



AI 內容最佳化與生成個案研究

探討以生成式 AI 技術最佳化與本地化內容的一項實驗

共測試 4 個 AI 工具

以 2 種語言撰寫和測試提示

以英文提示對德文內容進行最佳化後的平均品質分數：86%

此 LIONBRIDGE 實驗的幕後故事

Lionbridge 是使用自己內部的團隊、資源及內容進行這個實驗。具體地說，我們使用了自己的德文內容，也就是現有的部落格文章。儘管本個案研究是針對 Lionbridge 的內容（尤其是德文），但其目的是要建立程序並從中收集相關知識，以便應用到我們每個垂直市場和語言中數以千計的客戶。我們會運用這個實驗的結果與觀察，為內容需求與語言各異的客戶提供更好的服務。此外，我們也會將這些資訊應用於公司內部的行銷內容創作與最佳化流程。

面臨的挑戰

將內容最佳化，尤其英文以外的其他語言，往往既耗時又所費不貲。此外，將現有內容加以翻譯，也未必能在搜尋引擎最佳化 (SEO) 上展現優異的成效。這類翻譯工作有時只是單純的直譯，並沒有將語言的細微差異、地區特有的關鍵字，以及使用者行為和/或資料納入考量。而生成式 AI 工具或許可以協助我們最佳化多語言內容，同時又能降低成本和減少所需的時間與人力。

背景介紹

Amazon Bedrock、ChatGPT (Lionbridge Azure 執行個體，使用 ChatGPT 4)、Google Bard 以及 Meta 的 Llama 這四個生成式 AI 工具，都有能力創作和最佳化內容。

它們能分析搜尋模式、競爭對手資料以及使用者行為，再根據資料針對關鍵字和內容提出建議。

這些工具有助確保最佳化的內容盡可能符合使用者的期望，進而能吸引讀者並提供高搜尋引擎最佳化 (SEO) 價值。在某些案例中，這個技術根據要求所生成的文案，甚至可能比人類寫得還要好。

例如，對於必須遵守嚴格字數限制的內容（尤其為了 SEO 價值之故），GenAI 工具在撰寫和翻譯的表現上有時可以跟人類並駕齊驅。

解決方案

我們執行了三階段的 GenAI 工具實驗，以了解什麼因素對研究以及內容生成的影響最大，尤其是英文以外的其他語言。我們團隊著重在找出研擬提示的最佳實務做法，特別是多語言內容的提示；此外，我們也探索了哪些最佳實務做法，會有助以新語言最佳化舊有內容，進而提升網站的 SEO。

方法

Lionbridge 在這個研究中使用了四個 AI 工具：Amazon Bedrock、ChatGPT (Lionbridge Azure 執行個體，使用 ChatGPT 4)、Google Bard 以及 Meta 的 Llama。我們共進行了兩次分成三階段的實驗：研究、創作以及分析階段。

研究：在這個階段，我們要求每個 AI 工具完成關鍵字研究，並選擇兩個最合適的問題以使用德文創作內容。

我們將以下這些事項提供給各個工具：

- 角色提示，要求工具扮演德文內容寫作人員的角色，並「提供 3 個新的德文關鍵字和 3 個德文問題」。
- 要最佳化的來源內容
- 現有的關鍵字

此外，我們也選擇了兩個補充問題，稍後會用來生成小段落的内容。

在關鍵字研究方面，在過程中納入人類參與，也就是一般所謂的人機迴圈 (human-in-the-loop)，似乎非常關鍵。這一點對 Bard 來說尤其顯著，因為這個工具的研究工作雖然做得很好，但卻會提供太多選擇，因此需要由主題專家挑選進行內容生成所需的關鍵字或問題。



創作: 在這個階段, 我們研擬了兩組德文提示。這些內容創作提示中都含有前一個步驟找出的關鍵字, 其中一組是根據內容寫作人員這個角色, 另一組則是根據任務說明。

接著我們請一位 Lionbridge AI 部門的成員審閱並協助微調所有這些提示, 以獲致最大的有效性。這些提示一旦確定後 (列於下方「提示」一節), 我們便運用 Lionbridge 的資源將提示本地化至德文。然後, 我們將兩組完成處理的德文提示輸入所有四個 AI 工具中, 收集它們所建議的內容並加以評分。

我們使用了以下這些輸入來研擬提示:

- 研究階段所提出的問題與關鍵字 (德文)
- 我們希望模仿的優質內容範例 (德文)
- 來源文章的內文 (德文)
- 角色提示 (英文與德文)
- 任務提示 (英文與德文)

由於我們有兩組提示, 因此必須在各個 AI 工具中執行兩次專案。

分析: 在這個階段, 我們請十位通曉德文的商務和語言專家來審閱所得到的內容。為了確保審閱為「盲測」, 我們並未標示內容是由哪個工具所生成。

我們將要評估的輸出放入試算表中, 並使用兩種評分機制:

- 一個下拉式選單, 內有三個選項: 「無法使用」、「編輯後可使用」以及「不需編輯即可使用」
- 一個以 1-100 分為內容打分數的欄位

檢閱所收集的意見反應後, 我們便接著判斷:

- 哪一個工具效益最高?
- 在建立提示這方面有哪些最佳實務做法? 以其他語言建立提示時又有哪些最佳實務做法?
- 這些工具在關鍵字研究上的表現如何?
- 這些工具根據內容生成問題的表現如何?
- 這些工具在生成文章觀點上的表現如何?
- 這些工具在回答它們生成之問題上的表現如何?
- 我們可以如何確保享有最佳的 AI 內容創作流程?
- 如何研擬有效的提示?

規劃



選擇要使用的 3 個
GenAI 工具



選擇實驗中要加以
最佳化的 3 篇文章



尋找 10 位專家組成
審閱小組



針對關鍵字與
問題進行研究



研擬提示並將其
本地化至德文

使用的提示範例

角色提示

- 你是 Lionbridge 的內容寫作人員,要創作跟語言服務相關的內容。
- 你是通曉德文而且住在德國的人。
- 你要以德文創作內容,對象是住在德國、奧地利和瑞士的德語人士...

目標對象提示

- 從事行銷或傳播的人
- 擔任高階管理職位的人,例如資深經理、總監、副主席、VP、總裁、行銷長、執行長等
- 通曉德文而且住在德國、奧地利或瑞士的人...

創作提示

- 請另外創作 2 段文字以便附加到這篇文章...
- 每一段內容都要能夠回答這些問題...
- 這個 URL 的主要關鍵字是...
- 次要關鍵字則是...

跨領域團隊

- 10 位通曉德文的提示工程師
- 1 位 Lionbridge SEO 專家
- 1 位 Lionbridge 數位行銷專家

成果

AI 應該成為任何最佳化專案的重要元素。

我們的實驗證實了 AI 工具可以有效率並有效地輔助現有內容的最佳化與更新作業,甚至可以處理非英文的內容。

令人振奮的是,隨著實驗進行,輸出的品質也不斷改善。對任何內容最佳化或創作流程來說,納入人機迴圈機制非常關鍵,除了必須要有人從 AI 輸出中挑選關鍵字以利後續最佳化作業,最後的內容輸出也必須要經過審閱人員的檢驗 (但隨著輸出品質明顯不斷提升,這種做法日後可能會有所改變)。

以下是我們歸納出的一些最佳實務做法:

- 以目前來說,最後的 AI 輸出必須採用人機迴圈的機制加以驗證。
- 在整個實驗過程中,隨著我們學到新的最佳實務做法,輸入的品質也跟著進步。
- 如果採用以下方式送出要求給 AI,就更有機會產出高品質的生成內容:
 - 多個簡短的段落
 - 可輕易分類的內容 (例如十大清單等)
- 在與 AI 技術溝通這方面,目前還是以英文成效最好 (考慮到工具往往是以英文進行訓練,也就不令人意外了)。以德文提供提示並沒有因此讓 AI 工具產出品質較高的德文內容;反而是使用英文提示時,工具產出的德文內容比較好。但請注意:這個技術仍然有助我們以 (非英文的) 目標語言提供來源內容、關鍵字、問題和範例。
- 在內容創作流程中,應該充分完善地描述內容創作人員及目標對象的特色。此外,最好能建構提示資料庫,方便人員能重複使用提示,進而為客戶開發一致的內容。



在研究方面，輸出表現最好的似乎是 Bard。

它的結果跟 MarketMuse 這個工具產出的研究成果很接近，而後者是專為研究用途而開發的。但 Bard 的研究輸出顯然需要人工進一步處理，從中挑選適當的結果供流程的內容創作環節使用。使用德文指示時，Bard 在兩個階段的表現都很穩定；然而使用英文指示時，結果的分數卻從 82/100 下降到 66/100。由於這是個很小的資料集，因此 Bard (使用英文) 結果的平均分數為 74/100。

ChatGPT 生成內容獲得的評價稍微偏負面，不過有時也受到審閱人員的稱讚。Meta (Llama) 的表現不如其他兩者，可能還不足以可靠地用於最佳化或生成非英文內容。在這個為期一個月的短期專案中，輸出品質似乎會隨時間進展而改善。Bard 的研究輸出便有所改進，而 ChatGPT 的德文內容創作也有進步。根據德文指示生成的 ChatGPT 輸出一開始的分數只有 47/100。

但短短一個月後，它們的輸出分數就提高到 76/100。而根據英文指示生成的輸出，分數本來是 71/100，但也在一個月內上升到 86/100 (英文的平均分數為 78.5/100)。

Bard 及 ChatGPT 兩者都能熟練地產出簡短的內容 (例如只有一、兩個句子的段落)。審閱人員對它們產出的短文很少有負面評價。

表中顯示的分數，是實驗前兩階段中較長內容的分數。

第 1 階段分數

來源	無法使用	編輯後可使用	不需編輯即可使用	平均分數
ChatGPT (德文輸入)	9	9	0	47.61
Bard (德文輸入)	2	14	2	68.33
ChatGPT (英文輸入)	3	9	6	71.67
Bard (英文輸入)	0	9	9	82.22
Meta (Llama) (英文輸入)	7	8	3	46.50
總計	21	49	20	

第 2 階段分數

來源	無法使用	編輯後可使用	不需編輯即可使用	平均分數
ChatGPT (德文輸入)	1	7	0	76.25
Bard (德文輸入)	2	4	2	66.25
ChatGPT (英文輸入)	0	4	4	86.875
Bard (英文輸入)	3	4	1	59.375
總計	5	19	7	

附註：我們將 Meta (Llama) 從第 2 階段中移除，因為它還不足以用於最佳化或生成非英文內容。