【395】

氏　名（本籍）	柊（ひいらぎ）　和佑（わすけ）（東京都）
学位の種類	博　士（情報学）
学位記番号	博　乙　第　2545　号
学位授与年月日	平成23年3月25日
学位授与の要件	学位規則第4条第2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	**組織内での運用を指向したWebアーカイブの研究**

主　査	筑波大学教授	工学博士	杉本重雄
副　査	筑波大学教授	理学博士	石塚英弘
副　査	筑波大学教授	工学博士	海老原義彦
副　査	筑波大学教授	工学博士	川原崎雅敏
副　査	筑波大学准教授	博士（工学）	阪口哲男

論文の内容の要旨

WWWの出現と爆発的な発展の時期から、WWWに発信される情報資源をいかに収集保存するかは図書館にとっての大きな課題であった。WWWが社会的な情報基盤となった現在において、多数の国立図書館においてWWW上の情報資源の収集保存、すなわちWebアーカイビングが行われている。その一方、Webアーカイビングには、深層Webの問題や収集する情報資源の網羅性の問題があることが広く認識されている。こうした状況下において、本論文は、組織で運営するWebアーカイビングと言う視点から研究し、そのための新しい技術を提案し、その実現可能性を実証したものである。

本論文は全8章で構成されている。第1章で、本研究の背景、目的等の概説をした後、第2章において、従来より行われてきた一般的なWebアーカイビングに関して概説し、本論文が述べる組織内Webアーカイビングに関する基礎概念を述べている。一般的なWebアーカイビングについては、主として先進各国の国立図書館を中心に行われている網羅的収集、選択的収集といった収集方針に基づく取組み、Webアーカイビングを進める国際組織であるInternational Internet Preservation Consortium（IIPC）の取組み等を解説している。次に、一般的Webアーカイビングの問題点を述べ、それを解決するための手法として、情報発信者とアーカイブ運営者の連携に基づく組織内Webアーカイビングを提案し、その有効性について論じている。

第3章においては、第2章における提案を基礎として、組織内Webアーカイビングシステムのモデルとその構成方式について述べている。本論文が提案する組織内Webアーカイビングシステムの特徴は、情報資源提供者、あるいはWebサイトの運営者とWebアーカイブ運営者が協調する点である。本論文では、情報資源提供者が提供やクローリング受入れのためのポリシーを定義する一方、他方でWebアーカイブ運営者が情報資源をWebサイトからダウンロードしアーカイブするポリシーを定義することで、両者が協調して情報資源をアーカイブする仕組みを提案している。なお、本論文においては、システムを実現するためのWeb情報資源のアーカイブのためのデータ形式、ポリシー記述のための形式、そしてWebアーカイビングシステムの構成要素の実装について述べている。提案されたシステムにおいて、Web情報資源蓄積のデータ形式は、Resource Set、Resource、Componentという3階層からなっている。クローリングによって収集さ

れた情報資源の管理、アーカイブにおけるアクセス管理等はこの階層モデルをベースにして行われる。また、本研究においては、機関リポジトリ等でよく利用されるDSpaceを利用して提案したシステムの試作を行い、提案システムの実現可能性を実証している。

第4章においては、運用組織の異なるWebアーカイブを併合するための技術について述べている。Webアーカイブには、アーカイブ全体の長期利用性と格納された情報資源の長期保存性の確保が求められる。アーカイブの長期間の利用を考えた場合、Webアーカイブを運営する組織そのものが統合されたり、消滅したりすることを考えなければならない。本論文では、組織の消滅、統合、分割といった場合毎のアーカイブされた情報資源の管理方法と、組織構造の変化に伴うURLと組織資源提供サイトの対応関係の変化への対応方法について、組織内Webアーカイブのモデルとそれを実現するためのシステムを提案し、さらにそのシステムの試験的実装による実現可能性の検証を行っている。

公文書管理において、個人情報を含む資料は、袋とじや墨塗りといったマスキングによる閲覧制限がなされている。著者は、これと同じ機能がWebアーカイブの場合にも必要であるという観点から研究を進め、第5章において、Webアーカイブに蓄積された情報資源の閲覧制限技術について述べている。組織自体の変化や組織のアクセス管理ポリシーの変化、そして経年によるアクセス制限条件の変化などの理由によって、情報資源のアーカイブ時から実際の利用時の間にアクセス制限も変化していかなければならない。著者は、アクセス制限ポリシーを陽に記述することによってこうした問題に対処できるという立場から、本研究の組織内Webアーカイブシステムのためのポリシー記述に基づくアクセス制限のモデルを定義し、それを利用したマスキングによる閲覧制限のシステムの試作し、実現可能性を実証している。

第6章においては、前章までに述べた研究を基礎にして、組織内Webアーカイブに蓄積された情報資源の信頼性、データの整合性といった課題について検討している。ここでは、アーカイブを構成するための情報資源の収集ポリシーと、アーカイブ管理のための管理ポリシーの2点から検討を加えている。

最後に、第7章では、本論文で述べた研究全体を振り返った考察を述べ、第8章において研究全体のまとめを述べている。

審査の結果の要旨

WWWの爆発的な拡大と並行して、図書館界を中心としてWWW上の情報資源を保存することの必要性が論じられてきた。そうした背景の下、Webアーカイビングは先進各国の国立図書館を中心として研究開発が進められてきた領域である。こうした国立図書館中心の取組みは、人間の取捨選択の判断に基づく選択的収集、あるいはWeb情報資源を収集するソフトウェアであるクローラによる自動的、網羅的収集によって進められてきた。一方、Webの広がりの中で、深層Web（Deep Web）と言う名前で呼ばれる、一般的なクローラによる自動収集の不可能な情報資源が、Web情報資源全体の大きな割合を占めることも広く認識されてきた。基本的に、深層Webは、Web情報資源の提供者が、提供資源を外部からは見えなくする、あるいは機械的な収集を拒絶することによって形成されている。こうした深層Webが持つ情報資源の収集には、情報資源提供者がアーカイブ運営者と連携して提供資源を収集保存していくことが不可欠である。本研究はこうした視点から、情報資源提供者とアーカイブ運営者の連携を可能とする環境として組織内で行われるWebアーカイビングの研究を進めたものである。企業や自治体等においてイントラネット内でもWeb環境を構築して情報資源の発信と共有はよく行われる。そうした環境向けのWebアーカイビング技術の研究は有用なものであると考えられる。また、一般的なインターネット上でのWebアーカイビングが主流である中で、本研究はイントラネットを含む組織内環境でのアーカイビングの研究を進めたものであり、全体として独自性、新奇性に富む研究であると言える。

以下に本研究の特色と認められる点をいくつか列挙する。

1. アーカイビングのための提供ポリシーや収集ポリシーを、陽にメタデータ表現することにより、情報資源提供者とアーカイブ運営者の連携を可能にするのみならず、情報資源のアーカイブ後の環境変化に対してもメタデータを利用して対応することができ、情報資源の長期間に渡る管理性ならびに利用性を向上することができる。
2. 組織の消滅や統合などに対応するためのアーカイブ管理モデルを提案することで、アーカイブされたコンテンツが、当初の組織のポリシーに基づいて蓄積保存されるのみならず、組織構造の変化があった場合でも矛盾の起こりにくいアーカイブの構成方式を示している。従来のアーカイブは、組織ごとに個別のサーバを利用して実現されることが多く、組織構造の変化の場合にはサーバごと、あるいはアーカイブごと新組織に移行するということを仮定していたが、本研究は、組織構造の変化を分類した上で、それにあわせた収集、蓄積管理のポリシーを再構築することを提案しているので、組織構造の変化とそれに伴うネットワーク環境の変化に合わせたアーカイブ管理が可能になっている。今後クラウド化が進むにつれて、サーバと言った概念が不要になり、組織毎のアーカイブの実体も見えにくくなってくる。そうした環境化において、本研究が提案するアーカイブ構造の管理方法は有用性を増すであろうと期待できる。
3. 墨塗りなどのマスキング技術による閲覧制御は、公文書アーカイブにとって必須である。現在、電子文書であってもマスキングは人手によらざるを得ない作業であるが、本研究で開発された技術はマスキング作業の自動化に有用であり、将来のアーカイブサービスの効率化に資するところは大きいと考えられる。

本研究では、提案したモデルに関して、システムを実装して、実現可能性を実証しており、提案モデルに対する一定の信頼性を認めることができる。一方、研究室規模での評価を超えるには、今後、実務組織の協力を得てより実務環境に近い環境における評価を行うことが望まれる。また、本研究そのものは2004年頃から2009年ごろまでの時間をかけて行われたものであり、扱っているテーマもそれなりに広がりを持っている。そうした広がりを持つ研究であることは大いに評価することができる。その一方、論文全体を通じてそうした特色とその中における著者自身の経験を反映した考察をもう少し深めることが望まれる。しかしながら、こうした点は、前の課題とも合わせて、本研究をより発展させ、成果を社会に還元するための課題と言え、本論文の価値を損なうものではない。

以上のように、本論文の一部には今後の課題として残されている点も認められるが、本論文は、Webアーカイビングの大枠から、文書毎の閲覧制御といったアーカイブコンテンツに関する技術までを、組織内Webアーカイビングの視点から統一的にまとめたものであり、統合的なシステムとしてアーカイブをとらえている点は十分に評価できる。また、組織内Webアーカイビングに関する研究の先駆けとして、今後のWebアーカイビングの発展のために有用な知見を与えることが期待できる優れた成果を挙げた研究であると評価できる。

論文審査ならびに審議の結果、審査委員全員一致で合格と判定された。

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。