CSP を利用した類似症例検索システムの概要

土井俊祐[1]　十時章[1]　鈴木隆弘[2]　高林克日己[2]
千葉大学院工学研究科[1]　千葉大学医学部附属病院企画情報部[2]

背景と目的

近年、テキストマイニング技術を利用した類似症例検索が注目されている。我々も過去の連合大会で退院時サマリー等の診療文書を利用し、類似症例検索が行えることを報告してきた。しかし、それらは研究段階で利用されているだけであり、実際に医療現場で利用され、どのような評価を得るかは実証されていない。そこで本研究では、公開されている医療文書である学会の症例報告を利用して実際に検索システムを構築・実装し、利用者の評価を集めることを目的とする。

対象

対象としたデータは、日本内科学会地方会の症例報告約 15000 件で、これをデータベースとして類似症例検索システムを構築した。データベースソフトとして Cache、インターフェースとして CSP(Cache Server Pages)を利用し、Web ベースで検索を行うことができるシステムとした。

方法

検索方法としては、まず 15000 件の症例報告を 1 症例 1 テキストファイルとして用意し、その中から形態素解析により名詞を抽出する。抽出した名詞に対し、tf-idf 法を用いて重要度を算出し、重要度ベクトル空間を作成する。これを検索先となるデータベースとして Cache の中に保存する。ユーザが類似症例を検索する際には、Web ページに用意したフォームの中に検索したい症例の文書を入力すると、自動的に同様の手法で重要度ベクトルを作成し、データベースの中から内積値の大きい症例報告を類似症例として出力する。学会症例報告には個人情報が含まれないので、結果表示の際には症例報告の内容や、著者への連絡先等も表示できるシステムとした。

このシステムを試験段階として当院内に導入したので、本システムの詳細と評価について、デモを交え紹介する。

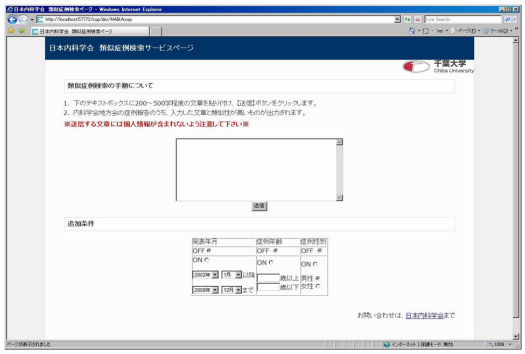

図 1　検索画面

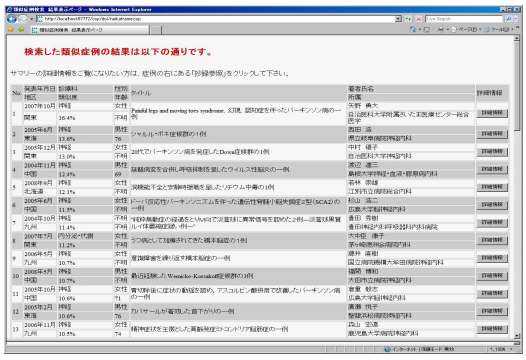

図 2　結果表示画面