海難審判庁裁決録のウェブデータベース化

海上安全研究領域 総合安全評価研究グループ　伊藤 博子

1. はじめに

海難審判庁裁決録 (以下裁決録) は、(財) 海難審判協会が海難審判の裁決書 3ヶ月分を一冊に編集して年 4 回刊行している書籍である。裁決書は、衝突、火災といった海難を事件として審判する際に作成される文書で、事件毎に関係者の資格や関係船舶の要目、事実、原因等の情報と行政処分が記載される。海上の重大事故の主要情報がまとめられており、要因分析や安全評価等の貴重な情報源である。ただし毎年数百件にのぼる中から必要な情報を収集するには労力がかかるので、事件区分や発生場所等の属性情報を簡単に扱えるよう、電子化が進められている。例えば、当所で研究を担当した機関室火災事故の安全評価では、国土交通省海事局により電子化されたデータ (平成 2〜10 年裁決分 7,038 件) が用いられた[1]。裁決録の電子データとしては、平成 7〜11 年度の裁決分が日本財団ウェブサイト[1]に置かれ審判庁や事件区分による分類と文字検索機能を備えており、それ以降の分は海難審判庁ウェブサイト[2]で担当した審判庁と言渡年で分類し、逐次最新情報を追加して掲載している。

当所では、裁決録の電子データを用いて、研究上必要な情報を効率良く取得することを目的に、条件設定による柔軟な検索機能と、大量の情報の中から必要事項を迅速に探す閲覧支援機能について具体的に検討し、ウェブを用いた裁決録データベースを設計、試作し、海難審判庁の了解を得てインターネットからアクセスできるようにしている[2]。

2. 裁決録データベースシステムの基本概念

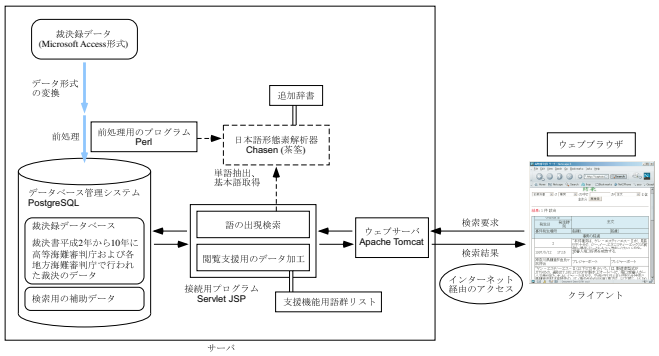
裁決録データベースでは、データを研究の情報源とするため、柔軟な検索機能と、必要事項を迅速に読みとるための閲覧支援機能を提供することを目指している。そこで、ウェブを用いて、インターネット経由でデータを提供することで、ユーザはウェブブラウザでアクセスすれば良いようにした。これにより、データの管理も一元的に行うことができ、ユーザによるデータ管理が必要なくなる。一方、サービス側 (サーバ) では、ウェブサーバを導入し、データアクセスのための仕組みを設計、構築してきた。使用したデータは、上述の海事局による電子データである。裁決書の中では、事件と裁決の過程により様々な項目を設けて記載しているが、この電子データには図表等を除く本文のテキストデータ (以下、これを裁決録データと呼ぶ) が格納されている。

全体の構成は図 1 に示すようになっており、ユーザからの検索要求 (クエリ) をウェブサーバが受け付け、接続用プログラムを経由してデータベースサーバにデータを要求し、得られた結果を閲覧支援等の必要に応じて表示用に加工してユーザ側に送信する。

2.1 サーバの構成

サーバは、Linux をインストールしたパーソナルコンピュータを使用し、インターネットからアクセスできるようにした。ウェブサーバとして、Apache[3] および Java Servlet や JSP (Java Server Pages) を使用するための拡張モジュール Tomcat[4] を、データベース管理システム用に、PostgreSQL[5] を使用した。この他、動詞や助動詞などの品詞活用による差異を吸収するために、自然言語処理用のモジュールである日本語形態素解析器の「茶筌[6]」を用いた。形態素解析では、使用される単語を辞書として登録しておくことで単語の解析処理を行うが、用意されているのは一般的な日本語の辞書であり、海事分野の専門用語まではカバーしていない (例えば、「主機」「他船」「受審人」「発航する」等の用語がない)。専門用語の周辺で語の扱いが不的確となり、文全体の検索精度に影響を与える。そこで、裁決録で頻繁に用いられる語について辞書を追加定義した。現在、使用頻度の高い名詞と動詞 317 語を登録してある。

[1] http://libdb1.nippon-foundation.or.jp/KAINAN/NF_KMST.nsf
[2] http://www.mlit.go.jp/maia/04saiketsu/04saiketsu.htm
[3] http://www.apache.org/
[4] http://jakarta.apache.org/
[5] http://www.postgresql.org/
[6] http://chasen.aist-nara.ac.jp/index.html.ja



図–1　システムの構成

2.2　検索の機能

検索機能は、指定された検索要求 (クエリ) に適合する裁決書の一覧を検索結果として表示させる。検索要求は、複数の単語と検索条件の指定によって行う。検索条件としては、検索語から論理式を作る際の AND/OR の選択や、事故の種類、裁決年度、海難 ID (データ毎に付けられた一意の ID) 等が指定できる。検索結果は、図 2(b) に示すように適合した裁決例の一覧として提供する。各裁決書の概要が分かるように、事件名、発生場所、発生日時、船名、主文が表示され、詳細表示のリンクを辿ることで (同図 (c))、事実の経過や原因の考察といった詳細情報が閲覧できる。事実の経過は、事故の状況に関する記述であり、背景に関する情報や経緯が述べられている。数千字から長いものでは一万字を越えるものもある。

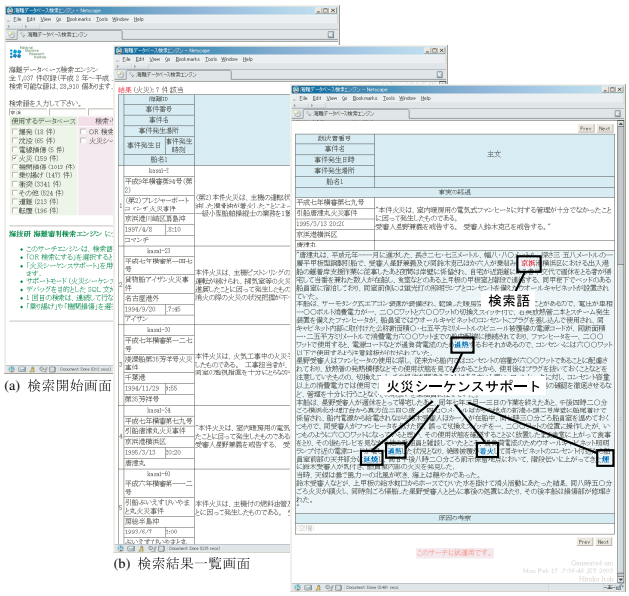
2.3　閲覧支援の機能

データとして多くの事例を集めるような場合には、長い詳細記述を読み、要求している事例であるかどうかを判断する必要がある。そこで文中の重要部分を迅速かつ見落としなく読みとれるよう、支援することを考えた。閲覧支援機能は、このような要求を満たすために、データを表示用に加工する機能である。ここでは、以下の 2 つを行えるようにした。

- 検索要求で与えた語の箇所を強調表示する。
- 特定の属性を持った語の箇所を強調表示する。

1 番目については、ユーザが入力した検索語を文中で、赤色表示するようにデータを加工する。2 番目については、一定の目的に対して共通に必要となる語群を青色表示するようにデータを加工する。

例として「火災の一般的シーケンスを調査する」という目的がある場合を想定して、「火災シーケンスサポート」という閲覧支援機能を考える。この場合、火災の過程を表す語として、「漏洩」、「火花」、「発



(a) 検索開始画面
(b) 検索結果一覧画面
(c) 詳細表示画面

図–2　ウェブブラウザでの閲覧

火」といった語に着目すると、文章からの情報抽出が行いやすくなると考えられる。そこで、「火災シーケンスサポート」を構成する語群を選定し、これを用いて強調表示を実現した。図 2(c) ではこれらの協調表示が行われている。同様にして、様々な閲覧支援機能を語群で定義できる。今後、調査目的と支援機能の関係について考える予定である。

3.　まとめと課題

ここでは、海難審判庁裁決録の電子データを用いて研究上必要な情報を効率良く取得することを目的に、設計、試作したウェブを用いた裁決録データベーについて説明した。検索機能として、事故の種類や裁決年といった属性や語句の指定、論理式の条件指定を含めた検索ができるようにし、閲覧支援機能として文中の重要語句を協調表示するようにした。今後、自然文処理のための専門用語の拡充や重要語句の定義について検討することで機能を充実させていく必要があると考えている。

参考文献

[1] (社) 日本造船研究協会 RR49: 船舶の総合的安全評価に関する調査研究 平成 13 年度報告書 (分冊: 安全評価事例), 2002

[2] (社) 日本造船研究協会 RR-S7: 船舶の総合的安全評価に関する調査研究 平成 14 年度報告書, 2003