로컬라이제이션의 미래

AI Business
전자책 시리즈

언어와 문화를 이용한 디지털 혁신 기법

공동 작성:

LIONBRIDGE

2

비즈니스 확장의 핵심

로컬라이제이션이란 제품, 대화, 서비스 등을 특정 지역에 있는 고객의 기대에 맞게 조정하는 것을 말합니다.

글로벌화 추세가 더욱 가속화되는 가운데, 기업은 로컬라이제이션을 통해 세계 시장에 진출함으로써 수익을 증대하고 그 과정에서 규모의 경제를 이룰 수 있습니다. 따라서 로컬라이제이션은 현지 제품만이 아닌 그 이상을 추구하는 모든 비즈니스 전략의 핵심에 있습니다. "캘리포니아주 소재의 Apple 디자인"을 제품의 모토로 하는 Apple의 경우에도 제 아무리 브랜딩과 특정 지역을 연계시킨 기업이라도 국경을 넘어 전 세계 고객의 공감을 불러일으키기 위해서는 로컬라이제이션을 수행해야 합니다.

물론 번역은 여전히 로컬라이제이션의 가장 중요한 부분입니다. 이 전자책에서도 다루겠지만, 이 작업의 일부는 여전히 사람에 의해 수행되며 다른 일부는 첨단 기계 번역(MT) 툴이 담당합니다. 그리고 남은 부분을 두 방식 모두를 이용하는 하이브리드 접근법, 즉 증강 번역을 사용해 수행합니다(*5 페이지 참고*).

하지만 진정한 로컬라이제이션을 위해서는 언어적 차원을 넘어서야 합니다. 서비스나 제품을 한 시장에서 다른 시장으로 옮겨 정착시키려면 각 지역만의 미묘한 특성을 이해할 수 있어야 합니다. NI(National Instruments)의 언어 서비스 부문 선임 관리자인 Inés Rubio가 AI Business에 밝혔듯이, "우리는 근본적으로 경험을 다른 언어로 재창조합니다. 언어 자체가 아니라 문화를 이해해야 하죠."

번역 및 로컬라이제이션 전문업체 Lionbridge가 후원하는 이 전자책에서는 국제 비즈니스가 직면하고 있는 몇 가지 과제와 로컬라이제이션을 원활히 수행했을 때의 이점을 살펴봅니다. 전문가 인터뷰를 포함해 Zynga, Royal Caribbean, Volvo, Wikimedia Foundation 등 여러 조직의 시각도 특집으로 담았습니다. 귀사의 글로벌 제품과 서비스 구축에 이 전자책이 유용하게 사용되길 기대합니다.

Luke Dormehl | 부편집자 | AI Business

전자책 공동 작성:

LIONBRIDGE

3 *현대의 로컬라이제이션 서비스 소개*
언어와 문화

4 *시장 개요*
로컬라이제이션, 주류가 되다

5 *작업에 AI 도입하기*
증강 번역이란?

7 *사례 연구*
Lionbridge와의 협력을 통한 Royal Caribbean의 로컬라이제이션 재정비

8 *전 세계 고객을 위한 지원*
로컬라이제이션을 통해 수백만 게이머들의 마음을 사로잡은 Zynga

10 *현지에 맞게 생각하기*
언어 서비스를 통한 Wikipedia 품질 개선

12 *기계 번역의 발전*
변곡점에 서서

14 사례 연구
*Volvo Cars*의 통합 온라인 시장 구축을 도운 *Lionbridge*

15 미래를 위한 준비
언어 기술의 미래

17 학계의 시각
동등성의 문제

언어와 문화

"언어 자체가 아니라 문화를 이해해야 하죠. 가령 스페인어에 대해 이야기할 때 우리는 어디를 염두에 두고 말하나요? 전 세계 모든 스페인어권인가요, 아니면 유럽 스페인어권인가요? 대상 고객과 공감대를 형성하려면 어떻게 해야 할까요? 단순히 단어를 번역하는 것만으로는 절대 부족하죠. 그보다 훨씬 심오한 부분을 다룰 수 있어야 합니다.

번역도 중요하지만 대상 고객을 이해하는 것이 더 중요합니다

로컬라이제이션은 일반적으로 웹사이트, 사용 설명서, 소프트웨어 애플리케이션과 같은 언어적 자산을 독자들이 읽고 이해할 수 있도록 번역하는 것을 말합니다.

Lionbridge 제품 부문 리더이자 전략가인 Kajetan Malinowski가 AI Business에 말했듯이 "로컬라이제이션의 목적은 제품 또는 서비스를 진정한 의미에서 현지화하는 것입니다. 최종 소비자의 언어로 말하는 것은 그중 큰 부분을 차지합니다." 그러나 언어는 로컬라이제이션의 한 부분일 뿐, 세계에서 가장 수익성 높은 일부 시장에 진입하려면 문화적 전문지식이 반드시 필요합니다.

번역을 넘어

"무엇을 말하는가가 아니라 어떻게 말하는가가 중요하다"라는 말을 들어본 적이 있을 것입니다. 로컬라이제이션을 단순히 번역과 동일시해서는 안 되는 이유가 바로 여기에 있습니다. 로컬라이제이션은 현지 커뮤니케이션의 다양한 측면을 고려하는 것을 의미합니다. 언어는 물론 지역적 차이를 고려하는 것도 로컬라이제이션의 한 부분입니다. 지역마다 다른 표준의 한 예로 측정 단위를 들 수 있습니다. 그뿐 아니라 규제나 기술적 측면, 심지어 '어조'나 어법 또한 지역마다 다릅니다.

NI(National Instruments)의 언어 서비스 부문 선임 관리자인 Inés Rubio는 AI Business와의 인터뷰에서 이렇게 말합니다. "우리는 근본적으로 경험을 다른 언어로 재창조합니다. 언어 자체가 아니라 문화를 이해해야 하죠. 가령 스페인어에 대해 이야기할 때 우리는 어디를 염두에 두고 말하나요? 전 세계 모든 스페인어권인가요, 아니면 유럽 스페인어권인가요? 대상 고객과 공감대를 형성하려면 어떻게 해야 할까요? 단순히 단어를 번역하는 것만으로는 절대 부족하죠. 그보다 훨씬 심오한 부분을 다룰 수 있어야 합니다."

로컬라이제이션에는 대상 고객의 취향에 맞게 인터페이스 아이콘과 같은 그래픽을 변환하는 작업도 포함될 수 있습니다. 이러한 아이콘에 보편적이거나 명백한 의미가 포함될 필요는 없습니다. 예를 들어, 전구 아이콘은 미국과 영국에서는 흔히 아이디어를 뜻하지만 다른 국가에서는 다른 의미로 받아들여집니다. '엄지척' 이모티콘은 미국과 영국에서는 긍정적인 의미를 지니지만 라틴 아메리카, 중동, 서아프리카에서는 모욕적인 의미가 됩니다. 색상이 지닌 숨은 의미 또한 지역마다 다릅니다. 인도에서 주황색은 사랑과 용기를 뜻하지만 중동 문화권에서는 상실이나 애도에 관한 경우가 많습니다.

로컬라이제이션은 올바로 수행하면 엄청난 이익이 되지만 현지의 중요한 관행을 잘못 이해하거나 놓치면 심각한 피해를 초래할 수 있습니다. 로컬라이제이션은 2020년대의 현대적인 기업들이 반드시 활용해야 할 가장 중요한 도구 중 하나입니다.

4

로컬라이제이션, 주류가 되다

"시작 단계에서는 로컬라이제이션에 대해 아예 생각도 못하죠. 거의 막바지에 이르러서야 떠올리게 되거든요. 로컬라이제이션은 전혀 고려의 대상이 아니라는 건데, 경험이 쌓이고 과거의 실수를 돌이켜보면서 나중에 변경을 한다는 것이 얼마나 큰 비용이 드는 일인지 깨닫게 되면 비로소 기업과 팀 전체가 로컬라이제이션의 중요성을 인정하게 되는 거죠."

로컬라이제이션 시장의 규모와 성장의 원동력

'l10n(숫자 10은 l과 n 사이의 문자 수)'으로 줄여 부르기도 하는 현대 로컬라이제이션의 탄생은 1980년대로 거슬러 올라갑니다. 정확히 일치하는 것은 아니지만 당시는 컴퓨터 소프트웨어가 전 세계 사무실과 가정의 필수품이 되기 시작한 시기이기도 합니다. 이제 막 성장하기 시작한 이 시장에 뛰어들기 위해 기업들은 신규 시장에 맞춰 새로운 문자 집합을 지원하거나 특정 기능을 추가 또는 변경하는 방식으로 로컬라이제이션을 시도해야 했습니다. 처음에는 사내에서 이러한 작업을 자체 수행했지만, 그 규모와 중요성이 점차 커지면서 로컬라이제이션 산업이 본격적으로 발전하기 시작했습니다.

역대 최고 성장률

지난 10년 동안 언어 서비스 시장은 교육, 번역 및 광범위한 로컬라이제이션 활동으로 거의 두 배 가까이 성장했습니다. 언어 서비스 업계를 선도하는 출판사인 Slator의 최근 보고서에 따르면, 2020년 현재 번역, 로컬라이제이션 및 통역 시장의 전체 규모는 238억 달러에 달하는 것으로 나타났습니다. 가장 빠르게 성장한 고객 분야는 기술, 게임, 미디어 및 생명과학이었으며, 전체 산업 규모는 2021년에만 최대 10%까지 성장할 것으로 예측되었습니다.

Statisa에 따르면, 유럽은 세계 인구의 10%도 안 되지만 전 세계 언어 서비스 시장의 절반 가량을 차지하고 있습니다. 매우 다양한 언어가 이 지역에서 사용되고 있기 때문이죠. 한편, 시장 규모 면에서는 여전히 미국이 선두를 달리고 있습니다.

국가별로만 격차가 있는 것이 아닙니다. 일부 업계에서는 로컬라이제이션이 거의 성숙 단계에 들어섰지만, 다른 업계(특히 신흥 업계)나 기업에서는 이제야 그 중요성이 부각되고 있는 실정입니다.

NI의 언어 서비스 부문 선임 관리자 Inés Rubio는 로컬라이제이션에 대한 노력 수준을 통해 그 성숙도를 가늠할 수 있다고 말합니다. "시작 단계에서는 로컬라이제이션에 대해 아예 생각도 못하죠. 거의 막바지에 이르러서야 떠올리게 되거든요. 로컬라이제이션은 전혀 고려 대상이 아니라는 건데, 경험이 쌓이고 과거의 실수를 돌이켜보면서 나중에 변경을 한다는 것이 얼마나 큰 비용이 드는 일인지 깨닫게 되면 비로소 기업과 팀 전체가 로컬라이제이션의 중요성을 인정하게 되는 거죠."

로컬라이제이션과 관련된 기술이 계속 발전하면서 이 산업을 움직이는 도구들도 꾸준히 개선되고 있습니다. 신경망 기계 번역(NMT)과 같은 새로운 혁신 기술은 등장한 지 몇 년이 채 되지 않았지만 이미 업계에 상당한 영향을 미치고 있습니다. 이러한 혁신은 고객사나 최종 사용자 모두에게 로컬라이제이션 산업의 지속적인 변화를 불러올 것입니다.

증강 번역이란?

인간과 기계의 관계, Lionbridge의 Will Rowlands-Rees 및 Jaime Punishill 과의 대담

Lionbridge의 최고 제품 책임자인 Will Rowlands-Rees가 AI Business와의 인터뷰에서 말합니다. "기계 번역은 AI의 기본적인 응용 프로그램 중 하나입니다. 최초의 기계 번역은 1950년대 초로 거슬러 올라갑니다. 그 당시에도 이제 더 이상 사람에 의한 번역은 필요하지 않다는 예측이 나왔었죠. 사실 그런 예측은 다소 빗나간 것 같지만요."

거의 75년이 지난 지금도 사람에 의한 번역은 여전히 중요한 부분으로 남아 있습니다. 이는 기계 번역을 적용하는 작업의 경우에도 마찬가지입니다. 어떤 상황에서는 공개적으로 사용되는 무료 번역기가 효과적일 수도 있겠지만 업무 용도로는 거의 적합하지 않습니다.

Lionbridge의 최고 마케팅 책임자 Jaime Punishill이 AI Business에 설명합니다. "번역 창에 어떤 문구를 붙여 넣었더니 말도 안되는 문장이 번역되어 나온 경험이 다들 있을 겁니다. 학습을 거치지 않은 무료 번역기는 아직 그런 수준입니다. 상용 환경에서는 조금 낫지만, 여전히 두 언어 사이에서 엉뚱한 언어적 문제가 발생하곤 합니다. 영어를 한국어로 번역하는 것은 쉽지 않죠. 독일어를 한국어로 번역하는 것은 말할 필요도 없고요. 이 번역기가 얼마나 웃긴 문장을 만들어내는지 모릅니다."

좋은 번역의 중요성

그저 내용의 골자를 파악하면 되는 상황이라면 번역된 뉴스 기사를 읽는 것으로도 충분할 수 있습니다. 하지만 '이 정도면 충분'한 것만으로는 사실 충분치 않은 경우가 많습니다. 예컨대 투자 상품 개요, 의학 연구 라벨 또는 법률 문건 등에는 완벽한 번역이 요구됩니다.

"최고급 스포츠카의 매뉴얼을 만들면서 잘못된 설명을 제공하면 건강과 안전상의 문제가 발생할 수 있지요." Rowlands-Rees는 말합니다. "임상 연구나 생사 문제와 관련된 다른 분야의 경우도 마찬가지이고요. 회사의 관점에서는 모든 것이 브랜드에 영향을 미칠 수 있으므로 번역이 정확한지 확인할 필요가 있습니다. 그렇지 않으면 브랜드에 심각한 영향을 미칠 수 있으니까요."

과거, 사후 교정으로 일컫던 이 프로세스에서는 먼저 기계 번역 툴을 이용해 초벌 번역을 수행합니다. 그런 다음 사람이 원시 기계 번역본에서 컴퓨터가 저지른 오류를 수정합니다. 이 접근법은 이론적으로는 시간을 절약할 수 있지만 프로세스의 마지막 단계에 이르러서야 비로소 번역사인 사람이 투입된다는 한계가 있습니다.

증강 번역의 세계로

증강 번역이란 원래 CSA Research에서 고안한 용어로, 언어 전문가인 사람이 우수한 품질의 번역을 보다 신속하게 수행하면서 일관성과 정확성을 높일 수 있도록 하는 다양한 기술을 활용한 번역 기술입니다. 증강 번역은 사람과 기술의 연결성을 증대시킴으로써 번역 경험을 개선합니다. 이전에도 이러한 여러 접근법이 탐구되었지만, 이제는 이 모든 기법을 하나로 연결함으로써 더 포괄적인 프레임워크를 만들 수 있게 되었습니다. 이 기술을 활용하면 가치는 낮으면서 시간이 많이 걸리는 작업은 자동화하고, 반면에 언어 전문가인 사람의 전문지식은 최대한 활용할 수 있습니다. ›

6

› 증강 번역이 적용되는 7가지 주요 분야:
번역 메모리
적응형 신경망 기계 번역
품질 평가
자동 콘텐츠 강화(ACE)
용어 관리
소등(Light-out) 프로젝트 관리
번역 관리 시스템(TMS)

이 모든 작업은 다 함께 구현됩니다. 예를 들어, 번역 메모리는 반복 번역 작업에 도움이 될 수 있습니다. Punishill이 설명했듯이, "매년 동일한 회사의 제품 설명서를 번역한다고 할 때, 대부분의 내용이 전년도의 것과 동일함에도 이전의 아날로그 시대에는 번역을 맡길 때마다 매번 비용을 지불해야 했습니다. 하지만 번역 메모리를 활용하면 원문이 일치하는 것이 있는지 확인할 수 있습니다."

정확히 일치하면 추가 수정 없이 텍스트를 넘기면 됩니다. 일부분만 다른 경우에는 플래그를 달아 번역사인 사람이 주변 단어와 문장을 확인하고 추가한 내용이 맥락에 영향을 주지 않는지 확인하면서 업데이트하면 됩니다.

증강 번역의 또 다른 이점은 특정 번역 작업에 가장 적합한 사람 및 기계 리소스를 자동으로 결정할 수 있다는 데 있습니다. 다시 말해, 증강 번역은 작업을 번역사인 사람에게 넘기거나, 아니면 어떤 기계 번역 엔진이 최상의 결과를 낼 수 있는지를 안다는 의미입니다.

장점의 극대화

"이 분야에서 활용할 수 있는 또 다른 특징은 기계마다 강점이 다 다르다는 점이죠." Punishill은 이어서 말합니다. "어떤 기계는 영어를 독일어로 번역하는 데 능하고 어떤 것은 스페인어를 한국어로 번역하는 데 뛰어납니다. 제약 콘텐츠나 마케팅 콘텐츠에 각각 특화된 기계도 있죠. 이러한 MT 엔진을 잘 섞어서 적절히 사용할 수 있어야 합니다. 번역 시스템이 제기능을 하기 위해서는 여러 가지 기능이 필요하거든요."

"Lionbridge에서는 기업이 이러한 시스템을 구축하고 운영하도록 지원합니다. 여기에는 엔진을 학습시키는 것부터, 브랜드/보이스 지침과 용어 사전 유지관리 및 이러한 작업이 원활히 구현될 수 있도록 모든 언어적 자산을 유지하는 일까지 포함됩니다. 이는 단순히 번역을 잘하는 것을 넘어, 예컨대 Apple의 보이스 또는 Microsoft의 보이스로 번역하는 것을 의미합니다."

Rowlands-Rees가 강조합니다. "우리는 고객에게 어떤 문서는 사람이 번역해야 하며 그 이유는 무엇인지, 어떤 문서는 MT로 라우팅하는 것만으로도 충분하며 왜 그런지 설명할 수 있습니다. 사람에 의한 번역이 더 나은 상황과 MT가 더 나은 상황에 대해 잘 알고 있으니까요. 우리는 이를 프로그램에 반영해 콘텐츠를 가장 적절하게 라우팅할 수 있습니다. 이 방식은 우리는 물론 고객에게 특히 효율적이죠."

증강 번역은 다른 방식으로도 번역 작업을 지원할 수 있습니다. 가령 텍스트가 얼마나 번역이 잘 될지 가늠하는 예측 분석 기법을 통해 콘텐츠 작성 작업의 업스트림과 다운스트림을 개선할 수 있습니다. 또한, 다른 언어에 대한 검색 엔진 최적화(SEO) 영역에서도 종합적인 평가를 수행하고 온라인 검색 용이성을 크게 개선할 수 있습니다.

증강 번역은 기계 번역의 최첨단 기술로서 결과물의 질을 현저히 향상시킬 수 있습니다. 번역과 품질 평가(기계 번역이 달성할 수 있는 품질 수준에 대한 통계적 추정치 제공), 그리고 최신 적응형 신경망 기계 번역 툴이 결합된 증강 번역 기술은 모든 로컬라이제이션 팀에게 강력한 무기가 될 것입니다. 따라서 앞으로 몇 년 동안 증강 번역이란 용어를 더 자주 접하게 될 것으로 예상됩니다.

Lionbridge 와의 협력을 통한 Royal Caribbean 의 로컬라이제이션 재정비

전 세계 고객을 공략할 방법을 모색하던 유명 크루즈선 운영사

전 세계 바다를 항해하는 크루즈선은 말 그대로 국제적입니다. 매년, 전 세계 약 530만 명의 여행객들이 Royal Caribbean과 함께 이국적인 목적지로 항해를 떠납니다. 여행객이 항해에서 바라는 경험을 선사하고 Royal Caribbean의 브랜드 품격을 유지하기 위해서는 회사 웹사이트인 royalcaribbean.com을 제대로 번역해야 했습니다. 그래야 고객이 이 웹사이트를 통해 Royal Caribbean의 다양한 크루즈선과 액티비티, 그리고 목적지를 검색하고 휴가 일정을 예약할 수 있을 테니까요.

해결 과제

Royal Caribbean과 같은 일류 브랜드의 경우, 콘텐츠 번역 워크플로를 통제 없이 주먹구구식으로 운영하게 되면 콘텐츠의 정확성과 커뮤니케이션 능력 모두에 부정적인 영향을 미칠 수 있습니다. 그래서 Royal Caribbean은 자사의 콘텐츠 플랫폼을 업데이트할 계획을 세우고 이를 지원할 로컬라이제이션 전문 업체를 찾던 중, 언어 서비스 분야에서 25년 이상의 경험을 보유한 Lionbridge와 협력하기로 결정했습니다. Royal Caribbean의 목표는 콘텐츠에 대한 중앙집중식 사이트를 구축함으로써 조회수와 판매를 모두 끌어올리고, 번역 워크플로를 개선하며, 탁월한 사용자 경험을 제공하는 것이었습니다.

솔루션

솔루션으로 Lionbridge의 Connector™와 Adobe Experience Manager™의 두 가지 툴로 구성된 서비스가 채택되었습니다. 이 서비스는 Royal Caribbean의 콘텐츠가 올바로 업데이트되도록 지원할 뿐만 아니라 로컬 검색에 최적화될 수 있도록 워크플로를 최적화하고 다양한 언어권의 전문지식을 제공합니다. 이에 따라 콘텐츠를 뛰어난 품질로, 필요한 언어에 맞춰 적절한 기간 내에 합리적인 비용으로 번역할 수 있었습니다. 이 서비스를 통해 기존 Royal Caribbean 콘텐츠의 80% 이상(약 1,500페이지)이 불과 4개월 만에 7개 언어로 번역되었습니다.

1,500

페이지를 7개 언어로 4개월 만에 Lionbridge에서 번역

"Adobe Experience Manager용 Lionbridge Connector를 사용하여 웹 번역 프로세스를 자동화함으로써 수천 개의 웹페이지를 손쉽게 번역할 수 있었죠." Royal Caribbean Cruise Lines의 상품 콘텐츠 최적화 부문 담당자인 Ariadna Castro의 말입니다. "Lionbridge가 제공하는 기술을 통해 우리는 효율적인 번역 워크플로도 구현할 수 있었습니다. 고객관리팀도 훌륭했고요. 그들은 우리의 목표를 지원하기 위해 언제든 뛰어들 준비가 되어 있었거든요."

로컬라이제이션 프로젝트의 최우선 목표는 고객에게 더 나은 사용자 경험을 제공하는 것입니다. 그러나 기업에 있어 로컬라이제이션이 왜 유익한 투자인지 보여주는 가시적인 성과도 중요하죠. Royal Caribbean의 경우, 그 효과를 거의 즉시 확인할 수 있었습니다. Royal Caribbean은 서비스 채택 후 처음 몇 년 동안 유기적 클릭수가 35% 증가했고, 유기적 검색 노출수는 37% 증가했으며, 글로벌 클릭수는 무려 3배나 향상되었다고 발표했습니다.

로컬라이제이션 정비를 통해 전 세계 고객에게 보다 맞춤화된 환경을 제공함으로써 고객은 물론 Lionbridge와 Royal Caribbean 모두가 큰 수확을 거둘 수 있었습니다.

8

로컬라이제이션을 통해 수백만 게이머들의 마음을 사로잡은 Zynga

"우리는 로컬라이제이션이라는 용어 대신 '컬처럴라이제이션'이라는 용어를 사용합니다. 즉, 언어별로 제품이 읽히는 방식을 바꾸는 것뿐만 아니라 게임 플레이 기능과 게이머의 참여 기능도 현지 시장에 맞게 역동적으로 조정하려는 거죠." - Zynga 퍼블리싱 부문 책임자 Bernard Kim

퍼블리싱 부문 책임자 Bernard Kim과 전략적 파트너십 부문 부서장 Vineet Rajosi Sharma 가 전하는 전 세계가 공감하는 게임 개발 팁

게임은 국경을 초월한 전 세계인의 사랑을 받고 있습니다. 오늘날의 게임 업계에서 성공하기 위해서는 대형 게임 개발사들의 접근법을 따를 필요가 있습니다. 2007년에 설립된 Zynga는 매일 전 세계 수백만 명이 즐기는 소셜 게임 업계의 세계 최고 개발사 중 하나입니다. 가장 잘 알려진 타이틀 중 하나인 FarmVille은 2009년 6월 Facebook을 통해 출시된 후 몇 주 만에 일일 활성 사용자 1000만 명을 달성했습니다. 그후로도 10여 년간 Zynga는 모바일을 중심으로 Zynga Poker, Words With Friends 2, Game of Thrones Slot Casino, Harry Potter Puzzles and Spells 등의 타이틀로 150여 개국에서 매월 수천만 명의 플레이어를 계속 유치하고 있습니다.

Zynga가 자사 게임을 지속적 자산이라 부를 만큼 "영원한 게임"으로 성공시킬 수 있었던 비결 중 하나는 자사 상품을 전 세계 시장에 맞춰 조정하는 일에 주력한 데 있습니다. 어차피 영어를 모국어나 제2외국어로 사용하는 사람은 세계 인구의 12분의 1도 되지 않는데 게임을 영어로만 즐길 수 있게 한다면 잠재 사용자 기반 중 상당수를 놓치게 되는 셈이니까요. 게임을 여러 언어로 출시하는 것은 상업적 관점에서도 꽤 일리가 있습니다. 그러나 Zynga 경영진이 AI Business에 말했듯이, 성공적인 로컬라이제이션을 위해서는 게임 속 텍스트 번역 그 이상이 요구됩니다.

'컬처럴라이제이션'

"우리는 로컬라이제이션이라는 용어 대신 '컬처럴라이제이션'이라는 용어를 사용합니다." Zynga의 퍼블리싱 부문 책임자 Bernard Kim의 말입니다. "즉, 언어별로 제품이 읽히는 방식을 바꾸는 것뿐만 아니라 게임 플레이 기능과 게이머의 참여 기능도 현지 시장에 맞게 역동적으로 조정하려는 거죠. 예를 들어, 한국 게이머들은 다른 지역의 게이머들보다 하루 중 더 많은 시간을 게임에 할애하고 우리 게임에 몰입하는 정도도 더 높습니다. 따라서 우리는 현지 시장에 컬처라이즈된 이러한 플레이어들을 위한 새로운 제안과 도전을 제시하기 위해 노력하고 있죠."

Zynga의 전략 파트너십 부문 담당자 Vineet Rajosi Sharma가 AI Business에 말했듯이, 게이머는 게임이 자기 자신만을 위해 특별히 설계되었다는 인상을 받을 수 있어야 합니다. "제가 사용자라면 앱에 접속했을 때 이 게임이 나를 위해 만들어졌다고 느낄 수 있어야 해요." ›

여러 나라의 수백만 사용자들을 동시에 만족시키면서 각 개인에게 맞춤화된 느낌을 주도록 게임을 만들기란 쉽지 않습니다. 게임 업계의 베테랑인 Kim은 전 세계적으로 IP 번역이 잘 맞지 않는 문제로 곤란을 겪은 적이 많다고 밝혔습니다. 그가 배운 한 가지 교훈은 타이틀을 특정 시장에 맞게 로컬라이즈하는 것은 더 이상 '부가 서비스'가 아니란 것입니다.

"게이머들은 자신이 이 게임을 얼마나 오래 즐기게 될지 처음부터 직감적으로 알죠." Kim은 말합니다. "느낌, 언어, 게임의 표현 방식 등, 특정 시장에 진출할 때 이들 요소 중 어느 하나라도 뒤늦게 이루어지면 게이머 참여 수에서 부정적인 영향을 미치게 될 겁니다."

이에 대비하여 Zynga는 새로운 프로젝트를 시작하는 바로 그 순간부터 로컬라이제이션(또는 컬처럴라이제이션) 전략을 고려합니다. 이런 것은 모두 '깔때기 넓히기', 즉 잠재적 참가자 수를 늘리기 위한 것입니다.

"우리의 영구적 프랜차이즈를 보면 알겠지만 이런 게임은 전 세계에서 통용되는 매력이 있습니다."라고 Sharma가 말하며, 어떤 느낌의 구상을 통과시킬지 선택할 때 "컬처럴라이즈가 가능한지 여부에 높은 우선순위를 두고 있죠. 해당 시장에서 이 앱이 어떻게 받아들여질지 모든 측면을 이해할 필요가 있으니까요."라고 덧붙입니다.

언어: 중요하지만 유일하지는 않은 요소

미국, 캐나다, 인도, 아일랜드, 터키의 사옥을 비롯해 전 세계에 스튜디오를 두고 있는 Zynga에게도 언어는 물론 중요한 고려사항입니다. Sharma는 로컬라이제이션 프로세스를 단순화할 수 있는 팁, 즉 텍스트 분량을 줄이는 방법을 공유했습니다. "게임에 텍스트를 너무 많이 넣으면 기본적으로 일이 더 복잡해지죠. 그리고 그렇게 많은 텍스트를 읽고 싶어하는 사람도 없고요." 화면이 작은 모바일 기기의 경우에는 더욱 그렇습니다.

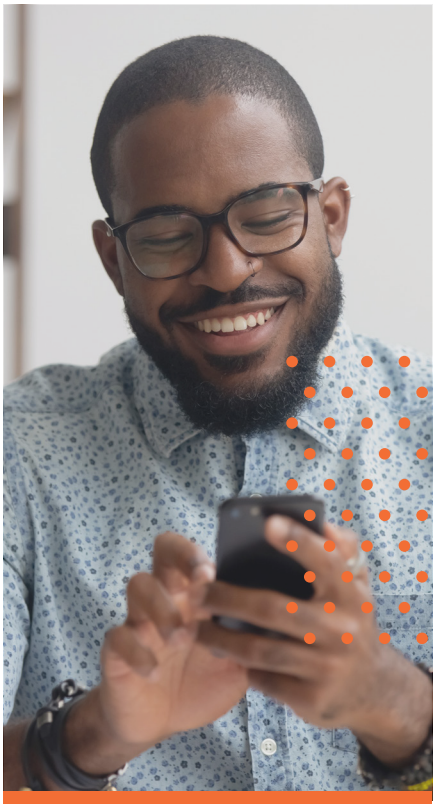
설상가상으로 텍스트를 번역할 때는 원본 타이틀의 '목소리'를 담아 내야 합니다. 이러한 텍스트에는 게임 내 텍스트와 표기법은 물론, 게임 속 가상 세계에서 통용되는 화폐, 측정 단위, 날짜 형식 등의 요소도 포함됩니다. 또한 앱스토어 설명, 마케팅 자료, 그 밖에 다른 첨부 텍스트도 번역해야 합니다. Sharma가 설명합니다. "번역은 사내 번역사를 고용하든지 아웃소싱하는 방법으로 해결할 수 있습니다. 하지만 우리가 최우선으로 삼는 것은 번역의 질이죠. 따라서 우리는 로컬라이제이션 QA(또는 내부적으로 '인터내셔널 QA') 팀을 두고 입고된 번역이 일관되고 적절한지를 확인합니다."

문어체나 구어체는 광범위한 컬처럴라이제이션 작업 중 빙산의 일각일 뿐입니다. 예를 들어, 특정 상징과 성별 표현 방식, 그리고 지위도 고려해야 합니다. 현지에서 인기 있는 소셜 미디어 플랫폼과의 통합, 마케팅 목적, 인증, 소셜 공유 등도 컬처럴라이제이션의 대상입니다. Kim은 시장마다 달라지는 중요 세부 사항 중 하나로 게임 내 '인앱 구매 패키지의 형태'를 언급하며, 이는 특정 지역의 플레이어가 수용할 수 있도록 조정되는 경우가 많다고 말합니다. 이전에 Zynga는 Harry Potter 타이틀의 특정 아시아 시장 출시 때 최적의 경험이 가능하도록 맞추기 위해 게임 출시를 보류하기도 했습니다.

"이런 이유로 우리는 번역가, 즉 사람이 번역합니다. 번역가들은 자신이 작업한 게임에 전문화되어 있으며 그 게임과 관련된 작업에 정말 오랜 시간을 함께하죠."라고 Kim이 말합니다.

무료 플레이 게임 시장에서 로컬라이제이션의 중요성

Zynga의 경영진들은 무료 플레이 게임(부분 유료화) 시장에서 로컬라이제이션의 중요성을 지적합니다. 정찰가를 부과하는 프리미엄(premium) 타이틀과 달리, '무료 프리미엄(freemium)' 게임은 플레이어들을 게임에 다시 불러옴으로써 수익을 얻습니다. 이는 현지 마케팅 캠페인에만 집중하고 게임 자체에는 최소한의 변화만 주는 겉핥기식의 로컬라이제이션으로는 부족하다는 것을 의미합니다. Kim도 강조하듯이, 수많은 타이틀로 혼잡한 게임 시장에서 게이머들은 자신이 가장



선호하는 타이틀로만 돌아가게 되므로 무료 프리미엄(freemium) 게임은 높은 품질을 보장해야 합니다. 다시 말해 이러한 게임은 로컬라이즈 전략을 바탕으로 개발될 필요가 있습니다. Statista에 따르면 전 세계 인구의 약 48.33%인 38억 명이 스마트폰을 소유한 상황에서, 로컬라이제이션은 모바일 게임 스튜디오를 위한 현명한 투자가 아닐 수 없습니다.

"전반적으로, 우리는 컬처럴라이제이션을 창조적 활동으로 봅니다. 광고 소재나 현지 마케팅 캠페인을 통해 사용자가 우리 게임에 대해 처음 접하는 순간부터 플레이어가 되기까지, 컬처럴라이제이션은 CRM(고객 관계 관리) 홍보 활동과 함께 계속되죠. 알림을 어떻게 보낼지, 가령 이메일로 보낼지 아니면 문자 메시지로 보낼지, 몇 시에 보낼지 등도 신중하게 고려해야 합니다. 자고 있는 사람에게 푸시 알림을 보내거나 잘못된 언어로 메시지를 보내면 안 되니까요. 이 모든 요소들이 어우러져 컬처럴라이제이션이 되는 거죠."라고 Kim은 말합니다.

10

언어 서비스를 통한 Wikipedia 개선

"기계학습은 독자와 편집자의 경험을 개선하기 위한 모든 곳에 사용되고 있습니다. 예를 들어, 우리는 편집 품질을 조사하고 이러한 편집이 해로운지 아닌지를 예측하기 위한 모델을 가지고 있죠." Albon의 말입니다.

Wikimedia Foundation의 기계학습 부문 부서장 Chris Albon이 생각하는 프로젝트의 임무와 현지 데이터의 중요성

세계 최대의 사용자 제작 백과사전이라면 가능한 한 많은 사람이 접근할 수 있어야 합니다. 이 귀중한 리소스를 전 세계 사용자에게 보다 원활히 제공하는 데 로컬라이제이션 서비스가 도움이 되지 않을까요? 300개의 언어, 4천만 개 이상의 항목을 자랑하며 매달 5억 명의 사용자에게 서비스를 제공하는 Wikipedia 의 제작자들이 궁금했던 것도 바로 이것입니다. 그 대답은, 당연히 도움이 됩니다.

Wikipedia 의 모든 버전에 가능한 한 포괄적인 정보를 담을 수 있도록 하는 가장 확실한 방법은 모든 Wikipedia 버전을 종합한 다음 자동 번역을 수행하는 것입니다. 예를 들어, 영문 문서(영어는 Wiki 버전 중 규모가 가장 크므로)를 사용자 수가 적은 다른 언어로 번역하는 방법이죠.

Wikimedia Foundation은 이러한 제한된 일부 측면을 통합하며, 먼저 Wikimedia에서 지원하는 콘텐츠 번역 툴을 사용해 Wikipedia 문서를 기계 번역한 후 편집자인 사람의 검토를 거쳐 필요에 따라 내용을 다듬은 뒤 게시합니다. 이때 자동으로 진행되는 기계 번역은 문제를 일으킬 소지가 있으므로 사람에 의한 검토 과정이 필수적입니다. 관련 오류의 한 예로, 영어 'village pump(마을 펌프)'가 포르투갈어로 '마을 폭격'으로 번역된 경우를 들 수 있습니다. 이러한 오류는 중간에 사람을 배치함으로써 최소화할 수 있습니다. 이 자동화 도구는 현재까지 50만 건 이상의 문서를 번역하는 데 사용되었습니다.

하지만, Wikipedia의 기계학습 본부장 Chris Albon이 AI Business와 인터뷰한 내용에 따르면, Wikipedia는 자동 번역을 최대한 활용하기보다는 "지역 사회의 실제 의견을 보다 잘 반영하는" 접근 방식을 채택했습니다.

로컬라이즈된 기계학습

Wikipedia는 많은 사용자들이 생각하는 것보다 훨씬 더 광범위한 영역에서 기계학습 모델을 사용합니다. " 기계학습은 독자와 편집자의 경험을 개선하기 위한 모든 곳에 사용되고 있습니다."라고 Albon은 말합니다. "예를 들어, 우리는 편집 품질을 조사하고 이러한 편집이 해로운지 아닌지를 예측하기 위한 모델을 가지고 있죠."

Wikipedia는 가능성 있는 문서 주제를 제안하는 것부터 반달리즘 행위를 단속하는 것에 이르기까지 페이지의 가독성 유지를 위해 개발된 수백 개의 다양한 복합 '봇'을 사용하고 있습니다. 하지만 Albon과 그의 동료들(Wikipedia 문서를 직접 작성하지는 않음)은 이 무료 백과사전을 최상의 상태로 유지하고자 할 뿐, 그들이 만든 AI 모델에 그들 자신의 편견을 끼워넣는 부담은 꺼려 합니다.

Wikipedia의 경우, 어떤 편견을 말하는 걸까요? 간단히 말해, 샌프란시스코에 사는 사람이 생각하는 Wikipedia에 필요한 것, 또는 좋은 Wikipedia 문서의 기준을 다른 문화적 기대치를 가진, 예컨대 케냐의 Wikipedia 사용자에게 자동으로 적용할 수 없다는 것입니다.

"제가 가장 많이 읽은 Wikipedia 는 분명 영어 Wikipedia 입니다." Albon이 설명합니다. "따라서 무엇이 좋은 문서인지, 문서의 주제는 어떠해야 하는지, ›

> 문서 구조는 어떤 형태가 바람직한지, 그 문서는 어떤 어조로 작성되어야 하는지에 대한 제 의견은 모두 Wikipedia 프로젝트의 한 지역 사회에 대한 저의 경험에 바탕을 두고 있죠. 그리고 다른 모든 지역 사회가 저와 같은 의견을 공유하지도 않고요. 우리 모두는 Wikipedia 프로젝트의 일부라는 폭넓은 공감대를 형성하고 있습니다. 하지만 무엇이 좋은 편집인지, 무엇이 해로운 편집인지, 또는 특정한 주제를 어떻게 다루어야 할지에 관한 문제는 매우 지역적입니다. 그것은 지역 사회에 의해 현지 언어로 관리되어야 하는 것이죠."

Wikipedia의 임무에 포함되는 로컬라이제이션

Wikipedia는 현재 오직 현지 데이터로만 훈련된 100여 개의 기계학습 모델을 운영하고 있습니다. 이 모델들은 약 30개의 서로 다른 Wiki에서 사용되고 있습니다. Albon은 Wikipedia의 목표가 전 세계 모든 사람에게 그들 자신의 언어로 된 무료 백과사전을 제공하는 것이라고 언급하며 "Wikipedia에게 로컬라이제이션이 매우 중요한 이유는 로컬라이제이션이야말로 전체 프로젝트를 포괄하고 있기 때문이죠."라고 말합니다.

각 모델을 현지 데이터로 학습시키는 이유는 국가별 사용자마다 좋은 Wikipedia 문서의 요건에 대해 완전히 다른 기대치를 가질 수 있기 때문입니다.

"이것이 중요한 이유는, 지역적 맥락으로 인해 스와힐리 공동체가 문서에서 중요시하거나 중요시하지 않는 요소, 중요한 관련 항목, 검토되지 않은 토픽 등이 영어 Wikipedia에서의 경우와 다를 수 있기 때문이죠. 예를 들어 영어 Wikipedia에서 마타투(개인 소유의 소형버스)는 별로 중요하게 간주되지 않습니다. 하지만 스와힐리어 Wikipedia에서는 매우 중요하죠. 마타투는 동아프리카, 남아프리카, 서아프리카 사람들의 일상 생활에 필수적인 운송수단이니까요 … 영어 Wikipedia를 토대로 모델을 만든 후 그것을 스와힐리어 Wikipedia에 적용하는 것은 우리의 견해를 강요하는 것이라 그것은 바람직하지 않아요."라고 Albon은 말합니다.

Wikipedia에서 사용되는 기계학습 모델 중 하나는 문서의 특정 단어가 다른 페이지로 연결될 필요가 있을 경우 이를 제안하도록 설계되었습니다. 이론적으로, Wikipedia의 대다수 명사는 자체적인 페이지를 가질 수 있고 문서 전체가 하이퍼링크로 가득 차게 될 수도 있으므로 페이지가 이러한 링크로 가득 차게 되면 주의가 산만해져 독자는 정작 자신이 찾으려는 링크를 놓칠 수도 있습니다. 이에 따라 Wikipedia는 거의 모든 방문객이 확실히 알 정도로 기초적인 '상식' 정보를 참조하는 링크는 배제하는 입장입니다. 하지만 여기에도 문제가 있는 것이 지역에 따라 상식이 다를 수 있기 때문입니다. 따라서 이 작업을 위해 배치된 AI 모델은 상식이 무엇인지에 대해 로컬라이즈된 접근 방식을 취할 필요가 있습니다.

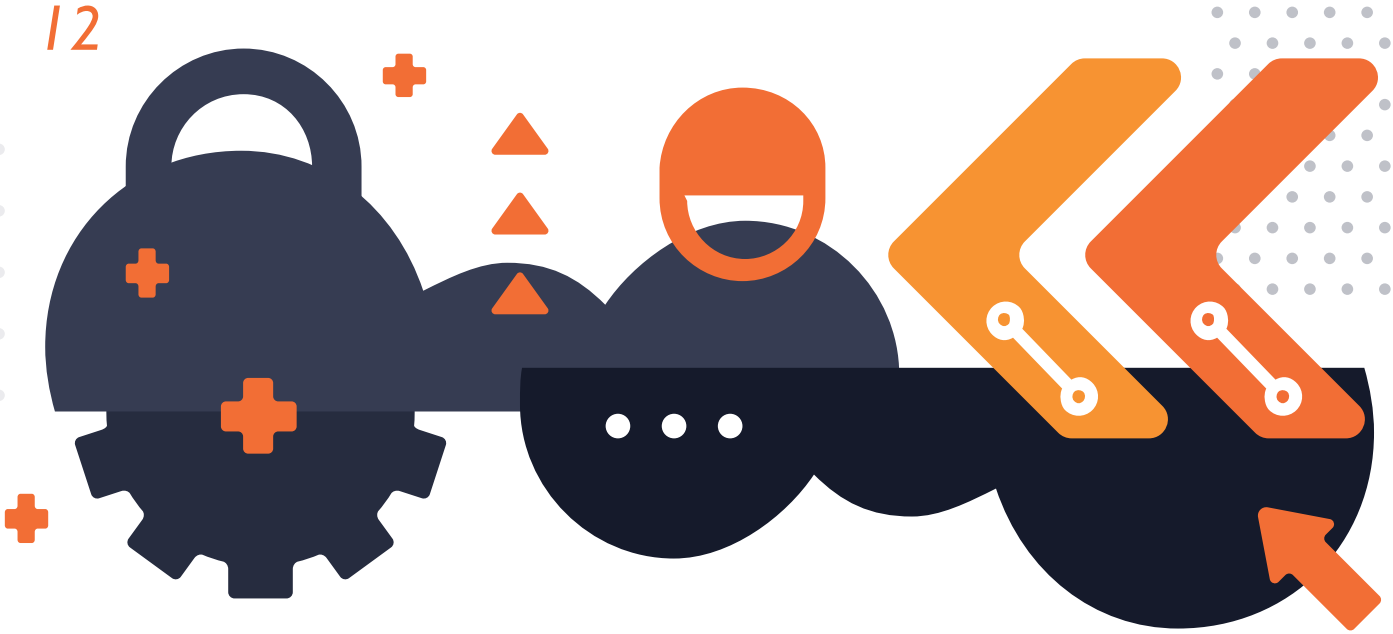
최적의 기준

"모델 제작에 대한 최적의 기준은 기계학습 모델을 훈련시킬 때 해당 모델이 서비스되는 지역의 데이터를 사용하는 것이죠." Albon의 설명입니다. "예를 들어, 만약 스와힐리어 Wikipedia에서 '편집 품질을 탐지하는 모델'을 원한다고 할 때 최적의 기준은 그 지역 사회에서 좋은 문서란 무엇을 의미하는지를 학습시킬 수 있는 데이터를 수집하는 것입니다. 이 데이터로 모델을 학습시킨 후 편집 품질을 탐지하는 거죠."

이처럼 여러 버전의 Wikipedia 품질이 향상되면 유용성도 더 높아지므로 결국 더 많은 사용자를 끌어들일 수 있습니다. 그런 후 이러한 새로운 사용자로부터 받은 피드백을 바탕으로 기계학습 모델을 더욱 스마트하게, 더 높은 수준으로 로컬라이즈할 수 있습니다. 이처럼 고유한 '선순환' 방식으로 접근한 로컬라이제이션에 따라 온라인 자원의 가치를 훨씬 더 높일 수 있습니다.

"그래서 지역 사회의 참여가 특히 중요합니다."라고 Albon은 말합니다. "이들의 참여가 없으면 Wikipedia는 Wikipedia가 다른 모든 이들에게 어떻게 보여야 하는지에 대한 특정 개념을 가진 특정 지역 사람들의 것이 되고 말기 때문이죠. 사람들은 기계학습 모델을 통한 편견 세탁에 대해 말하곤 하는데, 그것이야말로 정확히 우리가 피하고 싶은 것입니다. 우리는 지역 사회가 그들의 모델에 발언권을 가지고 그러한 모델을 만드는 과정에 참여하고 그러한 모델을 통제하는 주체가 되기를 원합니다. 결국, 이러한 AI 모델은 지역 사회 환경을 개선하는 데 사용되어야 하죠."

12



변곡점에 서서

기계학습과 언어의 관계: Lionbridge의 Rafa Moral 과 Kajetan Malionowski 와의 대담

번역은 오랫동안 인공지능의 한 부분이었습니다. AI가 독자적인 학문으로 확립되기 20여 년 전인 1933년, 소련의 과학자 Peter Troyanskii는 USSR 과학 아카데미에서 '단어를 선택 및 출력함으로써 한 언어를 다른 언어로 번역하는 기계'를 선보였습니다. 1964년 뉴욕 세계 박람회에서 IBM은 영어와 러시아어를 번역할 수 있는 초보적인 기계 번역 툴을 선보였습니다.

기계 번역을 AI 범주에 속하는 많은 하위 분야 중 하나로 여기는 경향이 있습니다. 하지만 Lionbridge 혁신 부문 부사장인 Rafa Moral은 기계 번역을 이 기술의 핵심 주제 중 하나라고 말합니다. "결국 AI는 인간의 지능을 모방하는 것이죠. 그리고 언어보다 더 인간적인 것은 없습니다. 따라서 AI 연구원이 사람처럼 행동하는 무언가를 만든다고 할 때 그것은 개념적으로 언어를 다루는 무언가를 의미합니다. 이것이 바로 처음부터 언어가 AI의 주제, 어쩌면 가장 중요한 주제라고 생각하는 이유죠."

더 강력한 기계 번역

AI의 다른 분야들처럼, 기계 번역도 그동안 많은 변화를 겪었습니다. 그 결과 기대되던 연구개발(R&D) 과정을 마치고 완전한 도구로 발돋움함에 따라 최근 몇 년간 전례 없는 혁신을 이루었습니다.

이러한 획기적 발전 중 하나가 바로 단어 임베딩 개념입니다. 기계학습과 딥러닝 툴셋의 일부인 단어 임베딩은 유사한 의미를 가진 단어는 유사한 방식으로 제시된다는 개념의 단어 표현 방식을 말합니다. 비록 의미가 비슷한 단어라도 서로 완전히 다른 방식으로 표현되었던 이전의 접근법과는 달리, 단어 임베딩 과정에서 자연스럽게 단어의 의미를 포착하므로 번역 프로세스가 크게 개선될 수 있습니다.

Moral은 이렇게 설명합니다. "임베딩 이전에 단어는 그저 텍스트 문자열이었습니다. 이제는 임베딩 개념에 따라 벡터화의 개념, 단어를 숫자로 바꾸는 방법, 그리고 그 숫자들에 어떤 의미론적 의미가 포함되는지 등의 개념이 나오죠. 임베딩은 문맥에 따라 단어와 연결할 숫자 또는 벡터를 식별하는 것입니다."

신경망 MT: 확실한 변곡점

최근 몇 년간의 가장 획기적인 혁신 중 하나는 Moral이 업계의 '확실한 변곡점'이라고 칭한 신경망 기계 번역의 출현입니다. 신경망 기계 번역(NMT)은 이전에 사용된 규칙 기반 MT(RbMT) 및 통계적 MT(SMT)와는 다른 완전히 자동화된 기계 번역 방식으로, 최근의 많은 기계학습 모델처럼 뇌에서 영감을 받은 도구인 신경망에 의존합니다. NMT는 모든 단어를 따로따로 분석하는 대신 단어가 사용된 맥락을 고려함으로써 더 자연스럽고 더 정확한 번역을 제공합니다. 각각의 단어뿐만 아니라 그 단어들 사이의 관계도 분석하는 것이죠.

Lionbridge는 문서 및 문헌의 고품질 번역을 위한 무기 중 하나로 NMT와 같은 기술을 사용해 대폭 강화된 로컬라이제이션 서비스를 제공할 수 있었습니다. ›

Lionbridge의 제품 부문 부서장이자 전략가인 Kajetan Malinowski는 AI Business와의 인터뷰에서 이렇게 말합니다. "번역해야 할 콘텐츠는 세상 모든 번역가를 동원해 처리할 수 있는 양보다 수백 배는 더 많습니다. 그런 점에서 기계 번역은 산업을 혁신하고 더 많은 콘텐츠를 관리하며 특히 인간이 보다 큰 가치를 더할 수 있는 곳에 집중하도록 기여할 수 있지요."

Moral과 마찬가지로 Malinowski는 NMT와 같은 접근법의 출현이 자동 번역에서 매우 결정적이었다고 말합니다. 그러나 고품질 번역을 위해서는 어떤 접근법을 사용해야 하는지 아는 것도 중요합니다. 특정 유형의 콘텐츠에는 특정 방식이 더 적합할 수 있으니까요.

최적의 결과를 위해서는 어떤 방법을 사용해야 하는가

"콘텐츠는 그 유형에 따라 각기 다른 목적을 가집니다." Malinowski의 말입니다. "수명이 짧고 가시성이 낮은 콘텐츠는 아주 우수한 품질의 번역을 요하지는 않죠. 반면에 마케팅 카피나 광고, 특정 분야의 법률, 의료, 규제 문서 등의 콘텐츠에는 매우 높은 수준의 번역이 요구됩니다. 이때 기계 번역을 사용하면 더 많은 콘텐츠를 처리할 수 있을뿐더러 사람은 '부가 가치'가 더 높은 작업에 집중할 수 있게 되죠."

예를 들어 Lionbridge에서는 번역에 들어가기 전에 AI 기반 툴을 사용해 어떤 방법을 사용하면 최적의 결과를 얻을 수 있을지 예측합니다.

"우리는 각 접근법의 장점을 최대한 활용하는 하이브리드 프로세스를 통해 어휘 밀도와 다양성, 복잡성과 가변성, 품질 등 이러한 모든 문제를 식별하고 있죠." Malinowski의 설명입니다. "우리의 작업은 여기에서 출발합니다. 고객의 콘텐츠를 명확히 파악하면 해당 콘텐츠 유형에 가장 적합한 프로세스인 워크플로를 식별할 수 있죠. 여기에는 사용하기에 가장 좋은 번역 메모리, 언어 또는 도메인에 사용할 수 있는 최고의 기계 번역 엔진, 그리고 특정 콘텐츠 유형에 적합한 툴셋 등이 포함됩니다."

고객의 콘텐츠를 명확히 파악하면 해당 콘텐츠 유형에 가장 적합한 프로세스인 워크플로를 식별할 수 있죠. 여기에는 사용하기에 가장 좋은 번역 메모리, 언어 또는 도메인에 사용할 최고의 기계 번역 엔진, 그리고 특정 콘텐츠 유형에 적합한 툴셋 등이 포함됩니다.

전체 주기에서 사람의 역할

5페이지의 증강 번역에 대한 글에서 설명한 바와 같이, 많은 경우 최선의 선택은 기계 번역과 사람의 전문지식을 접목하는 것입니다.

"기계 번역 엔진은 그 성능이 월등히 개선되어 그저 적당히 양호한 품질이면 충분한, 수명이 짧고 가시성이 낮은 웬만한 종류의 콘텐츠 유형을 자체 처리할 수 있습니다." 라고 Malinowski는 말하며 덧붙입니다. "좀 더 나은 품질이 필요하면 그때 사람이 투입되어 콘텐츠 품질을 향상시키죠. 이를 통해 사람은 좀 더 주의를 요하는 고부가가치 콘텐츠에 보다 집중할 수 있습니다. 예컨대 잘 읽혀야 하고, 어떤 흐름을 가지고 있어야 하며, 대상 고객에게 특정한 감정을 불러일으켜야 하는 종류의 콘텐츠 말이죠. 기계 번역은 이런 콘텐츠는 처리하지 못하거든요. 아직은요."

언어의 주관성, 그리고 좋은 번역의 구성 요건에 대한 모호성은 최고의 기계 번역 툴에서도 여전히 해결하기 어려운 과제이지만, 현대 번역은 이 분야의 발전으로부터 엄청난 혜택을 얻었습니다. Malinowski는 다음과 같이 말을 맺습니다. "인류는 불과 1, 2년 전에 비해서도 기하급수적으로 많은 콘텐츠를 생산하고 있으며 지금도 콘텐츠 양은 계속 증가하고 있습니다. 기계 번역, 그리고 지금 우리에게 주어진 모든 AI와 기계학습 툴은 우리가 더 많은 콘텐츠를 만들고, 더 많은 고객과 대화하고, 콘텐츠를 더 잘 처리할 수 있도록 도와주고 있지요."

14

Volvo Cars 의 통합 온라인 시장 구축을 도운 Lionbridge

"우리는 이처럼 산발적으로 분산된 브랜드 스토리에서 벗어나 단일 플랫폼을 통해 일관된 브랜드 목소리를 내고자 했습니다.

과거 약 100개의 다른 웹사이트를 운영했던 회사

Volvo Cars는 2030년까지 100% 디지털 온라인 판매 및 전기자동차로 전환하기로 결정한 후 Lionbridge와 비즈니스 서비스 및 클라우드 솔루션 업체인 Avanade와 함께 파트너십을 맺고 여러 변화를 모색했습니다. 여기에는 온라인에서의 입지를 정비하고 웹사이트를 단순한 마케팅 툴이 아닌 디지털 시장으로 재창조하는 과정이 포함되었습니다. 이를 위해서는 전 세계 잠재 고객에게 정보를 제공할 수 있는 만족스러운 사용자 환경을 구축해야 했습니다.

해결 과제

재정비 이전,Volvo Cars 웹사이트는 비록 점유율은 높았지만 극도로 분산되어 있었습니다. 이 회사는 총 45개 언어로 약 100개의 웹사이트를 운영하고 있었는데, 각 웹페이지는 하나의 중앙 팀에서 제작했지만 번역은 현지 팀이 담당했습니다. 그 결과, 시장마다 서로 다른 웹사이트 설정과 판매 전략을 채택하면서 전반적인 브랜드 일관성에 악영향을 미쳤습니다.

1,000

시간의 업무 단축 효과를 가져온 신규 시스템

Volvo Cars는 웹사이트를 중앙집중화함으로써 번역을 중앙에서 관리하고 이 작업을 진행하는 사람들과 직접 접촉하는 등 하향식으로 로컬라이제이션 작업을 통제하길 희망했습니다. Volvo Cars의 글로벌 온라인 디지털 생산 책임자인 Cecilia Ernby는 이렇게 설명합니다. "온라인 판매로 옮겨가면서, 우리는 자체적으로 메시지를 전달할 필요가 있었죠. 우리는 이처럼 산발적으로 분산된 브랜드 스토리에서 벗어나 단일 플랫폼을 통해 일관된 브랜드 목소리를 내고자 했습니다."

솔루션

Volvo Cars는 번역 프로세스를 최적화하기 위해 Smartling 번역 관리 시스템을 활용함으로써 번역 작업을 관리하는 데 걸리는 시간을 약 1,000 시간 단축했습니다. 이러한 작업에는 이전에 각 시장마다 독자적으로 사용한 번역 메모리를 중앙집중화하는 것도 포함되었습니다. 또한 번역 워크플로를 수정하여 현지의 검색 엔진 최적화 (SEO) 요건에 맞도록 콘텐츠를 적절하게 최적화하는 단계가 추가되었습니다.

중앙집중화와 로컬라이제이션: 무시할 수 없는 힘

이러한 변화의 결과로,Volvo Cars 웹사이트는 이제 여러 다른 시장 전반에 일관된 이미지를 선보입니다. 즉, 고객은 영국에서 제품 페이지를 보든, 독일 또는 벨기에에서 보든 관계없이 모두 동일한 사용자 여정을 경험할 수 있습니다. 모든 콘텐츠가 고품질 수준으로 번역된 것은 물론입니다. 이에 따라 Volvo Cars는 여러 지역에서 더 원활히 캠페인과 제품을 출시할 수 있게 되었습니다.

"굉장히 많은 캠페인을 동시에 출시할 수 있었어요. 예전에는 절대 불가능했을 일이죠."라고 Ernby가 말합니다.

이 사례 연구는 지역 고객 수요에서 오는 이점을 살리면서도 일관된 사용자 경험을 제공하기 위해 어떻게 로컬라이제이션과 중앙집중화를 함께 운영할 수 있는지 보여주는 귀중한 사례입니다.

언어 기술의 미래

웨어러블 기기, 증강 현실, 초대형 언어 모델의 잠재력에 주목하다

기계 번역의 한계에 대해 항간에 떠도는 오래된 이야기가 있습니다. 초기 MT 툴에 "The spirit is willing, but the flesh is weak(마음은 굴뚝 같은데 몸이 안 따라준다)"라는 영어 문장을 입력하고 러시아어로 번역하면 "위스키가 간절하지만 고기는 썩었다"라는 결과가 나온다는 것이죠.

기계 번역과 언어 기술은 산업 초기 이후 큰 발전을 이루었습니다. 하지만 그 지지자들이 항상 믿어온 것처럼 언어 기술은 우리 삶의 필수적인 부분이므로 앞으로 더욱 발전할 여지는 얼마든지 있습니다.

차세대 언어 기술의 가장 인상적인 데모 제품 중 일부는 하드웨어에 뿌리를 두고 있습니다. 빠르게 성장 중인 웨어러블 기술과 연관된 이러한 혁신은 번역 기술이 언제 어디에나, 최종 사용자가 눈치채지 못할 만큼 자연스럽게 우리 생활에 자리 잡을 것을 암시합니다. 과학 소설에서 오랫동안 꿈꾸어 온 것처럼, 어떤 언어든 즉시 번역해주는 '범용 번역기'가 실현될 날도 머지 않았습니다.

범용 번역기로 향하는 길

이 기술이 실현된 한 가지 예는 언어 번역기 이어버드로, 음성 처리와 기계 번역을 결합하여 구어 문장을 실시간으로 번역합니다. 대표적인 예로는 Pilot Smart Earbud(대화 녹취록을 제공하기도 함), Google Pixel Buds, WT2 Plus AI 실시간 번역 이어버드 등이 있습니다.

한편, 시각 활용 분야에서는 Google 번역의 증강현실(AR) 기능이 독보적인 위치를 차지하고 있습니다. Google Lens라고 부르는 이 제품은 사용자가 다른 언어로 된 표지판이나 메뉴 등의 텍스트에 스마트폰 카메라를 갖다대면 액정 화면 위에 번역문이 덧씌워 표시됩니다. 번역문은 보통 원본과 동일한 서체로 나옵니다. 이 기능은 기기에 탑재된 기계학습 툴을 사용하여 수행되므로 Lens는 인터넷에 연결되지 않아도 작동할 수 있습니다. 2021년 1월, Google Lens는 단독 앱이 Google Play 앱 스토어에서 5억 회 이상 다운로드되었다고 발표했습니다. Google Lens는 현재 100여 개의 Google 번역 언어 대다수를 지원하고 있습니다.

Google Lens는 상당히 진보된 기술처럼 보이지만, 이 기술은 실제로 AR 기반 번역 툴의 1세대를 대표합니다. Lens는 2017년 말에 처음 출시되었으며, 광범위한 언어를 지원하기까지는 한참 시간이 걸렸습니다. 하지만 스마트 안경과 같이 대중적인 헤드업 디스플레이가 실현될 경우 상황은 극적으로 바뀔 수 있습니다. AR은 전화기와 태블릿에만 국한되지 않고 이러한 웨어러블 장치에 도입되어 착용자는 언제 어디서나 더 자연스럽게 환경을 인식할 수 있습니다. 로컬라이제이션 관점에서, 이는 번역이 훨씬 더 간편하고 빠르게 이루어질 것임을 의미합니다.

소프트웨어 깊숙이

대중을 위한 이러한 데모 기술에서 우리는 언어 기술의 진일보를 확인할 수 있습니다. 하지만 고품질 번역이 필요한 사람들에게 진정한 혁신은 스마트 이어버드나 스마트 안경이 아니라, 소프트웨어 속 더 깊숙한 곳에서 빛을 발해야 합니다.

5페이지의 "증강 번역이란?" 글에서도 설명했듯이, 가까운 미래에 번역 기술에는 ›

AI와 번역사인 사람이 모두 참여하는 하이브리드 증강 접근법이 도입될 것이 분명합니다. 지적재산권, 법률, 생명과학 분야를 전문으로 하는 언어 서비스 제공업체인 Absolute Scientific Translation(ASTV)의 최고 경영자 Domenico Lombardini는 사람과 기계를 서로 비교하는 대중적 개념화 논리는 잘못되었다고 말합니다.

"새로운 기술로 사람의 개입이 어느 정도 억제되거나 감소할 수는 있지만, 이러한 기술은 또한 그만큼 새로운 전문적인 사람의 역할과 새로운 작업 방식을 만들어내고 있습니다. 현재 번역가가 하는 일 중 기계 번역 결과물을 편집하는 일의 비중이 계속해서 늘어나고 있지만, 기계는 여전히 오류를 범하기 쉽기 때문에 사람이 기여할 부분의 중요성은 근본적으로 바뀌지 않습니다. 따라서, 언어 전문가를 양성하는 교육은 최종 번역 품질을 보장하기 위한 필수 조건인 셈이죠."

번역 품질을 최종적으로 결정하는 것이 인간이라 해도 AI 툴은 증강 번역 작업에 여전히 매우 유용하게 사용될 수 있습니다. 예를 들어, AI는 특정 기술을 토대로 특정 프로젝트에 가장 적합한 번역사를 제안하는 추천 시스템을 만들 수 있습니다. 전문가는 알고리즘에 전체 번역 작업을 맡기는 대신 AI를 이용해 문서 내용을 분석한 후 키워드와 문체 관련 요구사항을 선별해 낼 수 있습니다. 또한 이러한 정보를 상호 참조하기 전에 텍스트에 태그를 달아 분류할 수도 있습니다.

차세대 문법 검사기나 철자 검사기처럼 언어 전문가인 사람의 번역 작업을 돕는 기계는 더욱 늘어날 전망입니다. 여기에는 오타 수정과 같은 단순 기능뿐만 아니라 문체에서의 비일관성이나 변칙, 검색 엔진 최적화 (SEO) 지점을 표시하거나 언어 전문가 (및 그들을 고용한 회사)가 궁금해하는 잠재적인 문화적 민감성 문제를 경고하는 기술도 포함됩니다.

대용량 언어 모델

AI 언어 모델은 로컬라이제이션 요건을 보조하기 위한 다른 작업에도 활용될 수 있습니다. 2020년 AI 분야에서 가장 중요한 발전 중 하나는 OpenAI에 의해 개발된 거대 언어 모델인 GPT-3의 등장입니다. GPT-3는 1750억 개의 매개변수를 가진 신경망을 자랑하며, 모델 그 자체도 45테라바이트의 데이터로 학습되었습니다. 이러한 분량의 데이터는 인터넷에서 사용할 수 있는 실질적인 콘텐츠 양과 동등한 수준입니다. 이처럼 많은 양의 데이터를 흡수함으로써 GPT-3는, 가령 '오렌지' 뒤에는 보통 '주스'가 뒤따른다는 것을 아는 등 통계적으로 각 단어 다음에 어떤 단어가 올 수 있는지 예측하는 작업을 매우 정확히 수행할 수 있습니다.

GPT-3는 이 방대한 훈련 데이터 말뭉치를 활용해 법률 텍스트를 알기 쉬운 영어로 바꾸거나 또는 그 반대로 바꾸는 것과 같이 서로 다른 '문체' 사이에서 언어를 변환하는 작업 등 광범위한 용도에 사용될 수 있습니다.

GPT-3와 같은 툴에서 우리는 언어 기술의 새로운 가능성을 엿볼 수 있습니다. 이제 기존 텍스트를 단순히 번역하는 것을 넘어 완전히 새로운 문서를 만들어낼 수도 있는 것입니다. Lionbridge의 제품 부문 본부장이자 전략가인 Kajetan Malinowski는 "AI는 우리가 단순히 번역을 넘어 언어로 더 많은 일을 할 수 있도록 도와줄 수 있습니다. 콘텐츠를 작성할 때도 AI의 도움을 받을 수 있죠. Lionbridge에서는 최근 GPT-3 모델을 테스트하고 있는데, 이 모델은 정말 다양한 용도로 사용될 수 있고 그중 가장 흥미로운 용도가 바로 콘텐츠 작성입니다."라고 말합니다.

이제 고객은 보다 정교하고 맞춤화된 메시지를 기대하고 있습니다. 이러한 때 대용량 언어 모델과 같은 혁신은 단지 국가나 지역을 위한 텍스트뿐만 아니라 잠재적으로 개별 고객을 위한 텍스트의 로컬라이즈에도 이용될 수 있을 것으로 기대됩니다.

언어 기술의 미래

기계는 인간의 언어 능력을 얼마나 따라잡았나?

2020년 주목할 만한 연구 논문인 "***언어 번역에서 인간-기계 동등성 평가를 위한 일련의 권고 사항*** *(웹 링크)*" 에서 저자들은 "지난 몇 년간 기계 번역의 성능이 현저히 개선되어, 다수의 실증적 조사에 따르면 전문가의 번역과 구별할 수 없을 정도로 발전한 것으로 나타났다."라고 주장합니다.

공동 저자 중 취리히 대학의 Samuel Läubli와 더블린 시티 대학의 Sheila Castilho가 기계 번역의 역량과 잠재력에 대해 AI Business와 이야기를 나누었습니다.

AIB: 논문에서 지난 몇 년간 기계 번역에서 큰 발전이 있었다고 언급하셨는데요, 이러한 발전은 주로 어떤 요인에 의한 것인가요?

Samuel Läubli (SL): 자원 민주화, 컴퓨팅 능력 향상, 상업적 관심 증가의 세 가지 요인을 생각해볼 수 있습니다. Paracrawl과 같은 시도 덕분에, 이제 기계 번역 시스템을 만들고자 하는 사람은 누구든 기본적으로 충분한 양의 학습 데이터를 사용할 수 있게 되었습니다. 기계 번역에 대한 신경망 접근법은 통계적 접근법과 같은 기존 방식보다 개념적으로 단순하고 다중 구현을 자유롭게 이용할 수 있어 학술용 또는 상업용으로 기계 번역 시스템을 구축하는 것이 그 어느 때보다 쉬워진 거죠.

대량 학습 데이터를 사용하여 신경망 모델을 학습시키려면 고도의 컴퓨팅 능력이 필요하지만, 이제는 몇 년 전과 비교해 같은 비용으로 훨씬 더 높은 컴퓨팅 성능을 얻을 수 있으므로 이 부분에 대해서는 Google, Microsoft, Amazon과 같은 대기업 이외의 조직과 개인에게는 더 이상 문제가 되지 않습니다. 그리고 상업적인 이익도 있죠. 기존 기술만으로도 번역 프로세스에서 이미 많은 시간과 비용을 절약할 수 있었지만, 신경망 모델링은 처음으로 진정한 의미에서 유창한 문장을 만들어내는 것을 보여줌으로써 의사결정자를 포함한 일반 대중에게서 감탄을 자아냈습니다.

AIB: 번역 정확도를 평가하는 가장 좋은 방법은 무엇인가요?

Sheila Castilho (SC): 대답하기 굉장히 어려운 질문이네요. 번역 평가에 대해서는 수천 편의 논문이 나왔지만 평가 방법은 물론 번역 품질의 정의에 대해서도 많은 저자들 사이에 이견이 있습니다. 제게 있어서 번역 품질을 측정하는 가장 좋은 방법은 먼저 해당 번역물에 요구되는 품질을 정의한 다음 거기서부터 방법론을 설계하는 겁니다.

SL: 저도 동의합니다. 번역의 목적이 노트북에서 발생한 문제를 사용자가 해결하게 하는 것이라면 전문 번역가에게 번역에서 언어적 오류를 찾아내고 분류하도록 하는 것은 적절한 방법이 아닐 테니까요. 하지만 한 가지 추가하고 싶은 것은, 어떤 작업이든 어느 시점에는 사람이 참여해야 한다는 점입니다. 번역 품질의 개선에서 확인된 것이 한 가지 있다면 그것은 ***BLEU*** ***(웹 링크)*** 와 같은 자동 번역 품질 평가법의 신뢰도가 점점 낮아지고 있다는 점입니다.

AIB: 조만간 선보일 가장 흥미롭고 유망한 번역 기술에는 무엇이 있나요?

SC: 다양한 접근 방식을 혼합한 방법은 무엇이든 흥미로워 보입니다. 가령 대화식 사후 편집과 발화의 결합 같은 거죠.

SL: 혼합 방식의 번역 외에도, 곧 두 가지 변화가 일어날 것으로 생각됩니다. 첫째, 기계 번역 시스템에서 문맥을 더욱 고려하게 될 것으로 보입니다. 즉, 문서를 문장 단위로 잘라서 각 문장을 따로 번역한 다음 다시 [연결]하기보다는 문서 전체를 하나로 번역하는 거죠. 이 문제는 기본적으로 연구 환경에서 이미 해결되었으며 ***(웹 링크)*** 상용화되면 일관성이 더욱 개선될 것으로 예상됩니다.

둘째로, 기계 번역 시스템의 생성물을 더 확실히 통제할 수 있게 될 것입니다. 기계는 이미 놀라운 품질로 번역 작업을 수행하고 있지만, 가령 문체와 관련해서는 사용자가 이를 결정할 수 없습니다. 하지만 앞으로는 텍스트를 격식체나 비격식체로 변역할지 여부를 넘어***(웹 링크)*** 만연체나 성별에 따른 특징 같은 것들도 통제할 수 있을 겁니다.

이 인터뷰는 명확한 의미 전달과 길이를 고려해 편집되었습니다.

AI Business
전자책 시리즈

LIONBRIDGE

Lionbridge는 다양한 브랜드와 협력하여 장벽을 허물고 전 세계를 연결하는 다리를 놓고 있습니다. 지난 25년간 Lionbridge는 여러 기업이 전 세계의 자사 고객과 직원들을 연결할 수 있도록 350개 이상의 언어로 번역 및 로컬라이제이션 솔루션을 제공해 왔습니다. Lionbridge는 세계적 수준의 플랫폼을 통해 전 세계에 걸쳐 열정적인 전문가 네트워크를 형성하고 있으며, 다양한 브랜드와의 파트너십을 통해 풍부한 문화 경험을 제공합니다. 또한 언어학에 대한 꾸준한 열정을 바탕으로 고객의 고객, 즉 최종 사용자와 공감대를 형성할 수 있도록 최상의 인간 지능과 기계 지능을 두루 활용합니다. 본사는 미국 매사추세츠주 월섬에 있으며 23개국에서 솔루션 센터를 운영하고 있습니다.

자세한 내용은 www.lionbridge.com
에서 확인할 수 있습니다.

추가 정보