

実世界オブジェクトを用いた Web 検索の検討

Study of Web Search by Real Objects as Ambiguous Queries

服部 峻

Shun Hattori

埼玉大学地圏科学研究センター^{*1}

Geosphere Research Institute of Saitama University

1. はじめに

近年, Web にはサイバー空間独自の情報だけでなく現実世界に関連する多種多様な情報も優先的にアップロードされるようになり, その情報量は日々爆発的に増え続け, 情報メディアとしての即時性も一二を争っている. そんな中, Web 検索は, 現実世界において必要な情報を効率的に素早く獲得するためのツールとして, 多くの一般ユーザにとって必要不可欠な存在になって来ている. 携帯電話などモバイル環境の整備が進み一般に広く普及したことで, 特にモバイル検索の存在意義が増している. 現状は今いる場とは無関係な暇潰しコンテンツの検索の割合が多く, 従来の固定環境における Web 検索と比べて時空間制約からの解放程度の変化ではあるが, 街中を移動中に出遭った実世界オブジェクトや必要が迫られた行動判断などを契機に生まれた情報要求に対して適時な情報を即時に検索できる技術が非常に重要になって行く.

従来の一般的なキーワード型 Web 検索エンジンでは, 実世界オブジェクトの情報を Web から精度良く検索するためには, 具体的なオブジェクト名やトピック語などの言語を用いて情報要求を正確に表現した検索質問を構成する必要があり, 検索者の能力への依存度は低くない. 検索対象の具体的な名称が不明であったり不正確であったり曖昧な検索質問になってしまうと検索結果の精度が著しく悪化してしまい, モバイル検索では諦めてしまう場合も少なくないため, 情報要求にできる限り合致した検索質問をより簡単に素早く構成できるように検索者を支援するシステムの重要性は大きい.

これまでの研究では, 検索者の現在地や周囲に在るモノなどを暗黙の意図として検索質問を修正する手法や, 検索対象の外観やクラス名などの曖昧な条件から具体的な名称を検索する手法を提案して来た. 本論文では, 別の支援方法として, キーワードだけではなく, 検索者自身が携帯していたり, 街中で出遭った実世界オブジェクトも活用して構成した検索質問による Web 検索について検討を行う. 情報要求に関連する実世界オブジェクトをバーコードや RFID, 写真等で直接的に指定するだけで適切なキーワードに変換されれば, 情報要求を言語で正確に表現する負担を軽減し, 検索者の能力や事前知識への依存度を低減でき, 実世界から Web 上の情報へのユニバーサルアクセスの実現に寄与すると考える.

2. 実世界オブジェクトに基づく検索質問パターン

従来のようにキーワードだけを用いて検索質問を構成するのではなく, 実世界オブジェクトも活用して検索質問を構成するパターンとして考え得るものを挙げる.

言語と実世界オブジェクトを組み合わせて検索質問を構成する場合と, 実世界オブジェクトだけを用いて検索質問を構成する場合とに大別できる.

(1) 実世界オブジェクト 検索質問

バーコードや RFID, 写真等を用いれば実世界オブジェクトと言語情報を関連付けることができ, 検索対象の実世界オブジェクトを直接的に指定するだけで, 検索者が具体的な名称やトピック語を知らなくても, 実世界オブジェクトに関連する情報を Web から検索することが可能である. 単一のキーワードだけを関連付けたり, 複数のキーワード集合をフラット (同等) に関連付けたり, クラス階層として関連付けたり, 公式サイトなどの Web ページを関連付けたり, 関連付ける言語情報の種類, 関連付ける方法は様々に考えられる. 単一のキーワードが関連付けられているだけの場合には単射変換するだけで良いが, ある実世界オブジェクトを検索質問として指定した検索者の情報要求は一般的に多様であるため, 単一のキーワードを関連付けておくだけでは不十分である. 複数のキーワード群や Web ページが関連付けられている場合には, より相応しいキーワードを選択して検索質問を構成する必要があるが, 同一の実世界オブジェクトが指定されたとしても検索者や状況に応じて変化する情報要求を正確に推定することは容易ではない. 例えば, 検索者が CASIO 製の腕時計である G-SHOCK を指定した場合, 「G-SHOCK」という検索質問に変換するのが一般的には妥当であると思われるが, その検索者の真の情報要求はより下位層の個体限定かもしれないし, より上位層の「腕時計」や「CASIO」かもしれない. もちろん, 指定された実世界オブジェクトに対する検索質問の候補をランキング表示し, 検索者に選択させるというインタフェースを採用すれば, 完全に正確に推定できなくても実用上は問題無い. 複数の実世界オブジェクト群を組み合わせて指定されることも考えられるが, 複数のキーワード群の組み合わせを重み付けする必要があるため, 検索者の真の情報要求を推定する困難度はさらに増す.

(2a) 部分文字列 × 実世界オブジェクト キーワード

携帯電話では文字入力時, 予測候補や変換候補を表示して, ユーザの文字入力を支援する機能は一般的である.

^{*1} 現在, 東京工科大学コンピュータサイエンス学部.

また、PC 環境での Web 検索においても、Google サジェストなど、検索質問のキーワードを入力する際に同様の候補を表示するキーワード入力補助が標準サービスとして提供され始めている。これら既存のキーワード補完では、あるキーワードを入力しようと既に入力済みの部分文字列に対するキーワード候補群を、ユーザ個々人のキーワード使用頻度や検索全体での人気度などに基づいてランキングしているが、ユーザが指定した実世界オブジェクト（の名称やトピック語など）に基づいてランキングすることも考えられる。例えば、検索者が「r」を部分文字列として入力済みの状況で、CASIO 製の腕時計である G-SHOCK を指定した場合、「G-SHOCK」に関連するキーワード候補を重要視すべきと判断すれば「ライズマン」「ラバーズコレクション」などを上位にするかもしれないし、「腕時計」に関連する「ロレックス」「ロンジン」などを上位にランキングするかもしれない。全く同じ「r」を入力済みの状況であっても、異なる実世界オブジェクトである GUCCI 製の財布を指定した場合には、「財布」に関連する「ルイヴィトン」「ラルフローレン」などが呈示されるかもしれない。

(2b) 検索質問 × 実世界オブジェクト 検索質問

ある検索質問（一般的にキーワードと AND/OR/NOT オペレータ等で構成される）が既に入力されている状況で、ある実世界オブジェクトを検索者が指定すると、その具体的な名称やトピック語、連想語によって質問拡張したり、或いは相応しいデータベースを選択し直して再検索したりすることが考えられる。例えば、「金閣寺」を検索質問として入力済みの状況で、検索者が携帯する CASIO 製のデジタルカメラである EXILIM を指定した場合、連想語である「画像」「写真」などで「金閣寺」という初期質問を拡張して Web 検索するかもしれないし、通常の文書データベースではなく画像データベースが相応しいと判断して文書検索エンジンではなく画像検索エンジンで「金閣寺」という初期質問のまま検索することを推薦するかもしれない。目の前にある地図案内板を指定した場合には、地図検索エンジンで「金閣寺」という初期質問のまま検索した結果が呈示されるかもしれない。

(2c) Web ページ × 実世界オブジェクト 検索質問

ある Web ページを閲覧している状況で、ある実世界オブジェクトを検索者が指定すると、その Web ページからトピック抽出したキーワード群と、その実世界オブジェクトに関連付けられているキーワード群との組み合わせの中から選定し、より相応しい検索質問を自動構成することが考えられる。また、閲覧している Web ページ自身の中に、ユーザが指定した実世界オブジェクトに強く関連する記述がある場合には、再検索するのではなく、表示箇所の移動だけを自動で行うことも考えられる。例えば、「金閣寺」の Web ページを検索して閲覧している状

況で、CASIO 製の腕時計である G-SHOCK や GUCCI 製の財布を指定した場合、その Web ページの主題語として抽出された「金閣寺」を「腕時計」の連想語である「時間」や「財布」の連想語である「料金」などで拡張した検索質問で再検索するかもしれない。主題語として抽出された「金閣寺」に特化した「参拝時間」や「参拝料」で質問拡張した方が良い検索結果になるかもしれない。

3. 実世界オブジェクトに基づく検索質問の構成モデル

実世界オブジェクトに基づく検索質問の構成パターンを分析すると、単純には図 1 のようなモデルが考えられる。キーワード補完や質問拡張においてキーワード候補をランキングするための重みは、例えば、キーワード候補 kw および実世界オブジェクト obj の関連度 $rel(kw, obj)$ と、部分文字列/初期質問/Web ページ q および実世界オブジェクトの関連度 $rel(kw, q)$ との積で定義できる。

$$weight(kw, obj, q) := rel(kw, obj) \cdot rel(kw, q)$$



図 1: 実世界オブジェクトに基づくキーワードの重み付け

4. まとめと今後の課題

従来の一般的なキーワード型 Web 検索エンジンでは、実世界オブジェクトの情報を Web から精度良く検索するためには、具体的なオブジェクト名やトピック語などの言語を用いて情報要求を正確に表現した検索質問を構成する必要があり、検索者の能力への依存度は低くない。検索対象の具体的な名称が不明であったり不正確であったり曖昧な検索質問になってしまうと検索結果の精度が著しく悪化してしまい、モバイル検索では諦めてしまう場合も少なくないため、情報要求にできる限り合致した検索質問をより簡単に素早く構成できるように検索者を支援するシステムの重要性は大きい。本論文では、実世界から Web 上の情報へのユニバーサルアクセスを最終的な目標として、キーワードだけではなく、検索者自身が携帯していたり、街中で出遭った実世界オブジェクトも活用して構成した検索質問による Web 検索を実現するための最初の検討として、実世界オブジェクトに基づく検索質問の構成パターンとして考え得るものを分類した上で、非常にシンプルな形式でのモデル化を行った。

今後は、検索者が指定した実世界オブジェクトに基づいてキーワードや検索質問の候補を重み付けする具体的な手法を考案し、ユーザ評価を行う必要がある。また、実世界オブジェクトを活用した検索質問の表現能力や、そもそも実世界オブジェクトに起因する情報要求はどの程度存在するのか、実世界オブジェクトを用いた Web 検索のニーズなど、より詳細な検討も続けて行く。