



AI 콘텐츠 최적화 및 생성 사례연구

생성형 AI 기반 콘텐츠 최적화 및 로컬라이제이션 실험

4가지 AI 도구 테스트

2개 언어로 프롬프트 작성 및 테스트

영어 프롬프트로 최적화한 독일어 콘텐츠의 평균 품질 점수: 86%

실험 기반

라이온브리지(Lionbridge)는 이 실험을 수행하기 위해 내부 팀, 리소스 및 콘텐츠를 활용했으며, 특히 라이온브리지 기존 블로그의 독일어 콘텐츠를 사용했습니다. 이 사례연구는 라이온브리지 콘텐츠(특히 독일어)를 중심으로 진행되지만 본 실험의 목적은 모든 업종과 언어에 걸쳐 수천 명의 고객에게 적용할 수 있는 프로토콜을 마련하고 유용한 정보를 수집하는 것입니다. 라이온브리지는 이 실험의 결과와 관찰을 바탕으로 고객에게 다양한 언어로 서비스를 제공하고 광범위한 콘텐츠 요건을 충족할 수 있을 것으로 기대합니다. 내부 마케팅 콘텐츠 제작 및 최적화 프로세스에도 이 정보를 활용할 수 있습니다.

과제

콘텐츠 최적화에는 많은 시간과 비용이 소요되며, 특히 다른 언어로 작성된 콘텐츠를 최적화할 때는 더욱 그렇습니다. 또한 기존 콘텐츠를 번역한다고 해서 검색 엔진 최적화(SEO) 측면에서 항상 최고 수준의 성과를 거둘 수 있는 것은 아닙니다. 특히 언어 및 지역별 키워드의 뉘앙스나 사용자 행동 또는 데이터를 고려하지 않고 단순 직역한 경우 SEO 성과가 낮을 때가 많습니다. 생성형 AI 도구를 사용하면 번역에 드는 비용, 시간, 인력을 줄이면서 다국어 콘텐츠를 최적화할 수 있습니다.

배경

Amazon Bedrock, 챗 GPT(Lionbridge Azure 인스턴스, 챗 GPT 4 사용), Google Bard, Meta 의 Llama 등, 이 4가지 **생성형 AI** 도구는 모두 콘텐츠를 작성하고 최적화할 수 있는 기능을 갖추고 있습니다.

이 도구는 검색 패턴, 경쟁사 데이터 및 사용자 행동을 분석하여 키워드와 콘텐츠에 대해 데이터를 기반으로 제안을 합니다.

이 도구들은 사용자의 기대에 최대한 부합하도록 콘텐츠를 최적화하여 많은 고객의 호응을 얻고 SEO 가치를 높이는 데 활용되는데, 때로는 사람보다 더 우수한 문구를 작성하기도 합니다.

가령, 글자 수가 엄격히 제한되는 콘텐츠(특히 SEO 가치 제고 목적)를 작성하고 번역하는 경우에는 생성형 AI 도구가 제법 사람과 동등한 성과를 보일 때도 있습니다.

솔루션

라이온브리지는 특히 다른 언어로 조사하고 콘텐츠를 생성하기에 가장 효과적인 생성형 AI(GenAI) 도구를 확인하기 위한 3단계 실험을 실행했습니다. 여기서는 특히 다국어 콘텐츠를 위한 프롬프트 개발 모범사례를 학습하는 데 중점을 두었습니다. 또한 오래된 콘텐츠를 새로운 언어로 최적화하여 웹사이트의 SEO를 개선하기 위한 모범사례도 살펴봤습니다.

방법론

라이온브리지는 이 연구에서 Amazon Bedrock, 챗 GPT (Lionbridge Azure 인스턴스, 챗GPT 4 사용), Google Bard, Meta의 Llama 등 네 가지 AI 도구를 사용했으며, 조사, 생성 및 분석의 세 단계를 각각 두 번씩 실행했습니다.

조사: 이 단계에서는 각 AI 도구마다 키워드 조사를 수행하고 독일어로 콘텐츠를 작성하는데 가장 적합한 질문 두 가지를 선택하도록 요청했습니다.

각 도구에 입력한 항목은 다음과 같습니다.

- 페르소나 프롬프트. 독일어 콘텐츠 작성자로 행동하며 "새로운 독일어 키워드 3개와 독일어 질문 3개를 제공"하도록 요청
- 최적화할 원문 콘텐츠
- 기존 키워드

또한 나중에 콘텐츠에서 소단락을 생성하는 데 사용하기 위해 두 가지 추가 질문을 선택했습니다.

키워드 조사에서는 일반적으로 '휴먼인더루프 (human-in-the-loop)' 라고 하는, 사람이 개입된 과정이 필수인 것으로 보입니다. Bard의 경우는 더욱 그렇습니다. 이 도구는 조사 단계에서 탁월한 능력을 보였지만 오히려 너무 많은 옵션을 제공하여 콘텐츠 생성에 필요한 키워드나 질문을 선택하기 위해서는 분야별 전문가가 필요합니다.



작성: 이 단계에서는 독일어로 두 가지 프롬프트 세트를 개발했습니다. 이 콘텐츠 작성 프롬프트는 이전 단계에서 식별된 키워드를 포함하고 있습니다. 한 세트는 콘텐츠 작성자의 페르소나를 기반으로 개발되었으며, 다른 세트는 작업 지침에 대한 것이었습니다.

다음으로 라이온브리지 AI팀의 지원을 받아 모든 프롬프트를 검토하고 최상의 성과를 낼 수 있도록 조정했습니다. 프롬프트를 확정한 후에는(아래 '프롬프트' 섹션에 나열됨) 라이온브리지의 리소스를 사용하여 프롬프트를 독일어로 로컬라이즈했습니다. 이렇게 완성된 독일어 프롬프트 두 세트를 네 가지 AI 도구에서 실행하여 이들 도구가 제안한 콘텐츠를 수집하고 표시했습니다.

프롬프트를 개발할 때 다음과 같은 입력을 사용했습니다.

- 조사 단계에서 얻은 질문 및 키워드(독일어)
- 모방하려는 좋은 콘텐츠의 예(독일어)
- 소스 문서의 텍스트(독일어)
- 페르소나 프롬프트(영어 및 독일어)
- 작업 프롬프트(영어 및 독일어)

프롬프트를 두 세트 마련했으므로 각 AI 도구에서 프로젝트를 두 번 실행해야 했습니다.

분석: 이 단계에서는 독일어권 비즈니스 및 언어 전문가 10명에게 콘텐츠 검토를 요청했습니다. 검토 과정이 눈가림(블라인드 테스트)으로 진행될 수 있도록 각 콘텐츠가 어떤 도구를 사용해 생성되었는지는 표시하지 않았습니다.

평가하려는 출력물을 스프레드시트에 추가했습니다. 점수 산정에는 다음 두 가지 기준이 사용되었습니다.

- 다음 세 가지 값을 가지는 드롭다운 메뉴: '사용 불가', '수정 후 사용 가능', '수정 없이 사용 가능'
- 콘텐츠 점수 필드: 1~100점

전문가의 의견을 수집하여 다음 항목을 기준으로 검토했습니다.

- 어떤 도구가 가장 효과적인가?
- 프롬프트 작성을 위한 모범사례는? 다른 언어로 프롬프트를 작성할 수 있는가?
- 이들 도구는 키워드 조사를 얼마나 잘 수행하는가?
- 이 도구들은 콘텐츠에 대한 질문을 얼마나 잘 생성하는가?
- 이 도구들은 문서의 맥락을 얼마나 잘 파악하는가?
- 이 도구들은 생성된 질문에 대해 얼마나 좋은 답변을 생성하는가?
- AI 콘텐츠 작성 과정을 최적화하기 위해서는 어떻게 해야 하는가?
- 효과적인 프롬프트 개발 방법은?

계획



사용할 3가지 생성형 AI 도구 선택



실험을 위해 최적화할 텍스트 3개 선택



검토를 수행할 전문가 10명 모색



키워드 및 질문 조사



프롬프트 개발 및 독일어로 로컬라이즈

사용된 프롬프트의 예

페르소나 프롬프트

- 당신은 라이온브리지의 콘텐츠 작성자로서 언어서비스 제공을 위한 콘텐츠를 작성합니다.
- 독일어를 사용하며 독일에 거주한다고 가정하겠습니다.
- 독일, 오스트리아, 스위스의 독일어 사용자를 대상으로 독일어 콘텐츠를 제작합니다.

오디언스 프롬프트

- 마케팅 또는 커뮤니케이션 관련 업무 담당자
- 임원진, 선임 관리자, 이사, 부사장, VP, 사장, CMO, CEO
- 독일어 사용자. 독일 또는 오스트리아, 스위스 등에 거주

작성 프롬프트

- 이 문서에 포함할 텍스트 단락 2개를 추가로 작성해 주세요.
- 각 단락은 다음 질문에 대한 답변이어야 합니다.
- URL의 기본 키워드는...
- 보조 키워드는...

통합 전문가 팀

- 독일어 사용 프롬프트 엔지니어 10명
- 라이온브리지 SEO 전문가 1명
- 라이온브리지 디지털 마케팅 전문가 1명



결과

AI는 모든 최적화 프로젝트의 핵심 요소가 되어야 합니다.

라이온브리지는 이 실험을 통해 기존 콘텐츠, 심지어 영어가 아닌 콘텐츠도 AI 도구를 통해 효과적이고 효율적으로 최적화하고 업데이트할 수 있음을 확인했습니다.

놀랍게도 출력 품질은 실험이 진행되는 동안 더욱 향상되었습니다. 어떤 콘텐츠든 최적화 또는 작성 프로세스에 사람이 참여하는 것은 매우 중요합니다. AI 출력에서 최적화를 위한 키워드를 선택하려면 반드시 사람이 있어야 합니다. 또한 최종 콘텐츠 출력의 유효성을 검사하기 위해서도 사람에 의한 검토가 필요합니다(출력 품질이 현저히 향상되고 있으므로 이러한 조건은 향후 변경될 수 있음).

다음은 확인된 몇 가지 모범사례입니다.

- 현재는 휴먼인더루프를 통해 최종 AI 출력의 유효성을 검사해야 합니다.
- 실험을 진행하는 동안 새로운 모범사례를 학습하면서 입력 자료의 품질이 향상되었습니다.
- AI는 다음과 같은 요청이 제출될 경우 고품질의 콘텐츠 생성 결과를 낼 가능성이 더 높습니다.
 - 짧은 분량
 - 쉽게 세분화할 수 있는 콘텐츠(예: 상위 10개 목록 등)
- AI 기술과의 상호작용은 현재 영어로 하는 것이 가장 효과적입니다(이들 도구는 영어로 학습되었으므로 놀라운 일이 아님). 프롬프트를 독일어로 작성한다고 해서 AI 도구가 독일어 콘텐츠를 작성하는 능력이 더 향상되는 것은 아니며, 특히 독일어 콘텐츠를 생성할 때도 프롬프트를 영어로 작성했을 때 품질이 더 높았습니다. 반면 원문 콘텐츠, 키워드, 질문 및 예시는 영어가 아닌 대상 언어로 제공하는 것이 더 좋습니다.
- 콘텐츠 작성 과정에는 콘텐츠 작성자와 대상 오디언스를 꼼꼼히 설명해야 합니다. 또한 프롬프트 라이브러리를 구축하는 것이 좋습니다. 이를 통해 사용자는 프롬프트를 재사용하고 고객을 위한 콘텐츠를 일관성 있게 개발할 수 있습니다.

조사 단계에 가장 우수한 출력을 보인 도구는 Bard로 나타났습니다.

Bard의 출력은 명시적으로 조사 목적을 위해 개발된 도구인 MarketMuse의 조사 결과와 유사했습니다. 특히, Bard에서 조사 결과를 출력하기 위해서는 콘텐츠 작성 과정에서 사람이 개입하여 콘텐츠 너겟을 선택해야 합니다. 독일어 지침 사용 시 Bard는 두 단계 모두에서 변화 없는 출력을 보였습니다. 그러나 영어 지침 사용 시 Bard의 결과는 100점 만점에 82점에서 66점으로 낮아졌습니다. 이는 데이터셋이 작기 때문으로, 영어 사용 시 Bard의 평균 점수는 100점 만점에 74점이었습니다.

챗GPT의 결과에 대해서는 부정적인 의견이 약간 더 많았으나 검토자들로부터 좋은 평가를 받은 콘텐츠도 있었습니다. Meta(Llama)는 성능이 좋지 않았으며, 아직 영어 이외의 콘텐츠를 안정적으로 생성하거나 최적화할 준비가 되지 않은 것으로 보입니다. 한 달간 실험을 진행하는 짧은 기간 동안 출력 품질이 개선된 것으로 나타났습니다. Bard는 조사 출력이 개선되었으며 챗GPT는 독일어 콘텐츠 작성 성과가 개선되었습니다. 처음에 챗GPT에서 독일어 지침에 따라 생성한 출력물은 100점 만점에 47점을 받았지만 불과 한 달 만에 76점을 기록했습니다.

영어 지침에 따라 생성한 출력물은 100점 만점에 71점을 받았으나 한 달 만에 86점으로 상승했습니다(영어에 대한 평균 점수는 100점 만점에 78.5점).

Bard와 챗GPT는 모두 한두 문장으로 이루어진 짧은 글을 작성하는 데 뛰어났습니다. 짧은 글에 대해서는 부정적인 의견이 거의 없었습니다.

이 실험의 두 가지 단계에서 장문 콘텐츠에 대해 얻은 점수를 표로 나타냈습니다.

1단계 점수

원문	사용 불가	수정 후 사용 가능	수정 없이 사용 가능	평균 점수
챗GPT (독일어 입력)	9	9	0	47.61
Bard (독일어 입력)	2	14	2	68.33
챗GPT (영어 입력)	3	9	6	71.67
Bard (영어 입력)	0	9	9	82.22
Meta (Llama) (영어 입력)	7	8	3	46.50
총점	21	49	20	

2단계 점수

원문	사용 불가	수정 후 사용 가능	수정 없이 사용 가능	평균 점수
챗GPT (독일어 입력)	1	7	0	76.25
Bard (독일어 입력)	2	4	2	66.25
챗GPT (영어 입력)	0	4	4	86.875
Bard (영어 입력)	3	4	1	59.375
총점	6	19	7	

참고: Meta(Llama)는 아직 영어 이외의 콘텐츠를 최적화하거나 생성할 준비가 되지 않았으므로 2단계에서는 실험 대상에서 제외됨

