# 関心が広がるキーワード自動生成システム

二井 俊也

私は、自らが自由にテーマを決めてそれについて調査や研究を行う際に、最後までモチベーションを保つことができないことが多かった。モチベーションを保つには、テーマへの関心が必要である。そこで、関心をもてるテーマを提案するシステムの作成を目指す。

本研究では、任意のテーマをシステムに入力することで、それに関連した物事へと関心が広がるようなキーワードを提示するシステムを構築した。システムの概要は、入力されたクエリに対して国立国会図書館件名標目表のデータと照合し、該当する日本十進分類記号（NDC）を取得する。入力されたクエリと NDC を条件に、国立国会図書館サーチからキーワード候補を含む書誌情報を取得する。そこからタイトルと主題を抽出し、MeCab で形態素解析を行い、キーワード候補を生成する。キーワード候補から関心が広がるキーワードを選定し、システム利用者に提示する流れになる。

システムを作成する際に、キーワードの選定重み付け手法について、MeCab で形態素解析する際に使用する辞書 mecab-ipadic-NEologd が持っている cost 値を利用して tf×cost 値で算出する手法 Cost、word2vec の二単語間の意味距離を利用して、キーワード候補同士における意味距離の平均による手法 Average Distance と直前に選定されたキーワードとの意味距離で算出する手法 Distance to Distance の 3 手法を提案した。

評価方法は、評価指標にパラメータ A を設けて各手法が提示したキーワードに対して、提示されたキーワードがクエリに対して関連しているかを見る適合性の観点（適合面）、提示されたキーワード自体やキーワードとクエリの関係性に関心が広がったかを見る関心度からの観点（関心面）、適合性と関心度の両方の観点（両面）を人手によって評価質問項目に回答する形で評価した。これらの評価指標によって、キーワードに評価スコアが付与される。手法の評価は、提示したキーワードに付与されている評価スコアの合計とした。

結果は適合面と両面で手法 Cost が最も有効であった。関心面では手法 Distance to Distance:0.6 が最も有効であった。結果から言えることは、システムとしては手法 Cost が最も有効ではあるが、「関心を広げる」という点では手法 Distance to Distance:0.6 が最も有効であった。その理由として考えられることは、選定されたキーワード同士の意味距離が離れていることが、結果に良い影響を与えられたことが考えられる。

今後の課題としては、ユーザーが実際にクエリを入力した場合にどのように評価するかといったことやキーワードの提示方法、キーワード候補の情報源を変更することが与える影響を検討する。

（指導教員　高久 雅生）