

## 異なる言語における検索クエリ形式の性能評価

松田 明梨

情報検索において、検索クエリはユーザーの情報要求を表現する基本的な手段であり、その形式が検索有効性に影響を与える可能性がある。同一の情報要求を表現する場合であっても、クエリに使用される単語や構文には多様性が見られる。例えば、「日本 首都」、「日本の首都の名前」、「日本の中心都市」といったクエリは、いずれも日本の首都に関する情報を求めているが、その表現方法は異なる。このような多様性の1つであるクエリ形式は、「質問クエリ」と「キーワードクエリ」の2種類に大別される。質問クエリは自然文で構成され、ユーザーの意図を明確に示す特徴がある。一方、キーワードクエリは簡潔な単語の組み合わせで構成され、入力効率に優れている。近年、質問クエリの利用が増加しており、特に複雑な検索タスクや特定のユーザー層において質問クエリが選択される傾向がある。一方、キーワードクエリはその簡潔さと効率性から広く使用され、検索システムにおいて主要な役割を果たしている。英語を対象とした既存研究では、語彙的な検索モデル（例：BM25）において、クエリ形式間の検索有効性に有意差は確認されなかったが、大規模言語モデル（LLM）ベースの検索モデル（mDPR）においては、質問クエリがキーワードクエリよりも高い検索有効性を示す傾向が報告されている。しかし、これらの知見が英語以外の言語にも当てはまるかは十分に検証されておらず、異なる言語におけるクエリ形式の影響を調査する必要がある。本研究では、18の異なる言語を対象に質問クエリとキーワードクエリの検索有効性を比較し、言語特性がその違いに与える影響を分析した。検索モデルとして、語彙的な検索モデル（BM25）およびLLMベースの検索モデル（mDPR, mContriever）を用いた。実験の結果、BM25では質問クエリとキーワードクエリの検索有効性に大きな差異は見られなかった。一方、LLMベースの検索モデルでは、多くの言語で質問クエリが高い検索有効性を示し、キーワードクエリの検索有効性が低下する傾向が確認された。特にmContrieverでは、この低下幅がmDPRと比較して顕著であった。また、リソースの規模がクエリ形式の検索有効性に影響を与える傾向が確認された。高リソース言語ではキーワードクエリの検索有効性が低下する一方、低リソース言語ではキーワードクエリが質問クエリを上回る検索有効性を示す傾向が観察された。本研究では言語特性が検索有効性に影響を与える可能性が示唆されたが、これらの影響を明確に解明するためには、より多くの言語を対象とした追加的な分析が求められる。

(指導教員 加藤 誠)