

Title	Knowledge Extraction from the Web and Its Application to Location-Based Information Systems(Abstract_要旨)
Author(s)	Hattori, Shun
Citation	Kyoto University (京都大学)
Issue Date	2009-03-23
URL	http://hdl.handle.net/2433/123845
Right	
Type	Thesis or Dissertation
Textversion	none

氏名	服部 峻
----	------

(論文内容の要旨)

本論文は、モバイル環境におけるWeb検索、および、実空間の地理案内のために、Webからの概念階層や外観情報などの知識の自動抽出手法とその位置情報システムに対する適用に関して研究を行ったものである。

近年、モバイル端末、特に携帯電話が広く普及し、モバイルアクセス網などのインフラの整備も着実に進み、従来のデスクトップPCによる情報アクセスだけでなく、モバイル端末による情報アクセスも恒常化しつつある。モバイル環境では音楽や動画コンテンツの利用が多い一方、街中で出遭うオブジェクトに関する情報の検索や、目的地まで移動するための道案内システムに対する社会的な需要も大きい。しかし、既存のモバイルWeb検索システムは、依然として、誰もが容易に精度良く検索できるようにはなっておらず、モバイルユーザの情報要求を的確に反映した検索質問を効果的・効率的に生成できるように支援する必要がある。また、既存のモバイル地理案内システムでも、実空間でのランドマークを容易に把握でき、道に迷ったり不安になったりしないようなナビゲーションは依然として実現できておらず、ランドマークの名称だけでなく外観情報で補完することが必要である。このような社会的な要求に対して、本論文における研究では、Webから概念階層や外観情報を自動的に抽出し活用することで、既存の位置依存情報システムを改善するというアプローチを提案している。

本論文では、まず、Webからの知識抽出手法として、Webから概念階層を自動的に抽出する手法、および、Webから外観情報を自動的に抽出する手法について研究を行い、その抽出精度の評価を行っている。次に、これらWebから抽出した知識の位置依存情報システムへの適用として、モバイルWeb検索のための位置依存検索質問修正システムへの適用、および、モバイル地理案内システムの外観情報による補完について研究を行い、その有効性の検証を行っている。

本論文は6章から構成され、各章の要旨は以下の通りである。

第1章は序論であり、本論文の研究の背景、本論文の研究を行うに至った動機、および、本論文の研究の全体の概要について述べている。

第2章では、Webから概念階層（上位下位関係および部分全体関係）を抽出する手法を提案している。従来手法の多くは上位下位関係の十分条件に基づくが、提案手法は上位下位関係の必要条件である「性質継承」に基づくことで、抽出精度、特に再現率の改善を図っている。さらに、下位語群から上位語への「性質集約」により、上位語の性質（部分全体関係）の抽出を修正する手法も提案している。これは、上位語から下位語への性質継承と下位語集合から上位語への性質集約を繰り返すことで、再帰的に上位下位関係および部分全体関係をともに修正して行く手法である。

氏名	服部 峻
----	------

第3章では、Webから外観情報を抽出する手法を提案している。外観記述の出現パターンを分析し、外観修飾語と名詞（対象物名、上位語、あるいは、部分語を想定）との対を用いてモデル化し、外観修飾語と名詞の連続パターンを含む文書数に基づく単純手法、対象物名に対する名詞の相応度を加重した手法、複合語ペアを除去するために構文パターンを用いた手法、概念階層知識を活用した手法の4種類を提案し、比較実験によりこれらの手法の評価を行っている。

第4章では、モバイルWeb検索のために位置依存型の検索質問修正手法を提案している。Webから抽出した概念階層およびWebから抽出した外観情報を適用することにより、既存システムの問題点の改善を図っている。モバイルWeb検索における情報アクセス過程を、対象オブジェクトの（名称の）認識、検索質問の主トピック語の入力、検索質問の副トピック語追加による絞り込み、関連する代替質問への置き換えの4つのステップに分け、各ステップに対して、上位語・外観・位置に基づくオブジェクト名検索、位置に基づく検索語確定支援、位置に基づく質問拡張語の発見、位置に基づく代替質問（同位語）の発見手法を提案し、その有用性を示している。

第5章では、モバイル地理案内システムにおいて、Webから抽出した外観情報を適用することによって、既存システムの問題点の改善を図っている。道案内システムにおいて、次の目印となるランドマークが名称や方向だけで指示されるとユーザは把握し難いという問題に対して、Webから抽出した外観情報を追加することでランドマークの認識性を改善している。また、移動中に画像情報などを見るのが危険であるような場合に対処するために、検索された画像を、Webから抽出した外観記述テキスト情報で置き換え可聴化するシステムを提案している。

最後に、第6章では、本研究で得られた研究成果全体としての学術的意義および社会的意義についてまとめ、今後の研究課題や展開について述べている。

(論文審査の結果の要旨)

本論文は、Webから概念階層および実世界の地物等の外観情報を自動的に抽出し、これら抽出された知識を活用することで既存のモバイルWeb検索や地理案内といった位置依存情報システムを改善する手法について検討したものである。

モバイルWeb検索は、ユーザ個々人の能力に依らず、精度良く素早く検索可能にする必要がある。本論文ではこのために、Webから概念階層や外観情報を抽出し、これらを活用した位置依存検索質問修正手法を提案している。また、効果的・効率的にモバイル地理案内を実現するために、本論文では、Webから抽出した外観情報を活用する手法を提案している。

具体的には、本論文の研究によって得られた成果は以下のように要約される。

1. 概念階層における上位下位関係の必要条件として性質継承に着目し、ある語が与えられた場合にその下位語を網羅的にWebから抽出する手法について提案した。また、上位語からの性質継承による下位語の重み付けと、下位語群からの性質集約による上位語の性質（部分全体関係）の修正とを繰り返すことで、概念階層抽出を再帰的に改善して行く手法についても提案し、評価実験によってその有効性を示した。
2. Webから実空間のオブジェクトの外観情報を精度良く抽出するため、外観記述を外観修飾語と名詞のペアでモデル化して集約するというアプローチのもとで、4種類の抽出手法を提案し、比較実験を行った。その結果、概念階層知識を活用して名詞部を対象物名、その上位語や部分語に限定する抽出手法が最良であり、上位10件適合率で約0.7という十分に実用性のある高い精度を実現した。
3. 名称不明なオブジェクトに関する情報も精度良く検索できるように、位置だけでなく上位語や外観記述に基づくオブジェクト名サーチ手法を提案している。評価実験により、同一のオブジェクトに対する記述であっても多様になり易い外観に基づいてオブジェクト名をランキングする場合、Web、図鑑、Wikipedia、Flickrタグという4種類の文書コーパスの中で、より多様なWebを用いて共起計算するのが最良であるという知見も明らかにした。
4. 外観情報が無く、名称や方向だけから成る道案内（文）を、Web抽出した外観情報で補完する手法、および、文書中の画像をWeb抽出した外観記述テキスト情報で置換して可聴化する手法を提案し、その有効性を議論した。

以上、本論文は、Webから抽出した知識を活用してモバイルWeb検索や地理案内といった位置依存情報システムを改善するという新たなフレームワークを提案し、その中で概念階層や外観情報をWeb抽出する新たな手法、モバイル検索質問修正を支援する新たな手法、視覚的アクセスだけを想定して作成されたマルチメディア文書の可聴化手法を提案しており、学術上、實際上寄与する所が少なくない。よって、本論文は、博士（情報学）の学位論文として価値あるものと認める。また、平成21年2月9日に論文内容とそれに関連した事項について試問を行った結果、合格と認めた。