

モバイル・ユビキタス環境のための 実世界コンテキストに基づく検索質問修正

服部 峻[†] 手塚 太郎[†] 田島 敬史[†] 田中 克己[†]

[†] 京都大学大学院 情報学研究科 社会情報学専攻 〒 606-8501 京都府京都市左京区吉田本町

E-mail: †{hattori,tezuka,tajima,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし WWW の情報量の爆発的な増加、及び、モバイル環境の整備によって、日常生活のいつでも、どこでも、大量の Web ページを情報検索しアクセスすることが可能になって来た。モバイル環境における情報検索では、デバイスの入出力性能が乏しく、また、実世界で何らかの活動をしながらの場合も多いため、検索結果をゆっくりとブラウジングしていく時間的な余裕もない。従って、モバイル環境では、検索結果をより一層に洗練する必要がある。しかしながら、一般的な検索エンジンではキーワード式を検索質問として入力する必要があるが、多くのユーザにとって自身の情報要求を複雑なキーワード式で的確に表現することは困難であり、短く、曖昧な検索質問によって、ユーザの情報要求に適合しない文書も混合した大量の検索結果が返されてしまう。このような問題に対して、WWW などの検索対象の文書データベースにおける文脈や、閲覧中または作成中の文書における文脈といったユーザ個々人の情報世界コンテキストを活用して検索質問を拡張することで、固定環境における情報検索を支援する様々な手法が提案されている。これらの従来手法に対して、本稿では、モバイル・ユビキタス環境における情報検索を支援するため、時々刻々と変化するモバイルユーザの状況を表す実世界コンテキストとして、現在の地理的な位置や周囲に在るモノ群に関する情報を活用することで、コンテキスト・アウェアに検索質問を修正する手法について提案を行う。

キーワード モバイル情報検索, 質問拡張, コンテキスト・アウェア, 地理情報.

Query Modification Based on Real-World Contexts in Mobile and Ubiquitous Computing Environment

Shun HATTORI[†], Taro TEZUKA[†], Keishi TAJIMA[†], and Katsumi TANAKA[†]

[†] Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida-Honmachi, Sakyo-ku, Kyoto, 606-8501, Japan

E-mail: †{hattori,tezuka,tajima,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract With the exponentially growing amount of information available on the WWW and the improvement of mobile computing environments, we have been able to access information via mobile Web search engines, anywhere at any time in our daily life. Because mobile devices have the restriction of input/output performance and we have little time for browsing information while moving or doing some activities in the real world, it is necessary to refine the retrieval results better in mobile computing environments. However, since most mobile users' queries are too short and ambiguous for information retrieval systems to guess their information demands accurately, too many retrieval results would be often returned. There exist various approaches for these problems, by utilizing a user query's context in the corpus of Web pages to be retrieved or the user's Virtual-World contexts such as profile, access history, query history or text in the document being read or written. In this paper, for mobile and ubiquitous computing environments, we propose a novel method of query modification based on Real-World contexts, such as user's geographic location or the real objects surrounding the user, aiming to enhance location-awareness, and moreover, context-awareness to the existing location-free information retrieval systems.

Key words Mobile Information Retrieval, Query Expansion, Context-aware, Location-aware.

1. はじめに

近年、インターネットの利用が一般にも広く普及し、特に、WWW (World Wide Web) には新しい情報が日々拡充されており、日常において我々が何らかに関する情報を得たいと思った際に、最初に利用する情報源 (メディア) である場合が多くなって来ている。WWW の情報量は依然として爆発的に増加し続けており、この膨大な情報の中から、ユーザの情報要求に合った Web ページを探すためには、検索エンジンは必要不可欠なシステムであり、その重要性は今後も増すばかりであろう。

一般的に利用される検索エンジンの多くでは、索引語とブール演算子から成るキーワード式を検索質問として入力すると、その条件に合致した内容を持つ Web ページを求め、さらにランキングを行ってリスト形式で表示してくれる。ある著名な人、企業、製品、作品などの公式サイトを閲覧しただけであれば、その名称をそのまま検索質問とした結果の上位にランキングされている場合が多く、ユーザにとって労力はほとんどない。しかし、その対象のサブピックや特定の側面など、より詳細に限定した内容を持つ Web ページを検索したい場合、ユーザは自分の情報要求を正確に表現した検索質問を作成することができないと、ユーザの情報要求に適合しない大量の Web ページも玉石混交した検索結果が返されてしまう。これらの中から本当に必要な情報だけを取捨選択するためには、上位にランキングされた Web ページから順に長時間の労力を掛けて確認していったり、検索結果をさらに絞り込むために、ある程度の専門的な知識や経験を活かして、検索質問を修正したりする必要がある。例えば、「ハリーポッター」は、原作の小説シリーズ、映画、DVD、ビデオゲームなど、様々なサブピックを持っており「ハリーポッター」という検索質問の結果には当然、これらが混合されている。どれか一部に関する情報をユーザが要求している場合には、「ハリーポッター」という検索質問に何らかのキーワードを追加して限定する必要がある。ユーザが入力する検索質問の大半は、単一の索引語そのまま、あるいは、それに準じる程度であり、ユーザの情報要求の複雑さに依って、複数の索引語から成る詳細なキーワード式を作成する必要があることは、多くのユーザにとって非常に大きな負担となっている [1]。

このような従来の一般的な検索エンジンの問題点に対して、情報検索というタスクを支援するための様々な手法が提案されている。例えば、検索結果の俯瞰的な理解を支援するために、検索結果の文書集合を自動的に分類して、各クラスターの代表語や関係性を可視化したり、検索結果をより洗練するために、リランキングや質問拡張を行ったり、さらには、キーワード式ではなく自然文による情報検索などがある。しかしながら、多くのユーザの検索質問は短過ぎるため、検索対象の文書データベースにおける文脈だけから、ユーザの情報要求を正確に推定することは非常に困難である。そこで、検索エンジンに対して不特定多数のユーザ群がこれまでに入力して来た検索質問の履歴を参考にして、検索結果を絞るために追加すると良い索引語の候補をソーシャル・フィルタリング的に推薦したり、ユーザ個々人の状況を表すコンテキストとして、プロフィール情報や

アクセス履歴、ユーザが閲覧中または作成中の文書の内容なども活用することで、検索質問の修正をパーソナライズ (個人適応化) する手法も提案されている。

一方で、デバイスの小型・高性能化、及び、無線通信インフラの整備も進んでおり、モバイル端末を常に携帯しさえすれば、いつでも、どこでも、ロケーションフリーに情報検索を利用することができる環境が整って来た。モバイル環境では、従来の固定環境と比較して、操作性や画面サイズ等のデバイス性能に物理的な制約がより強く、また、実世界で何らかの活動しながらである場合が多く、ゆっくりと Web ブラウジングする時間的な余裕も少ない。従って、モバイル環境では、検索質問の修正を支援するシステムの必要性はより一層大きい。

本稿では、モバイルユーザが置かれている状況が動的に変化することに着目し、ユーザの実世界コンテキストとして、現在いる場所や周囲に在るモノ群に基づいて検索質問を修正する手法について提案する。我々は本手法により、既存のロケーションフリーな情報検索システムに対して、ロケーション・アウェアネス、さらには、コンテキスト・アウェアネスを拡張することを目指している。例えば、「ハリーポッター」という同じ索引語で検索質問したとしても、ユーザが「映画館」にいる場合には、「ハリーポッター」の「映画」に関する情報を要求していると推測できる。同様に、ユーザが「書店」にいる場合には、特に「本」に関する情報を要求していると推測できる。もちろん、ユーザの情報要求が常に状況に依存するとは限らない。センサー技術の発展やユビキタス環境の整備により、大量の実世界コンテキストを取得することが可能になって来ているが、その全ての実世界コンテキストが検索結果を洗練するために有用であるわけではないので、ユーザの検索質問の状況依存性を評価して、その検索結果を洗練するために有用であるか否か、実世界コンテキストの重み付けを行う必要がある。従って、モバイルユーザの置かれている状況を表すコンテキスト語 (現在いる場所語や周囲に在るモノ語) と検索質問との間に何らかの強い関連性が認められる場合に限り、その語を検索質問に追加することで、検索結果の適合率の改善が期待できる。

2. 関連研究

本稿に関連する研究分野として、質問拡張とモバイル検索について紹介し、本研究との類似点および相違点を述べる。

2.1 質問拡張

質問拡張とは、ユーザの初期質問に対して、何らかの関連性を持つ語を追加することによって、その曖昧性を低減し、検索結果の精度向上を目的としており、手製の概念辞書における類義語、反対語、上位・下位語を用いたり [2]、検索対象の文書データベース全体をグローバル解析して構築したシソーラスにおける関連語を用いたり [3]、ユーザの初期質問の検索結果だけをローカル解析したりする手法が提案されている。

適合フィードバック [4],[5] では、まず、ユーザの初期質問の検索結果の上位数件の文書に対して、そのユーザが手作業で適合・非適合の評価を下し、システムにフィードバックする。次に、ベクトル空間モデル上において、検索質問のベクトルを適

合文書群の方向により近付け、非適合文書群からより遠ざけた検索質問ベクトルに更新し、再検索を行う。このようなインタラク션을繰り返すことで、ユーザの情報要求に合った検索結果へと漸次的に改善していく。一方、擬似適合フィードバック [6], [7] では、ユーザの初期質問の検索結果の上位数件の文書群を全て適合文書であると仮定することで、適合評価という負担をユーザに対して掛けることなく、自動的に質問拡張を行うことができる。本稿では、検索質問とコンテキスト語との関連性を測る尺度として、文書データベース全体においてコンテキスト語を含む文書の割合と、初期質問の検索結果においてコンテキスト語を含む文書の割合とを比較することにより算出しており、グローバル解析とローカル解析を併用している。

一方で、検索対象の文書データベースを解析して検索質問に関連する語を抽出するだけでなく、ユーザの置かれている状況として、ユーザが閲覧中あるいは作成中の文書における検索質問の周辺テキスト [8], [9] や検索質問ログ [10] を活用する手法なども提案されている。ユーザのコンテキストを情報検索において考慮している点は本稿と類似している。しかし、これらの手法は情報世界コンテキストだけを活用しており、本稿では、現在位置や周囲に在るモノ群といった実世界コンテキストを活用している点が異なる。実世界コンテキストに基づいた情報配信システムも数多く提案されてはいるが、現在地の周辺の場所 [11], 周囲にいるヒト [12], 周囲に在る展示物 [13] に関連する内容を持つ情報を自動的に検索してユーザに提供するシステムであり、実世界コンテキストを考慮することで、ユーザの検索質問の曖昧性を低減するという本研究の目的とは異なる。

2.2 モバイル検索

モバイル環境に特化した情報検索としては、モバイル端末は画面サイズが小さく計算リソースも乏しいといった性能の低さに対して、モバイル端末での閲覧に適した Web ページだけを選択的に検索したり [15], 任意の Web ページをモバイル端末での閲覧に適した形に変換したり [16], 一方で、モバイル環境の特徴である移動性に着目し、モバイルユーザの現在地に基づいた情報検索を可能にするシステムなどが提案されている。

モバイルサーチエンジン WithAir [15] は、ページタイプ判定技術によりモバイル端末での閲覧に適したページを選択的に収集する機能、モバイルページ中の位置情報（住所や電話番号）を抽出・分類しタウン情報や観光情報といった地域情報を検索できる機能、リンク構造解析による人気度に基づいたランキングとページ要約文抽出により全文検索結果を高精度に表示する機能、人気のキーワードを手軽に入力できる機能、ユーザの検索目的を先読みして優良サイトを呈示する機能などを特徴として有する。ユーザが入力済みである読みから始まる候補キーワード群の中で、検索質問ログに基づいて利用頻度の高い人気キーワードを推定結果として表示するキーワードナビ機能は、モバイル環境における情報検索を支援するタスクの一つとして我々が本稿で取り扱っている入力キーワード推定と非常に類似している。しかし、我々の提案手法は、検索質問ログにおける人気度だけでなく、現在いる場所が何処であるか、周囲にどんなモノ群が在るかといった実世界コンテキストとの関連性の強

さも考慮して、ユーザが入力しようとしているキーワードを推定しており、実世界で置かれている状況に適応したキーワード候補を呈示することが可能である点が異なる。

モバイルインフォサーチ [14] は、インターネットから収集した Web ページの内容から住所表記を抽出して地理空間の経度・緯度にマッピングしておくことで、Web ページが記述している地理的位置とモバイルユーザの現在位置との距離に基づいて、ユーザの状況に適応した Web ページを検索することができる。ユーザの現在位置という実世界コンテキストに基づいて情報検索を行う点は類似しているが、本稿では、検索対象の Web ページとの距離に基づくランキング手法ではなく、ユーザの初期質問と関連性が強いコンテキスト語（場所名やモノ名）により質問拡張する手法について提案している点が異なる。

以上の二つのモバイル検索システムは、地域情報（住所などを内容に含む Web ページ）の検索を主な目的としているが、本稿では、任意の Web ページの検索において、実世界コンテキストを考慮することでユーザの情報要求を推定し、検索質問の曖昧性を低減することによって検索結果の適合率を改善することを主な目的としている点が大きく異なる。

3. コンテキスト・アウェアな質問修正

モバイル環境では、モバイルユーザの置かれている状況は動的に変化するが、その状況を表す最も重要な実世界コンテキストは位置情報（移動性）であり、モバイル環境で利用されるアプリケーションには基本的にロケーション・アウェアネスが求められる。既存のサービスとしては、現在地を地図上に表示して目的地まで道案内するナビゲーション、現在位置の周辺に在る施設や店舗などの最寄り地域情報検索、特定の地理的位置に限定した情報配信サービスなどが実現されている。さらに今後、ユビキタス環境の整備が一層に進めば、あらゆる物に RFID タグなどを貼付することで、モバイルユーザの周囲にどんなモノ群が在るか、様々な無線通信システムを搭載したデバイスが遍在することで、ユーザの周囲にどんな情報が出力されているかといった実世界コンテキストも収集することが可能になる。本稿では、モバイル・ユビキタス環境における情報検索を支援するために、モバイルユーザの現在地や周囲に在るモノ群といった実世界コンテキストを活用することで、ユーザの情報要求をより詳細に推定して、検索質問を修正する手法を提案する。

3.1 システムの概要

本章では、モバイル・ユビキタス環境における情報検索のユーザ支援として、次の二種類のタスクに対して取り組み、実世界コンテキストを活用する手法について順に述べる。

(1) 質問拡張：

ある検索質問（キーワード式）をユーザがシステムに投げると、その検索結果をより絞り込むために追加するのに適当な索引語を求め呈示する。

(2) キーワード推定：

ある索引語をユーザが入力しようとしている途中で、既に入力された部分文字列（読み）から、そのユーザが入力しようとしている索引語を推定する。

本稿で我々が提案する実世界コンテキストに基づく検索質問修正システムは以下の四段階の処理プロセスから構成される。このシステム概観を図1のようになる。

Step 1. 実世界コンテキストの取得：

モバイルユーザが何処にいるかを表す実世界コンテキストとして、GPSなどの測位システムにより地理的な位置（経度・緯度）を、或いは、その場所に埋め込まれたRFIDタグにリンクされた場所情報を取得する。また、モバイルユーザの周囲にどんなモノ群が在るかを表す実世界コンテキストとして、実世界のモノに埋め込まれたRFIDタグをリードすることで、リンクされたモノ情報を収集する。

Step 2. コンテキスト語への変換：

実世界から取得したコンテキスト・データを、地理情報システム（GIS）やシソーラス、オントロジーに通すことで、場所語やモノ語といったコンテキスト語に変換する。

Step 3. コンテキスト語の重み付け：

モバイルユーザの入力した検索質問、或いは、入力済みの部分文字列に前方一致する各キーワード候補（ q ）に対するコンテキスト語（ c ）の関連（依存）性 $R_{cq}(c, q)$ 、及び、モバイルユーザ（ u ）とコンテキスト語（ c ）との関連性 $R_{cu}(c_i, u)$ を統合することで、各コンテキスト語の重要度を評価する。

$$W_{cuq}(c, u, q) = F(R_{cq}(c, q), R_{cu}(c, u))$$

但し、結合関数としては、線形結合や積など様々に考えられ、各値が一定以上などの制約を課すこともできる。

Step 4. 質問修正の実行：

質問拡張では、モバイルユーザの検索質問と強い関連性を持つコンテキスト語を質問拡張の候補語として呈示する。一方、キーワード推定では、全てのコンテキスト語とより強い関連性を持つキーワード候補を推定結果として呈示する。

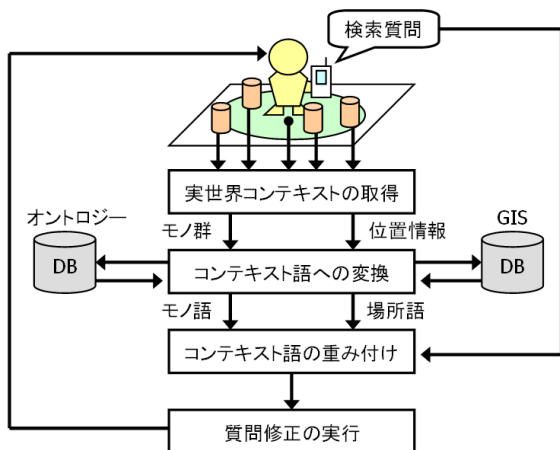


図1 実世界コンテキストに基づく質問修正システムの概観
Fig.1 Overview of Context-aware Query Modification System

3.2 コンテキスト語と検索質問との関連性

コンテキスト語（場所語とモノ語）および検索質問（あるいはキーワード候補）は共に、複数の索引語から成るキーワード式である。まず、二つの単語間の関連度を評価する指標は様々な提案されており、次のように分類できる。

1. 個人や不特定多数ユーザ群の検索質問ログ（履歴）
2. 手製シソーラス辞書における関連枝に基づく距離
3. 検索対象の文書データベースにおける特徴量

第一の手法は、ユーザ個人が過去に行った検索質問の履歴、検索エンジンが蓄積している不特定多数ユーザ群が過去に行った検索質問の履歴を参照することで、検索質問の履歴の中で共に頻出する度合いに基づいて、単語間の関連度を評価することができる。検索質問を拡張するためのコンテキスト語の重み付けに利用する場合、検索者としてのユーザが実際に過去に入力した検索質問の履歴は、ユーザの情報要求を直接的に反映しているため、非常に有用である。既存の検索エンジンにおいても、検索質問を入力している最中や検索結果ページに関連語のリストを呈示するサービスも提供され始めている。

第二のシソーラス上での距離は、節点が単語であり、複数の単語から成るキーワード式ではないため、検索質問やコンテキスト語が単一の索引語の場合には適用することが可能であるが、キーワード式の場合にはそのまま適用することはできない。

第三の手法としては、共起頻度、Jaccard 係数、TF-IDF 法、相互情報量 [18]、 χ^2 検定値 [20]、代表性 [19]、分節性 [21]、RSV (Robertson's Selection Value) [22] など、様々な指標が提案されている。例えば、Jaccard 係数は、

$$Jaccard(c, q) = \frac{|D(c \wedge q)|}{|D(c \vee q)|}$$

により算出される。但し、 $|D(q)|$ は検索対象の文書データベースに対してキーワード式（ q ）によって検索した結果の件数を表す。また、TF-IDF 法による関連性評価は、

$$tfidf(c, q) = -Pr(c|q) \cdot \log_2(Pr(c))$$

$$Pr(c|q) = \frac{|D(c \wedge q)|}{|D(q)|}$$

$$Pr(c) = \frac{|D(c)|}{N}$$

と計算される。但し、 N は検索対象の全文書集合の要素数を表す。一方、相互情報量、 χ^2 検定値、分節性などは、単語（ t_i ）の単語（ t_j ）に対する各々の特徴量を $w(t_j; t_i)$ と表すと、あるコンテキスト語（ c ）の重要度は次式のように記述できるが、

$$R_{cq}(c, q) = \frac{|D(c \wedge q)|^2}{|D(c)|} \cdot \sum_{t_j \in c} w(t_j; c)$$

後半の項は基本的にあらゆる単語とコンテキスト語（ c ）との特徴量を計算する必要があり、計算量が非常に大きく、また、検索質問にも全く依存していない。

本稿では、検索対象の全文書においてコンテキスト語（ c ）を含む割合 $Pr(c)$ と、検索質問（ q ）の検索結果においてコンテ

キスト語 (c) を含む割合 $Pr(c|q)$ との比および差を掛け合わせることで、検索質問 (q) のコンテキスト語 (c) に対する関連 (依存) 性 $R_{cq}(c, q)$ を算出する。

$$R_{cq}(c, q) = \frac{Pr(c|q)}{Pr(c)} \cdot \{Pr(c|q) - Pr(c)\}$$

もし、検索質問 (q) とコンテキスト語 (c) とが互いに完全な独立 (無相関) であるとすれば、

$$Pr(c|q) = Pr(c) \Leftrightarrow R_{cq}(c, q) = 0$$

となる。正に十分に大きいコンテキスト語 (c) ほど、検索質問 (q) に対して特徴的であると言えるという仮説に基づいている。

3.3 コンテキスト語とユーザとの関連性

モバイルユーザ個々人の情報要求をより正確に推定し、検索質問に対して各コンテキスト語の重要度を評価するためには、コンテキスト語と検索質問との関連性を評価するだけでは不十分である。モバイルユーザが検索質問を行うか否か、或いは、その検索質問の内容に依らず、現在置かれている状況に対して、ユーザがどの程度関心を持っているか否かも評価する必要がある。検索質問と強い関連性を持つコンテキスト語であっても、ユーザが関心をあまり持っていないのであれば、そのコンテキスト語で質問拡張したとしても、ユーザの情報要求には近付かない。一方、あるキーワード候補がコンテキスト語と強い関連性を持っていたとしても、そのコンテキスト語自体がユーザにとって重要でないのであれば、そのキーワード候補を入力しようとしていると推定することは適切でない。

場所語とユーザとの関連性を評価する判断材料としては、その場所自体を、或いは、その傍を通過したことがある頻度に対して、その場所に実際に留まった頻度の割合が大きいほど、ユーザは強い関心を持っていると推測することができる。もちろん、ユーザのプロファイルやアクセス履歴、デバイスにローカル保存されている文書群の内容なども活用することができる。

一方、モノ語とユーザとの関連性を評価する場合、ある場所にユーザが留まっていたとしても、ユーザの周囲に在る全てのモノ群が重要であるわけではなく、その場所とモノとの関連性を考慮する必要がある。場所とモノとの関連性は、そのモノが実世界でユーザの周囲に存在する出現頻度 (OF: Object Frequency) と、場所語とモノ語という語間の関連性を基に評価する。前者は、ユーザの周囲に存在する数が多いモノ語ほど重要であることを意味し、後者により、それが置かれている場所との関連性が強いモノほど、その場所にいるモバイルユーザにとって重要であると推測する。例えば、「本屋」という場所であれば、その場所に強く関連するモノである「本」というクラスに属するインスタンスは重要であると、その場所に訪問者が持ち込んだバッグなどが周囲にたくさん存在しても、多くのユーザにとっては重要ではないと判断する。もちろん、ユーザ個々に依っては、その場所との関連性が弱い、訪問者などが持ち込んだモノに対して非常に強い関心を持っていることも有り得る。しかし、そのモノが在ったのは偶然であり、その場所に訪れた元々の目的とは掛け離れているため、本稿では考慮しない。

3.4 コンテキスト・アウェアな質問拡張

モバイルユーザが、ある場所で、あるモノ群に囲まれながら、自身の情報要求を表現したキーワード式を検索質問として入力すると、その現在地や周囲に在るモノ群を変換したコンテキスト語 (場所語やモノ語) に対して、検索質問との関連性、および、ユーザ自身との関連性に基づいて重要度を評価し、重要度の高いコンテキスト語によって質問拡張を実行する。

例えば、「紀伊国屋書店 梅田店」内にいるモバイルユーザが検索質問として「ハリーポッター」と入力したとする。まず、現在の地理的位置、或いは、その場所に埋め込まれた RFID タグからリンクした場所情報を取得する。次に、地理情報システムに通すことにより「紀伊国屋書店 梅田店」という場所語に変換し、単純に緩和した「紀伊国屋書店」という場所語も得られる。さらに、タウンページ情報などを参照することで、その場所の業種である「書店」や「本屋」という親概念を次々と辿って行き、場所語のリストを取得する。そして、各場所語 (p_i) に対して、「ハリーポッター」という検索質問 (q) との関連性 $R_{cq}(p_i, q)$ 、及び、ユーザ (u) との関連性 $R_{cu}(p_i, u)$ を算出し統合することで、検索質問を拡張する候補である場所語それぞれの重要度を評価する。図 2 の例では、場所語リストの中で「本屋」の重要度が最も大きくなったため、検索質問である「ハリーポッター」に追加して質問拡張を行っている。「本屋」にいるユーザは特に「本」に関する情報に興味があると推測できるので、「ハリーポッター」に「本屋」という場所語を追加することで、映画やビデオゲームに関する情報をフィルタリングした検索結果を得ることができる。「ハリーポッター」はサブピックとして、本 (書籍) の他に、映画や DVD、ビデオゲームなども持つため、モバイルユーザが「映画館」や「ゲームショップ」にいる場合には、各場所語と検索質問との関連性が高くなり、検索質問に追加することで、それぞれ「映画」や「ビデオゲーム」に特化した検索結果に洗練できる。一方、「ハリーポッター」との関連性が低い「レストラン」や「コンビニ」などの場所においては、モバイルユーザがどのサブピックに関する情報を特に要求しているのか推測することができないため、質問拡張は行わず、初期質問そのままを検索した結果を呈示する。

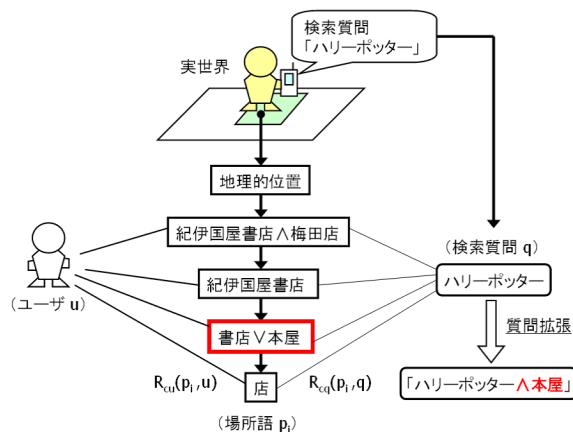


図 2 場所語に基づく検索質問の拡張

Fig. 2 Query Expansion Based on Place-Names

上述の場所語に基づく検索質問の拡張と同時に、モノ語に基づく検索質問の拡張も処理される。まず、周囲に在るモノ群に添付された RFID タグをリードしてモノ語を取得する。また、ある場所にどんな人や物が存在するかという存在物データも保持するシソーラスである常識場所判断システム [17] に場所語を通したり、業種を表す場所語から接尾の「店」「屋」「館」などを除去したりすることで、ユビキタス環境が整備されていない段階でも、ある程度のモノ語を取得できる。例えば、「本屋」から接尾の「屋」を除去することで「本」というモノ語が得られ、さらに「本」の下位語として「漫画」や「小説」などのモノ語も収集できる。そして、各モノ語 (o_i) に対して、検索質問 (q) との関連性 $R_{cq}(o_i, q)$ 、および、ユーザ (u) との関連性 $R_{cu}(o_i, u)$ を算出し、これら二つの関連性を統合することで、検索質問を拡張する候補であるモノ語それぞれの重要度を評価する。図 3 の例では、モバイルユーザの周囲から全部で四つのモノ（インスタンス）が収集され、各モノ語の重要度を評価した結果「本」というモノ語が最大値となったため、検索質問の「ハリーポッター」に追加する索引語として選択されている。

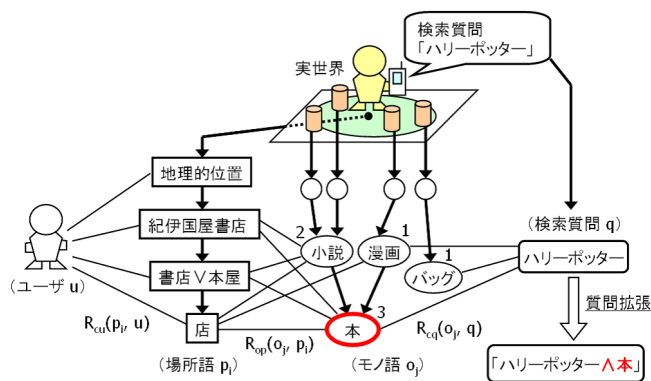


図 3 モノ語に基づく検索質問の拡張
Fig. 3 Query Expansion Based on Object-Names

3.5 コンテキスト・アウェアなキーワード推定

モバイルユーザが、ある場所で、あるモノ群に囲まれながら、ある索引語を入力しようと、その一部分（読み）まで既に入力済みである時、その部分文字列に前方一致する各キーワード候補に対して、その所在地や周囲に在るモノ群を変換した各コンテキスト語（場所語やモノ語）との関連性、および、そのコンテキスト語とユーザ自身との関連性に基づいて重要度を評価し、重要度の高いキーワード候補をそのユーザが入力しようとしている索引語と推定する。実世界コンテキストは考慮しないが、ユーザのプロファイルや検索質問の履歴に基づいてキーワード推定を行うシステムは、携帯電話の入力インターフェースなどとして既に広く採用されている。本稿では、さらに、実世界で置かれている状況も考慮することで、コンテキスト・アウェアなキーワード推定を実現している。

例えば、モバイルユーザが検索質問を作成する際、それを構成する索引語として「ハリーポッター」と入力しようとしていたとする。その索引語の一部分である「はり」まで入力し終わっている時点では、実世界コンテキストを考慮しなくても、「ハ

リ」に前方一致する「ハリアー」「ハリーポッター」「ハリウッド」「ハリケーン」「ハリソン」「ハリガネムシ」などのキーワードを候補として推測することができる。Google サジェスト [23] では、検索質問の履歴における利用頻度や全文書中で各語を含む割合などに基づいてランキングして呈示している。

まず、コンテキスト・アウェアな質問拡張と同様に、実世界コンテキストを取得し、それらをコンテキスト語（場所語とモノ語）に変換する。次に、各コンテキスト語 (c_i) に対する各キーワード候補 (q_k) との関連性、および、ユーザ (u) との関連性を統合することで、各キーワード候補の各コンテキスト語に対する重要度 ($Weight_{cuq}(c_i, u, q_k)$) を算出する。そして、全てのコンテキスト語に対する重要度の総和を取ることで、ある状況 (l) 下のユーザ (u) に対する各キーワード候補の重要度 $W_{qui}(q_k, u, l)$ を評価する。

$$W_{qui}(q_k, u, l) = \sum_{c_i \in l} W_{cuq}(c_i, u, q_k)$$

最後に、重要度が一定値以上のキーワード候補について、その値が大きい順に呈示することで、モバイルユーザが検索質問を作成する労力を低減することができる。図 4 の例では、「書店」に対して「ハリーポッター」の関連性が最も強いために、「書店」にいるモバイルユーザに対して「ハリーポッター」を索引語の推定結果として呈示している。一方、映画に関連する場所である「映画館」にいる場合であれば「ハリーポッター」や「ハリウッド」が、自動車に関連する場所である「ディーラー」にいる場合であれば「ハリアー」の重要度が高く評価され、キーワード推定の結果として呈示される。

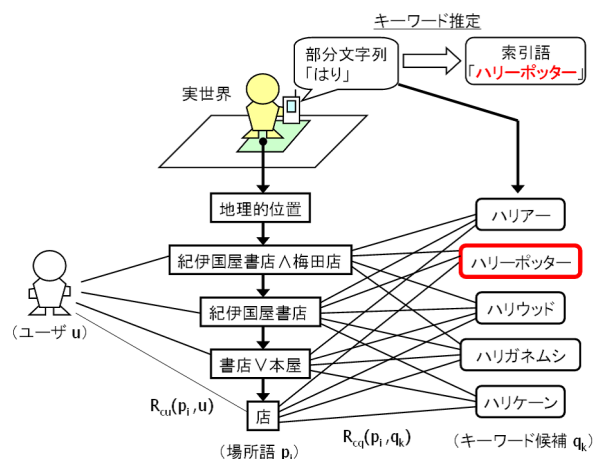


図 4 場所語に基づくキーワード推定
Fig. 4 Keyword Inference Based on Place-Names

4. 実験結果

本章では、実世界コンテキストに基づくコンテキスト語（場所語とモノ語）が、質問拡張やキーワード推定という二種類の情報検索のユーザ支援タスクに有用であることを確認するためにに行った実験結果を示し、その考察を述べる。まず、コンテキスト・アウェアな質問拡張において、本稿で用いているコンテ

キスト語と検索質問との関連性の指標が有効であること、重要度を高く評価されたコンテキスト語で実際に質問拡張することで適合率が改善されることを確認する。次に、コンテキスト・ウェアなキーワード推定において、場所語およびモノ語とキーワード候補との関連性を評価することで、その状況に適したキーワード候補を推定することができることを確認する。

4.1 実験 (1) 質問拡張

「本」「映画」「DVD」「ゲーム」などのサブピックを持つ「ハリ Potter」「家電」「ゲーム」「音楽」「映画」「出版」「銀行」などの子会社を持つ「ソニー」「自動車」「動物」「ミックジャガー」などのサブクラスを持つ「ジャガー」という三種類の検索質問に対して、様々な場所語との関連性を算出すると表 1 のようになる。但し、検索件数として、MSN サーチ [24] で日本語ページ限定検索した結果の件数を用いている。

表 1 検索質問と場所語の関連度

Table 1 Relevance between Query and Place-Name

	ハリ Potter	ソニー	ジャガー
映画館	(1)1.298	0.115	0.123
書店	(2)0.806	0.152	0.075
本屋	(3)0.568	0.124	(2)0.137
図書館	0.367	0.042	0.019
大学	0.224	(2)0.285	0.048
コンビニ	0.169	(3)0.216	0.136
小学校	0.136	0.038	0.058
病院	0.085	0.051	0.029
動物園	0.071	0.025	(3)0.132
レストラン	0.059	0.064	0.048
公園	0.053	0.022	0.060
駅	0.043	0.053	0.031
銀行	0.039	(1)0.538	0.035
ディーラー	0.017	0.082	(1)1.198
ゲームショップ	0.013	0.029	0.009
CD ショップ	0.012	0.052	0.024
ビデオショップ	0.008	0.006	0.001
電器店	0.003	0.017	0.001
カーディーラー	-0.000	0.002	0.031

まず「ハリ Potter」では、「映画館」や「書店」「本屋」との関連性は期待通りに高くなっているが、「ビデオショップ」や「ゲームショップ」とは無関係になってしまっている。また、児童向け小説であるために「図書館」「小学校」などとの関連性も高くなっており、「図書館」から「ハリ Potter」という検索質問を投げた場合、「図書館」を質問拡張の候補語として呈示してしまう。実際に追加してみたが、「図書館」にいるモバイルユーザにとって望ましいと思われる検索結果は得られなかった。次に、「ソニー」では、「銀行」との関連性は期待通りに高くなっているが、「映画館」や「書店」とはやや低く、「電器店」「ゲームショップ」「CD ショップ」とは無関係になってしまっている。また、就職関連のためか「大学」との関連性も高くなっている。最後に、「ジャガー」では、「ディーラー」や「動物園」との関連性は期待通りに高くなっているが、「CD ショップ」と

の関連性は低くなってしまっている。この原因は「ミックジャガー」が昨今のアーティストではないためであろう。また、全体を通して「コンビニ」との関連性がやや強く出ている。以上から、検索質問の場所依存性を測る尺度として、本稿で用いているコンテキスト語と検索質問との関連度では必ずしも期待通りの結果が得られておらず、今後、改善する必要がある。

次に、「ハリ Potter」という検索質問をコンテキスト語で質問拡張した場合の検索結果の上位 20 件における各サブピックの含有率を手作業で計算すると表 2 のようになる。モバイル環境における情報検索では、上位数件しかチェックする時間的な余裕がないと考えられるため、上位 20 件における適合率の改善を確認している。但し、サブピックを網羅する Web ページやファンサイトは「その他」に分類している。

表 2 検索質問「ハリ Potter」のコンテキスト語での質問拡張による上位 20 件の適合率の改善

Table 2 Improving Approximate Precision by Adding Contextual Word (Name of Place or Object) to User Query ("harry potter")

	関連度	本	映 画	D V D	ゲーム	その他
なし	—	0.20	0.30	0.05	0.10	0.35
本屋	0.568	0.80	0.15	0.00	0.00	0.05
映画館	1.298	0.00	0.90	0.00	0.00	0.10
ビデオショップ	0.008	0.00	0.25	0.20	0.00	0.55
ゲームショップ	0.013	0.00	0.00	0.00	0.50	0.50
小学校	0.136	0.15	0.15	0.00	0.00	0.70
本	3.914	0.80	0.00	0.00	0.00	0.20
映画	3.547	0.05	0.75	0.10	0.00	0.10
DVD	1.251	0.05	0.05	0.75	0.00	0.15
ゲーム	1.790	0.05	0.05	0.05	0.45	0.40
自動車	0.025	0.25	0.20	0.00	0.00	0.55

各サブピックに直接的に関連する場所語やモノ語で質問拡張すると、多くの場合、上位 20 件における適合率が大きく改善されている。しかし、各サブピックに直接的に関連しないが「ハリ Potter」との強い関連性を持つと上述で評価された「小学校」で質問拡張した場合、「小学校」にいるモバイルユーザがどのサブピックに対して情報要求しているかは定かではないが、どのサブピックについても適合率が悪化してしまっている。また、コンテキスト語で質問拡張すると、「その他」に分類される割合が増加してしまう傾向も見受けられる。以上から、検索質問とコンテキスト語との関連性と、そのコンテキスト語で質問拡張した場合の適合率の改善性との相関が必ずしもあるとは言えない結果となっており、各サブピックに直接的に関連しないコンテキスト語で質問拡張してしまうと適合率は逆に悪化してしまうため、各コンテキスト語が検索質問に関連するか否かをより正確に判定する必要がある。

4.2 実験 (2) キーワード推定

「ハリ」という同じ部分文字列を入力したとしても、現在地に依って推定結果が変わる。「書店」にいれば「ハリ Potter」を、「映画館」にいれば「ハリウッド」や「ハリ Potter」を、「ディーラー」にいれば「ハリアー」を、「病院」にいれば「ハリガネムシ」を推定結果として呈示することになる。ユーザの

実際の情報要求としては、「映画館」にいる場合、「ハリウッド」よりも「ハリポッター」の方が上位に来るべきであるかもしれない。検索対象の文書データベースを解析して得られた関連度は、情報を提供する側の意図を反映しているため、情報を利用する側の意図とは必ずしも一致しない。従って、ユーザ側の検索質問の履歴を活用する必要があるであろう。

表3 索引語「ハリ...」と場所語との関連度

	書店	映画館	ディーラー	学校	病院
ハリポッター	0.806	1.298	0.017	0.314	0.085
ハリア	0.018	0.032	2.758	0.008	0.003
ハリウッド	0.243	4.123	0.044	0.521	0.193
ハリガネムシ	0.126	0.064	-0.000	0.681	0.758
ハリケーン	0.140	0.271	0.050	0.547	0.281

表4 索引語「ハリ...」とモノ語との関連度

	書籍	映画	自動車	動物	椅子
ハリポッター	9.936	3.547	0.025	0.431	0.067
ハリア	0.003	0.073	3.045	0.019	0.016
ハリウッド	0.198	11.430	0.126	0.412	0.237
ハリガネムシ	0.380	1.301	0.017	5.024	0.103
ハリケーン	0.158	1.376	0.321	0.586	0.237

5. まとめと今後の課題

本稿では、モバイル環境における情報検索を支援するため、現在いる地理的位置や周囲に在るモノ群といった実世界コンテキスト・データを、場所語やモノ語といったコンテキスト語に変換し、Webマイニングした語間の関連性に基づいてモバイルユーザの検索質問を状況に適應して修正する手法を提案した。本手法により、従来のロケーション・フリーな情報検索に対して、コンテキスト・ウェアネスを拡張できることを確認した。

今後は、検索質問の状況依存性の尺度の改良、及び、モバイルユーザの情報要求と合致しているか否かの検証を行っていく。

謝 辞

本研究の一部は、21世紀COEプログラム「知識社会基盤構築のための情報学拠点形成」、文部科学省科学技術振興費プロジェクト「異メディア・アーカイブの横断的検索・統合ソフトウェア開発」(代表:田中克己)、および、平成17年度科研費特定領域研究(2)「Webの意味構造発見に基づく新しいWeb検索サービス方式に関する研究」(課題番号:16016247,代表:田中克己)によるものです。ここに記して謝意を表すものとします。

文 献

[1] D. Butler: "Never trust a human", *Nature*, Vol.405, pp.112-115, May 2000.
 [2] E. M. Voorhees: "Query Expansion Using Lexical Semantic Relations", In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.61-69, July 1994.
 [3] Y. Qiu and H.-P. Frei: "Concept Based Query Expansion", In *Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.160-169,

June/July 1993.

[4] J. J. Rocchio: "Relevance Feedback in Information Retrieval", *The SMART Retrieval System: Experiments in Automatic Document Processing*, pp.313-323, Prentice Hall, 1971.
 [5] G. Salton and C. Buckley: "Improving Retrieval Performance by Relevance Feedback", *Journal of the American Society for Information Science*, Vol.41, No.4, pp.288-297, June 1990.
 [6] J. Xu and W. B. Croft: "Query Expansion Using Local and Global Document Analysis", In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp.4-11, August 1996.
 [7] J. Xu and W. B. Croft: "Improving the Effectiveness of Information Retrieval with Local Context Analysis", *ACM Transactions on Information Systems*, Vol.18, No.1, pp.79-112, January 2000.
 [8] J. Budzik and K. J. Hammond: "User Interactions with Everyday Applications as Context for Just-in-Time Information Access", In *Proceedings of the 2000 International Conference on Intelligent User Interfaces*, ACM Press, 2000.
 [9] T. Kawashige, S. Oyama, H. Ohshima and K. Tanaka: "Context Matcher: Improved Web Search Using Query Term Context in Source Document and in Search Results", In *Proceedings of the 8th Asia Pacific Web Conference*, pp.486-497, January 2006.
 [10] H. Cui, J.-R. Wen and W.-Y. Ma: "Probabilistic Query Expansion using Query Logs", In *Proceedings of the 11th International Conference on World Wide Web (WWW2002)*, pp.325-332, ACM Press, May 2002.
 [11] G. D. Abowd, C. G. Atkeson, J. Hong, S. Long, R. Kooper and Mike Pinkerton: "Cyberguide: A Mobile Context-Aware Tour Guide", *ACM Wireless Networks*, Vol.3, pp.421-433, 1997.
 [12] B. Rhodes: "Using Physical Context for Just-in-Time Information Retrieval", *IEEE Transactions on Computers*, Vol.52, No.8, pp.1011-1014, August 2003.
 [13] J. Rekimoto, Y. Ayatsuka and K. Hayashi: "Augment-able Reality: Situated Communication through Physical and Digital Spaces", In *Proceedings of the 2nd International Symposium on Wearable Computers*, pp.68-75, October 1998.
 [14] 三浦 信幸, 横路 誠司, 井上 香織, 高橋 克巳, 高橋 健司, 島 健一: "位置指向の情報構造化と情報フィルタリング ~ モバイルインフォサーチ 3 実験 ~", 情報処理学会 研究報告「モバイルコンピューティングとユビキタス通信」, Vol.1999, No.97, pp.39-44, 1999年11月.
 [15] 河合 英紀, 赤峯 享, 喜田 弘司, 松田 勝志, 福島 俊一: "モバイルサーチエンジン WithAir の試作と評価", 情報処理学会 研究報告「情報学基礎」, Vol.2001, No.86, pp.71-76, 2001年9月.
 [16] 近藤 宏行, 灘本 明代, 田中 克己: "モバイル環境における検索エンジンの出力結果の再構成と呈示", 情報処理学会 研究報告「データベースシステム」, Vol.2000, No.69, pp.199-206, 2000年7月.
 [17] 杉本 二郎, 渡部 広一, 河岡 司: "概念ベースを用いた常識場所判断システムの構築", 情報処理学会 研究報告「自然言語処理」, Vol.2003, No.004, pp.81-88, 2003年1月.
 [18] K. W. Church and P. Hanks: "Word association norms, mutual information, and lexicography", *Computational Linguistics*, Vol.16, No.1, pp.22-29, 1990.
 [19] T. Hisamitsu, Y. Niwa, S. Nishioka, H. Sakurai, O. Imaichi, M. Iwayama and Akihiko Takano: "Extracting Terms by a Combination of Term Frequency and a Measure of Term Representativeness", *Terminology*, Vol.6, No.2, pp.211-232, 2000.
 [20] 松尾 豊, 石塚 満, "語の共起の統計情報に基づく文書からのキーワード抽出アルゴリズム", *人工知能学会論文誌*, Vol.17, No.3, pp.213-227, 2002.
 [21] 若木 裕美, 正田 備也, 高須 淳安, 安達 淳, "検索語の曖昧性を解消するキーワードの提示手法", 情報処理学会 研究報告「データベース」, Vol.2005, No.68, pp.269-276, 2005年7月.
 [22] S. E. Robertson: "On Term Selection for Query Expansion", *Journal of Documentation*, Vol.46, No.4, pp.359-364, 1990.
 [23] Google サジェスト: <http://www.google.co.jp/webhp?complete=1>.
 [24] MSN サーチ: <http://search.msn.co.jp/>.