

論文の内容の要旨

論文題目 ジオワード・マイニングを用いたローカルサーチの研究
氏 名 高橋克巳

本論文は「ジオワード・マイニングを用いたローカルサーチの研究」と題し、コンテンツ中のジオワード(地理位置情報)を用いてウェブコンテンツを地理的に検索するローカルサーチシステムを構成する方法を明らかにし、プロトタイプ検索システムでインターネット実証実験を行いその有効性を示すとともに、アクセスログから地理的なルールを発見するジオワード・マイニング法を提案した。

「ジオワードを中心とした情報の統合」に関し、様々なコンテンツに緯度経度情報を持たせるための方法を提案した。イエローページデータに対し、ジオワードを用いてデジタル地図データを統合し緯度経度情報を持たせる手法、さらにウェブコンテンツに対し、イエローページデータを統合し同様に緯度経度情報を持たせる手法を述べている。前者は住所属性を持つコンテンツの検索結果を地図に表示する仕組みとして実システムで長く利用され有効性が実証された。また、後者は次章で述べる多様なローカルサーチ実現の基盤技術をなすものと言える。

「ジオワードを用いたウェブローカルサーチの実現」に関し、前章の提案手法の有効性を明らかにすべく、ウェブ文書から抽出したジオワードを介し文書に緯度経度情報を付与することにより空間索引付けを可能とし、更にウェブ文書に対する空間問合せの実現手法について詳述した。実ウェブ文書データを用いた評価実験により、地理的検索条件の領域と住所が示す領域の差異により従来の住所方式では25%程度の検索もれが発生する場合があったのに対し、提案方式では当該問題を解消出来ることを明らかにした。実証システムは1998年6月から2003年までインターネット上において実験サービスとして公開され、世界に先駆け、いち早く大規模なローカルサーチの利便性を体感できる場を提供した。

「アクセスログから地理特性を提示するためのジオワード・マイニング」に関し、アクセスログ内のジオワードに注目した新しいマイニング手法を提案した。即ち、ジオワードの地理的關係を用いることにより、空間的相関ルールをマイニングする方法を、地域に関連する単語を推薦することを目的に考察した。検索回数が少ない地域に対し、ジオワード間の地理的距離を用いジオワードのクラスタリングを行いサポート値の調整を可能とする一般化空間相関ルールのマイニング手法を提案した。9000万件の実検索ログを用いた実験により、提案手法は従来の単純な住所階層を用いる手法に比べ、多様な地域でより多くの興味深いルールの抽出が可能であることを明らかにした。

「ジオワード・マイニングのための固有名詞解析手法」に関し、固有名詞を処理する際に生じる曖昧性の解消法に関して、住所の省略および固有名詞の表記のゆれの問題について論じている。省略のある住所文字列の正式住所名への変換は、異地名住所の場合は困難であるという問題があった。アクセスログを詳細に解析し、頻出ジオワード10000の3%に達する異地名住所問題を、セッション情報を用いた上位住所推定手法により解消可能となることを明らかにした。更に、固有名詞の表記ゆれに関しては、ゆれの同値規則から自動的に正規化規則を作成する方法を考案し、日本人姓のかな表記で分析した結果、約9万通りのゆれの単位に分類可能であること、加えて、完全一致検索時に1検索あたり15%存在していた検索もれを、93%という高い適合率を達成しつつ解消できることを示した。