

関連情報を用いた法令検索システムの開発

金井 貴

北陸先端科学技術大学院大学 知識科学研究科

E-mail kanai@jaist.ac.jp

1. はじめに

近年の IT 技術の普及はめざましいものがあり、法学教育においても IT 技術の導入が徐々に浸透しつつある。法学教育において最もよく用いられている IT 技術としては、法令および判例の検索があげられるが、従来の情報検索手法では良い検索結果を得るために適切なキーワードを吟味して入力しなければならないという問題が指摘されている。また多くの検索システムでは複数のキーワードを用いて複雑な論理式風の検索文を使用することができるが、利用方法が類雑等の理由であり用いられていない。そこで本報告では法学学習者、実務家および法学研究者等が検索したい内容を含む文章をマウスで範囲指定することで、関連する法令や判例等法律関連コンテンツを検索することができる情報検索手法を提案する。現在開発中のシステムでは文章をマウスで範囲指定するのみで文章から検索に適したキーワードを自動生成し、また生成されたキーワードと AND, OR, NOT 等を組み合わせた検索文を自動生成することで、既存の情報検索システムを活用可能でかつ検索の利便性を向上させることを目指し研究を行っている。また既存の情報検索システムのフロントエンドとして本報告で提案するシステムを用いることにより、従来のデータベースをそのまま用いることができ、かつ自動生成された検索文を用

いて効率的かつ精度の高い情報検索を行うことができる。

2. 関連情報を用いた法令検索システムの開発

本節ではこれまで行った研究のうち検索文生成技術の開発、情報検索支援システムインタフェースの開発、および法令検索エンジンの構築についてその概略を述べる。

法令検索システムは法令検索支援インタフェース、法令データベース、法令検索エンジンの3モジュールから構成される。ユーザは法令検索システムユーザインタフェースを用い、法律関連情報を関連文書としてマウスで選択する。次にコンテキストメニューから法令検索を指示すると関連文書中から検索に適すると思われるキーワードを抽出する。ユーザは提示されたキーワードから実際検索に用いるキーワードを選択し、法令検索エンジンを用いて法令の検索を行う。この検索プロセスにおいてキーボードを使用することなく検索を行え、かつ検索キーワードの絞り込みを計算機により自動的に行うため検索精度の向上が期待できる。

本研究で提案する法令検索システムは、本年度末までにマウスのみを用いた情報検索支援システムに関してはプロトタイプを稼働する予定であり、来年度にはプロトタイプシステムを用いて予備実験を行う予定である。

2. 1 検索文生成技術の開発

本研究では法律関連コンテンツの検索支援システムとしての利用を目的としているため、まず初めに法律分野における単語の出現頻度や偏りに関する情報等の統計情報の収集を行った。広範囲の分野における単語の出現頻度や偏りに関する情報等の統計情報等の収集データとして、第一法規出版が販売している現行法規 CD-ROM から第一法規出版の許可を得て、第6編民事編のテキストデータを抽出し、法令テキストの分析に用いることとした。使用した第一法規出版の現行法規 CD-ROM には2003年4月までに施行されている法令すべてが収録されており、第6編 民事編だけでも410の法令が収録されている。そのため民事法に関する広範囲のテキスト情報と考えると問題は無いと考えられる。

次に本研究以前に開発された関連文書による情報検索支援システム[1][2][3][4]の検証を行い、法律検索エンジンに関連文書を用いた情報検索を適用した際の問題点を検証した。すでに開発済みの情報検索支援システムでは、一般の Web 上にある情報を検索することを念頭に置いて、Web 情報に関するメタ情報を基に検索キーワードの自動抽出および検索式の自動生成を行っている。検証の結果、本研究以前に構築された関連文書を用いた情報検索エンジンでは検索の再現率のが高いが、検索の精度に関してはばらつきがあることが明らかになった。そのため本研究で開発する情報検索支援システムでは再現率よりも精度を向上させる手法を開発する必要があることがわかった。そこで北陸先端科学技術大学院大学の学生に対して検索エンジンの使用方法についてインタビューを行い、検索エンジンのユー

ザがどのように検索精度をあげているのかを調査した。インタビューの結果として、一般に2から3語のキーワードを用いた AND 検索で検索を行っており、検索キーワードには複数の基本的な単語からなる複合語を用いることで検索精度が向上することがわかった。このインタビュー結果を基に、単語をスペースで分けて検索した検索結果と複合語をそのまま用いた場合の Google における検索結果を比較したところ検索結果が異なることが確認され、一般に複合語を用いた検索の方が経験的に検索精度が高いことがわかった。そこで本研究では関連文書から複合語を取り出して検索キーワードとして用いることにより、検索精度の向上を計ることとした。

しかし複合語を検索キーワードとして用いたところ、多くの法律概念等の専門用語は計算機の辞書に含まれていないため、専門用語の多くを複合語キーワードとして認識することができないことがわかった。そこで既存の字句解析エンジン Chasen を基に単語の品詞と単語の前後関係から複合語を推定するアルゴリズムを開発した。また複合語はあらかじめ作成した辞書に現れないため複合語を検索キーワードとして用いることの良否が推定できないという問題があるが、複合語の検索キーワードとしての良否を判定する方法として、複合語に含まれる部分単語の中で IDF 値と関連文書中の出現頻度の積が最も大きいものをもとに複合語の良否を判定する手続きを考案した。

次に検索キーワードの集合から類似のキーワードを OR でまとめることで検索の再現率を上げる手法の開発を行った。OR 検索式の生成については、重要語として抽出された複合語の中に出現する単語をもとに重要語の類似性を判断し OR で結びつけることで、類似キー

ワードによる検索を可能にした。

上記の技術を新規開発した上それらを統合し、検索したい内容が記述されている文章をマウスで範囲指定するだけで AND と OR を組み合わせた和積標準型の検索文を自動生成し、関連文書から自動的に法令の検索を行えるシステムを構築した。

2-1.情報検索支援システムインタフェースの開発

前節では計算機が関連文書から自動的に和積型の検索文を生成し、情報検索システムへ自動的に質問を行う手法について説明したが、前節で説明したのシステムを用いて関連文書からの法令検索に関して簡単な予備実験を行ったところ、計算機による検索キーワードの選択は概ね問題ないが、ユーザの意向を反映していないキーワードを選択することがしばしばあることがわかった。また予備実験の被験者の声として、NOT 式を用いた検索文の自動生成に関しては否定的な意見が多く、自動的に NOT 式を付加することは困難であることがわかった。

そこで自動生成された検索文をユーザとのインタラクションの中で変更することで検索精度の向上を図れるよう、ユーザの指示により半自動的に検索文を生成するインタフェースを開発した。開発したインタフェースでは、関連文書から自動的に検索を行うのではなく、ユーザがマウスにより検索キーワードを指示し、計算機により提示されたキーワードを付加した(あるいは取り除いた)検索文をインタラクティブに生成する方式を考案した。また抽出した検索キーワードのみではなく法令から抽出した関連キーワードも同時に提示し、検索キーワードとして使用できるようにした。

また抽出した検索キーワードのみではなくベクトル空間法を用いて検索キーワード群と法令中の各条文との間の類似度を求め、検索キーワードに関連する条文を推定し、関連条文と関連条文中に現れる単語のうち、ユーザが与えた関連文書に現れなかったキーワードを関連キーワードを提示することで、関連カテゴリのキーワードの付加や関連キーワードを含まない条文を検索できるようにした。図1-3に本研究で構築した情報検索インタフェースおよび本研究で開発された検索エンジンを用いた検索結果を示す。

図1はWWW上の法律関連コンテンツを例としてマウスで選択している様子である。ここで右クリックを行うことでコンテキストメニューが表示され、法令検索を行うことができる。



図1 関連文書をマウスで選択

図2ではマウスで選択したテキストから検索に適したキーワードを抽出し、重要度に従い表示している。テキスト上のキーワード以外にも関連文書として与えられたテキストに類似した法令をベクトル空間法により検索し、関連文書内に現れていない、法令内の重要キーワード上位5個を選択できるように表示している。ユーザはテキスト内から抽出された重要キーワードと関連法令内の重要キーワードから検索に用

いるキーワードを選択し、情報検索を行うことができる。



図2 関連文書から重要キーワードと関連条文からキーワードを提示

図3は選択されたキーワードを用いて法令を検査した結果である。関連文書に用いたテキストは遺失物に関する民法の解説であったため、遺失物法や民法が検索結果として表示されている。



図3 選択したキーワードから法令を検索

2-3. 法令情報検索システムの構築

本研究で構築した法令情報検索システムは、一般のWeb検索エンジンと同様にキーワードから関連する文書を検索することができるが、検索対象を法令に特化していることが特徴としてあげられる。一般のWeb情報における情報

検索では数億以上のWebページが検索対象となっているため、再現性の低下よりも精度の低下がボトルネックとなることが多い。しかし法律検索システムにおいては、法令は8000強しかないため一般のWWWと比較して検索対象は比較的限定されており、また判例に関してはその数は多いが、一般に検索結果として含まれるべき判例数はそう多くはない。そのため精度を重視する一般のWeb検索とは異なり、精度と再現率のバランスがとれた検索結果となるようにしなければならない。

そこで検索エンジン側では検索支援システムにより生成された検索文中のキーワードのうち、複合語として認識できるキーワードについては、複合語を基本単語に分割した個々のキーワードも複合語と共に検索キーワードとして用いることで、検索結果の再現率向上を行った。これは基本単語を検索キーワードとしてOR検索を行うことと類似しているが、検索キーワードには元の複合語を含んでいるため、検索精度が著しく落ちないことが期待され、また基本単語に分割したキーワードも検索キーワードとすることで、検索結果の再現率の低下を防ぐことが期待できる。

3. 関連研究との比較

本節では本研究で構築した法令情報検索支援システムおよび法令検索システムと、他の検索支援システムおよび検索システムとの比較を行う。本研究の目的は法令検索を容易にする検索インターフェースおよび検索エンジンの開発であるため、比較対象として現在WWW上で利用されている検索エンジンや他の法律情報検索システムを中心に比較を行う。

本研究のシステムは文章を元に検索支援を行うが、WWW検索エンジンの中にもAskJeev

esのように文章を入力させて検索することが可能な検索エンジンが存在する。しかしこれらの検索エンジンでは検索質問を文章としているため入力された文章から検索結果を予想することが難しく、利用が難しいことが問題点としてあげられる。また本研究の手法では検索質問の自動生成を行うため既存の情報検索システムを有効活用することが可能であり、検索質問をインタラクティブに変更することで予想される検索結果を改善することが可能である点も従来のWWW検索エンジンより使いやすいと言える。

次に法律情報検索の関連研究としては、統計的情報を用いた法情報マネジメントシステム FLEXICON[5]と法情報検索システムに事例ベース推論を組み込んだRisslandらの研究[6]があげられる。

FLEXICONでは、法情報検索手法としてベクトル空間モデルを用い、関連フィードバックの手法等を用いより検索精度の高い判例検索を実現している。判例検索と法令検索の点で相違はあるが、本研究でも関連文書と類似した法令内の条文を検索する手法としてベクトル空間を用いている。本研究の手法は関連フィードバックとは異なり、ユーザが検索結果を理解して関連を判断するのではなく、情報検索を行う前に、計算機より提示された関連キーワードを選択するというインタラクションを行う。本研究の手法ではユーザがキーワード選択において試行錯誤することによりユーザ自身の判断において検索結果に影響を与えることができ、ユーザからのフィードバックはキーワード選択の試行錯誤により行うことができる。我々の考案したキーワード選択インタフェースでは、キーワードを入力する必要がないため、従来のキーワード入力を伴う検索手法より容易に検

索を行うことができる。また本研究の特色として、文章から検索質問を自動生成するという手法を用いているため既存の情報検索システムを変更することなく使用することが可能である。そのため本研究によりキーボードを使ってキーワードを入力することなく検索が行えるようになり、法学部学生等のコンピュータに不慣れな初心者に使いやすく、かつ従来の情報検索よりも関連する情報を素早く入手することが可能となる。また初心者にありがちななどのキーワードを用いて検索すれば良いかわからないという問題を克服することができる。

次にRisslandらの事例ベース推論を用いた情報検索では、FLEXICONと同様に、関連フィードバックを用い検索精度を向上させる手法をとっているが、関連フィードバックは事例ベース推論により自動的に行われる。Risslandらのシステムでは、まず検索したい事例を入力することで事例ベース推論を行い、事例ベース推論エンジンにより関連あると判断された判例が特定される。次に情報検索を行うときに、事例ベース推論により関連あると判断された検索結果が検索結果の上位にくるよう関連フィードバックが自動的に行われる。

本研究のアプローチは、情報検索の前にキーワード選択を行うことである種のフィルタリングを行っていると考えられるが、我々の研究では事例は基本的にマウスで文章の範囲指定を行うだけで精度の高い情報検索が可能となる点がRisslandらの研究と異なる。またRisslandらの手法では事例ベース推論を用いるために事例の入力を予め行わなければならないが、我々の手法ではキーワード選択の良否を判断するために必要な情報としては法令や判例等のテキスト集合で良いため、データを用

意する手間がほとんどかからないという利点がある。

4. おわりに

本報告では、検索したい内容を含む法律分野に関連するテキストをマウスで範囲指定することで、関連する法令や判例等の法律関連コンテンツを検索することができる情報検索手法についてその概要を述べた。また法令検索エンジンの構築について説明し、関連研究との比較を行った。

今後の課題としては、現在のデータベースでは現行法規のすべてをデータベース化していないため、データベースが不完全な状態であることがあげられる。現在法令を中心に検索対象の拡大を進めている。

また現在の法令検索システムでは、類義語として、関連文書として与えられたテキスト情報中の単語および法令文書中の単語を提示しているが、法学で一般に使用されている単語を類似語として提示することはできないことも問題点としてあげられる。そこで法律学の辞典等から法令文書中に表われていない単語を抽出して関連付けるデータベースを構築し、検索中に提示する機構を構築する必要があると考えられる。

<参考文献>

- [1]Takashi Kanai, Li Jian, Susumu Kunifuji, Related Document based Information Filtering Applied to the Association Model In formation Retrieval System. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, pp.225-228, Aug. 2000
- [2]金井貴, 齊藤主税, 國藤進, 文書による情報検索を用いた対話場における創造性支援, 日本創造学会論文誌, Vol.5, pp.122-132, 2001
- [3]竹端和歩, 金井貴, 國藤進, ノウファー・マネジメン支援システムのフレームワークに関する新提案, 社会情報学会全国大会研究発表論文集, Vol.17, No.1, pp.183-188, 2002
- [4]竹端和歩, 金井貴, 國藤進, 組織内の口コミ情報を活用した Know-Who マネジメント手法の提案, 情報処理学会第 65 回全国大会 pp. 4-445 - 4-446, 2003
- [5] Daphne Gelbart, J. C. Smith, FLEXICON: An Evaluatioin of Statistical Ranking Model Adapted to Intelligent Legal Text Management, Proceedings of the fourth international conference on Artificial intelligence and law
- [6]Edwina Rissland and Jody J. Daniels, A Hybrid CBR-IR Approach to Legal Information Retrieval, Proceedings of the fifth international conference on Artificial intelligence and law