【 68 】

氏名	木本晴夫
博士の専攻分野の名称	博士（工学）
学位記番号	第 10971 号
学位授与年月日	平成 5 年 11 月 24 日
学位授与の要件	学位規則第 4 条第 2 項該当
学位論文名	Automatic Indexing of Japanese Documents and its Application to Information Retrieval （日本語文書の自動索引とその情報検索への応用）
論文審査委員	（主査） 教授 菊野 亨 （副査） 教授 嵩 忠雄　教授 橋本 昭洋　教授 都倉 信樹

論文内容の要旨

本論文では，日本語文書の自動索引とその応用について述べる。前半で日本語文書の自動索引について述べ，後半でその応用について述べる。

最初に，文書の主題とほとんど関係が無い不必要なキーワードを効果的に削除する日本語文書自動索引技術について述べる。この方法は，文章解析情報，専門家が用いる索引付けの知識と語の出現位置情報や出現頻度情報などの統計情報を利用して自動索引をおこなう。この方法に基づいてキーワード自動抽出システム，INDEXER を実現した。INDEXER は 2 つの主要な機能を持っている。ひとつは，フリーキーワード抽出法や統制キーワード抽出法によって抽出されたキーワードの中で不要なキーワードを削除する機能である。もうひとつは，キーワードをその重要度評価点で順位付けする機能である。自動索引システムの評価尺度はキーワードの再現率と適合率である。従来からの方法であるフリーキーワード抽出法を用いた場合は，再現率が70%であり，適合率が10%であった。本稿で提案する方法を利用した場合は，再現率と適合率がともに50%となる。適合率を従来方法の10%から50%まで高められたのは，非常に大きな改良である。また，キーワードを重要度順に評価して順位付けする機能については，実験の結果，人がキーワードとした必要な語の95%を順位付けの10位以内に評価することができた。

一方，自動索引技術を利用したさまざまな応用分野が有る。それらは，情報検索，テキスト型データベースの加工，日本語教育支援などの分野である。本稿の後半部分では，これらの応用分野の中で，特に情報検索への応用について述べる。

新しい情報検索の方法として連想型情報検索について述べる。この方法は，動的シソーラスと呼ばれるキーワードのネットワークを導入して，その学習効果を利用する。連想型情報検索のメカニズムは次のとおりである。まず，個々のユーザによって検索のために学習用として与えられたサンプル文書から，①キーワード，②サンプル文書中でのキーワードの重要度，③サンプル文書中でのキーワード共起関係など，を自動抽出する。この時，キーワードの自動抽出とキーワードの重要度評価のために，先に述べた自動索引技術を用いる。次に，キーワードの重要度をネットワーク上のノードの重みとし，共起関係をネットワークの新しいリンクとして付加して動的シソーラスを構築する。そして，

ユーザが検索用に入力したキーワードから動的シソーラス内でのリンクをたどって，重みの小さいものを削除しながら，検索用キーワードを選択生成してゆき，生成されたキーワードを利用して検索をおこなう。この方法では，個々のユーザ毎に動的シソーラスを生成して，検索時に利用するので検索の結果が個々のユーザに適合したものとなる。プロトタイプシステムを作成して，評価実験をおこなったところ，従来の情報検索方法と比較して，飛躍的に高精度に検索ができることを確認した。

論文審査の結果の要旨

近年，データベースの巨大化に伴い，データベースへの効率的なアクセス技術へのニーズは益々高まりつつある。アクセスの速度と精度を向上させる目的で，自動インデクシング技術，及び，それに基づく検索技術が提案されてきている。本論文は日本語文書を対象に知識処理，統計処理，言語処理を統合することによって，不必要なキーワードを大幅に削除することが可能であることに注目し，重要度評価も取り入れたキーワード自動抽出システム INDEXER を開発し，適合率を従来の10%から50%にまで高め，自動索引システムの実用化に成功した。本論文はこの高精度自動索引システム開発の一連の研究成果をまとめたものである。

本論文ではまず文章解析情報，専門家の索引付けの知識，そして語の出現位置や出現頻度などの統計情報を利用することによって再現率と適合率を共に50%にできることを，実際にキーワード自動抽出システム INDEXER を開発することによって実証した。更に，キーワードの重要度順の順位付け機能の導入により，人手によるキーワードの95%を順位付けの10位以内に評価可能にするという成果も挙げた。次に，検索者から与えられるサンプル文書中のキーワードの重要度，共起関係に基づいて重みつきネットワーク構造をした動的シソーラスを構築した。引続き，検索者の入力キーワードから開始して，動的シソーラス上の枝をたどり重みの大きいものから順次，検索用キーワードを選択生成することにより検索者の意図を反映したキーワードの生成を可能とした。これらの研究成果に基づいたプロトタイプシステムの作成によって，飛躍的に高精度の検索が出来ることを確認した。

以上のように，本研究は高速，高精度の自動索引システムを開発し自動インデクシング技術の発展に大きく寄与すると共に，高精度の連想型情報検索技術を開発し情報検索技術に大きな進歩をもたらしており，博士（工学）の学位論文として価値有るものと認める。