審査の結果の要旨

氏名　　森　純一郎

本論文は「Entity Information Extraction from the Web Using Search Engine: Methodology and Application（検索エンジンを利用したWebからのエンティティ情報抽出手法と応用に関する研究）」と題し，英文で記されており，９章から成る．

第１章「Introduction（序論）」では，WWW（Web）が社会の重要な情報インフラになってきており，膨大なWeb情報を集合知によるコーパスと見なせ，また最近のBlogやSNS（social networking service）に見られるように，Webが実社会の状況を反映するメディアになってきているという背景を述べている．その Web からの情報抽出，特にエンティティ（具体的には人物や組織）に関連する情報抽出が，Web 情報の新しい活用法に向けて価値あるものであることを述べている．

第２章「Background and Related Work（背景と関連研究）」では，Webからの情報抽出と Web マイニングの関連研究，コンピュータが Web コンテンツの意味を把握できる次世代Webに向けてのSemantic Webに関する関連研究，Web関係の社会ネットワークにおける関連研究について纏めている．そして，既存研究に対する本研究の特徴を述べている．

第３章「Modeling Entity Information From Web（Webからのエンティティのモデル化）」では，Web上のエンティティを表現するための基本となるモデルと，検索エンジンを利用してそのモデルを構築するための方法を述べている．

第４章「Entity Information Extraction from Web（Webからのエンティティ情報抽出）」では，人物に関するキーワードを Web から抽出する手法を示している．この機能はSemantic Webのメタデータ作成の観点からも重要となる．この手法の核となっているのは，人物名と語の共起の統計情報を用いることであり，人物としては主に研究者を対象にした場合について具体的に提示している．例えば，検索エンジンのAND検索で”Alfred Kobsa AND User Modeling”のヒット数が 3100 で，”Alfred Kobsa AND Software Engineering”のヒット数が450であれば，この研究者は”User Interface”により関係があると判断できる．複合語を含むキーワード候補の切り出しにはターム抽出ツールを用いる．人物名との共起に基づくスコアリングの尺度としては，共起の割合を表す Jaccard 係数を用いている．ある人物が複数のコンテクストでのWebに出現する場合，例えば研究者でありかつ芸術家としても活動している場合の問題への対処法も提示している．キーワード抽出法として広く用いられるTFIDF（term frequency ＊ inverse document frequency）による方法と比較し，本手法によれば優れた結果が得られることを実験による数値で示している．

第５章「Entities Relation Extraction from Web（Webからのエンティティ間関係の抽出）」では，Web上のエンティティの内で特に人物を中心とする事柄について，関係の種別を抽出する手法を示している．この手法では基本的に各エンティティ対が現れるコンテクストを単語ベクトル（bag-of-words）で表し，このコンテクスト集合をボトムアップにクラスタリングし，得られた各クラスタについてエンティティ間関係の種別を表す共通的な語彙を記述ラベルとして抽出する．例えば，“小泉純一郎，日本”や“森喜朗，日本”などを含むクラスタから，“首相”が関係を表す記述ラベルとして抽出されることになる．これは教師なし学習法によるものであることから，事例データを必要としない利点を有する．政治家と地理的エンティティ間の関係，人工知能分野の研究者間の関係（会議論文やジャーナル論文の共著者，本の共著者，本の共編者，同一研究プロジェクトの研究者，同一所属機関の研究者など）について実験を行い，コンテクストの範囲は30用語程が適当であることを見い出し，

本提案手法により良好な結果が得られることを示している．

第 6 章は，「Social Network Extraction from Web（Web からの社会ネットワーク抽出)」で，具体例として人工知能学会大会の論文著者，参加者を対象とする人間関係ネットワーク生成について記している．人物をノード，関係をアークとしてネットワークを構成するのだが，関係の有無は Web での両人物名の共起の割合を用い，Web に現れる回数が少ない場合は検索エンジンで両人物名が共起する上位 10 ページを精査して関係の有無を判定する．関係の種別の付与は第 5 章の手法により，また人物には第 4 章の手法によるキーワードを付与してその属性が判るようにする．このネットワークに Web の PageRank アルゴリズムのように隣接ノードに権威者度の伝播を繰り返すアルゴリズムを適用し，権威者度の高い人物を見い出せることを示している．このような人間関係ネットワークシステム Polyphonet を共同研究者と共に構築し，実際に人工知能学会大会（2003, 2004, 2005, 2006, 2007）と国際会議 Ubicomp2005 で，参加者に関連情報を伝える支援システムとして運用している．特にその研究者検索機能では，ある研究者が行っている研究トピックを検索できたり，ある研究者から他の研究者への知人関係経路を提示できたりする．

第 7 章「Information Sharing using Social Networks（社会ネットワークを用いる情報共有)」では，社会ネットワークを用いることによる情報公開・共有の範囲を適切に制御する方法を提案している．この方法では，ユーザは自身に関連する社会ネットワークを Web，電子メール等から抽出し，それを編集することによって自身の情報種別毎に公開の範囲を指定する．また，ネットワーク分析の手法により各人物の中心性等の指標を提示し，情報公開の範囲に入る中心性が高い人物などを分けるようにする．具体例として，研究者に関するネットワークについてこの機能を実現している．

第 8 章「Expert Finding using Social Networks（社会ネットワークを用いるエキスパート発見)」では，料理法についてのオンラインコミュニティを具体例対象にして，各人物のプロファイルや料理法情報と共に社会ネットワークを形成し，特定の条件下で適切な料理法を持ち，かつ関係の近い人物を見い出す方法を提案している．単にその料理法に詳しいと言う観点だけでなく，社会ネットワークにより関係の近い人物を探すことにより，必要に応じて尋ねることも可能になる．この方法の実験システムを作成し，限定的ではあるものの，問題点や評価を示している．

第 9 章は「Conclusion（結論)」であり，本論文の成果を纏めている．

以上を要するに，本論文は Web 情報活用の新側面を拓くことに向けて，検索エンジン機能を利用して，Web 上のエンティティ情報，具体的には人物に関するキーワードを抽出する手法，Web 上のエンティティ（具体的に扱っているのは人物）間の関係を種別も含めて抽出する手法を考案し，人間関係ネットワークを構成する方法を提示している．これらの手法を含む具体的システムとして人工知能学会大会の論文著者，参加者を対象とする人間関係ネットワークを構築し，実際に運用することによる実証的研究により，その実現性と効用を示している．また，このような人間関係ネットワーク（社会ネットワーク）を用いることにより，あるコミュニティで権威度が高い人物を見い出す機能，情報公開の適切な範囲を指定できる機能，ある事柄について詳しくかつ関係の近い人物を見い出す機能を実現できることを示し，これらの機能も具体的に実現し評価している．これらの研究成果により，本論文は電子情報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）の学位論文として合格と認められる．