

OPAC 利用ログを用いた文献検索システムの構築

小幡 将司

大学において履修している科目で課題が出された時、参考文献を探索するために履修者は大学図書館を利用することがある。大学図書館において文献を探索するためには OPAC を利用することが多いが、時間がかかり見つからないことがある。それに対し、先輩が学んだ経験を反映できれば適切な資料を検索できるのではと考えた。そこで、過去同じ科目を履修した学生たちが使用した文献の情報を参考に検索を行うことにより、学生の学習が促され、課題をよりよいものにできると考えた。本研究では過去の利用情報を反映した文献検索システムについて研究を行う。

本研究では OPAC で検索を行う際、①入力したクエリと②検索を行った時期、③OPAC の利用ログの情報の 3 点を利用する手法を提案する。科目の課題等のために OPAC で検索を行う際、入力クエリで検索するだけでなく、利用ログから過去の同一科目が実施されていた時期によく閲覧された文献に関する情報を用いてリランキングすることで、授業に関連する文献を効率よく提供することができるようにすることを目指す。

本研究において、リランキングに用いるスコアとして 6 つのスコアを提案する。スコアとしては、①クエリとの文献類似度のスコア、②月ごとの各文献へのアクセス回数によるスコア、③クエリがどの科目に関連するかのスコア、④関連科目のスコアを用いて期間の重みづけを行うアクセス回数のスコア、⑤検索を行った時期を開講期間に含む科目のみを抽出し、該当科目の開講期間での重みづけを行うアクセス回数のスコア、⑥科目のキーワードを用いたクエリ拡張に基づくスコアの 6 つを用いる。

評価実験では、国立教育政策研究所教育図書館の OPAC 利用ログを用いた。レファレンス協同データベースのレファレンス事例を 15 件抽出して、評価を行った。抽出したレファレンス事例を元にクエリを作成して、評価用クエリとした。そのうえで、上位 10 件、上位 20 件、上位 100 件における検索結果の精度を測った。評価の結果、上位 10 件、上位 20 件、上位 100 件において、文献類似度手法による検索結果の精度がそれぞれ 0.30, 0.26, 0.11 であったのに対し、月ごとのアクセス回数に基づく手法による検索結果の精度はそれぞれ 0.08, 0.07, 0.03 であった。検索結果の上位 10 件では 15 クエリ中 1 件で、上位 20 件では 15 クエリ中 2 件で、上位 100 件では 15 クエリ中 4 件が利用ログを反映した手法が文献類似度のみの手法よりも精度の高い検索を行うことができた。

利用ログの傾向分析と利用ログを元にリランキングを実現した。しかし適合性評価では、精度の向上があまり見られなかった。失敗事例について分析したところ、今回の評価用クエリと利用ログの関係においては本研究の手法はあまり適していないと考えられる。また、期間ごとの閲覧文献の偏りからスコアの付け方を改善する必要がある。

(指導教員 高久雅生)