

Title	AUTOMATIC INFORMATION RETRIEVAL BASED ON GRAPH-THEORETICAL CONCEPTS
Author(s)	伊藤, 哲郎
Citation	
Issue Date	
Text Version	ETD
URL	<a href="http://hdl.handle.net/11094/943">http://hdl.handle.net/11094/943</a>
DOI	
rights	
Note	

*Osaka University Knowledge Archive : OUKA*

<https://ir.library.osaka-u.ac.jp/>

Osaka University

氏名・(本籍)	伊 藤 哲 郎
学位の種類	工 学 博 士
学位記番号	第 4 3 4 3 号
学位授与の日付	昭 和 53 年 6 月 21 日
学位授与の要件	学位規則第 5 条第 2 項該当
学位論文題目	グラフ理論的概念に基づく文献情報検索の自動化
論文審査委員	(主査) 教 授 木 澤 誠 (副査) 教 授 田 中 幸 吉 教 授 藤 澤 俊 男 教 授 嵩 忠 雄 教 授 都 倉 信 樹

### 論 文 内 容 の 要 旨

入手による主題索引付けを主要課題とした伝統的な検索システムでは、多量の文献・質問の高速処理や、文献主題の質問に応じた柔軟な取り扱いが困難となっていた。本論文では、要求された文献が、質問との内容的関連性に従い、順序付けられて取り出せるような文解析法及び文献ファイル編成法を、述語ネットワークと文献集合の順序付けの考えに基づき定式化し、計算機実験を通して、文献処理に関するこれらの方法の妥当性を示す。

述語ネットワークは、自然言語で記述されている文献を、自動的な方法で処理するために提案されたもので、文献中の文内容を述語パターンと語適用領域と呼ばれる2種の属性を用いて表わすことにより、構成される。各語の文献中における使用のされ方を示す語適用域は、自然言語のもつ意味概念としての同義性・多義性を考慮しながら、グラフ理論的手法で求められる。なかでも、自然語文の形式的取り扱いに重要となる多義語は、グラフ中のクリーク概念に基づく簡単で有効な方法によりみつけられることを示す。また、語を関係付ける述語パターンを求めるには、文中、述語の項として表われる語の代りに、その語の適用域を与えるだけでよいことから、述語ネットワークは、従来の自然言語処理にみられるような人手による語の意味表現を必要とせず、多量のデータを対象とした計算機による言語解析のための有効な手段となっている。

検索システムの効率を上げるためには、文献間の関連性を新しく、文献集合の作るグラフ中で、それらを結ぶmax-minパス上の最小線重みとし、検索結果を質問との新しい類似度に従い、順序付けて取り出す方法を提案する。文献集合は最初、文献間の類似度に従い、一次元配列上に並べられた後、大きな類似度を示す文献対が、できるだけ近くにあるような部分を1つのクラスタとして分割される。

次に、これらクラスタは、そこに含まれている文献間の類似度によって階層的に関連付けられ、全体として文献ファイルに蓄積される。ここで提案したファイル編成法によれば、上位レベルでのクラスタを下位レベルでのクラスタのインデックス域と考えることができ、又、各クラスタ内では、蓄積項目が類似度の順に並べられているため、要求された文献集合を効率よく取り出すことが可能となる。文献クラスタを1つのファイルと考え、それらをここで述べた方法で階層的に関連付けることにより構成されたファイル群は、文献データベースとして扱うことができる。

文献集合を類似度の順に並べる操作は、又多量の類似情報を圧縮して記憶する方法やグラフ理論的概念を取り出すアルゴリズムの定式化に有効であることも示されている。

### 論文の審査結果の要旨

本論文は情報検索における質問と情報との適合判断過程にグラフ理論的な考え方を導入し、その自動機械化の効果の向上を図ることを目的としている。全篇は英語文で記述された文書情報に対して語(word)間、文(sentence)間及び文書(document)間の関係の3観点に着目して論じられている。語間の関係については、著者は異語間の類似度を定義し、これを数値的に表現して、これから同義性及び多義性を考慮したシソーラスを自動的に作成して利用する方法を示した。文間の関係については単文を動詞と前置詞とを中心にした述語と名詞を中心とした項とに分解し、これらおのおのに文集合中で同義的な使われ方を示すクラスを逐次的に対応させて述語ネットワークを求め、これの上で文間の類似性を判定する方法を試みた。また、文書間の関係については、文書間の類似度を述語ネットワークを通じて求め、これに基づいて文書集合を一行に配列してファイル編成を行い、探索に際しては与えられた質問に対する類似度の大きい順に効率よく取出すようなアルゴリズムを案出した。これらのいずれの方法に対しても Proc. IEEE及びComm. ACM誌所載の抄録を題材として電子計算機を用いた実験結果を示し、これらが情報検索に応用して有効であることを報告している。

上記のように本論文は自然言語（この場合は英語）で表現された文書情報の検索に有効な手段を与えるものであり、博士論文に値するものと認める。