



소프트웨어 다국어 확장을 위한 솔루션

17개 추가 언어

25개 프로젝트

모든 구성 요소를 처음부터

고객은 소프트웨어 구현 프레임워크를 다국어로 확장할 수 있는 파트너가 필요했습니다. Lionbridge는 25개의 프로젝트를 진행하면서 17개의 언어를 추가로 제공했습니다.

과제

고객은 다국어로 소프트웨어를 구현할 수 있는 프레임워크를 구축했습니다. 이 프레임워크는 감독 및 비감독 기계 학습 방법뿐만 아니라 토큰화와 품사 태그 지정을 포함한 전/후처리 절차 및 텍스트 분석을 활용합니다.

고객은 이를 통해 여러 언어에서 작동하는 구성 요소를 구현할 일련의 도구와 프로세스를 개발하고 테스트했습니다. 고객은 프레임워크를 문서화하고 개발 및 관리하며 구성 요소를 다른 언어로도 구현할 파트너를 찾았습니다.

고객사 관련 정보

세계 최대 규모의 엔터프라이즈 소프트웨어 공급자로 손꼽히는 이 고객은 자연어 텍스트 기술의 선두 주자로서, 여러 제품군을 통해 다른 애플리케이션에서도 사용할 수 있는 효율적인 텍스트 처리 알고리즘을 구축합니다.



핵심 용어

개체명 인식기

주어진 텍스트에서 개체명을 찾고, 그것이 어떤 사전 정의된 범주 (예: 사람 이름, 장소 이름)에 속하는지 식별합니다.

예: Peter is an engineer and lives in Toronto
Peter = 사람, Toronto = 장소

주요 구문 추출기

주어진 텍스트의 맥락에서 관련성이 높은 핵심 문구나 키워드의 목록을 추출합니다.

예: 위 문장의 경우, 1. 주요 구문 추출기 2. 핵심 문구 목록 3. 주어진 텍스트

결과

Lionbridge는 고객이 17개 추가 언어의 확장 솔루션을 개발하는 데 효과적으로 지원했습니다. 대부분의 언어 소프트웨어가 처음부터 새로 구현되었습니다. 고객의 프레임워크를 사용하여 처음부터 소프트웨어를 새로 구현하고, 기존 소프트웨어를 개선하는 데 필요한 모든 문서, 로드맵 및 보조 도구를 개발했습니다. 또한 소프트웨어의 성능을 만족스러운 수준으로 끌어올리기 위해 취한 모든 조치를 기록하는 구조도 추가적으로 개발했습니다. 마지막으로, 고객에게 프레임워크와 구성 요소를 더욱 향상시킬 수 있는 방법에 대해서도 조언했습니다.

Lionbridge는 25개 프로젝트에서 이러한 성과를 달성했으며, 고객은 그 결과로 나온 솔루션을 클라우드 기반 API에 통합하여 기본 텍스트를 통한 자연어 처리 서비스를 제공하게 되었습니다.

자세한 내용:

[LIONBRIDGE.COM](https://www.lionbridge.com)

해결책

Lionbridge는 고객이 17개 추가 언어의 확장 솔루션을 개발하는 데 효과적으로 지원했습니다. 대부분의 언어에서 소프트웨어가 처음부터 새로 구현되었습니다.



1. 효율적인 워크플로우 구축

Lionbridge는 고객과의 협업 과정에서 멘토링 시간을 거쳐 고객의 프레임워크로 시험용 언어에 대한 소프트웨어를 구현했습니다. 이후 내부에서도 시험용 언어뿐만 아니라 여러 추가 언어들에 대한 소프트웨어 구현을 계속 진행했습니다. 그 결과, 시의적절하게 소프트웨어가 실행되고 프로세스가 개선되었습니다.



2. 언어별 과제 극복

한 언어에서 다른 언어로 전환하는 과정에서 Lionbridge 팀은 다음과 같은 언어별 과제를 성공적으로 수행했습니다.

개체명 인식: 대문자 표기 문제

- 일부 언어(예: 아랍어, 일본어, 한국어)에는 개체명 인식의 주요 특징인 단어 대문자 표기가 없습니다.

주요 구문 추출: 과잉 생성

- 주요 구문은 텍스트에 자주 등장하는 특정 단어를 포함하므로 올바르게 추출될 수 있습니다. 그러나 이와 동시에, 자주 등장하는 동일한 단어가 포함된 다른 주요 구문은 잘못 추출될 수도 있습니다. 단, 이렇게 추출된 주요 구문은 주어진 텍스트와는 관련이 없습니다.

Lionbridge 팀은 언어 기반의 규칙 작성 스크립트를 구현하고 이를 프레임워크에 통합하여 이러한 과제를 해결했습니다. Lionbridge 팀은 또한 소프트웨어 개발을 지원할 텍스트 처리 보조 도구와 개선이 필요한 영역을 쉽게 식별할 수 있는 분석 도구를 구현했습니다.



3. 인프라 제공

Lionbridge는 언어 전문가, 프로젝트 관리자, 데이터 엔지니어, 그리고 자연어 처리 솔루션 및 기타 언어 자산 개발을 위한 전 세계 언어 전문가 네트워크로 구성된 세계 최대 규모의 자연어 처리 팀을 구성했습니다. Lionbridge에서 배포한 클라우드 기반 플랫폼 덕분에 글로벌 팀은 전 세계 어디에서나 Lionbridge 도구를 사용하여 Lionbridge 프로젝트 관리자와 코디네이터의 지도와 감독을 받을 수 있습니다. Lionbridge가 말하는 글로벌 접근성이란, 하루 24시간 내내 전 세계 고객을 대상으로 생산성을 발휘한다는 뜻입니다.

고객은 일반적인 로컬라이제이션 프로젝트에서 요구하는 서비스 그 이상을 기대했습니다. 기계 학습 방법에 대한 심층적인 지식을 갖추고 언어학적 기반의 규칙을 작성하는 동시에 텍스트 전/후처리 절차를 개발하여 발전시킬 수 있는 파트너를 원했습니다. 일반적인 번역사에게 이러한 능력을 기대하기는 어렵습니다. Lionbridge는 200개 이상의 언어에서 기계 학습에 대한 이해와 시설을 갖춘 전문가 네트워크를 양성하여 배포합니다.