JAIST Repository
https://dspace.jaist.ac.jp/

Title	電子化辞書を利用した、概念に基づくクエリーの拡張に関する研究
Author(s)	太田，千晶
Citation	
Issue Date	1998-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1113
Rights	
Description	Supervisor:奥村 学，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

電子化辞書を利用した
概念に基づくクエリーの拡張に関する研究

太田千晶

北陸先端科学技術大学院大学 情報科学研究科

1998 年 2 月 13 日

キーワード： クエリー拡張, 概念, 情報検索, 電子化辞書.

本論文では、情報検索システムにおけるユーザ支援の一手法として、クエリー (検索質問) を “概念” に基づいて拡張する “クエリー拡張” の方法を提案する。

近年、様々な情報の電子化が急速に進み、膨大な数の情報の中から必要な情報をいかに効率良く かつ正確に発見するかという「情報検索」技術への期待が高まって来ている。また一方、ここ数年で 情報機器が一挨に一般家庭に普及したことにより、情報検索に不慣れなエンドユーザがこのような大量の電子化情報にアクセスする機会が急増し、先の「情報検索」に加え、「検索支援」技術のニーズも高まって来ていると言えるだろう。

しかし現在、一般の情報検索システムでは、基本的にクエリーと検索文書との単純なキーワード・マッチングにより正解文書を導き出しているため、その検索精度はユーザが検索に有効なクエリーを生成できるか否かに大きく依存してしまうという問題が生じている。

このような問題に有効な方法として “クエリー拡張” がある。クエリー拡張とは、ユーザがシステムに入力したクエリーを基に検索に有効なクエリーの生成を行うもので、検索精度の向上を目的とする。

本研究では、従来の研究で使用されてきた「表記」だけでなく、そのクエリーが持つ意味的な情報である “概念” に着目し、EDR 電子化辞書中の様々な概念関係・概念記述を利用したクエリー拡張手法を統合して、クエリー拡張システムを試作した。

既存のデータベースにおける “概念” を用いることによって、確実な意味的関連性を伴った拡張を行うことができ、それによって意味的な関連性の高い拡張タームが獲得できると考えられる。

Copyright © 1998 by Chiaki Ohta

まず本研究では、以下の 2 種のクエリー拡張手法を提案を行った。

(手法 1) 名詞と動詞の概念間関係を利用したクエリー拡張手法

ここでは、尤もらしい共起を為す名詞と動詞間の意味的な関連性に注目し、EDR の概念記述辞書を用いて、初期クエリータームのもつ概念と意味的に尤もらしい共起を為す動詞的 もしくは 名詞的概念を関連概念として獲得する。本手法によって生成される「初期クエリーターム AND 関連ターム」というクエリーは 初期クエリータームを含む文書集合の中から、正解文書を効率良く獲得する役割を担うものとして期待できる。

(手法 2) 概念説明文を利用したクエリー拡張手法

本手法は、異なる 2 つの初期クエリータームを含むような概念説明文を検索し、その説明文の意味する “概念” を 初期クエリーの関連概念として獲得するものである。本手法によって獲得できる関連タームは、初期クエリーと意味的な関連性を持つが、従来の方法で獲得してきた初期クエリータームの { 上位・同義・下位 } 語とは全く異なるタイプの新しい関連語を獲得することが可能である。

また、次に本研究では (手法 1) で概念の拡張を行う前に、クエリータームの語義の曖昧性解消を行うことを提案し、それによって検索精度の向上が得られると仮定した。本研究では、概念を利用したクエリータームの語義曖昧性解消のための手法を 3 種提案している。

以上で 提案した手法により生成された関連タームの検索精度への効果を調査するために、本研究では 情報検索システム評価用ベンチマーク BMIR-J1 を用い、検索文書 全 600 件、クエリー 全 40 種を用いた実験を実施した。なお検索結果は recall と precision によって評価している。

まず (手法 1) の実験の結果では、拡張前と比較して平均値においては わずかな preicision 値の向上しか得られなかったが、分析の結果 40 クエリー中 9 クエリーでは 本手法による拡張クエリーの方が拡張前のクエリーよりも高い precision 値を得ることができていることがわかった。

また、動詞と名詞の共起を利用していることに注目し、検索文書において 初期クエリータームと その関連ターム間の距離が 2 文以内である場合にのみ その文書を正解としたことにより、recall 値全般的における precision 値の向上を見ることができた。

次に (手法 2) の実験結果より、本手法によって全 40 クエリーのうち 20 クエリーにおいて 平均 10 関連概念、12 個の関連タームを獲得することができた。また それらを用いた検索実験により、いくつかのクエリーにおいては本手法で得られた関連タームによる検索精度の上昇が見られた。

また、クエリータームの語義の曖昧性解消実験の結果、人手で語義決定した場合の精度

が最も高く、これにより初期クエリータームの語義の曖昧性解消は 検索精度の向上に貢献することがわかった。本研究で提案した手法については、3 手法を統合して用いることによって、全多義語の約半数に対して、正しい語義を割り当てることができた。

今後は、より正確な処理を行うために係り受け解析等を行う必要があると考える。また今回は 今回の実験セットはサンプル数が非常に少なかったため、大規模な実験セットにおける実験・評価を行う必要があると思われる。