

## 書誌情報検索におけるゼロヒット問題の類型化とその解決手法の提案

堀田 優介

これまで書誌情報検索における目録システム (OPAC) の研究では、OPAC の利用行動調査が行われてきた。OPAC の利用行動調査で、検索結果がゼロ (以下ゼロヒット) である場合や、検索結果件数が多すぎるといったことが分析され、検索システムの改善策が提示されてきた。しかし、ゼロヒットに関するこれらの研究ではゼロヒットを引き起こしたクエリの特徴を見出すよりも情報探索行動を知る手段の一つとして着目することがほとんどであった。

そのため本研究では、ゼロヒットを起こす原因となったクエリの特徴を調査し、その結果を類型化、そして各々の特徴にあった改善策を提案する。

対象としている筑波大学附属図書館の OPAC 利用ログは、2014 年 3 月～2018 年 1 月の 46 ヶ月分の期間におけるアクセス約 3 億 6 千万件を対象とした。そこから、クローラアクセスの除去とセッション抽出により得られたセッション単位の、のベクエリ数は 1,747,568 件であり、OPAC を経由して得られた、のベクエリ数は 1,229,698 件を対象とする。ゼロヒットを引き起こしているクエリを抽出するために CiNii Books の API を用いて筑波大学附属図書館に限定し問い合わせた結果、OPAC を経由して検索されたクエリのうち 6.97% である 85,659 件がゼロヒットであることがわかった。

サンプリング調査で 500 件のクエリを無作為に抽出したが、このうち 128 件がゼロヒットではないクエリであることが判明したため残りの 372 件を対象として調査した。各クエリの特徴をもとに 7 つに類型化した。1) 変換ミスによるクエリ (6%)、2) システム的なクエリ (6%)、3) その他 (88%)。さらに 3) を細かく類型していき、4) 日本語を含むクエリ (72%)、5) 英語のみのクエリ (16%)。また 4) を 2 つに類型した。6) 筑波大学附属図書館以外ならヒットするクエリ (40%)、7) 筑波大学附属図書館以外でもヒットしないクエリ (32%)。その後、それぞれの改善策を提案した。各特徴の中で、6) がゼロヒットを引き起こしたクエリの約 40% を占めていたことから、このクエリに対する改善が重要ではないかと考えた。また、1)、5)、7) においてユーザーの意図しないクエリが打ち込まれた場合にシステム側が誤っている箇所を修正または、自動的に再検索を行うことでゼロヒット問題を改善できるのではないかと考えた。

今後の課題としては、それぞれの類型化したゼロヒットを引き起こしたクエリの特徴に沿った改善策をシステム化し実装することがあげられる。そのためには、ユーザーが誤って打ち込んだクエリを同定し指摘や修正ができるような正しいクエリのデータセットの収集やクエリを単語ごとに分割して検索し適合度順に提示するシステム構築が必要である。また、本研究では著者のみで類型化を行なったため各類型化における基準の安定性には検討の余地がある。さらに、類型化の種類において今回の研究で用いた 7 種類以外も考えられる。

(指導教員 高久 雅生)