

氏名(国籍)	李 ^り 穎 ^{えい} (中国)
学位の種類	博士(情報学)
学位記番号	博甲第4140号
学位授与年月日	平成18年3月24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	構造化デジタルオブジェクトの概念に基づく学術情報表現およびWeb資源システム構成方式の研究
主査	筑波大学教授 石塚英弘
副査	筑波大学教授 石川徹也
副査	筑波大学教授 杉本重雄
副査	筑波大学教授 田畑孝一
副査	(財)東洋文庫客員研究員 山本毅雄

論文の内容の要旨

現在、大量の学術情報がWebによって得られるようになってきているが、Webにも三つの課題、すなわち、Webにおける情報の表現の問題、インターネット上に個別に存在するWebサーバシステム間の相互連携、検索ノイズが多い全文検索の欠点の改善、が存在すると本論文の著者は同論文の冒頭で指摘している。そして著者は、これらの課題に対する一つの解を提案することを目的として、Web資源の情報表現のために「構造化デジタルオブジェクト」の概念を考案し、その概念に基づいてプロトタイプシステムを開発して、ノイズの少ない検索が可能であることを示し、さらにWeb Servicesの技術とTopic Mapsの技術を加えることによって、構造化デジタルオブジェクトに基づいたWeb資源のサーバシステムであれば、システム間の相互連携性を高められることをプロトタイプシステムにより示した。

本論文は、第1章序論において、まず、Webにおける三つの課題に対する一つの解を提案することを目的として、構造化デジタルオブジェクトの概念を考案し、その概念に基づくプロトタイプシステムを開発すると述べている。次いで、研究の背景として、既存の情報表現のデータモデル、関連システムを概観し、本研究の必要性を主張している。なお、本研究では、システムの利用者として、学術情報を調査・検索し、得られた結果から自分が得たい情報を抽出し、再構成してまとめる仕事を行う者—その典型は科学または工学の研究者—を想定している。これらの研究者は、増大し多様化する学術情報を前にして、特定の情報源に限定されることなく、広く情報を調査し、よりノイズの少ない状態で有用な情報を得て、得られた結果を自らの観点から再構成しようとする。そのため、この種の利用者にとっては、ノイズの多い検索結果を一々読んで必要な部分を取り出すという現状は満足できる状況とは言いがたく、もしも利用者が自分の得たい情報をノイズの少ない状態でピンポイントにWeb資源システムから得ることができ、その情報を再構成・再利用できるようになれば、仕事の効率は大いに向上するだろうと著者は述べている。

第2章では、本研究で考案した構造化デジタルオブジェクト(Structured Digital Object, SDO)の概念、それをXMLで記述するSDO/XMLモデル、そしてそれに基づく学術情報表現のモデルについて述べてい

る。SDO では、情報として意味を持つ最小単位（EDO）を基本の単位とし、この最小基本単位を要素とする順序付き階層構造として情報を表現する。SDO により、様々な粒度の情報に対応できるようになると述べ、また学術情報を SDO/XML によって記述することにより、コンピュータ処理可能になり、その結果、個々の利用者の異なる情報粒度要求を満たし、利用者が望む情報を少ないノイズで検索できるようになると主張している。

第3章では、化学・薬学・医学分野の典型的なファクトデータベースである Merck Index のデータを例として採り上げ、元の Merck Index のデータベースの場合よりも細かい単位で EDO を設定し、SDO のスキーマを設計し、サンプルレコードについて SDO/XML でデータを記述して Web 資源とし、その検索システムのプロトタイプを ASP.NET を用いた Web サーバシステムとして開発したと述べている。また、このプロトタイプによって、2章に述べたモデルが実現可能なことを示し、全文検索に比べてノイズの少ない検索が可能なこと、利用者が必要とするピンポイントの情報提供が可能になることを示した。

第4章では、SDO/XML の電子ジャーナルへの適用可能性を示し、また、SDO/XML に基づくファクトデータベース、電子ジャーナル等を Web 資源として持つデジタル図書館システムを考案し、それを「SDO/XML に基づくデジタル図書館」と名づけた。ただし、このデジタル図書館は、その館で定めた SDO に基づくため、そのままでは他の館の利用者には使いにくいと考察している。

第5章では、Web Services と Topic Maps を補完技術として導入することにより、個々の SDO/XML に基づくデジタル図書館の欠点を改善した相互連携性を持つデジタル図書館が実現しようと述べ、これを「SDO/XML に基づく Global Digital Library」と名づけた。また、Web Services とその関連技術である SOAP を用いたプロトタイプを開発し、SDO/XML に基づく Global Digital Library (DGL) の実現可能性を示した。また、この DGL に Topic Maps を用いたトピックからの検索機能を用意することによって、SDO のスキーマを知らない利用者でも利用可能になることを示した。なお、SDO/XML に基づく GDL の有効な例として、薬品の化学・薬学情報は豊富であるが治療法に関する情報は足りない Merck Index のサイトに、治療法に関する情報が豊富なサイトを組み合わせることを考案し考察している。治療法に関する情報源として、日本医薬品医用機器情報提供機関が提供する情報と、米国の FDA（食品医薬品局）が提供する情報を考え、前者の方が細かく項目が分かれているため SDO/XML 化しやすいと分析している。

第6章 考察では、各専門分野において SDO が設定可能であることを述べ、大規模な SDO 制作の実行可能性については、それを行う専門分野の担当者の人件費が問題になると指摘し、その解決策として、担当者が論文から EDO を抽出する際に自然言語処理技術を用いて支援する方策、多くの専門家が欧米よりも安い人件費で得られる国、たとえば中国で行う方策を述べている。また、コスト回収方法についても言及している。

第7章 結論では、第1章から第6章までを概観し、当初の目的を達したと述べると共に、今後は著者の祖国：中国において本研究成果の実用化を目指したいと述べている。

審査の結果の要旨

現在、利用者は多くの情報を Web によって得ているが、Web にも三つの課題、すなわち、情報の表現の問題、インターネット上に個別に存在する Web サーバシステムの間の相互連携、検索ノイズが多い全文検索の欠点の改善、が存在すると著者は指摘している。この三つがセットで課題になる理由は、単に Web を検索するだけの利用者には分かりにくいかもしれないが、大量の学術情報が存在する分野、たとえば、化学、医学、薬学分野で、文献を調査する必要がある研究者には理解できるであろう。検索ノイズが多ければ、その分、料金を支払って読む文献が増えること、これらファクトを記述する分野では情報が細かな項目に分か

れた構造になっているため、その項目を選択できれば検索ノイズが減らせること、独立して多数存在するサイトを調査するため、相互連携性が高い方が便利なのことがその理由である。

著者は、これらの課題に対する解として、Web 資源の学術情報表現のために「構造化デジタルオブジェクト」(SDO) の概念を考案した。SDO は情報の最小基本単位: EDO を構成要素とする順序付きの階層構造として情報を表現する。この概念は化学、医学、薬学等ファクトを記述する分野の情報の構造に適合していると考えられる。審査委員会は、分野共通の SDO が設計しえるかについて議論し、その結果、その分野の専門家ならば、その分野の情報構造の枠組みを理解しているため、設計可能であると判断した。また、分野を越えた SDO の設計は可能かについて議論し、化学、薬学のような関連分野であれば共通の EDO が存在すること、それぞれの分野の専門家が協議して共通の SDO を設計しうること、分野によって異なる SDO が存在してもマッピングは可能であると考えられることから可能と判断した。なお、審査委員山本と石塚は情報学の研究者になる以前に化学の研究者であった経歴を持っている。また、著者は化学の研究者であった経歴を持つことから、この経験を活かして SDO を考案したと考えられる。

著者は SDO の記述形式として SDO/XML を考案した。XML は順序付き階層構造の表現に適し、コンピュータ処理可能であるため、本研究に適しており、また、XML で記述するため、Web Services の技術が使用可能となり、そのためシステム間の相互連携性を高めることが可能になるなどの効果があると認められた。

本著者が SDO/XML による学術情報の記述を論文で発表した後に、情報検索システムで有名な DIALOG が「検索結果を XML 形式で提供するサービスを開始する」との方針を発表した。これは XML でデータを利用者に提供することによって、利用者によるデータの再利用を支援しようとするものである。一方、本著者は単に XML で表現するだけでなく、SDO の概念に基づいた XML 記述を先に考案し、プロトタイプを開発している。この事実は本著者の先見性を示すものと言えよう。

第 3 章に述べられている Merck Index のデータを例として開発された検索システムのプロトタイプについて、審査委員会は、SDO のスキーマ設計の妥当性、ノイズの少ない検索の実現可能性を検討した。前者については、Merck Index のデータ項目に関する詳しい記述の分析と実際のデータ例の分析に基づいていること、Merck Index よりも細かい EDO を設定したことを評価し、妥当と判断した。後者については、EDO を細かく設定した結果、検索対象項目が限定可能になるため、検索ノイズが減少することが定性的に認められると判断した。

著者は、SDO/XML に基づくデジタル図書館の相互運用性を高めることによって、SDO/XML に基づく Global Digital Library (GDL) が可能だと述べ、相互運用性を高めるための技術として Web Services と Topic Maps を導入している。HTML に基づくデジタルライブラリの場合、ハイパーリンクによって他のデジタルライブラリの Web 資源に跳ぶことによって相互運用性を保つが、得られる結果は人間が読むためのものであって、コンピュータ処理には適していない。一方、SDO/XML に基づくシステムでは得られる結果が SDO/XML であるため、その結果をコンピュータで処理することができる。そこで、異なるシステム間で SDO/XML でデータを交換できれば、相互運用性を高めることができる。著者はプロトタイプを開発し、Web Services を用いることによって、異なるシステム間の相互運用性を高められることを示した。SDO/XML に基づくデジタル図書館同士はもちろん、既存のデジタル図書館であっても、Web Services による SDO/XML に基づく外部への情報提供機能を持てば、SDO/XML に基づく GDL になりえるとの著者の指摘は SDO/XML に基づくシステムの展開に有用と考えられる。また、Topic Maps の導入も補完技術として有効と判断された。

第 5 章の考察では、SDO/XML に基づく GDL の可能性として、薬品の化学・薬学情報は豊富であるが治療法に関する情報は足りない Merck Index のサイトに、治療法に関する情報が豊富なサイトを組み合わせることを考えている。治療法に関する情報源として、日本医薬品医用機器情報提供機関が提供する情報と、米

国の FDA（食品医薬品局）が提供する情報を考え、前者の方が細かく項目が分かれているため SDO/XML 化しやすいと分析している。この考察は、SDO/XML に基づく GDI の実現可能性を示しており、評価できる。

第 6 章 考察では、大規模な SDO 制作の実行可能性については、それを行う専門分野の担当者の人件費が問題になると指摘し、その解決策として、多くの専門家が欧米よりも安い人件費で得られる国、たとえば中国で行う方法を示している点は注目される。中国は PC 製造だけでなく、ソフトウェア開発、デジタルライブラリのための文化資料のデジタル化を大々的に進めていることから、SDO 制作を中国で行うとの案は可能性があるためである。

以上、本研究は当初の目的を達成するとともに、技術的に新規性があり、有用性が高い。また、化学など専門分野の知識を活かして細かく EDO を設定して SDO のスキーマを設計することに始まり、プロトタイプ開発と考察を経て、SDO/XML に基づく GDL という雄大な構想に至るなど概念的にも見るべきものがある。従って学位論文審査委員全員により論文審査合格と判定された。

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。