

オープンナレッジを用いたクエリ拡張法に関する研究

大塚 淳史

膨大なWeb ページから所望のページを探し出すWeb 検索エンジンの必要性は、ますます高まっている。ユーザが調べたいことである情報要求が具体的に言語化されている場合、ユーザが検索エンジンに入力するクエリのキーワードも具体的であり、検索エンジンは、ユーザの望む情報を的確に検索することができる。しかし、ユーザが情報要求を上手く言語化できない場合は、具体性に欠けるキーワードしか検索エンジンに入力できず、満足できる情報を探すことは難しい。このため、クエリを補完または置換する候補をユーザに提示するクエリ拡張に関する研究が盛んに行われている。

本研究では、質問回答サイトに投稿された質問記事をユーザの情報要求と結びつけることで、ユーザに情報要求自体の候補を提示する拡張クエリを提案する。質問回答サイトは、誰もが参加できる知識共有の場であることからオープンナレッジと呼ばれている。提案法は、検索に使用する拡張クエリだけでなく、拡張の元となる質問記事をユーザに提示するという特徴がある。質問記事は自然言語で記述されているため、質問記事からクエリを作成するというプロセスは、ユーザ自身が行っていた、情報要求を言語化し、クエリを作成するというプロセスをシステムが代替することに相当する。クエリを作り出す過程をシステム側で再現することでユーザの情報要求を的確に反映した検索結果を得ることを目指す。

提案法では、質問回答サイトのカテゴリ別に拡張クエリを作成し、それらを統合してユーザに提示することで、幅広い観点からクエリを拡張する。また、カテゴリとキーワードを1 つ追加した拡張クエリと、質問記事とキーワードを2 つ追加した拡張クエリの2 種類の拡張クエリを、段階的にユーザに提示する手法を提案し序々に具体的な検索を行えるように工夫した。追加するキーワードは、潜在的意味インデキシングにより、概念的に関連のある語を検索している。

提案手法をシステムに実装し、提案法の有効性を検証した。その結果、“ウイルス” という語では、異なるカテゴリに存在する病原菌のウイルスとコンピュータウイルスの同音・同字異義語に対して、それぞれの観点から拡張クエリが生成されることを確認した。また、質問記事ごとに異なる拡張クエリが作成されており、質問記事から適切にキーワード組を作成できていることを確認した。

今後は、投稿の季節性を考慮したクエリ拡張を考えている。オープンナレッジは、定常的に質問が投稿されていることから、投稿時期や季節によって質問の傾向が変化すると考えられる。季節性を考慮することで、よりユーザの要求に合った拡張クエリを生成する手法へと発展させたいと考えている。

(指導教員 佐藤哲司)