

検索キーワードに応じて適切な検索 API を選択するメタ検索システムの開発

小西 隆太† 上原子 正利‡ 小柳 滋‡

†立命館大学大学院理工学研究科 ‡立命館大学

1 はじめに

WWW におけるデータの急激な増加と優れた検索エンジンの登場によって、さまざまな情報の入手が可能になった。Google や Yahoo! などの検索エンジンは、クローリング技術やランキングアルゴリズムの改良により、入力されたキーワードを含む適切なウェブページを非常に高い精度で検索できる。また、ブログ、画像、人物などの特定のコンテンツに特化した検索エンジンも現れ、多種多様な情報を効率的に検索することができる。

しかし、現在の検索エンジンはコンピュータリテラシーの高いユーザーにとって十分であるとは言えない。なぜなら、ユーザーに一定水準の選択能力を要求するからである。どの検索機能を使うかの選択、適切な検索キーワードの選択、検索結果からの欲しい情報の選択、これらはコンピュータリテラシーの高いユーザーにとってはそれほど問題とならないが、そうでない人たちにとっては大きな障害となる。

本稿では、ユーザーがさまざまな検索エンジンを自身で判断して使い分けるのではなく、入力されたキーワードに応じて、システムが適切な検索 API を自動的に選択するようなユーザー負担の小さいメタ検索システムを提案する。

2 提案方式

現在の検索エンジンでは、ユーザーに入力された検索キーワードに曖昧性がある場合、検索結果にも曖昧性が生じて、ユーザーの意図しない結果となることがある。そこで、本システムは、入力された検索キーワードが曖昧であっても、ユーザーの求めている情報の提供を可能にするために、ウェブ上の利用可能な多種多様なデータおよびサービスを用いて曖昧性の解消を行う。これは、クエリ拡張を用いた検索キーワードの曖昧性の解消と、検索 API を用いた検索結果の曖昧性の解消である。

まず、本システムはユーザーが入力した検索キーワードにクエリ拡張を行い、曖昧性を自動的に解消することで、ユーザーが本当に欲しい情報の推測を行う。例えば、「横浜」という検索キーワードが入力された場合、ユーザーが本当に欲しい情報としてさまざまなものが考えられる。横浜の中華街に関する情報、横浜の夜景画像、あるいは横浜ベイスターズの試合結果などユーザーの興味によってさまざまであろう。そこで本システムでは、ユーザーが入力した「横浜」というキーワードの代わりに、「横浜 中華街」や「横浜 夜景」、「横浜 試合結果」などのキーワードをクエリ拡張によって生成して、検索に利用する。

次に、本システムは特定のコンテンツ検索に特化した検索エンジンの API を自動的に選択して検索に用いることで、ユーザーが真に求めている情報を本システムの検索結果ページ上に直接表示する。具体的には、「横浜 中華街」という検索キーワードに対して、飲食店検索 API を用いて店舗情報、「横浜 夜景」に対して、画像検索 API を用いて画像情報を検索して、結果ページ上に直接表示する。結果のイメージを図 1 に示す。

これらの検索キーワードと検索結果の曖昧性解消機能によって、検索キーワードの選択や検索結果から欲しい情報を探すというユーザー負担の軽減を目指す。本システムの処理の流れをまとめると図 2 のようになる。ユーザーが検索キーワードを入力すると、システムはそのキーワードの曖昧性を解消した複数の検索キーワードを生成する。そして、これらの検索キーワードごとに適切な検索 API を選択する。最後に、選択された複数の検索 API による検索結果を 1 ページ上にそれぞれ個別に表示する。

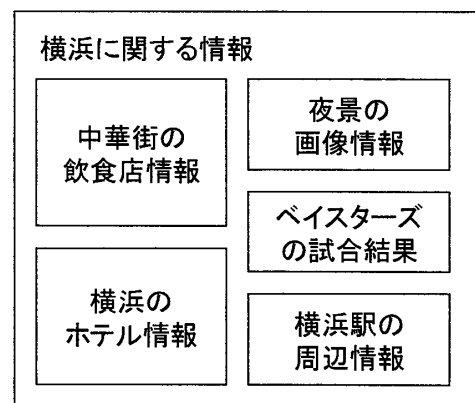


図 1: 検索キーワード「横浜」による結果のイメージ

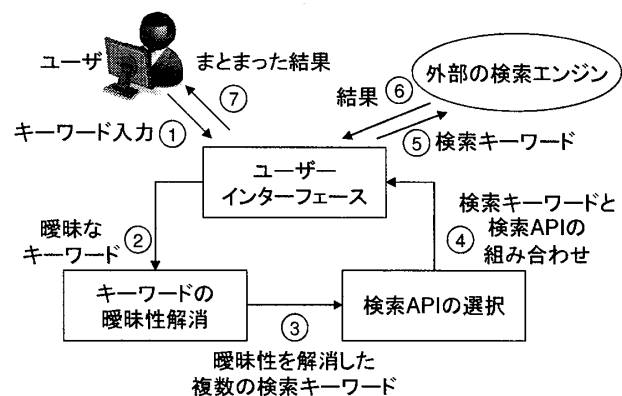


図 2: 処理の流れ

3 実現方法

3.1 検索キーワードの曖昧性解消

検索キーワードの曖昧性を解消するためにクエリ拡張技術を用いる。本システムでは、クエリ拡張を実現するためにウェブ上の利用可能なさまざまなデータを用いてデータベースを作成する。具体的には、オンライン百科事典である

Meta Search System that Selects Appropriate Search APIs according to Keyword

†Ryuta KONISHI ‡Masatoshi KAMIHARAKO ‡Shigeru OYANAGI

†Graduate School of Science and Engineering, Ritsumeikan University

‡Ritsumeikan University

Wikipedia*のエントリとそれに含まれる見出し語、Yahoo!関連検索ワード†などの多くのユーザによって用いられているキーワードの組み合わせ、そして、検索結果のスニペット解析によって抽出された関連語を用いる。これら3つの情報を組み合わせることで精度の高い曖昧性解消機能を目指す。

3.2 検索 API の選択

システムが検索 API を自動的に選択するには、検索キーワードがどの検索 API に適しているのかを調べる必要がある。そこで、本システムでは検索 API とそれに関連の強い単語のデータベースを作成する。本システムでは、さまざまな検索 API を採用予定であり、それらをシステムに組み込む際に各 API に関連の強い単語をデータベースに手動で登録する。具体的には、飲食店検索であれば「店、中華、フランス料理」、画像検索であれば「画像、写真、景色」などである。クエリ拡張によって生成されたキーワードとこれらの単語が一致すればシステムはその検索 API を選択する。

3.3 検索結果の提示

検索結果の提示は、システムが自動的に選択した複数の検索 API による検索結果を1ページ上に個別に表示することで行う。ただし、検索 API によって得られた各検索結果の上位数件のみを提示する。

4 関連技術との比較

4.1 検索キーワードの候補提示

現在の主要な検索エンジンは、多くのユーザによって用いられている検索キーワードの組み合わせを提示する機能を持っている。この機能は入力されたキーワードに関連する複数の検索キーワード候補を提示することによって、ユーザのキーワード決定を助ける。

本システムでは関連する検索キーワードをユーザに提示するのではなく、ユーザが入力した検索キーワードの曖昧性解消に用いる。ユーザに入力された検索キーワードにシステムが自動的に関連語を追加することで曖昧性を解消して、検索キーワードがどの検索 API が適しているかの選択を行う。

4.2 クラスタリングを行う検索エンジン

検索結果のページから目的の情報を探すことを支援するものとして、Clusty[‡]などの検索結果のクラスタリングを行う検索エンジンがある。Clusty はユーザが入力した検索キーワードを他の複数の検索エンジンに問い合わせるメタ検索エンジンで、複数の検索エンジンによって得られた検索結果を分類することで目的の情報をユーザが見つつけやすくしている。

これに対して、本システムでは検索結果をクラスタリングによって分類するのではなく、クエリ拡張によって生成された複数の検索キーワードを用いてさまざまな結果を獲得して、それらを個別に表示することで同等の効果を実現している。

4.3 質問応答

質問応答とは自然言語で表現された質問文の答えを、新聞記事やウェブ上の文書などの大規模なテキストから解答箇所のみを抽出して、直接回答を行う技術である。

本システムでは、質問の入力を自然言語ではなく単なるキーワードに限定しており、自然言語で表現された質問文が入力された場合に、それを意味解釈することはしない。

4.4 メタ検索

メタ検索エンジンは非統合型と統合型の2種類に分けられる。非統合型は検索エンジンの玄関口のようなもので、ユーザが検索する際にどの検索エンジンを利用するかを選択する。統合型は複数の検索エンジンの検索結果を統合するもので、検索エンジンによる結果の違いを比較できる。

これらに対して、本システムは、非統合型のようにユーザが使いたい検索エンジンを選択するのではなく、キーワードに適した複数の検索エンジンをシステムが自動で選択する。これによって、ユーザが検索対象にあわせて検索エンジンを使い分けるために費やす負担の軽減が期待される。

4.5 ユニバーサル検索

ウェブサイトや画像、ニュース、ブログなどさまざまな情報検索を一度の検索で可能にするものがユニバーサル検索である。ユニバーサル検索は大きく2種類に分けられる。検索対象を固定にしているものと、検索対象がキーワードによって自動で選択されるものである。前者は入力された検索キーワードに関わらず、常にウェブサイト、ウィキペディア、ニュース、ブログ、動画などの検索を同時に行い、結果を1ページ上に個別に提示する[§]。これは異なるコンテンツを対象にした統合型メタ検索であるとも考えられる。後者は入力された検索キーワードに応じて、適した検索機能を選択して、その結果を通常のウェブ検索の結果に追加する[¶]。検索キーワードに地名が入力されると、その地名の地図情報を検索結果に表示するといったものである。

また、Yahoo! Glue^{||}はこれら二つの機能をあわせたようなもので、検索キーワードに関係の強い複数の分野を同時に検索して、1ページ上に貼り合わせて表示する。これによってユーザはさまざまな情報検索を同時に、そして的確にできるようになる。

本システムと Yahoo! Glue の結果イメージは非常によく似ている。しかし、本システムと Yahoo! Glue はその目的が異なる。本システムが、検索 API を用いてユーザの求める情報を結果ページ上に表示することで、ユーザ負担を軽減することを目的としているのに対して、Yahoo! Glue は、さまざまな検索機能を同時実行可能にすることによるユーザ体験の向上を目的としている。

5 まとめと今後の課題

本稿では、入力された検索キーワードに応じて、システムが適切な検索 API を自動的に選択することでユーザ負担を軽減するメタ検索システムを提案した。これに基づいて現在開発中である。

今後の課題として、クエリ拡張技術によってどの程度正しく適切な検索 API を選択できるのかを確認しなければならない。また、検索 API の選択方法を手動によるデータベースだけに頼るのではなく、異なる方法も考案したい。

参考文献

- [1] 若木裕美, 正田備也, 高須淳宏, 安達淳. 検索語の曖昧性解消のためのトピック指向単語抽出および単語クラスタリング. 情報処理学会論文誌. データベース, Vol. 47, No. 19, pp. 72-85, 2006/12/15.
- [2] 中山浩太郎, 原隆浩, 西尾章治郎. 大規模 web 事典からのシソーラス辞書構築. 日本データベース学会 letters, Vol. 5, No. 4, pp. 41-44, 2007/03.
- [3] 眞哉村田, 浩之戸田, 由美子 他松浦. 検索結果中のアクセス集中サイトを利用したクエリ拡張法の提案. 日本データベース学会 letters, Vol. 6, No. 4, pp. 45-48, 2008/3.

* <http://ja.wikipedia.org/wiki/>

† <http://developer.yahoo.co.jp/search/webunit/V1/webunitSearch.html>

‡ <http://clusty.jp/>

§ BIGLOBE まどめんとす - <http://matome.biglobe.ne.jp/>

¶ Google - <http://www.google.co.jp/>

|| <http://glue.yahoo.com/>