

ディスカバリサービスにおけるクエリとの関連度を考慮した 情報源提示システムの構築

大木 隆寛

ディスカバリサービスには、その機能と利用者のメンタルモデルが合っていないこと、簡易検索のみを用いる利用者が大半を占めていることの2つが問題点として指摘されており、これらからディスカバリサービス利用者の多くは目的の文献に辿り着いていないのではないかという課題が考えられる。特に、「ある主題に関する文献を数多く見つけたい」という利用者に関しては、既存のディスカバリサービスの機能のみでは解決できない課題である。この課題に対して、情報源提示が有効ではないかという指摘がある。

そこで、本研究では、ディスカバリサービスにおいて、利用者が目的の文献に辿り着くことを支援するために、学問分野という観点からクエリに最適な情報源提示システムの構築を行う。

提案システムでは、学問分野に関連の深いタイトル、抄録、主題、キーワードの4つのフィールドを検索対象として、それぞれのヒット件数から関連度スコアを算出し、関連度スコアを基に情報源提示を行った。対象情報源について、CiNii Articles、社会老年学文献データベース (DiaL)、アジア経済研究所学術研究リポジトリ (ARRIDE)、AgriKnowledgeの4つを対象とした。関連度スコア算出手法について、3つのパターンを提案した。それぞれのパターンについて、パターン1はヒット件数の合計を総レコード数とフィールド数で割ったもの、パターン2は各フィールドにおいて当該フィールドにデータを含むレコード数とそのフィールドのヒット件数の比の合計をフィールド数で割ったもの、パターン3はパターン1の算出式に対して各フィールドに重みづけを行ったものを関連度スコアとしている。

システム評価では、日本十進分類法を基にした、クエリ数がそれぞれ100個、80個からなる2つのクエリセットに対して、関連度スコア算出手法の各パターンでクエリを実行し、各情報源における平均順位と一要因分散分析による関連度スコア同士の有意差の比較を行った。この比較結果から、提案システムにおいて、どの関連度スコア算出手法が適切かの評価を行った。

評価結果は、2つのクエリセットにおいて、各フィールドに対する重みづけを行った関連度算出手法が最も適切な情報源提示を行っていた。このことから、学問分野の観点からクエリに最適な情報源提示のためには、学問分野に大きく作用するフィールドを適切に選定し、それらに対して適切な重みづけを行うことが必要であると考えられる。本研究における情報源提示について、クエリと情報源の関連性を測る人手判定を通じた評価や、利用者の情報探索時にどのような効果を与えるか、同分野の情報源ではどのような結果が得られるか、分野の近接関係によって情報源提示の結果がどのように変化するかを検討することが、今後の課題となった。

(指導教員 高久 雅生)