

氏 名 (本籍)	久 保 順 子 (千 葉 県)
学 位 の 種 類	博 士 (情 報 学)
学 位 記 番 号	博 甲 第 5559 号
学位授与年月日	平成 22 年 9 月 30 日
学位授与の要件	学位規則第 4 条第 1 項該当
審 査 研 究 科	図書館情報メディア研究科
学 位 論 文 題 目	情報資源アクセスを支援するための語彙の構築手法に関する研究

主	査	筑波大学教授	杉 本 重 雄
副	査	筑波大学教授	石 塚 英 弘
副	査	筑波大学教授	川原崎 雅 敏
副	査	筑波大学准教授	阪 口 哲 男
副	査	慶応義塾大学教授	岸 田 和 明

論 文 の 内 容 の 要 旨

著者は、情報資源アクセス支援のための語彙構築の観点から、我が国の公共図書館が Web 上に発信するデジタル情報資源に関する研究と、学術情報資源からの専門用語自動抽出に関する研究を進め、それを本学位論文としてまとめたものである。本論文は、全 6 章から構成されており、以下、本論文の各章について示す。

「第 1 章 序論」では、本論文で述べる研究の位置づけとして、インターネットの登場による情報環境の変化と新しい情報環境の中での情報資源へのアクセス支援への要求について述べ、また、本論文を構成する主たる二つの研究内容に関して概略を述べている。

「第 2 章 情報資源アクセスを支援する語彙に関する基礎概念」では、本論文における研究の基盤であるメタデータに関して、メタデータとメタデータスキーマに関する基礎概念、メタデータを構成する語彙、特に主題等による索引を構成するために用いられる語彙（索引語彙）に関して述べている。また、第 2 章では、第 3 章と第 4 章で述べる研究内容に関する基礎概念を示すとともに、こうした概念の本論文における位置づけならびに関連研究を示している。

「第 3 章 公共図書館の情報資源の特徴分析と主題語彙の構築」は、2003 年度から 2005 年度にかけて著者自身が進めた我が国公共図書館が Web 上で発信する情報資源に関する調査研究をまとめたものである。著者は、我が国の公共図書館の Web サイトを網羅的に調査し、事前に設定した基準に基づき情報資源を選択的に収集した。さらに、著者は収集した情報資源の内容分析に基づいて、情報資源の内容を簡易なメタデータとしてまとめ、作成したメタデータを分析して主題を表す語彙を構築した。

著者は、本研究において、できるだけ網羅的な情報資源の収集を行うことで、我が国の公共図書館からどのような情報発信がなされているかを総合的に見ようと試みている。公共図書館からの情報発信には、大規模なデータベース開発のケースもあれば、1 点あるいは数点だけの情報資源の提供のケースもある。本研究では、公共図書館が Web 上で提供する情報資源を、Web サイトの規模の大小にかかわらず収集している。著者は、情報資源に関する事前の調査によってあらかじめ定めた基準に従って情報資源を収集し、最終的に、

国内 121 館の公共図書館の Web サイトから約 5,000 件の情報資源を収集し、それら情報資源毎のメタデータ（リソースメタデータ）を作成している。さらに、リソースメタデータの分析結果を基に、Web サイトごとの情報資源の集まり（コレクション）を対象とする 194 件のメタデータ（コレクションメタデータ）を作成した。

著者は、コレクションメタデータを基礎として、主題、時代、地域、および情報資源タイプに関する 4 つの統制語彙を作成している。これらの語彙を用いたコレクションに関する調査から、自然科学、社会科学などの学問分野の情報資源が少ないこと、地域の歴史や文化に関係した情報資源が多くかつそれらには江戸時代以降のものが大半を占めること、さらに原資料を完全に電子化した情報資源よりも、原資料の一部を電子化し説明を加えて作ったものが多いといったことを報告している。

「第 4 章 学術文献からの専門用語自動抽出による主題語彙の作成」では、専門用語の自動抽出手法に関する研究について述べている。専門用語の自動抽出は、効率的に専門用語を同定し、それを基にした辞書やシソーラス等の作成により、情報資源アクセスを支援することのできる重要な研究課題である。こうした視点に基づき、本研究では、従来の専門用語自動抽出手法に関する課題の分析を行った上で、ある特定の専門領域を対象とする主要な学術雑誌の論文からコーパスと、当該領域外のコーパスにおける用語の出現率の差を利用した新しい自動抽出手法を提案している。

著者は、提案手法を女性学分野における英文の学術論文を対象として適用し、既存の 3 つの専門用語抽出手法と比較することで、提案手法の有用性を示している。また、本提案手法では、対象専門分野のコーパスに加えて、対象分野以外の 39 分野の文献から作られたコーパスを利用して、提案手法の有用性を示した。さらに、対象分野外コーパスを作るための対象分野数を減らすこと、そして、どのような対象分野を選択すればよいかに関する研究を行い、その結果、対象分野に隣接する 3 領域のコーパスを利用することにより比較対象とした従来手法より優れた結果を出せることを報告している。

「第 5 章 考察」においては、はじめに、第 3 章と第 4 章で述べた二つの研究を通じて得た知見について、語彙構築のコスト、情報資源の幅、情報資源の記述の 3 つの視点から考察している。また、語彙構築に関する課題として、対象領域の多様化、語彙の維持管理、フォークソノミー、そしてオントロジーの観点から述べている。最後に、二つの研究に関する総合的な考察を示している。全体として、実際に公開されている情報資源を用い、その情報資源の特徴に合わせたアプローチをとることで、新しい知見を示すことができたことを述べている。

「第 6 章 結論」では、全体のまとめおよび感想を述べている。

審 査 の 結 果 の 要 旨

著者は、情報資源アクセスを支援するための語彙構築の視点から、我が国の公共図書館が Web 上に発信するデジタル情報資源の特徴分析と主題語彙の構築に関する研究と、学術文献からの専門用語自動抽出による主題語彙の作成に関する研究を行い、それらを本学位論文として総合的にまとめている。本審査委員会では、本論文の中心となる二つの研究の内容、ならびに博士論文としてまとめるに際しての著者の論点等を中心に博士論文の審査を行った。以下では、(1) 公共図書館の情報資源の特徴分析と主題語彙の構築に関する研究、(2) 学術文献からの専門用語自動抽出による主題語彙の作成に関する研究、(3) 論文全体の構成に関する評価を示す。

(1) 公共図書館の情報資源の特徴分析と主題語彙の構築に関する研究に関して

本研究は、予備的な調査を含めると 2003 年度から 3 年をかけて著者が中心となって行ったものである。著者は、我が国の公共図書館が Web サイトを網羅的に調査し、そこから基準に合わせて選択的に情報資源

を収集している。この調査において、著者は、121 館の Web サイトから約 5000 件の情報資源を収集し、さらに、収集した情報資源のメタデータを作成し、それを用いて内容分析およびそれに基づく主題語彙構築を行った。

調査の結果から、公共図書館が Web 上で発信する情報資源の組織化に関しては、原則としてコレクションを単位として行うこと、そしてコレクションのパターンとして、電子化した資料の提供（電子化資料）、電子化していない資料に関する情報の提供（非電子化資料）、部分的に電子化した資料を利用した資料情報の提供（半電子化資料）に分けることを提案している。さらに、こうした組織化の方針の下に、コレクションを内容によって分類することを試み、主題、時代、地域および情報資源タイプによる分類のための語彙を作成した。この分類を用いて、我が国の公共図書館から発信される情報資源に関する分析を行った。

この研究は、情報資源収集と分析、組織化によって情報資源の内容を表すために用いることのできる語彙構築を手で行ったものである。蓄積したメタデータの品質、構築された語彙の評価に関する論文の記述が物足りないところが一部あるものの、事前の調査でメタデータ記述に関する検討も加えており、また語彙構築に際しては、類義語辞書等の適切な補助資料を用いてシステムティックに作業を行うこと、そして共同作業者との合意を基本とすることといったガイドラインを設けることによって主観性をできるだけ排除しており、構築された語彙の信頼性については評価できる。本研究での調査は、手で行ったものとしては大規模であり、加えてこの研究が行われた時点においては同規模の調査やメタデータ蓄積は見受けられない。なお、調査終了時(2005 年度)における日本図書館協会の公共図書館の Web サイトに関する提供データと比較して、デジタル情報資源を提供する我が国公共図書館 Web サイトの半数以上をカバーしていることがわかっていいる。以上のことから、質的、量的に信頼できる調査研究であると理解できる。

(2) 学術文献からの専門用語自動抽出による主題語彙の作成に関する研究に関して

本研究は、MDP (Minimum of the Difference between Population Proportions) と呼ぶコーパスからの用語抽出手法を提案し、それを女性学領域に適用して手法の有効性を確かめたものである。専門用語の自動抽出は、学術情報資源、特に新しい領域の学術情報資源へのアクセス性を高める上で重要な技術である。そのため、従来から専門用語の自動抽出に関する研究は多く行われてきている。従来の手法では主として対象領域のテキストコーパスのみが利用されてきたのに対し、本研究では、対象領域のコーパスと対象領域以外のコーパスを利用し、それらに現れる用語の出現率の差を考慮して専門用語語彙を自動生成する新奇性のある手法を提案している。また、対象領域以外の領域のコーパスとして 39 領域のコーパスを用いた実験を行い、従来の手法に対してより優れた結果を得られることを示した。一方、本研究で提案した手法では、対象領域とそれ以外の領域のコーパスを必要とすることになるため、対象領域以外の領域のコーパスを最小化することが求められる。こうした課題に対して、本研究では先の研究を発展させて、1～3 領域のコーパスを利用した領域外コーパスを利用する方法について実験を行い、対象領域に近い領域のコーパスを用いることにより、比較対象とした従来手法より優れた結果を出せることを示した。

本論文では、有用性の評価手法に関する記述にやや甘さがあるとの指摘もあったが、有用な専門用語自動抽出手法を作り出したことは十分評価できる。新しい学問領域がしばしば作り出される現代において、専門用語の自動的な抽出技術は新しい学問領域の情報資源の索引付けや分類に大いに役立つ。本論文が述べる専門用語抽出技術はこうした情報資源アクセス支援にとって有用な技術であると大いに評価することができる。

(3) 論文全体に関して

論文全体として、上記の二つの研究を通じた基盤となる知識や関連研究、二つの研究とそこから得られた知見、さらに全体を総合して得られた知見についてまとめたものである。著者が行った性質の異なるアプローチの研究について、情報アクセス支援のための語彙とその構築手法の観点からまとめたものであり、博士論

文として十分な幅を持つものであるといえる。考察や考察の記述に関する課題も見受けられるが、博士論文としての価値を損なうものではなく、全体として十分な内容と価値をもった博士論文であると結論できる。

よって、著者は博士（情報学）の学位を受けるに十分な資格を有するものと認める。