

審査の結果の要旨

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ており、7章から構成される。

第1章は序論であり、本研究の背景および目的について概観し、本論文の構成を述べている。

第2章は「関連研究」と題し、ローカルサーチ実現技術の問題点を指摘するとともに、データマイニングにおける空間マイニングとログマイニングの研究を纏めている。

第3章は「ジオワードを中心とした情報の統合」と題し、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ている。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。

第4章は「ジオワードを用いたウェブローカルサーチの実現」と題し、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ている。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。

第5章は「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」と題し、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ている。即ち、ジオワードの地理的関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ている。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ている。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ている。

第6章は「ジオワード・マイニングのための固有名詞解析手法」と題し、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、異地同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%に達する異地同名住所問題を、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ている。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ている。

第7章は結論であり、本研究の成果と今後の課題について総括している。

以上、これを要するに、本論文は、ジオワードを用いてウェブコンテンツを地理的に検索する方法、および空間領域を適的に調整することにより興味深いルールの抽出を可能とする新しいジオワード・マイニング方法を提案すると共に、その有用性を多様な実験により実証しており、情報理工学上貢献するところが少なくない。

よって本論文は博士（情報理工学）の学位請求論文として合格と認められる。