

氏 名	野村 早恵子
学位の種類	博士 (情報学)
学位記番号	情博第91号
学位授与の日付	平成16年1月23日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	情報学研究科社会情報学専攻
学位論文題目	Analyses of Academic Communities Using Information on the Web (Web情報を用いた研究コミュニティの分析)

論文調査委員 (主査) 教授 石田 亨 教授 上林 彌彦 教授 片井 修

論 文 内 容 の 要 旨

本論文は、様々なコミュニティにおけるコミュニケーションの中核メディアである World Wide Web が、どの程度現実社会を反映しているのかを定量的に明らかにするため、「計算機科学分野の研究コミュニティ」を分析対象とし、そのコミュニティの知的構造や研究者間の社会的関係性のあり方について分析したものである。論文は、6つの章により構成されている。

第1章は序章であり、本研究の目的とアプローチ、また研究の背景、そして本論文の構成を述べている。研究の目的は、いかに Web が研究コミュニティの知的構造や研究者同士の社会認知的紐帯などを反映した構造になっているかを定量的に明らかにすることである。研究のアプローチとして、研究者の個人 Web サイト上で発信されている情報やサイト間のリンク関係を分析することが述べられている。また、研究の背景として、Web を代表とする CMC 技術 (computer-mediated communication technology) の発展による研究者間の情報共有のあり方や研究活動のあり方の変化に関し、先行研究のレビューを行っている。

第2章では、本研究で用いる分析手法について述べている。本研究では、文書や書誌などを解析するための既存の分析手法を Web に応用し、Web 上での研究コミュニティを分析している。本研究で用いた分析手法は、Web 内容解析と Web 引用解析である。前者は、Web ページ上に掲載される情報を分類し、各分類における情報の量や特徴を定量的・定性的に分析する手法である。後者の Web 引用解析は、書誌引用解析を Web に応用したものである。本研究では Web 著者共引用解析および Web 直接引用解析手法を用いている。Web 著者共引用解析は、個人 Web サイト間の引用の共起関係を用い、研究コミュニティの知的構造を抽出する手法である。一方、Web 直接引用解析手法は、個人 Web サイト間における直接的な引用のあり方を分析し、研究者間の社会認知的紐帯、および研究コミュニティにおけるオーソリティを検出する手法である。

第3章では、Web 内容解析を用い、研究者が個人 Web サイト上にどのようにアイデンティティを表現しているか、つまり個人 Web サイトを自己アピールのためにどれほど有効活用しているかを分析している。Web がどの程度、実社会を反映するかを明らかにするため、個人 Web サイトは、日本、米国、ドイツ、韓国の4カ国から集められている。4カ国200大学に所属する教授、助教授、講師、助手、博士課程の学生、1,700名の Web サイトが調査されている。例えば、米国は研究成果を、日本は個人情報、また韓国は学歴を他国に比べてより詳細に発信する傾向が明らかにされている。このように、Web 上に表現されているアイデンティティには、国により特徴があることが示されている。実社会での自己アピールの方法が各国で異なることが、Web にも反映されることが示されている。

第4章では、研究コミュニティにおける知的構造 (研究分野を表すクラスター) を抽出するため、Web 著者共引用解析を用いられている。分析対象は、CiteSeer から入手した8,811本の論文の著者、3,878名の個人 Web サイトである。Web 著者共引用解析により抽出される研究コミュニティの知的構造がどこまで正確なものであるかを確認するため、8,111本の論文に対する文献引用解析結果と比較されている。すなわち、Web/文献の両方で被引用数の高い上位200人の研究者を対象と

して、Web著者共引用解析およびWeb相互引用解析が行われた。その結果、Web著者共引用解析により得られるクラスタが、文献著者共引用解析により得られるクラスタのスーパーセットとなるという強い傾向を得られている。さらに、Webサイト間の相互引用性から、研究者同士の社会認知的紐帯を抽出できることが示されている。

第5章では、Web直接引用解析のためのHITSアルゴリズムを研究コミュニティ抽出のために利用している。HITSアルゴリズムの持つトピックドリフト問題を修正するために、1) オーソリティ尺度を算出する固有値計算時のフィルタリング法、2) オーソリティ抽出のもととなるページ集合生成時のフィルタリング法を開発・適用し、これらが研究コミュニティの抽出に有効に働くことを確認した。さらに、アルゴリズムを3,878名の研究者Webサイトに適用し、固有ベクトル毎に抽出された研究者が各研究分野において著名な研究者であることを確認した。HITSアルゴリズムでオーソリティとして抽出される研究者には、論文ではもはや引用されない著名な理論やプログラム言語を開発した研究者、またどの大学でも利用するような著名な教科書の執筆者が含まれることが明らかにされている。

第6章は結論であり、本文で得られた成果について要約している。また、今後の研究の方向性を示唆している。

論文審査の結果の要旨

本論文は、様々なコミュニティにおけるコミュニケーションの中核メディアとなっている World Wide Web が、どの程度現実社会の様相を反映しているのかを定量的に明らかにするため、「計算機科学分野の研究コミュニティ」を分析対象とし、そのコミュニティのあり方について論じたものである。得られた主な成果は次の通りである。

(1) これまで定量的に分析されることがなかった、研究者の個人Webサイト上に表現されるアイデンティティを、2年間にわたり調査・定量分析した。調査されたデータは、日本、米国、ドイツ、韓国の200大学に所属する合計1,700名の計算機科学研究者のWebサイト（1万ページ以上）である。結果として、研究者がWeb上で行う自己アピールの方法は、国や慣習の違いを如実に反映していることを明らかにした。さらに、研究者の個人Webサイト情報を用いると、研究コミュニティの様々な構造を分析できることも明らかにした。

(2) 書誌著者共引用解析を3,878名の研究者Webサイトおよび同研究者による8,811本の論文引用関係に適用し、両者を比較した。この目的は、研究者の個人Web間の共引用関係が、論文間の著者共引用関係のように、研究コミュニティの知的構造（研究分野を表すクラスタ）を表しているかどうかを確認することである。結果として、Web著者共引用解析結果は、文献のそれよりは多少粗いものの、研究コミュニティの正確な知的構造を抽出することを確認した。さらに、Webサイト間の引用関係から、研究者同士の社会認知的紐帯を抽出する手法として、引用の相互性が有効に働くことを明らかにした。

(3) 書誌直接引用解析の応用であるHITSアルゴリズムの問題点を分析し、Webのリンク情報のみを用いて修正を行った。問題点を明確にするために、アルゴリズムの可視化ツールを開発した。本研究で行ったHITSアルゴリズムの修正は、オーソリティ尺度を算出する際の固有値計算手法、およびオーソリティ抽出の起点となるページ集合フィルタリング手法である。いずれの手法もWebページの内容を一切考慮することなく、ページ群のリンク構造のみを抽出する方法であり、特に後者は計算量削減にも役立っている。結果として、HITSアルゴリズムの問題といわれているトピックドリフト問題を修正した。また、研究者の個人Webサイト間にHITSアルゴリズムを適用し、各研究分野で著名な研究者とその特徴を明らかにした。

以上、本論文ではWeb情報を用いた研究コミュニティの分析手法を提案すると共に、実際に大量の情報を分析し、Webサイト間のリンク解析結果を論文間の引用解析結果と比較するなど、一般性の高い結果を導いている。社会科学と計算機科学との境界領域研究として、学術上、実際上寄与するところが少なくない。

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15年11月19日に実施した論文内容とそれに関連した試問の結果合格と認めた。