText를 통해서 구현하였습니다. FastText를 통해 원문의 정답 “경복궁에는 어떻게 가나요?”와 사용자의 발화 간의 유사도를 점수로 사용합니다. 그렇기 때문에 FastText의 특성을 통해 음성인식에도 강점을 지니고 있습니다. 가령, “경복궁에는 어떻게 가나요?”와 같이 음성인식 오류가 발생한 문장도 충분히 높은 점수를 받을 수 있습니다. 하지만, “창경궁에는 어떻게 가나요?”와 같은 문장도 높은 점수를 받을 가능성이 있습니다. 왜냐하면 경복궁과 창경궁은 모델 상에서 비슷한 위치에 있기 때문입니다. 그렇기 때문에 TF-IDF를 통해 점수를 보정하여 제공합니다.

## 나. 어순정확도

어순정확도는 사용자 발화의 어순이 정확한지를 판별하는 과정입니다. 입력된 문장의 문장성분을 분석하여주는 ADAMS.ai의 구문분석 api를 사용하여 문장의 성분과 문장성분간의 연결 관계를 분석합니다. 하지만 이 api를 통해서만 어순의 정확성까지 판별할 수 없었습니다. 그래서 어순의 정확성을 판단하기 위해서 직접 올바른 한국어 어순을 분석하는 알고리즘을 제작하였습니다.

한국어를 사용한 문장은 크게 5가지의 구조로 나눌 수 있습니다.

- 1.Np\_sbj + vp(주어 + 서술어)
- 2.Np\_sbj +Np\_cmp +Vp(주어 + 보어 + 서술어)
- 3.Np\_sbj +Np\_obj +Vp(주어 + 목적어 + 서술어)
- 4.Np\_sbj +Np\_ajt +Vp(주어 + 부사어 + 서술어)
- 5.Np\_sbj +Np\_멧 + Np\_obj +Vp(주어 + 부사어 + 목적어)

이 문장성분들에 수식하는 단어들이 와서 문장의 의미를 더 다양하게 해 줍니다. Np(명사)나 -mod(관형어) AP(부사구)들이 수식을 해 줄 때 문장의 어떤 성분의 앞, 뒤에 와서 수식을 해 주는지 알고리즘 내에서 규명을 합니다. 이를 통하여 다양한 한국어의 어순을 분석할 수 있습니다. 하지만 정형화된 문장의 구조가 있다고 하더라도 예외사항은 발생하게 됩니다. 그래서 형식을 벗어나는 한국어 어순을 가진 문장을 예외처리 해 주기 위하여 예외 문장형식을 따로 만들어 놓아서 예외처리를 해 줍니다. 예를 한 가지 들자면, ‘만나서 반가워요’ 같은 경우는 문장이 (Vp + Vp)로 구성이 되어있습니다. 어떠한 체언도 들어가 있지 않는 문장입니다. 이러한 경우를 예외처리 해 주면서 다양한 한국어 문장의 어순 정확도를 판별 할 수 있습니다.

## 5. 결론

### 1. 기대효과

#### 1) 개인적 측면

- 한국어 습득 비용 절감 : 강사나 책이 필요 없음.
- 편의성 증대 : 웹서비스이기에 시공간적 제약이 존재하지 않음.
- 회화능력 상승 : 한국어 활용능력 상승.

#### 2) 사회적 측면

- 데이터 수집 : 자주 틀리는 문법 정보를 수집하여, 한국어 연구에 활용가능.

- 한국어 사용자 증가 : 향후 세계적 언어의 도약의 발판역할.
- 한국어 긍정적 인식 확대 : 쉽게 배울 수 있는 언어로 인식 확대
- 한국어 연구에 활용 : 한국어 어순과 인공지능 연구의 원동력 제공