

意味グラフのマッチングによる 事故問い合わせ文からの判例検索システム JCare

原田研究室 鈴木 亮 (35599009)

今日ある検索システムは、索引語を用いたキーワード検索が主流となっている。しかし、一般にキーワードによる絞り込みは難しく、検索結果からまさに必要とする情報に絞り込むには、その内容の説明文などを検索要求と比べる必要があった。

本研究ではこの問題を解決するために、日本語文書で記述した「問い合わせ文」を検索要求として受け付ける判例検索システム JCare を開発する。検索対象は「判例」のなかでも「交通事故関連の判例」に絞り込む。また、この判例の中でも「当事者間に争いのない事実」文のみを検索対象とする。この検索要求と判例の内容を照合するのに、日本語意味解析技術を用いる。本検索システムでは、文章を意味解析し、語が格納された節点と語間の関係(深層格)を表す弧からなる意味グラフに変換する。この結果得られる判例文と問い合わせ文の意味グラフ間における位相同型部分の大きさをもとに、両者間の内容類似度を算出する。しかし意味グラフ同士をそのまま比較するのは計算量的に見てコストが高いため、意味グラフを視点 (View) により 11 個の部分グラフ (View グラフという) 群に分割する。View とは「事故現場の状況」「当事者の行動」「当事者の静的特徴」など、事実文中に記述されている概念の集まりを特徴付ける“視点”である。View グラフごとの類似度算出においては、節点同士の語意類似度および対応する節点間の最短路上にある深層格リストの類似性を評価し格類似度を求める。ここで視点によって、深層格の類似度算出値を変化させることで、視点毎の深層格の重みに応じた類似度算出を行えるようにした。

本研究では、語の持つ語意による類似度に加え、語と語の間にある深層格群の類似度も類似性評価値として取り込むことで、文章の内容の類似性に基づく検索システムを実現した。そして実際に 161 字からなる問い合わせ文に対して 3 つの判例の事実文(平均 700 文字)との類似度を算出したところ、それぞれ 138, 202, 264 となり、その値は「専門家による類似性の程度をよく表わしている」という評価を得た。