

## 課題情報シート

課題名：	XPath を利用した『Web ページ上の情報抽出ソフトウェア』の制作		
施設名：	中国職業能力開発大学校		
課程名：	専門課程	訓練科名：	情報技術科
課題の区分：	総合制作実習課題	課題の形態：	制作

### 課題の制作・開発目的

#### (1) 課題実習の前提となる科目または知識、技能・技術

工場内ネットワーク、システムプログラム、ソフトウェア設計、コンテンツ制作、Web アプリケーション

#### (2) 課題に取り組む推奨段階

コンテンツ制作実習および Web アプリケーション実習終了後

#### (3) 課題によって養成する知識、技能・技術

課題を通して、主に Web クライアント/Web サーバプログラミングおよび連携プログラミング技術の実践力を身に付けます。

#### (4) 課題実習の時間と人数

人数：1人

時間：360時間

Web 上におけるアプリケーションは、Web サーバ上に仕掛けられますが、本課題は Web クライアントと Web サーバが連携し、書籍の ISBN から書籍情報を Web 上から抽出し Web クライアントが表示するシステムを制作しました。

今回、このようなシステムに注目し、“システムづくり”の一連の流れを理解し、専門技術の向上を図ることを目的として、XPath を利用した『Web ページ上の情報抽出ソフトウェア』の制作に取り組みました。

### 課題の成果概要

Web クライアント側では、フォームのテキストエリアに入力された ISBN を JavaScript で受け取り、ISBN のチェックと校正を行います。処理後、その ISBN と選択された検索先の Web サイトの情報をサーバー側の CGI に Ajax を利用してデータを渡します。また、送信後は非同期で待機しているため、結果を待たずに次の ISBN を入力・送信することができます。

Web サーバー側から処理した結果が戻ってくると、JavaScript でテーブルを動的に追加します。また、これらで使用した JavaScript のライブラリの prototype.js はブラウザの違

いを吸収するのでInternet Explorer®やFirefox®など、ほとんどのブラウザで動作可能です。

サーバー側では、3つのPerlプログラムが動作しており、データの受け渡しを行うメインプログラムと、そこから呼び出され各検索サイトにあわせて処理を行うライブラリに分かれています。そのうち2つはクライアントからデータを渡され、処理しやすい形に修正するプログラムとISBNから検索するURLを作成するプログラムです。残りのプログラムでHTMLから必要な情報の抽出を行っています。

検索サイトの情報は、ライブラリ化し、メインプログラムから分離することによって検索サイトの修正・追加が容易になり、またメインプログラムもほとんど手を入れなくてもいいようにしました。

メインプログラムでは、クライアントから受け取ったISBNとWebサイトの情報を各ライブラリと受け渡ししながら処理し、結果をクライアント側に返します。

URL作成ライブラリでは、検索サイトのURLが検索単語などを含められるようになっていたため、その部分をISBNに置き換えて、取得したい書籍情報が含まれたWebサイトのURLを作成します。

XPathライブラリでは、抽出を行うXPathを利用しており、そのためのモジュールとしてTreeBuilderとLibXMLを利用しています。抽出の際、うまく抜き出せない可能性もあるため、どちらのモジュールでも最小の手直しで利用できるようにプログラムを作成しました。ブラウザの検索画面を図1に示します。

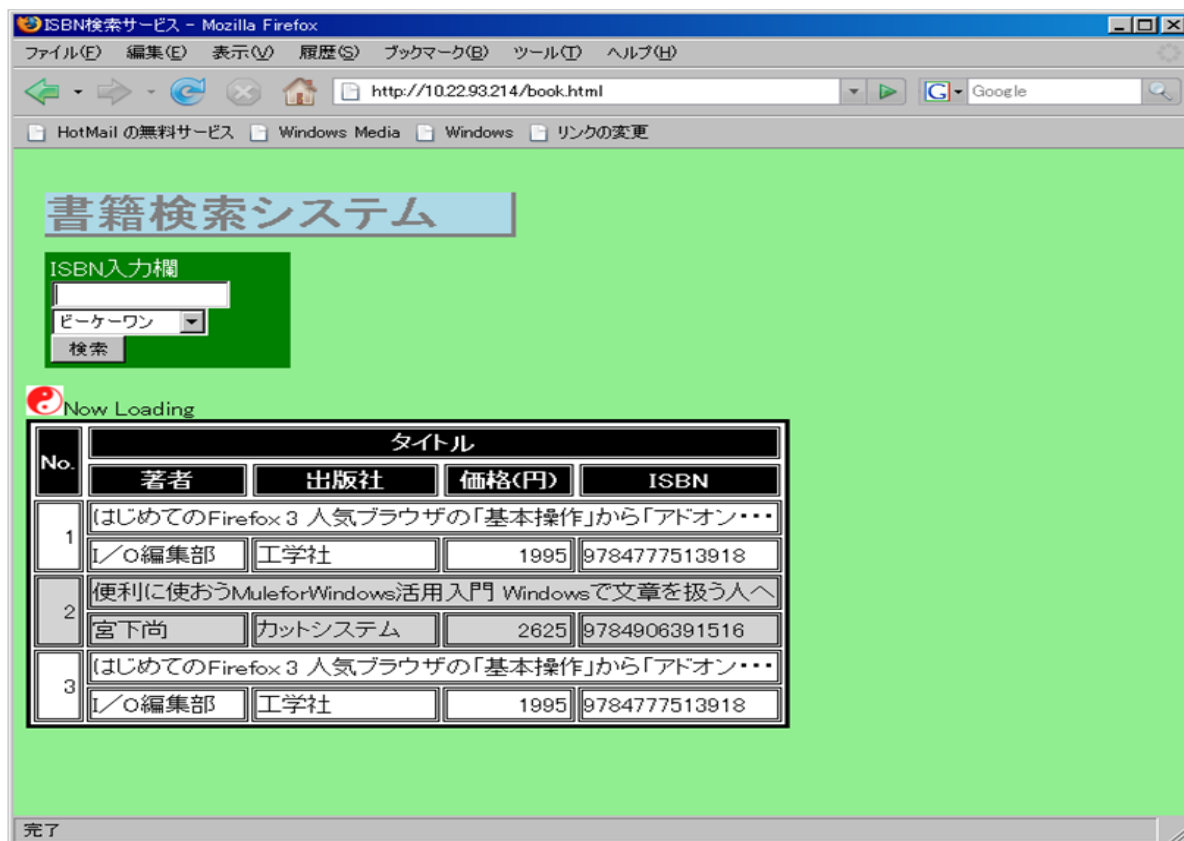


図1 書籍検索画面

## 課題制作・開発の訓練ポイントおよび所見

本課題についての訓練ポイントおよび所見を以下に紹介します。

養成する能力 (知識、技能・技術)	課題制作・開発のポイント	訓練（指導）ポイント
<p>○ Web クライアント側におけるプログラミングおよびコンテンツ制作、スタイルシート設定の各技術および各々の独立した形での設定手法が習得できる。</p> <ul style="list-style-type: none"> <li>・ JavaScript プログラミング技術</li> <li>・ DOM スクリプティング技術</li> <li>・ HTML タグ技術</li> <li>・ スタイルシート(CSS) 技術</li> <li>・ Ajax プログラミング技術</li> </ul>	<p>◇ データ入力時に利用する JavaScript プログラミング</p> <ul style="list-style-type: none"> <li>・ HTML タグによるページ作成</li> <li>・ JavaScript の Event 処理</li> <li>・ DOM 利用による動的ページ作成</li> </ul> <p>◇ 想像するほど難しいプログラミングではありませんが、Ajax を利用時のプログラミングに工夫が必要です。</p>	<ul style="list-style-type: none"> <li>● Event 特有な処理の理解</li> <li>● DOM 利用の難しさの理解</li> <li>● 動的なページ切り替えの理解</li> <li>● Ajax の使いどころの理解</li> </ul>
<p>○ Web サーバー側におけるプログラミング</p> <ul style="list-style-type: none"> <li>・ CGI プログラミング技術</li> <li>・ モジュールの利用技術</li> <li>・ XPath の利用技術</li> <li>・ モジュールプログラミング技術</li> </ul>	<p>◇ Perl を利用した CGI プログラミング</p> <ul style="list-style-type: none"> <li>・ データ形式の設計</li> <li>・ データの取得方法</li> <li>・ 利用可能なモジュールの選択</li> <li>・ モジュールを簡単に切り替え可能とするプログラミング</li> </ul>	<ul style="list-style-type: none"> <li>● 利用可能なモジュール選択に手間がかかることがあります。</li> <li>● 利用に合うデータの形式を考えさせます。</li> <li>● データの保存の工夫と保存の際の問題点の洗い出しと解決方法の模索</li> </ul>
<p>○ Web クライアント側と Web サーバー側との連携技術</p> <ul style="list-style-type: none"> <li>・ Ajax プログラミング技術</li> </ul>	<p>◇ この処理に適したデータ形式を設計し、利用してみて、再度データ形式の設計に戻る形での開発、工夫。</p> <p>◇ Ajax を利用した非同期通信による連携</p>	<ul style="list-style-type: none"> <li>● 比較的新しい技術で、親しみやすい。</li> <li>● 最終的には、システムが出来上がったところで、新たに Linux®をパソコンにインストールし、Web サーバーWeb クライアント</li> </ul>

養成する能力 (知識、技能・技術)	課題制作・開発のポイント	訓練（指導）ポイント
		の設定、システムの再構築 までを目標としました。

### <所見>

本課題は Web クライアント側の技術、Web サーバー側の技術およびこれらの連携技術が要求されます。まず、クライアント側の Web ページおよびプログラムの作成から始めましたが、比較的初めの段階から取り組みやすく、興味もてる内容です。DOM 利用によるページ作成あたりまでをまず目的としました。次に、Web サーバー側とその連携に取り組みましたが、プログラムの量的には多くはないのですが、Perl モジュールの利用とその抽象化までを目的としました。

本課題における一連の取り組みを通じて、作る楽しさを感じていることがうかがえました。その後の実習作業においても、より積極的に議論、制作していく姿が見られました。学生はこの経験を踏まえより実践的な能力が養成されたと考えます。

### 課題に関する問い合わせ先

**施設名** : 中国職業能力開発大学校  
**住 所** : 〒710-0251  
 岡山県倉敷市玉島長尾 1242-1  
**電話番号** : 086-526-0321  
**施設 Web アドレス** : <http://www.ehdo.go.jp/okayama/pc/index.html>