

제안서



과 목 명: 오픈소스SW기여

담당 교수: 송 인 식 교수님

전 공 명: 산 업 보 안

학 번: 32200492

이 름: 김 명 현

<목차>

1. 프로젝트 배경 및 목표
2. 요구사항
3. 추진 계획
4. 위험 평가 및 대책
5. 주요 용어 설명
6. 참고문헌

1. 프로젝트 배경 및 목표

한글을 영문으로 번역하는 과정에서 짧은 문장이나 단어는 정확하게 번역이 가능하다. 정확하게, 기존에 사람들이 많이 번역했던 텍스트는 많은 양의 데이터가 학습되어 있기 때문에 정확한 번역 품질을 보여주지만 이해와 관련하여 번역 오류가 발생한다. 예를 들어, 기존에 많이 번역되지 않아 축적된 데이터의 양이 적은 희귀언어의 경우, 번역물의 정확성이 매우 떨어진다는 한계를 지닌다.[1] 이러한 경우는 새로운 창작물이나 논문을 영작할 때 구글 번역을 사용하는 경우이며, 주로 문단이나 긴 문장의 전체적인 흐름을 통한 이해가 수반되어야 한다.

따라서 이 프로젝트의 목표는 단순 기계 번역이 아닌 사전지식이 요구되는 어느정도 학문적인 내용의 글을 자연스럽게 번역하기 위해 Google 번역과 ChatGPT 오픈소스를 활용하여 기존보다 정확성이 높은 번역기를 만들고자 한다. 이로 인해 기대할 수 있는 효과는 긴 문장이나 하나의 문단을 원어민이 사용할 만한 표현을 사용하여 매끄럽게 번역을 할 수 있고 영어가 아닌 타 언어로 된 논문도 번역을 통해 읽을 수 있을 것이다.

2. 요구사항

프로젝트에서 제공하고자 하는 기능은 다음과 같다.

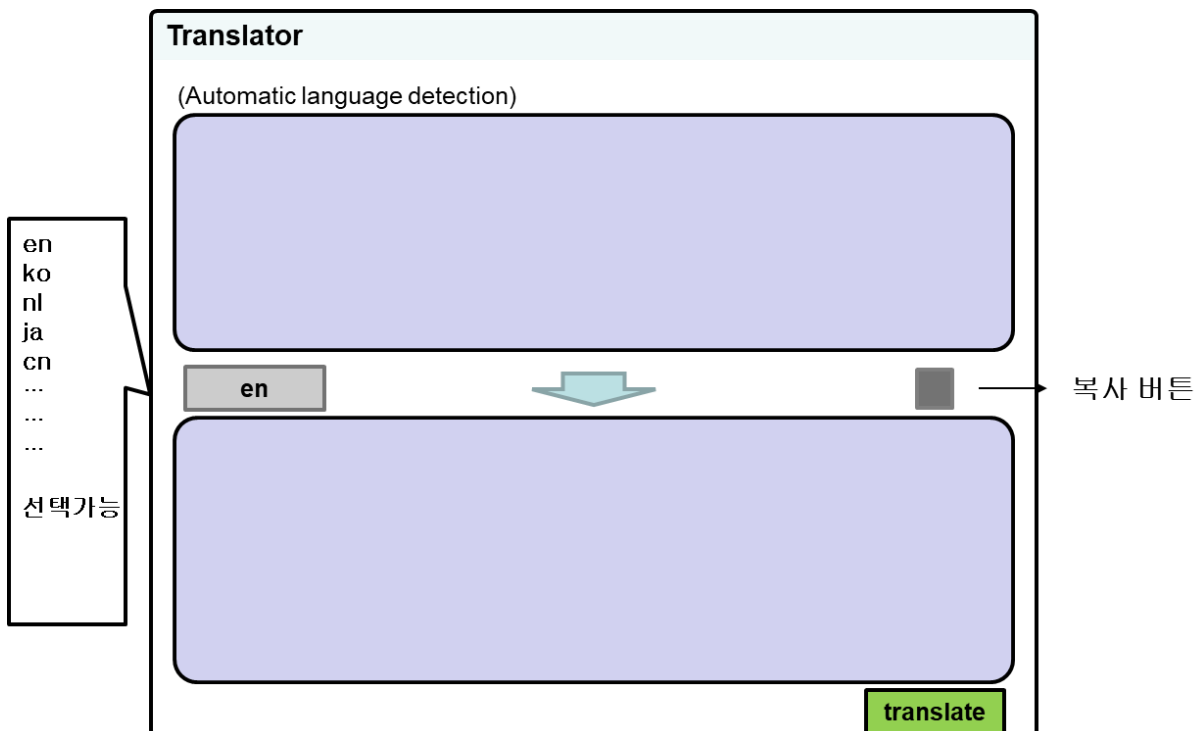
기능	내용						
자동 언어 탐지	<p>사용자가 따로 입력한 글의 언어가 무엇인지 선택하지 않아도 입력된 텍스트의 언어를 탐지할 수 있다. 언어 선택창이 입력과 결과창 둘 다 있으면 선택 가능한 언어가 많기 때문에 찾는데 번거롭고 귀찮다. 따라서 자동 언어 탐지 기능이 있다면 사용자에게 편리함을 제공할 수 있다.</p> <p>translator.detect (lang, confidence)</p> <table><tr><th>요소</th><th>의미</th></tr><tr><td>lang</td><td>감지한 언어</td></tr><tr><td>confidence</td><td>결과의 신뢰도 (0~1.0)</td></tr></table> <p>감지 가능한 언어 리스트</p> <pre>LANGUAGES = { 'af': 'afrikaans', 'sq': 'albanian', 'am': 'amharic', 'ar': 'arabic', 'hy': 'armenian', 'az': 'azerbaijani', 'eu': 'basque', 'be': 'belarusian', 'bn': 'bengali', 'bs': 'bosnian', 'bg': 'bulgarian', 'ca': 'catalan', 'ceb': 'cebuano', 'ny': 'chichewa', 'zh-cn': 'chinese (simplified)', 'zh-tw': 'chinese (traditional)', 'co': 'corsican', 'hr': 'croatian', 'cs': 'czech', 'da': 'danish', 'nl': 'dutch', 'en': 'english', 'eo': 'esperanto', 'et': 'estonian', 'tl': 'filipino', 'fi': 'finnish',</pre>	요소	의미	lang	감지한 언어	confidence	결과의 신뢰도 (0~1.0)
요소	의미						
lang	감지한 언어						
confidence	결과의 신뢰도 (0~1.0)						

	'fr': 'french', 'fy': 'frisian', 'gl': 'galician', 'ka': 'georgian', 'de': 'german', 'el': 'greek', 'gu': 'gujarati', 'ht': 'haitian creole', 'ha': 'hausa', 'haw': 'hawaiian', 'iw': 'hebrew', 'he': 'hebrew', 'hi': 'hindi', 'hmn': 'hmong', 'hu': 'hungarian', 'is': 'icelandic', 'ig': 'igbo', 'id': 'indonesian', 'ga': 'irish', 'it': 'italian', 'ja': 'japanese', 'jw': 'javanese', 'kn': 'kannada', 'kk': 'kazakh', 'km': 'khmer', 'ko': 'korean', 'ku': 'kurdish (kurmanji)', 'ky': 'kyrgyz', 'lo': 'lao', 'la': 'latin', 'lv': 'latvian', 'lt': 'lithuanian', 'lb': 'luxembourgish', 'mk': 'macedonian', 'mg': 'malagasy', 'ms': 'malay', 'ml': 'malayalam', 'mt': 'maltese', 'mi': 'maori', 'mr': 'marathi', 'mn': 'mongolian', 'my': 'myanmar (burmese)', 'ne': 'nepali', 'no': 'norwegian', 'or': 'odia', 'ps': 'pashto', 'fa': 'persian', 'pl': 'polish', 'pt': 'portuguese',
--	---

	<pre> 'pa': 'punjabi', 'ro': 'romanian', 'ru': 'russian', 'sm': 'samoan', 'gd': 'scots gaelic', 'sr': 'serbian', 'st': 'sesotho', 'sn': 'shona', 'sd': 'sindhi', 'si': 'sinhala', 'sk': 'slovak', 'sl': 'slovenian', 'so': 'somali', 'es': 'spanish', 'su': 'sundanese', 'sw': 'swahili', 'sv': 'swedish', 'tg': 'tajik', 'ta': 'tamil', 'te': 'telugu', 'th': 'thai', 'tr': 'turkish', 'uk': 'ukrainian', 'ur': 'urdu', 'ug': 'uyghur', 'uz': 'uzbek', 'vi': 'vietnamese', 'cy': 'welsh', 'xh': 'xhosa', 'yi': 'yiddish', 'yo': 'yoruba', 'zu': 'zulu' } </pre>						
번역	<p>내가 입력한 언어를 바꾸고자 하는 언어로 바꿔주는 기능이다. 단순히 번역 결과만을 알려주는 것 보다는 번역 결과의 발음까지 알 수 있는데, 이는 발음이 생소한 언어로 번역할 때 유용하게 사용할 수 있다. 번역 버튼을 누르면 번역+교정이 이루어져 최종 번역본이 결과창에 뜨게 된다.</p> <p>translator.translate(src, dest, text, pronunciation)</p> <table border="1"> <thead> <tr> <th>요소</th><th>의미</th></tr> </thead> <tbody> <tr> <td>src</td><td>번역 전 언어</td></tr> <tr> <td>dest</td><td>번역 후 언어</td></tr> </tbody> </table>	요소	의미	src	번역 전 언어	dest	번역 후 언어
요소	의미						
src	번역 전 언어						
dest	번역 후 언어						

	text	번역 결과
	pronunciation	번역 결과의 발음 (영어로 표기)
	<p>Ex) KOR to ENG</p> <p>Translator.translate('안녕', src='ko', dest='en').text</p> <p>>>> 'Hi'</p> <p>번역하고자 하는 언어 또한 지원하는 내에서는 자유롭게 선택하여 번역할 수 있다. 언어 선택 버튼을 누르면 번역 가능 리스트 중에 사용자가 원하는 언어를 선택하고 번역을 진행하면 된다.</p>	
교정	<p>구글 번역에게서 받은 초벌 번역본을 ChatGPT를 통해 교정을 진행한다. 사용자가 따로 커스텀할 수는 없는 기능이지만 번역을 위해 선택된 언어로 교정이 다시 진행되며 이는 기계 번역의 한계를 보완해줄 수 있어 사용자에게 정확하고 세련된 작문을 제공할 수 있다.</p>	
복사	<p>결과로 나온 최종 번역본을 ctrl + c, ctrl + v하지 않고도 결과창 옆에 있는 복사 버튼만 누른다면 최종 번역본이 자동으로 복사가 되어 필요한 곳에 ctrl + v만 하면 붙여넣기가 가능하다.</p>	

3. 사용자 인터페이스



4. 추진 계획

업무 내용	4				5				6			
	1	2	3	4	1	2	3	4	1	2	3	4
coding												
testing												
최종제작												
최종보고서												

4. 위험 평가 및 대책

API 를 통해 서버에 요청을 하기 때문에 빠르게 번역을 진행하게 되면 오류가 발생할 수 있을 것 같다. 따라서 번역 버튼을 누르면 3~5 초정도는 딜레이를 두고 다시 버튼을 누를 수 있게끔 설계가 필요할 것 같다.

5. 주요 용어 설명

초벌 번역본: 구글번역으로 먼저 번역을 돌리고 ChatGPT 에게 교정을 맡길 대상

최종 번역본: ChatGPT 에 의해 교정이 된 번역본

6. 참고 문헌

[1] 임순정(Lim Soon jeung). (2012). 구글 번역기를 통해서 본 통계기반방식 번역기의 특징. 통번역 교육연구, 10(3), 183-210.

[2] <https://platform.openai.com/account/api-keys>

[3] [Googletrans: Free and Unlimited Google translate API for Python — Googletrans 3.0.0 documentation \(py-googletrans.readthedocs.io\)](#)