

審査の結果の要旨

氏名 山本 泰智

本研究は、利用者の興味とは無関係に検索結果を表示していた従来の静的な文献検索システムに対して、利用者の視点を検索結果の文献集合の動的なクラスタリングとクラスに対するラベル付けに反映する文献検索システムを提案し、そのために必要な基礎技術、および、それらを統合した生命科学のための文献検索システムを実際に開発したものである。実ユーザとの緊密な共同研究を行うとともに、利用者の視点を意味空間に反映する独自性の高い基礎技術を開発している。

論文は、以下の7つの章から構成されている。

第1章では、生物・医学の文献情報の急激な増加にともない、知的な文献検索システムが不可欠となっている背景を簡潔に整理し、既存の文献検索システムの欠陥、特に、MeSHターム（領域固有統制語）による絞込みという静的な手法では、潜在的に有用な情報を落とす可能性を指摘し、この問題の対処には、利用者の視点を検索時点で動的に反映できるシステムが不可欠であることを指摘して、本論文の研究背景とその意義を述べている。

第2章では、テキストクラスタリングの従来手法を概観し、利用者との対話によって、クラスタリングを動的に変更する試みがないこと、また、クラスタリング結果を利用者に理解可能な形で提示する手法、すなわち、クラスのラベル付けの技術が十分でないことを指摘し、これらが本研究の目指す動的な情報検索の基本技術となることを指摘している。

第3章では、生命科学における知的情報検索の研究を行う上で、各種の資源が整備されてきていること、とくに、米国 NLM で開発されている PubMed/MEDLINE、文献情報中の語を正規化するための領域固有辞書 UMLS LEXICON、利用者の視点やクラス概念を表現するための領域固有ソーラスである UMLS Semantic Net と MeSH タームが、本研究の主要な資源となっていることを説明し、それぞれの特質とそれらが本研究で果たす役割とを詳述している。

第4章は、本研究での主要課題である利用者の視点を反映したクラスタリングの実現手法を系統的に展開している章である。既存手法、すなわち、ベクトル空間モデルの利用や、群間平均法による階層的クラスタリング、及び並列化 buckshot アルゴリズムによる k-means を利用した非階層的クラスタリング等を基本とし、本研究の独創である利用者の視点を表す MeSH タームを意味空間の変換に活用する手法を定義している。また、この種の対話を前提とした文献検索システムの評価の難しさを議論し、本研究で用いた定量的・定性的評価の方法を詳述している。

第5章は、本研究での要素手法が統合された全体システムの構成、および、第4章で定義した評価手法を適用することで得た評価結果を報告している。文献情報の処理方針（領域

固有辞書の利用の有無等) と処理対象に関する条件を設定し、クラスタリングの効果、クラスタ・ラベルの効用、および、利用者視点がクラスタリング結果に及ぼす影響度合いの 3 項目について、定量的評価を行っている。その結果、1) 領域固有辞書の利用、2) 取得文献情報中の頻出語と低頻度語の削除が文献検索に及ぼす効果を定量化し、第 3 章で導入した言語資源が非常に有効であったことを確認している。

また、定性的評価については、関連研究との比較を行い、生成されるクラスタの特徴を示す概念ラベルが内容理解の容易さの向上に大きく寄与することを確認し、生命科学者とともに具体的事例に基づく評価を行っている。これは、類似研究には見られない大きな成果で、文献情報を複数視点からクラスタリングすること、また、クラスタリングを対話過程で動的に変化させることの有効性が、実際の利用者により確認されたことは、本研究の大きな成果である。

第 6 章は、本研究は、MeSH タームの使用、論文抄録など、さまざまな条件が固定された中でのシステム開発という側面もつ反面、開発された手法は、これらの特殊条件に依存しない一般性を持つものであることを議論している。特に、本研究で中心的な役割を果たした Mesh タームは、任意のオントロジーに置換可能であり、本研究が、文献検索と知識処理との融合の一つの形態を示していることを指摘している。

第 7 章は、本研究の貢献を、1) 文献検索結果の概観取得をクラスタリングで行う統合的な対話型検索システムで実現したこと、2) 利用者の視点をクラスタリングに反映させる手法を開発したこと、3) クラスタに既存オントロジーを用いてラベル付けを行う手法を開発したこと、の 3 点に簡潔に整理し、フルペーパーの処理、文脈解析、クラスタリングの高速化など、今後の研究課題を整理している。

以上のように、本論文は、利用者視点を意味空間の変換で実現する文献クラスタリングの手法、および、既存オントロジーによるクラスのラベルづけ手法という、対話型の文献検索システムのための独創的な手法を提案するとともに、その有効性を統合システムの開発で実証したものであり、審査委員会は、その独創性。有効性は、博士号に十分値するものと判断した。