

ユーザ情報に基づいた Web 検索支援システムの提案

朝倉 理浩[†]村山 隆彦[‡]多田 好克[†][†] 電気通信大学 大学院情報システム学研究科[‡] 日本電信電話株式会社 NTT 情報流通プラットフォーム研究所

1 はじめに

ネットワーク上には膨大な数の Web サイトが存在する。そのため、検索エンジンを用いて所望の情報を含むページを探そうとすると、検索結果には様々な内容のページが大量に混在して表示される場合がある。そうすると、ユーザは自らでページを選択して、それに目を通さなければならなくなり、取捨選択に時間と労力を費やすこととなる。

本研究ではユーザの所望の情報を含むページを検索結果の上位に表示するシステムを提案する。

2 検索エンジン

一般的な検索エンジンは、クエリが同じならユーザに関わらず同じ検索結果が表示される。しかし、クエリが同じでも求めている情報はユーザによって異なる場合もある。Google Personalized Search[1] はユーザ情報をユーザの検索履歴から取得し、検索結果を修正して表示する。しかし、検索コンテキストの変化への対応は困難である。

また、検索結果から所望のページが見つからない場合、検索エンジンに与えるクエリを追加して再検索を行い、検索結果を絞り込むのが一般的である。しかし、絞り込みのためのクエリを自らで選択するには、検索しようとしている対象分野についての知識が必要となる。さらに、追加するクエリの選択を誤ると検索結果に表示されるページの内容が偏ってしまい、所望のページを逆に検索結果から除外してしまうこともある。大石ら [2] は、Web 検索を行う際にクエリに関連のある語をユーザに提示することで検索を支援する方法を提案している。しかし、この方法は提示された複数の語の中からユーザが自らで選択し、AND 検索を行わなければならない。

3 研究アプローチ

本研究では以下の点に着目してシステムを設計する。

ユーザ適応 本研究では、ユーザの嗜好や属性などのユー

A proposal of web retrieval support system based on user information

Masahiro Asakura[†], Takahiko Murayama[‡], Yoshikatsu Tada[†]

[†]Graduate School of Information Systems, The University of Electro-Communications

[‡]NTT Information Sharing Platform Laboratories, NTT Corporation

ザ情報とクエリとの両方から連想される語を関連語と定義する。ユーザが所望の情報を得るために検索を行った時、所望の情報を含むページにはクエリと関連のある語が含まれている。また、ユーザの興味に適したページにはユーザの嗜好や属性などと関連のある語が含まれている。よって、クエリとユーザ情報の両方と関連のある語 (関連語) が含まれているページの方が、ユーザの所望の情報を含むページである可能性が高い。そのため、関連語を順位付けの基準として用いることで、ユーザに適した検索結果を得ることができると考える。

検索対象を狭めない 例えば、「Java」で検索しても所望のページが見つからない場合、「Java」と「Windows」の AND 検索をして絞り込みを行う。そうすると、検索結果には「Java」と「Windows」を含むページが表示されるが、「Windows」は検索結果を絞り込むための語であって、必ずしも所望のページに含まれているとは限らない。そのため、「Java」と「Windows」を同列に扱うと検索結果から所望のページ除外してしまう可能性がある。よって、「Java」のみで検索して得たページを対象に、順位付けを行うのが望ましいと考えられる。

以上より、本システムではユーザが入力したクエリのみで検索して得たページを対象に、ページ内に含まれている関連語を考慮して新たに順位付けを行い、検索結果を並び替える。

4 Web 検索支援システム

図1に本システムの概要図を示す。本システムではユーザがユーザ情報を直接入力することで、検索履歴がなくてもすぐに使用できるようにした。本システムの処理の流れは以下の通りである。

Step1 本システムにユーザ情報を登録する。

Step2 ユーザが入力したクエリで Web 検索をする。

Step3 クエリとユーザ情報の両方に関連のある語 (関連語) を Web で収集し、関連の度合いを算出してその語に重みとして与える。

Step4 Step3の関連語を基に Step2の検索結果の各 Web ページに重みを与え、その重みに基づいて検索結果を並び替える。

