

ユーザの検索履歴を用いた情報検索システムの提案

三浦典之

山名早人

早稲田大学理工学部情報学科

1 はじめに

現在、Google は約 30 億の Web ページを収集しているが、それは現在 WWW に存在する Web ページの約 40% ほどに過ぎない。しかも、WWW に存在する Web ページは今後ますます増えつづけると予想される。Google や Yahoo! などの検索エンジンが、世界中の Web ページを網羅することは困難である、という見方が強い。そこで、専門分野のみの情報を網羅した、専用検索エンジンが今後重要な役割を持つようになる。それらの検索エンジンをユーザにとってより使いやすくしよう、という動きも活発になるはずである。

本稿では、専用検索エンジンの一例として、東大とマイクロソフトとの共同研究によって開発された、大規模なテキスト知識ベースに対する対話的な問い合わせシステム「ダイアログナビ[1][2]」を挙げる。そしてダイアログナビの検索履歴を解析した上で、ダイアログナビの問題点を提示する。さらに、その問題点に対する解決策を提案し、実装することによって、ダイアログナビがユーザにとってより使いやすくなることを実証する。

2 ダイアログナビの概要

ダイアログナビとは、東大とマイクロソフトとの共同研究によって開発された、大規模なテキスト知識ベースに対する対話的な問い合わせシステムである。

検索システムに対話の要素を入れる必要性を、東大の黒橋は次のように述べる。「膨大な量の WWW を調べても専門家に聞くような便利さが無いのは、ユーザの質問が完全でない場合が多く、またその質問を完全にするための対話的な能力が計算機にないからである。ユーザの疑問を明確化するためのプロセスとして対話が重要となる[3]」

A Proposal of the Information Search System using User's Search History

Noriyuki Miura, Hayato Yamana

Department of Information and Computer Science, School of Science and Engineering, Waseda University

ダイアログナビは“曖昧な質問に対する聞き返し”という対話の要素を取り入れた情報検索システムである。ダイアログナビは、曖昧な質問に対する聞き返しを、“対話カード”という「ある特定の質問に対して、システムが聞き返す内容をあらかじめ書き留めた」カードを用いることによって実現している。

例えば、ユーザが「エラーが発生した」というような曖昧な質問を入力した場合、システムは対話カードを参照し、「エラーはいつ発生しますか?」という聞き返しを行い、選択肢を提示する。これを繰り返すことによって、ユーザは自らの疑問を明確に把握し、適切な質問文での検索を行うことができる。

3 ダイアログナビの問題点

ダイアログナビは、マイクロソフトのサポート技術情報などのテキスト文書がそのまま検索対象となっている。そのため、テキスト文書に書かれていないキーワードが検索語として入力された場合、ダイアログナビは「検索しましたが何も見つかりませんでした」と出力し、ユーザの検索が行き詰まってしまう。

ダイアログナビの検索履歴を解析した結果、マイクロソフトの製品に関する質問をしているにも関わらず、その質問に対する文書がデータベースに存在しないため、何も検索結果を得ることができなかったユーザや、ダイアログナビの検索対象を理解していないのか、マイクロソフトとあまり関係のない質問を入力し、何も検索結果を得ることができなかったユーザが多数見られた。

4 システムの提案

上記の問題を解決するため、ダイアログナビの検索を行うと同時に、Web ページの検索も行うシステムを提案する。Web ページを検索するのは、WWW にはパソコンに関する質問集やメーリングリストなど、数多くの優れたページが存在するからである。図 1 にシステムの流れを示す。

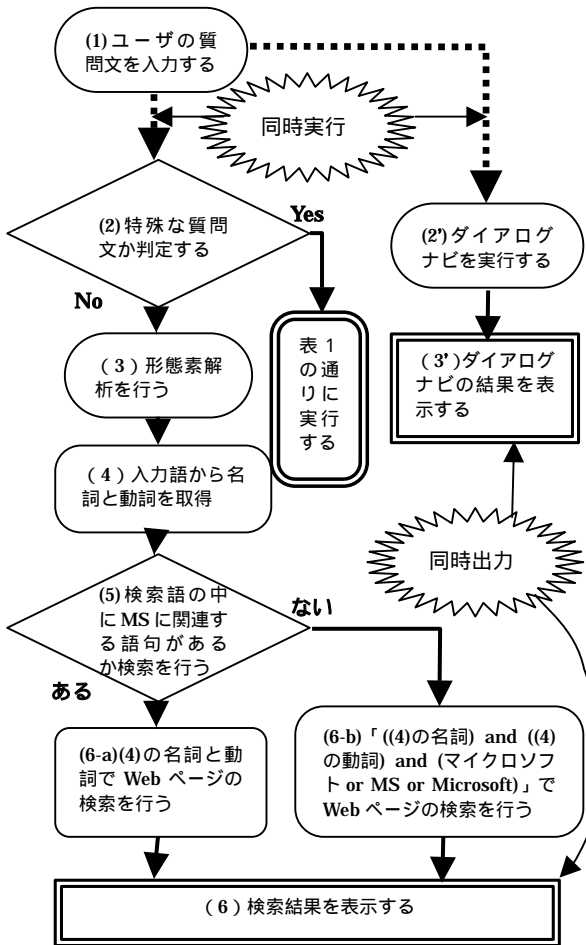


図1 提案するシステムの流れ

以下で図1の補足説明を行う。(1)と(2)、(1)と(2)の間の点線は、その二つを同時実行していることを意味する。(2)の特殊な質問文と、それらへの対応は以下の表に示す。

表1 特殊な質問文への対応

Query 形式	一般の検索エンジンに投入する Query	検索エンジンに投入する Query を変える理由
って何?	「とは」	「何」では良い検索結果が得られない
について...	「とは」	形態素解析の結果が検索に適した形にならない
を教えてください がわからない	「とは」	「教えてください」では良い検索結果が得られない
ができない	「できない」	形態素解析の結果が検索に適した形にならない

(5)の検索語の中に MS に関連する語句があるかは、あらかじめ MS に関連する語句を集めたファイルと、入力された名詞を照らし合わせることによって判断する。

5 システムの実装

図1で示したシステムを実装する。プログラミング言語は Perl を使用した。ユーザ側入出力画面は HTML で作成し、「検索」ボタンをクリックすると CGI プログラムが起動する仕組みになっている。ユーザの質問文に対する形態素解析は「JUMAN[4]」を使用した。以下にシステムの実行画面を示す。

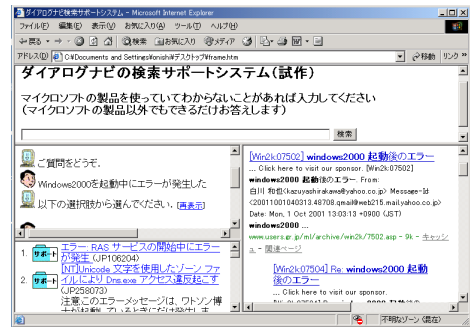


図2 システムの実行画面

6 まとめ

Web ページの増大により、検索エンジンの役割は今よりもさらに重要になる。専門分野のみの情報を網羅した検索エンジンも数多く開発され、ユーザにとってより使いやすくなるよう、様々な検索サポートが行われると考えられる。そこで最も重要なことは、検索履歴を解析するなど、検索エンジンの特徴を十分に理解した上で、検索サポートを行うことである。

謝辞

本研究を行うにあたり、ダイアログナビの情報を提供していただいた東京大学の清田陽司さん、マイクロソフト(株)の木戸冬子さんに深く感謝致します。

参考文献

- [1] “ダイアログナビ”: <http://www.microsoft.com/japan/navigator/>
- [2] 黒橋 禎夫, 清田 陽司, 木戸 冬子: “自動質問応答システム・ダイアログナビの現状と課題”, 情報処理学会研究報告(SIG-SLP), Vol.2002, No.43, pp.19-24.
- [3] 黒橋 禎夫: “大規模テキスト知識ベースに基づく自動質問応答”, 情報処理学会研究報告(SIG-SLP), Vol.2001, No.39, pp.147-152.
- [4] “JUMAN”: <http://www.kc.t.u-tokyo.ac.jp/nl-resource/juman/juman-form.html/>