



為多語言軟體打造彈性調度解決方案

17 種額外語言

25 個專案

從頭製作元件

客戶需要尋找一名合作夥伴，協助他們進行多語言的軟體實作架構擴展作業。我們總共交付了 25 個專案，提供額外 17 種語言的版本，協助他們達成目標。

面臨的挑戰

我們的客戶建置了一個可用多種語言實作軟體的架構。這個架構採用了監督式及非監督式的機器學習方法，以及文本分析和多種前處理與後處理程序，包括分詞 (tokenization) 和詞性標記等。

客戶開發並測試了一套工具和流程，並用它們為多種語言實作工作元件。客戶想要找的合作夥伴，除了要能幫他們記錄、開發與維護架構，也要能以額外的語言實作相關元件。

客戶簡介

我們的客戶是全球規模數一數二、同時也是自然語言文字技術先驅的企業軟體供應商，提供一系列多樣產品，包括建置高效演算法來處理之後可供其他應用程式使用的文字。



重要詞彙

命名實體辨識程式

在指定文字中，找出命名實體並辨識它們屬於哪個預先定義的類別 (例如人名、地名等)。

例句：志明是住在台北市的工程師
志明 = 人名；台北 = 地名

關鍵詞擷取程式

根據字詞在指定文字脈絡下的關聯性，擷取出關鍵詞或關鍵字清單。

以前述句子為例：1. 關鍵詞擷取程式 2. 關鍵詞清單 3. 指定文字

成果

Lionbridge 協助客戶針對 17 種新增語言，有效地開發了可彈性擴展的解決方案。由於軟體大部分的語言版本是從頭開始實作，因此我們建立了所有必要的文件、路徑圖和輔助工具，以利從頭實作軟體；此外我們也改善了一個現有軟體，並將客戶的架構應用於任何語言。我們另外開發了一個結構，詳實記錄為了讓軟體達到令人滿意的效能而採取的所有步驟。最後，我們也與客戶進行諮商，討論如何進一步改善架構及其元件。

我們以 25 個專案達成所有成果，客戶並將這些解決方案整合至他們的雲端式 API 中，提供以文本為基礎的自然語言處理服務。

如需深入了解，請造訪

LIONBRIDGE.COM

解決方案

Lionbridge 協助客戶針對 17 種新增語言，有效地開發了可彈性擴展的解決方案，其中大部分的語言都是從頭開始實作軟體。



1. 建立高效率的工作流程

Lionbridge 與客戶協力合作，先在輔導階段針對某個先導語言，使用客戶的架構展開軟體實作，接著再於內部繼續進行先導語言的軟體實作，並同時展開多個其他語言的軟體實作。這種作法除了有助專案按照時程執行，更可同時調整與改善流程。



2. 克服與特定語言相關的挑戰

在處理不同語言時，我們的團隊成功解決了個別語言的相關挑戰，包括：

命名實體辨識：大小寫問題

- 大小寫是命名實體辨識的主要特色，但某些語言 (例如阿拉伯文、日文、韓文等) 並沒有大小寫。

關鍵詞擷取：過度生成

- 正確的關鍵詞會因為含有經常出現在文本中的某個特定字詞，而由系統擷取到；但於此同時，另一個含有同樣特定字詞、卻與指定文本不相關的關鍵詞，也可能會錯誤地被系統擷取出來。

Lionbridge 團隊實作了基於語言的規則編寫程式碼，並將其整合至架構中，以解決這些跨語言的挑戰。我們的團隊也實作了文本處理的輔助工具來支援軟體開發，並運用不同分析工具輕鬆找出需改進的地方。



3. 提供基礎設施

Lionbridge 建立了一支全球最大的自然語言處理團隊，囊括優秀的語言專家、專案經理、資料工程師以及遍及全球的語言專業人才，專責為不同語言創建自然語言處理解決方案與資產。我們採用雲端式平台，因此無論工作地點為何，全球團隊只要使用 Lionbridge 的工具，就能在 Lionbridge 專案經理與協調人員的指引與督導下協同工作。此外，我們的營運範圍遍及全球，更名為世界各地的客戶提供 24 小時全年無休的優質服務。

客戶需要的是能提供超越傳統本地化專案服務的合作公司，他們需要的合作夥伴，不但要對機器學習方法具備深入的了解，更要熟知如何開發與增進文本的前後處理流程及基於語言的規則編寫。一般的譯者通常沒有這些技能背景。為此 Lionbridge 特地培養並部署了一個具備機器學習知識與能力的專業人才網路，可處理超過 200 種語言。