

取得したWEB上の情報をAPIとして 提供するシステムの検討

M0113025 新井良太

研究の背景

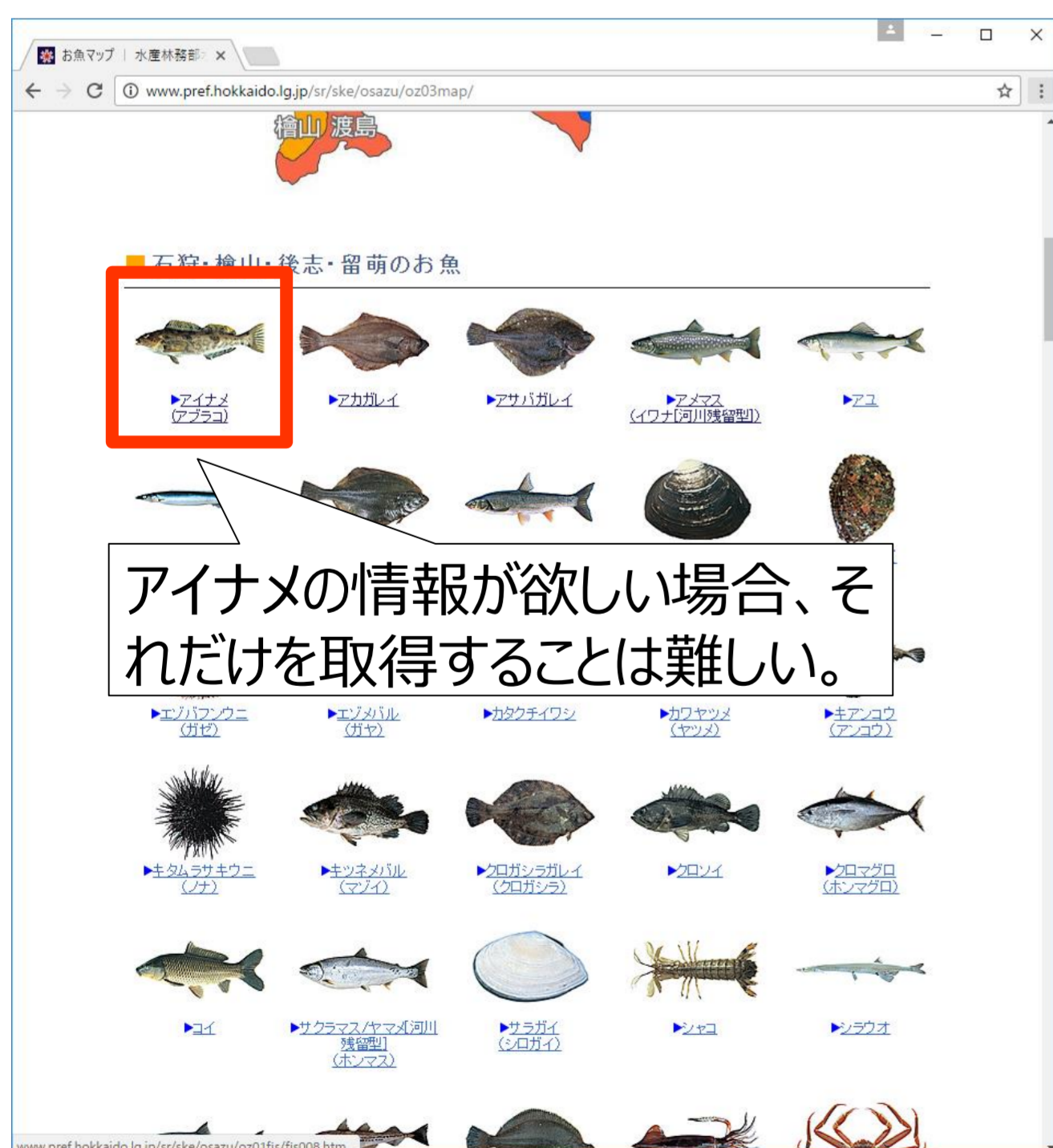
現在、WEB上には様々な人や企業が作成したデータが存在する。それらは閲覧することは簡単だが、自分で使用することは難しい。そこで、WEB上の欲しいデータをAPIとして提供できないかと考えた。

研究の目的

ユーザーが必要なデータが存在するサイトからデータを抜き出し、データの利用を容易に行うことができるAPIを作成、提供するシステムを開発する。また、そのシステムによってどこまで正確にデータを取得できるのかを検討する。

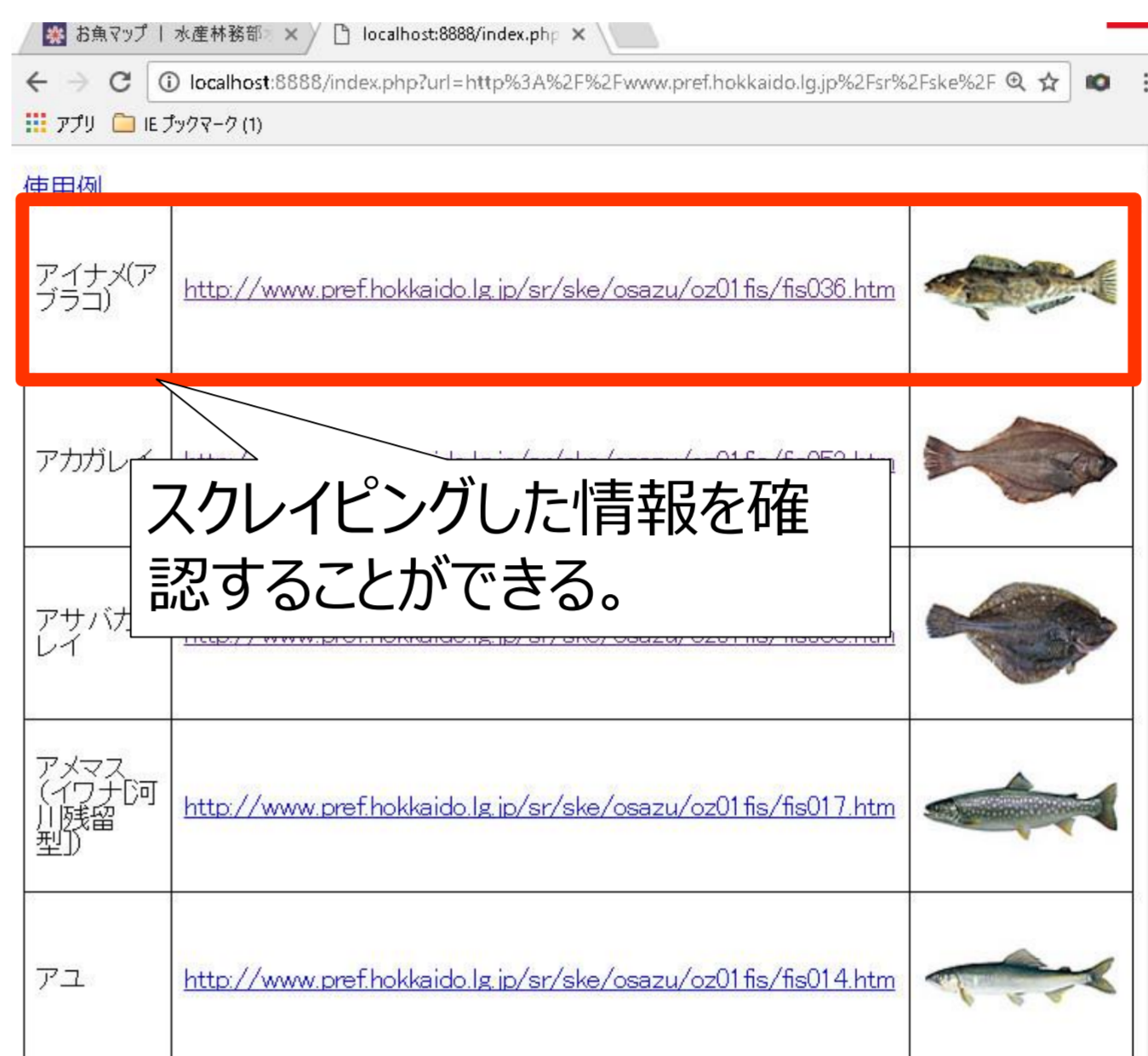
システムの概要

現在、WEBサイト上の一部のデータを取り出すことが難しいので、データを使えるようにするためのAPIを提供するシステムを作成する。

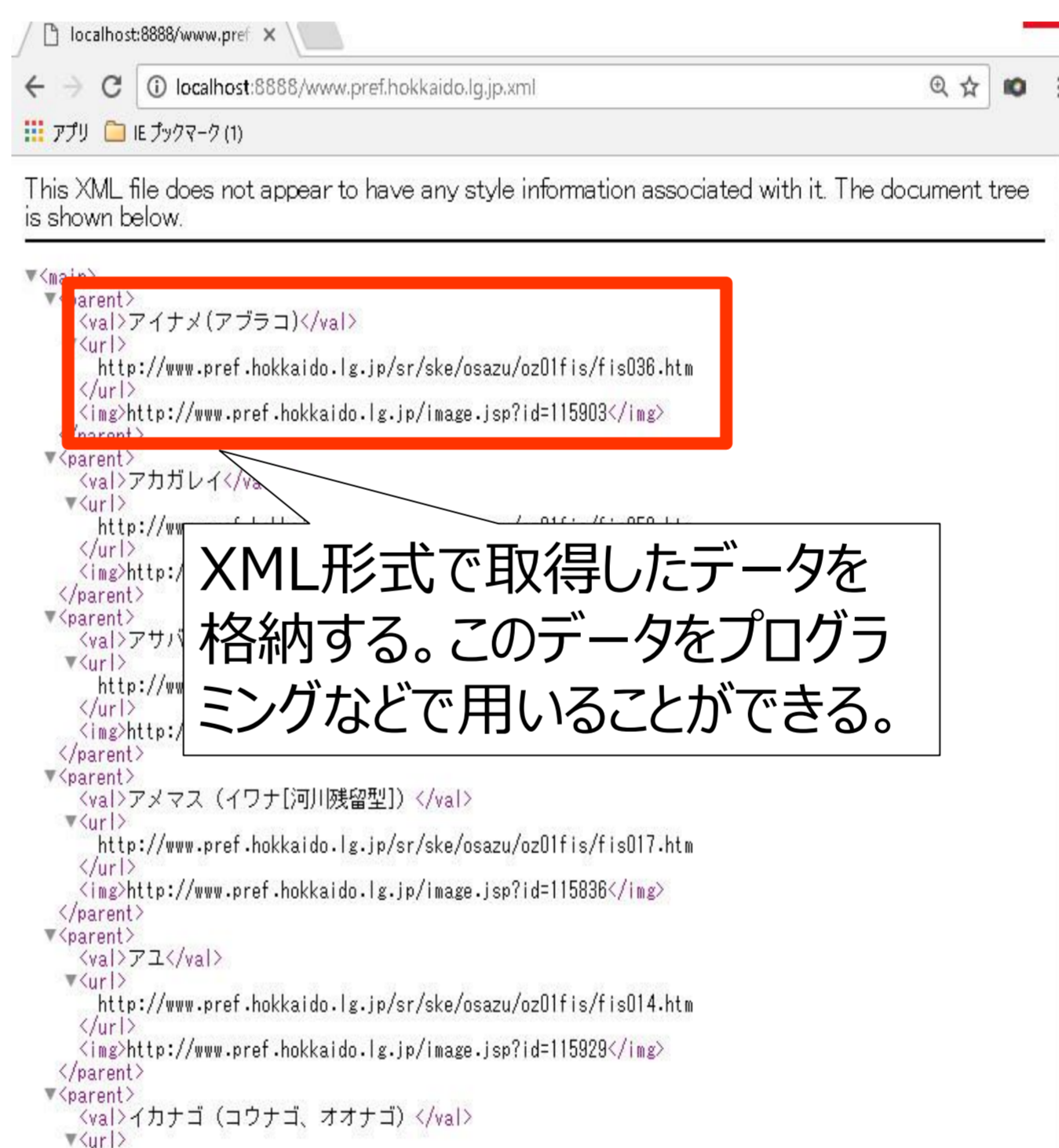


最初に必要なデータがあるWEBサイトのURLを入力し、そのURLを元にスクレイピングを行う。

魚の情報が必要な場合このシステムを利用することで、魚ごとの情報を取得することができる。



入力されたデータを元にWEBサイトから必要なデータだけを抜き出す。例では魚名とURLと画像を取得している。



XMLファイルが作成されるので、URLからほしい情報を取得し、自分で作成しているプログラムに使用することができる。現在、表示することは可能だが、検索や加工して使用することが困難なので、今後検索、加工する機能作成していく必要がある。

現状と今後の課題

現在はWEB上にある一部のデータのみ取得し、利用することが可能である。そのため、今後はPDFなど他のデータにも対応できるようにしていきたいと考えている。また、現時点では精度が悪く、ユーザーにXPathを入力してもらわなければならない欠点もあるので、その点も改善したいと考えている。

出典:北海道お魚マップ
URL:<http://www.pref.hokkaido.lg.jp/sr/ske/osazu/oz03map/>

