

平成 22 年 6 月 14 日現在

研究種目：基盤研究 (C)

研究期間：2007～2009

課題番号：19500087

研究課題名 (和文) 新聞記事を用いた対話型訴訟相談支援システム

研究課題名 (英文) A Framework of a interactive consultation system for suit using newspaper articles

研究代表者

樽松 理樹 (KUREMATSU MASAKI)

岩手県立大学・ソフトウェア情報学部・准教授

研究者番号：00305286

研究成果の概要 (和文)：

新聞記事を利用した訴訟相談支援システムに対し、次に示す基本アルゴリズムを構築した。

- (1) 新聞記事に対し、複合語の一部の省略をそれ以前の文を元に復元を試みる省略の解消を行う。
- (2) 記事を、見出し、リード、1段落目、2段落目以後に分割し、各部分中の語句を抽出する。
- (3) それらの語句に対し、時系列を遡り検索する。この時、記事のブロック間での語彙の一致率と占有率に着目し、閾値以上の場合、同一事項とみなし、記事クラスとしてまとめる。この記事クラスに対し、キーワード検索を行い関連する文を表示する。本アルゴリズムに対し、交通事故事案を対象に机上評価を行った。その結果、検索能力に課題があることが明確になった。

研究成果の概要 (英文)：

I proposed the following algorithm for a lawsuit consultation support system using newspaper articles. This algorithm has 3 Steps. In step.1, it tries to decode clipped words using reference terms. In order to do so, it focuses phrases which have same words in a clipped word and appear before the clipped word. It replaces the clipped word by the nearest phrase. In step.2, it divides an article into title, lead, first paragraph and rest of the paragraph. Then, it extracts words from each block. In step.3, it retrieves articles from previous articles based on words of each block. When it retrieves articles, it calculates the coincidence rate and the share rate of words in each block. If both rates of an article over these thresholds, it returns the article and adds it to a collection of articles about same topics. When a user asks this system, it extracts some sentences which include words in a question. I evaluated the performance of this algorithm about traffic accident cases on with desk check. Consequently, I have to improve the retrieval performance of this algorithm.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	270,000	1,170,000
2008年度	600,000	180,000	780,000
2009年度	700,000	210,000	910,000
総計	2,200,000	660,000	2,860,000

研究分野：総合領域

科研費の分科・細目：情報学・メディア情報学・データベース

キーワード：情報検索、法情報

1. 研究開始当初の背景

我々と法律との間の隔たりは大きい。その理由としては、法律は専門性が高いこと、弁護士などの専門家の数が少ないことなどがあげられる。

このような問題に対し、利用環境が整いつつあるコンピュータでの支援が考えられる。すなわち、国民が求める法サービスを提供する法律を対象とした情報システム、法情報システムが開発し、利用環境を整えることで、司法と国民との壁を取り除くことの支援が可能になる。さらに裁判員制度における一般の裁判員への支援にもつながる。このことから法情報システムの必要性は高くなると考えられる。

法情報システムは、これまでに人工知能分野を中心に研究されている。代表的なものとしては、Ashlay らによる HYPO、ICOT で開発された新田らの HELIC-II などがあげられる。また判例、法令、法律用語の電子化、データベース化が進み、商用化されている。さらに近年、領域オントロジーの一つとして、法律知識の体系的定義である法律オントロジーの構築も行われている。法律オントロジーの研究はセマンティック Web で利用可能な法律 RDF 辞書の開発へと発展している。

しかし、これら法情報システムの多くは専門家向けのものであり、非専門家には利用しにくい。司法と国民の壁を取り除くには、非専門家向け法情報システムの構築が必要である。

そのようなシステムに要求されるものの一つとしては、「人の行為に対する刑罰、量刑」を示すことである。この点に着目すれば、専門性が高い判例よりも、求める点が端的に記載されている新聞記事と照合し、検索するほうが有用であると考えられる。また新聞記事は判例に比べると身近であり、情報が早いという利点がある。また法令が法律世界の言葉で記載されているのに対し、現実世界の言葉で記載されているため、実際の事案との照合が行いやすい。さらに新聞記事に関しては、現在 WWW の普及とともに研究が進められている情報検索、文章検索の技術が利用できる。現在の検索は、条件に該当する記事を検索するものが主流であるが、それを拡張することで、より柔軟な新聞記事の検索が可能である。

2. 研究の目的

本研究では「新聞記事を用いた対話型訴訟相談支援システム」の開発を目指す。本システムは、与えられた事例に対し、新聞記事内の訴訟例と比較し、考えられる刑罰と量刑を提示する。この際、類似度に基づき確信度を付与する。また照合しなかった点をユーザに問うことで、ユーザが忘れていた点、見逃していた点を想起させるとともに、推論結果の精度を高めることを試みる。

3. 研究の方法

上記のシステムを構築するために、(1)新聞記事データベースと(2)訴訟推論エンジンの構築を行う。

(1)新聞記事データベースでは、初めに新聞記事に対し、指示詞や“同文”、“右文”といった繰り返しを省略する照応関係に対し、省略された語句を復元する解析を行う。次に、同一事件の記事を抽出し、一つの事例としてまとめる。最後に事例検索の効率を高めるために事例を特徴づけるインデックスを作成する。事例とインデックスをまとめたものを事例データとし、訴訟推論エンジンにて活用する。

(2)訴訟推論エンジンでは、初めに入力事例と事例データとのインデックスを照合する。このとき既存の汎用概念辞書を使うことで検索精度を高める。次に検索で見つけられた事例と入力事例との語句の対応付け抽出を行う。対応付けの抽出を行うのは、行為が記載されている部分である。このときにも汎用概念辞書と法律オントロジーを活用する。次に対応付けした語句を事例の刑罰・量刑の部分に適合し、入力事例に対する刑罰・量刑を推論する。最後に入力事例と適合した事例結果をまとめ、互いに不足している部分と合わせてユーザに提示する。ユーザが不足部分に対応する行為などを追加入力した場合、再度対応付け抽出、適合を行い、出力を更新する。

本研究期間内においては、上記のシステムを構築するため、それぞれのブロックの基本アルゴリズムの確立、プロトタイプの開発、評価実験を行う。具体的には(1)省略を復元する照応解析技術、(2)同一事件の新聞記事を事例としてまとめる技術、(3)新聞記事から生成する事例からのインデックス抽出技術、(4)概念辞書を用いた事例検索技術、(5)事例間の

語句の対応付け抽出技術、(6)語句の適合に基づく推論手法の開発を行う。

4. 研究成果

(1)省略を復元する照応解析技術に対しては次の事を行った。“同〇〇”という表現に対し、それ以前に出現する文において、同じ続く部分と一致する直近の語句を用いることで、省略の解消を試みた。本手法に対し、アルゴリズムを構築し、データにて実験を行った。正解率は約70%であった。また、相対表記を絶対表記、漢数字をローマ数字に候補を提案し、対話的に変換するアルゴリズムを構築した。

(2)同一事件の新聞記事を事例としてまとめる技術、(3)新聞記事から生成する事例からのインデックス抽出技術、(4)概念辞書を用いた事例検索技術、(5)事例間の語句の対応付け抽出技術、(6)語句の適合に基づく推論手法の開発について実施したことは、互いに関連が強いため、以下にまとめて示す。

最初に、参考文献を元に新聞記事の文書構造として、見出し、リード、本文・1段落目、本文・2段落目以降の4つの構造を洗い出した。これらのうち、ポイントとなる語句が出現頻度、記事での有無に基づき、前半部、リード、見出し、後半部の重要度順位を決定した。また同一事項に関する記事であっても、同一新聞内の掲載面や異なる新聞間において意見の部分に差があることに着目し、検索の上での一つの指標として捉えた。

これらの部分を利用し、記事の関連付けを行うアルゴリズムを設計した。アルゴリズムとしては、(1)種の記事を決定、(2)種の記事のブロック毎の語句を抽出、(3)これらの語句に対し、時系列を遡る形で検索する。この時、記事のブロック間での語彙の一致率と占有率に着目し、閾値以上の場合、同一記事とみなし、一つの情報、記事クラスにまとめる。一致率、占有率の計算においては編集距離を活用し、次に示す計算を用いる。一致率では、はじめに記事内に出てくる単語とその後に続く助詞を一つにまとめた句とする。この句単位での編集距離を求める。各ブロックに対し、全部変更した場合の編集距離を100とし、それに対する割合を求める。ただし、移動が多いほうがより編集距離が小さくなりように変更と移動とでは重みを変える。この編集距離に、句単位での編集距離に基づく類似度を求め、この値で割ったものを一致率とする。これが一定の値以上の場合、一致率が高いとする。また占有率は、両方の記事に含まれる語句の総数を検索対象の記事内に含まれる語句の総数で割ることによって求める。なお、このとき総数には、出現個数も反映する。これにより、記事においてどれだけ語句が重

なっているかが明確となる。ただし、語句は名詞、動詞、未知語に限定する。

この記事クラスに対し、キーワード検索を行う。キーワードを含む文に関連する文とし、表示することで、訴訟相談支援を試みる。処理の流れとしては、(1)キーワードにより記事クラスを選択。(2)記事クラスの中から、記事クラスを絞り込むのに有用と思われるキーワードをユーザに提示。(3)ユーザの回答に基づき、記事クラスを絞り込む。(4)記事クラス数が一定値以上であれば(2)に戻る。それ以外の場合は、絞り込んだ記事クラス内にある回答(判決など)部分を表示する。

本アルゴリズムに対し、交通事故事案を対象に机上評価を行った。その結果、検索能力に課題があることが明確になった。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計0件)

〔学会発表〕(計2件)

① 樽松理樹、鈴木奈緒、“判例間の連関係数を用いた法的推論支援システムの提案”、平成20年度電気関係学会東北大会、2007/8

② 樽松理樹、“複数記事の融合・再構築による情報抽出手法の提案”、平成20年度電気関係学会東北大会、2008/8

〔図書〕(計0件)

〔産業財産権〕

○出願状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況(計0件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]

http://p-www.iwate-pu.ac.jp/~kure/Study/kiban_c_07_09.html

6. 研究組織

(1) 研究代表者

樽松理樹 (KUREMATSU MASAKI)

岩手県立大学・ソフトウェア情報学部・

准教授

研究者番号：00305286

(2) 研究分担者

()

研究者番号：

(3) 連携研究者

()

研究者番号：