

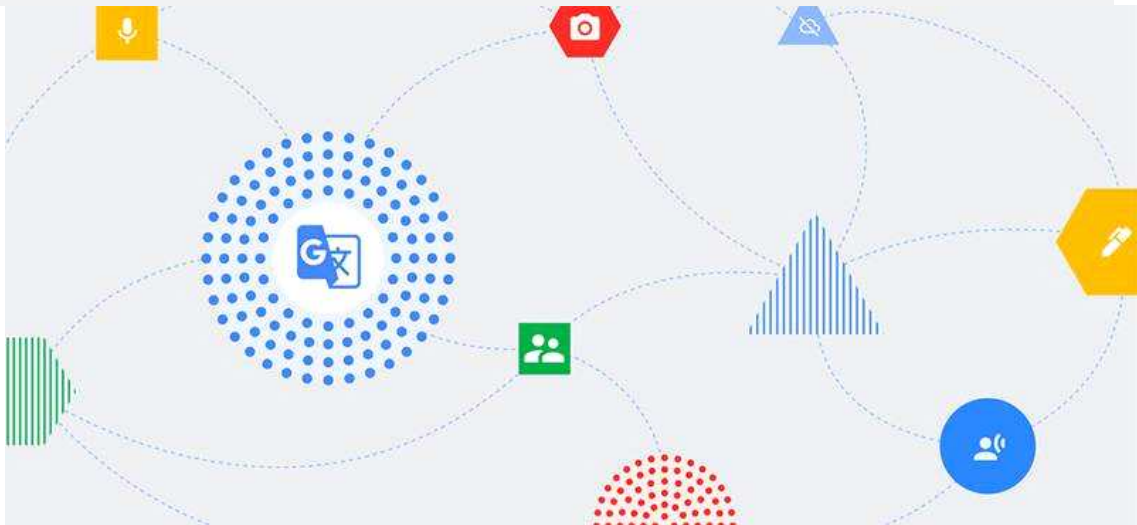
## 1. 구글 번역으로 PDF 전체 번역하는 방법

#IT 하는법 #PDF #PDF 번역 #구글번역 #번역 #자동번역

남시현 sh@itdonga.com

2022.08.25.

[IT 동아 남시현 기자] 인공지능경망과 빅데이터 기술이 발전하면서 기계번역의 품질도 빠르게 향상하고 있다. 통계 기반의 데이터로 번역하던 2016 년까지만 해도 구글 번역의 품질은 단어 사전에서 찾은 내용을 줄줄이 늘어놓는 수준에 불과했지만, 인공지능경망이 도입된 시점을 기준으로 문장이 부드럽게 전환되고, 문맥까지 반영할 만큼 발전했다. 여전히 전문가가 번역한 결과물과 비교할 정도의 품질은 아니지만, 이제는 일상 회화나 전문적인 내용, 기사 정도를 이해하는데는 무리가 없다. 덕분에 간단한 작업이라면 구글 번역을 활용해 진행하는 업무도 많아졌다.



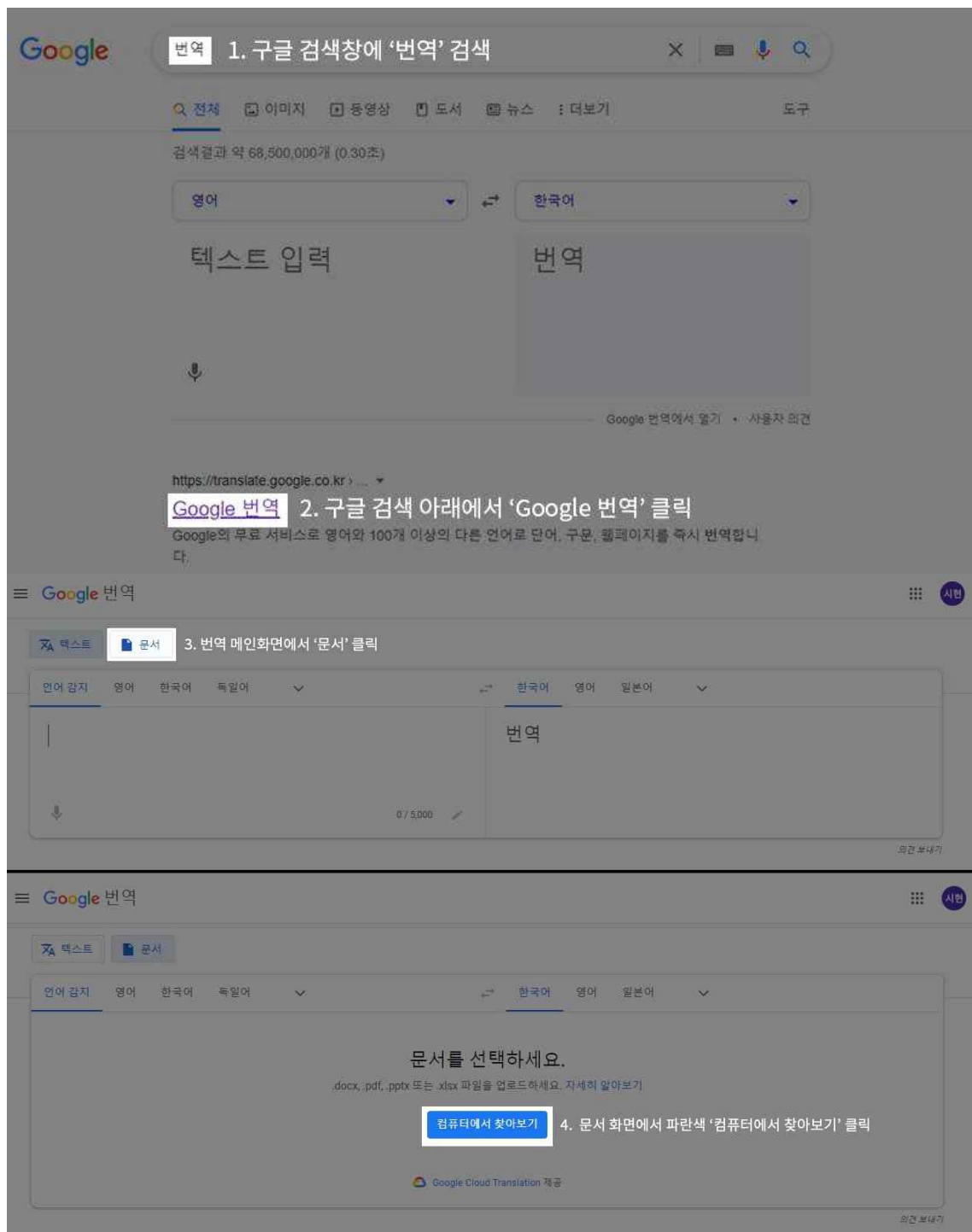
출처=구글

그런데 구글 번역에 포함된 기능 중에는 단순히 문장이나 웹페이지를 번역하는 기능 이외에도 문서를 번역하는 기능까지 포함돼있다. 말 그대로 구글 번역에 파일을 업로드하면 문서 내용의 언어를 인식해 자동으로 번역한 다음 파일로 다시 만들어서 제공하는 기능이다. 참고용 자료 수준이라면 번거로운 과정을 모두 건너뛰고 손쉽게 번역된 PDF 파일을 획득할 수 있다. 어떤 방식으로 구글 번역의 PDF 변환이 이뤄지는지 상세히 소개한다.

### 1) 일반 번역 이외에 '문서'까지 번역 가능

가장 먼저 구글 번역 메인 페이지로 진입한다. 출처=IT 동아

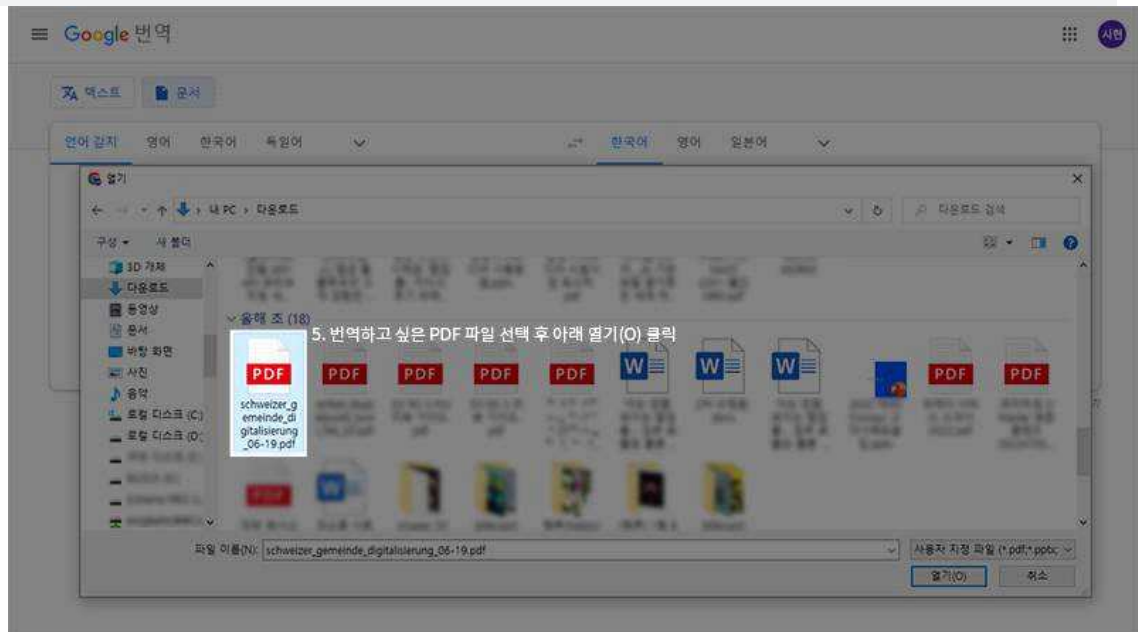
구글 번역은 100 여 개 이상의 언어를 번역해주는 구글의 무료 서비스로, 구글 검색 기능에서 바로 활용하거나 웹 페이지를 번역하는 등으로 활용할 수 있다. 기본 활용은 문장을 복사해서 번역 기능에 붙여넣고 원하는 언어를 선택하면 자동으로 번역되는데, 파일 전체를 번역하는 기능도 제공된다. 첫 단계에서는 구글에서 '번역'을 검색한 다음, 아래 'Google 번역'으로 진입해 구글 번역 메인 페이지로 접속한다.



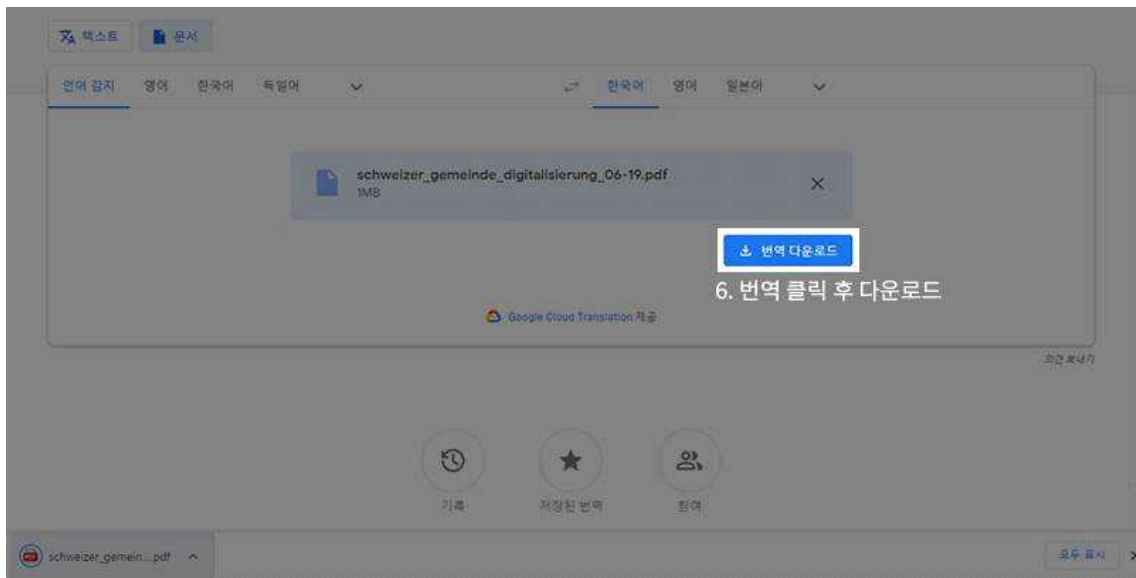
메인 화면에서 문서 탭으로 이동 후 문서를 업로드한다. 출처=IT 동아  
 구글 번역 메인 페이지에 들어오면 기본 상태에서는 바로 텍스트 번역  
 페이지가 있는데, 측면에 '문서'를 눌러 문서 페이지로 접속한다. 그  
 다음 아래 컴퓨터에서 찾아보기를 눌러 파일을 찾는다. 이때 아무  
 파일이나 모두 업로드할 수 있는 건 아니고, 마이크로소프트 워드  
 문서인 .docx 확장자와 파워포인트 pptx 확장자, 엑셀 파일인.xlsx  
 확장자,어도비 PDF 파일만 가능하며, 용량은 10MB 이내에 300 페이지

이하만 지원한다. 한글과 컴퓨터용 포맷인 HWP 파일은 지원하지 않는다.

번역 시 기본적으로 언어를 자동 감지하지만, 정확하게 어떤 언어인지 안다면 언어 감지 측면에 있는 ▼를 눌러 언어를 지정해준다. 또 반대 창에서는 어떤 언어로 선택할지 선택해준다. 기본적으로는 한국어로 돼있지만, 영어나 다른 언어로도 변환할 수 있다.



파일은 docx, pptx, xlsx 와 PDF 를 지원한다. 출처=IT 동아  
파란색 '컴퓨터에서 찾아보기'를 누르면 내 컴퓨터 내에 저장된 파일이  
뜬다. 이때 10MB 이하의 번역할 파일을 선택하고 확인을 누른다.  
참고로 드래드 앤 드롭으로 파일을 넣으면 브라우저가 파일을 그대로  
인식해버리므로, 클릭을 해서 파일을 지정해주어야 한다.

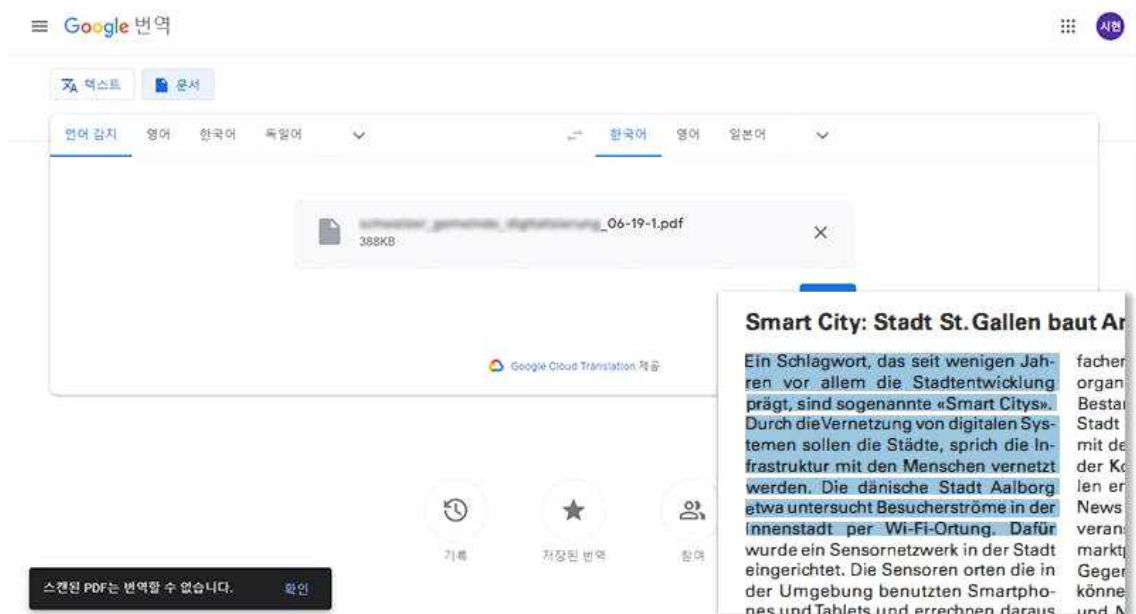


파일 용량은 10MB 300 페이지 내외만 가능하고, 업로드되면 '번역'을 눌러서 작업을 진행한 다음 다운로드한다. 출처=IT 동아지정이 끝나면 전환할 언어를 다시 한번 확인하고,파일 아래에 있는 '번역' 버튼을 클릭하면 전환을 시작한다. 전환은 파일 용량마다 다르지만 대개 20 초~2 분 이내로 끝난다. 번역이 완료되면 '번역 다운로드'를 눌러 파일을 다운로드한다. 기본적으로 파일은 '다운로드'폴더에 저장되지만, 웹 브라우저의 파일 다운로드 경로를 다르게 지정한 경우 해당 경로로 다운로드된다. 또한 파일 경로에 동일한 파일이 있더라도 (1) 등의 숫자를 통해 구분할 수 있으므로 원본 파일과 중복되지 않는다.



번역된 파일은 원본과 동일하지만 파일 형태로 저장된다. 출처=IT 동아지정이 끝나면 전환할 언어를 다시 한번 확인하고,파일 아래에 있는 '번역' 버튼을 클릭하면 전환을 시작한다. 전환은 파일 용량마다 다르지만 대개 20 초~2 분 이내로 끝난다. 번역이 완료되면 '번역 다운로드'를 눌러 파일을 다운로드한다. 기본적으로 파일은 '다운로드'폴더에 저장되지만, 웹 브라우저의 파일 다운로드 경로를 다르게 지정한 경우 해당 경로로 다운로드된다. 또한 파일 경로에 동일한 파일이 있더라도 (1) 등의 숫자를 통해 구분할 수 있으므로 원본 파일과 중복되지 않는다.

우측 아래 예시처럼 텍스트가 드래그되지 않는 문서는 OCR 이 안된 스캔본이라 번역이 먹히지 않는다. 출처=IT 동아



구글 문서 번역은 워드나 엑셀, 파워포인트 파일은 자동으로 번역하지만, PDF 파일은 번역 가능한 파일이 정해져 있다. 바로 OCR(광학 문자 인식) 판독이 기록돼있는 파일에 한해서다. OCR 처리된 문서는 웹브라우저 등으로 확인할 때 텍스트를 드래그 앤 드롭으로 복사해서 붙여넣기 할 수 있는 반면, 이미지 그 자체가 PDF 파일인 경우에는 드래그 앤 드롭이 안된다. 만약 디지털 문서를 PDF로 인쇄한 파일 형태라면 번역이 가능하나, 프린터나 카메라 등으로 책이나 문서를 스캔해서 만든 PDF는 '스캔된 PDF'는 번역할 수 없습니다'는 문구가 뜨며 번역되지 않는다.

구글이 자동 번역한 문서는 완벽한 형태는 아니지만, 문맥을 이해하는 데는 부족함이 없다. 특히 전문 번역을 활용하거나 세부 내용을 확인할 필요가 없는 수준의 가벼운 내용이라면 자동 번역 기능이 큰 도움이 된다.

글 / IT 동아 남시현 (sh@itdonga.com)