

コミュニティ連動型情報集約サイトの構築

廣瀬研究室 4年 渡部雄生

平成 28 年度 1 月 19 日

概要

現在、東北公益文科大学では、学生生活に有用な情報が紙媒体や電子媒体を通して多数提供されている。しかし、多くの情報の中から必要な情報を見つけることは時間と労力を要する。そこで、必要な情報だけを取り出して再構築した媒体を設ける。構築した媒体によって、学生が効率的に情報を集める手助けをし、時間の節約や労力の削減に繋げる。

また、本学は地域の活性化に重点を置いている。そこで、地域に根ざした情報を集め、その中から必要な情報を取り出す。本学の学生にとって有用な情報や、地域の企業や組織が提供する情報を学生の視点で収集し、Web サイトとして情報を公開する。この Web サイトによって、学生と地域住民の交流を効率化し、地域住民と学生の連携の強化を図る。

本論では、Web スクレイピングとデータベース、MTA を利用したプログラムを作成する。また、作成したプログラムを組み合わせることで、上記 2 点の要件を満たすことのできるコンテンツを持つ、地域連動型情報集約サイトの構築を目指す。(439 字)

目次

第 1 章	はじめに	4
1.1	背景	4
1.2	提案	4
1.3	キュレーションサイトサービスの先行事例と効果	6
1.3.1	グノシー	6
1.3.2	antenna	7
1.4	キュレーションサイトによる弊害	8
第 2 章	利用手段および開発環境	9
2.1	利用手段	9
2.1.1	Web スクレイピング	9
2.1.2	データベース	10
2.1.3	MTA(Message Transfer Agent)	11
2.2	開発環境	11
2.2.1	Java	11
2.2.2	Eclipse	12
2.2.3	Apache Tomcat	12
2.2.4	MySQL	12
2.2.5	qmail	12
第 3 章	構築する Web サイトの概要	13
3.1	各手段の役割	13
3.2	コンテンツの内容の決定	13
3.2.1	サイトレイアウト	17
第 4 章	システム解説	18
4.1	HTTP 通信に基づく Web ページの表示	18
4.2	サーブレットを用いた Web アプリケーション	19
4.3	システムの概観	20
4.4	各プログラムの実行の流れ	21
4.4.1	Web スクレイピングプログラム	21
4.4.2	データベースを利用したプログラム	21
4.4.3	MTA を利用したプログラム	22
第 5 章	構築した Web サイト	23
第 6 章	結論	28
6.1	結論	28
6.2	今後の課題	28
6.2.1	Web スクレイピングプログラムについて	28
6.2.2	データベースを利用したプログラムについて	28
6.2.3	MTA を利用したプログラムについて	28
6.2.4	全体を通して	29

第1章 はじめに

1.1 背景

東北公益文科大学 (以下、本学) は地域に根ざした活動や学習を推進し、学生たちはそれに倣うことで地域の活性化を図っている。例えば、学生たちが講義の一環として地元の商店街を訪れて交流を図ったり、共同スペースを設けて地域住民と共に活動したりといった工夫がなされている。また、本学は地 (知) の拠点整備事業 [1] を採択しており、地域コミュニティの中核的存在としての大学の機能強化を図っている。しかし、本学を取り巻く環境として、少子高齢化が進む地域であることや、都市部に比べて人的資源や物的資源が少ないという問題がある。例えば、本学の学生が地元の商店街を訪問する際には、市が運営する巡回バスを利用する。しかし、この巡回バスの1日あたりの本数は9便であり、気軽に商店街に行けるものではない。また、本学の属する山形県は、人口に対する高齢者の割合が高い地域である。内閣府の平成28年度版高齢社会白書 [2] によると、山形県の高齢化率は約29%であり、全国平均の26.7%を上回っている。年齢に伴って情報資源の利用率も下がる傾向にあり、平成27年度通信利用動向調査 [3] によると、「年齢階層別インターネット利用状況」は、平成27年度末において、20~29歳は約99%であるのに対し、60~69歳は約68%という結果であった。情報化による利益の拡大が進む昨今、高齢者の割合が高い山形県において現状が続くことは望ましくない。以上の問題を受け、限られた人的資源や物的資源を、豊富な情報資源を有効活用して最大限活用する方法を考えた。また、本論では若者の情報資源の利用率が高齢者より高いという結果から、学生を対象として情報資源の利用手段を考案する。本学の学生の日常生活に有用な情報を提供すること、地域に関する情報を提供することで、学生が時間を有効に活用できるように、地域住民との交流を促進する手助けになると考えた。

1.2 提案

前述の課題を解決するためのアプローチとして、本学の学生の利用を柱とした情報集約サイトの構築を提案する。個人がインターネットから情報を収集するための方法として、Google や Yahoo! などの検索エンジンを利用したり、ブックマークによって見たい Web ページに誘導したりする機能が考えられる。また、近年では Facebook や twitter といった SNS も発達し、Web サイトとユーザの動的な結びつきが強まっている。しかし、日ごとに増え続ける情報の中から効率的に情報を集めることは未だ多くの時間と労力を要する。この状況を改善するために、情報集約サイトとして、キュレーションの考え方を基にした Web サイトの構築が挙げられる。キュレーションとはインターネット上の情報を収集してまとめること、あるいは収集した情報を分類し、繋ぎ合わせることで新しい価値を持たせ、共有することをいう。キュレーションを行う人をキュレータといい、博物館の管理者や図書館の館長を意味する言葉が語源となっている。インターネット上のあらゆるデータを、キュレータ独自の視点で取捨選択し、整理されたものをキュレーションメディアという。

また、同じく情報を収集するための技術としてRSS[4]がある。RSSは、RSSリーダを用いて、提供されたRSSフィードを読み取ることで、Webサイトの更新情報を集めることができるものである。一定の書式に基づいて要約して記述されているため、すべてのWebサイトの更新情報を統一された形式で確認したり、広告や余計な情報を排したりすることができる。RSSもキュレーションも、必要な情報のみを集めてくるという点において有用な技術であるが、RSSは登録したサイト数に応じて情報が届けられるので、登録サイトが増えるほど集めてくる情報が肥大化する。利用者が有益と考える情報は増えるが、肥大化した情報の中からもう一度有益な情報を探すということにも繋がるため、手に入る情報量は長所にも短所にもなる。一方、キュレーションは的確に情報を取得する方法として有用であるが、情報の偏向に陥りやすい。以下の表1.1は、RSSとキュレーションを比較したものである。本論では、本学を取り巻く環境に存在する多様な情報の中から必要と考えられる情報を再構築したものをユーザに提供することを考えているため、主にキュレーションの考え方を基にWebサイトの構築に臨む。

表 1.1: RSS とキュレーションの比較

	RSS	キュレーション
概要	ソフトウェア	設計思想
ユーザ	受動的	能動的
情報量	登録したサイト数に比例	収集する情報に比例
特徴	<ul style="list-style-type: none">登録したWebサイトの更新情報を、Webサイトにアクセスしなくても確認できる統一された型式で読むことができる	<ul style="list-style-type: none">必要な情報のみ取得できる集めた情報を加工できる

1.3 キュレーションサイトサービスの先行事例と効果

以下の先行事例をはじめとして、キュレーションサイトは経済活性化や地域活性化に貢献している。

1.3.1 グノシー

「グノシー」[5]は、「株式会社 Gunosy」が提供している情報キュレーションサービスである(図1.1)。ユーザに対して多くの情報を提供する傍ら、ユーザ動向の分析によって、各ユーザに合わせた最適な情報を提供することを目的としている。コンテンツ内容として、エンターテインメントやスポーツ、政治、技術情報など、ユーザの生活に役立つ情報を、日々更新されるニュースという形で提供している。グノシー利用者数は年々増加傾向にある。2013年にサービスを開始してから2016年までに1500万ユーザを獲得しており、他のキュレーションサイトの先頭に立つメディアとなっている。「Gunosy」は「グノシー」を広告事業と連動する形で収益を上げており、2015年から2016年にかけて約46億円の売上を計上している[6]。



図 1.1: グノシーの Web ページ

1.3.2 antenna

「antenna」[7] は、「株式会社 antenna」が提供するキュレーションサービスである。ユーザをテレビや新聞、音楽など様々なメディアとつなげるための情報を提供することを目的としており、ユーザ主体の情報とメディア主体の情報を合わせて発信している。同サービス内の「ヨナナにっぽん」(図 1.2) は、各都道府県ごとの情報を提供している。コンテンツ内容として、食べ物や観光、地域独自の有名なイベントなどについての情報がある。

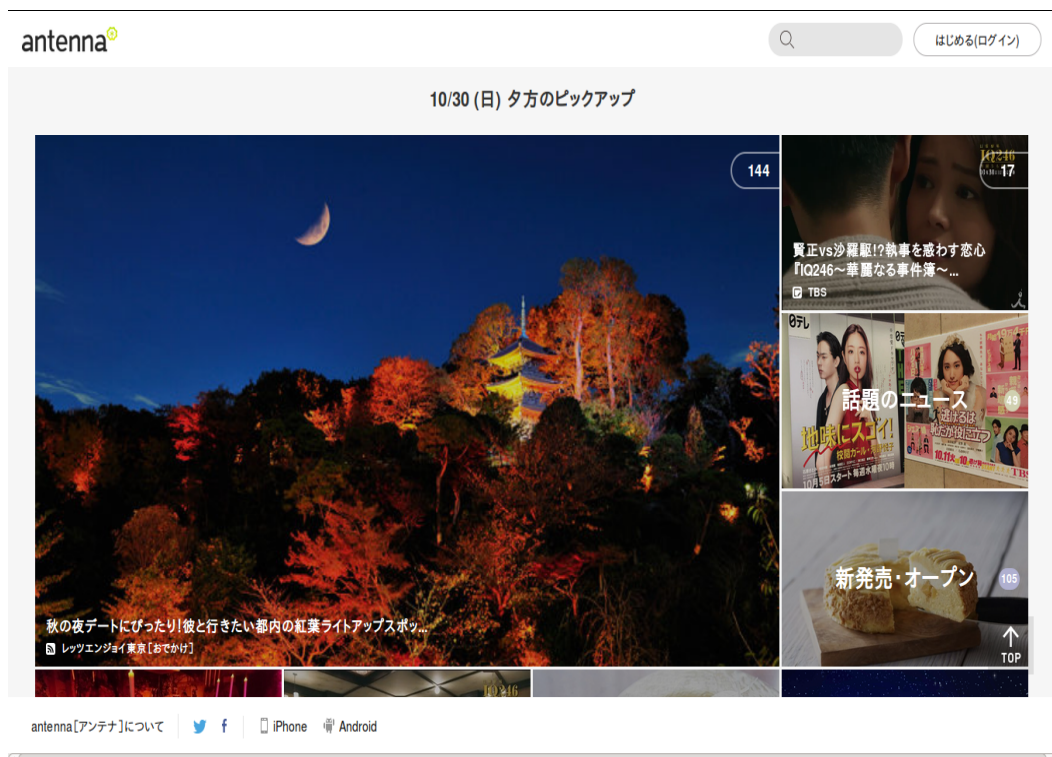


図 1.2: ヨナナにっぽんの山形県の Web ページ

1.4 キュレーションサイトによる弊害

キュレーションサイトによる問題も存在する。2016年11月29日に、株式会社DeNAが運営するキュレーションサイト「Welq」において、医療関係者から該当サイトの扱う情報の信憑性が疑われたために、Webサイト内の全記事を非公開にする事件[8]が起こった。この事件をきっかけに今後キュレーションサイトによる逆風は強くなると考えられる。このように、キュレーションメディアが引き起こす問題がいくつか考えられる。1つ目は、知的所有権の問題である。知的所有権には、著作権、特許権、意匠権、商標権などがある。例えば、キュレータが、Webスクレイピングの対象とするWebページの画面デザインやイラスト、文章などを、著作者の許可無く利用することは著作権を侵害することに繋がるおそれがある。2つ目は、Webスクレイピング技術が、対象となるメディアの衰退を招く可能性があると考えられる。キュレーションサイトは元のWebサイトに直接アクセスしなくても必要な情報だけを手に行うことができるため、元のWebサイトへのアクセス数が減る。例えばSEO対策¹と称して意図的に検索エンジンの検索結果の上位にキュレーションサイトを配置すると、元のWebサイトが相対的に検索結果の下位になる。検索結果の下位になったWebサイトはアクセス数が減ってしまう。元のWebサイトへのアクセス数が減ることで運用の継続をしなくなれば、貴重な情報リソースが失われることになる。3つ目は、情報の信憑性である。Webスクレイピングを利用したコンテンツは、情報源がWebページであることから、情報の正確性に難がある。対象サイトの所有者との連携に加え、第三者機関の監視の下で運用がなされるのであれば、情報の正確性は確保できる。また、ユーザに信頼し続けてもらえるように、できるだけ正確な情報をユーザの役に立つ形で届けるように努力が必要である。4つ目として、広告技術との連携がある。日本の企業で運用されているキュレーションサイトの多くは広告媒体との繋がりが、利益を上げているケースがある。これは、他人の情報を利用し収益を上げることと同義であるため、倫理観やモラルの点で人びとの反発を招く。キュレーションは社会に有用な技術であると同時に、利用方法によっては社会に損害を与えるものであることを理解する必要がある。本論では、プログラム作成の際に、情報発信元が社会的な信用と責任を持つと考えられる企業や組織の運用するWebサイトに対してWebスクレイピングを利用することで、情報の正確性の確保に努める。また、ビジネスモデルとしてでなく、地域の活性化のための公益の精神²に基づいて活動することを念頭に置いてWebサイトの構築に臨む。

¹検索エンジンによる検索結果の中で、対象とするWebページがより速い段階で表示されることを目的とした一連の取り組みのことである。

²ここでいう公益の精神は、個人の利益追求よりも他者との利益共有を優先する意思のことを指す。

第2章 利用手段および開発環境

本論の Web サイト構築にあたり、利用した手段と開発環境について記述する。

2.1 利用手段

本論で構築する Web サイトのコンテンツを作成するにあたって、以下の技術を利用する。本論では、電子媒体と紙媒体に分けてアプローチを変えてプログラムを作成する。

2.1.1 Web スクレイピング

Web スクレイピング [9] は、Web ページから必要な情報だけを取得したり、取得したものに加工を施すことでより役立つものへと改善したりすることを目的とした技術である。この技術を利用することによって、人間がブラウザを開いて、欲しい情報だけを多くの情報の中から探す作業をより効率的に改善することができる。Web スクレイピングのメリットとして以下の点が挙げられる。

- ファイルの効率的なダウンロード
- 欲しい情報の抽出
- 取得した情報を元にした加工
- 情報の集約
- 複数の Web サイトの組み合わせ
- 複雑な作業の自動化

ただし、この技術を利用する際には、知的所有権を侵害しない、Web ページへの過度なアクセスを避ける、利用規約を守って行うなどの注意が必要である。Web スクレイピングの本論における役割は、電子媒体を扱う仕組みのうち、Web ページとして公開されているデータを基に新規に Web ページを生成するものである。概観は、図 2.1 のようになる。

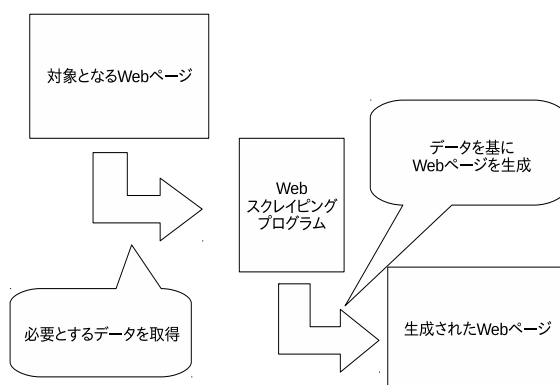


図 2.1: Web スクレイピングプログラムの概観

2.1.2 データベース

データベースとは、複数の主体でデータの共有や利用を行ったり、用途に応じて加工や再利用がしやすいように一定の形式で作成や管理がなされたりしたデータの集合のことを指す。データベースの本論における役割は、電子媒体を扱う仕組みのうち、Web ページ以外で表現されているデータを基に新規に Web ページを生成するものである。概観は図 2.2 のようになる。

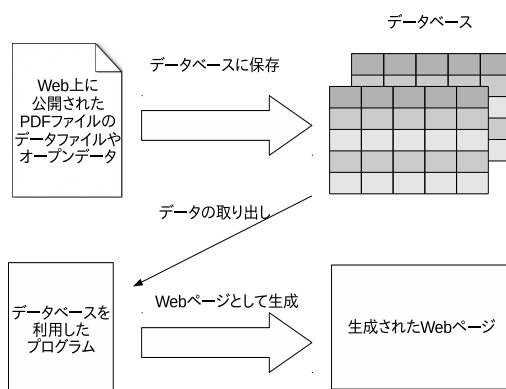


図 2.2: データベースを利用したプログラムの概観

2.1.3 MTA(Message Transfer Agent)

MTA とは、ネットワーク上でメールの転送や配送を行うソフトウェアである。MTA の主な役割は、メールを受け取って次の適切な配信先を決定することである。実際に配信先へのメールの送信や転送はメールサーバの機能を担っている MDA(Mail Delivery Agent) が行う。MTA と MDA をまとめて MTA と呼ぶこともある。MTA の本論における役割は、紙媒体として提供されているデータを基に Web ページとして生成するものである。概観は図 2.3 のようになる。

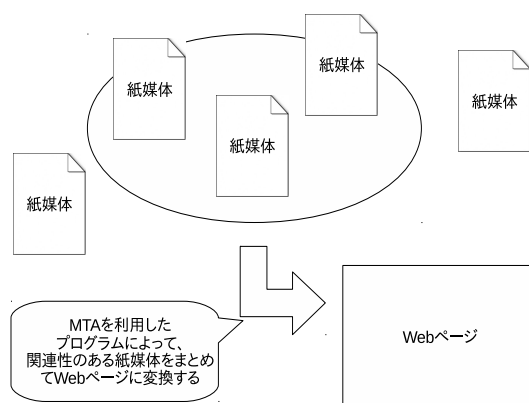


図 2.3: MTA を利用したプログラムの概観

2.2 開発環境

本論の Web サイト構築は以下の環境を基に行う。

2.2.1 Java

Java[10] はプログラミング言語の一種である。特徴として以下の点が挙げられる。

- オープンソースであり、誰でも利用できる
- 機種依存性が少なく、JVM をインストールしていればどのようなプラットフォーム上でも動作する
- 強い静的型付けにより、開発者が誤りを犯しにくい
- ガーベジコレクションによって、自動的に不要なメモリを解放できる
- オブジェクト指向言語であり、プログラムの簡潔性や可読性、グループ化がしやすい
- マルチスレッドが利用でき、並行プログラミングに向いている

本論では「Java SE Development Kit version 8」を利用する。

2.2.2 Eclipse

Eclipse[11]とは、IBM社によって開発された統合開発環境 (IDE) の1つである。本論では「Eclipse version 4.6 neon」を利用する。

2.2.3 Apache Tomcat

Apache Tomcat[12]とは、サーブレットプログラムを動かすために必要なソフトウェアである。サーブレットプログラムについては後述する。TomcatのみでもWebサーバとしての役割を持つが、WebサーバとしてTomcatよりも優れているApacheと連動させることで、お互いの長所を活かしてサーブレットプログラムを動かすことが出来るようになっている。本論では「Apache Tomcat version 8.5.8」を利用する。

2.2.4 MySQL

MySQL[13]とは、リレーショナルデータベースの一種である。リレーショナルデータベースは、2次元の表形式で表現されることや、表に対して、データの定義や操作をするSQLを用いて、データの検索や追加や変更、削除などを行うことができるという特徴がある。

データ分析が可能なトランザクション処理¹を実現している

本論では「MySQL Ver 14.14 Distrib 5.6.31」を利用する。

2.2.5 qmail

qmail[14]は、Daniel J. Bernsteinが開発したMTAである。従来、MTAの標準となっていたsendmailに比べ、設定の容易さやセキュリティの強固さ、パフォーマンスにおいて優位性を持つ。本論では学内に用意されているメールサーバを利用する。学内で利用されているqmailは「netqmail version 1.05」[15]である。

¹情報の一貫性を保つために、関連する複数の処理を連結し、処理が成功した場合のみ処理を完了する。

第3章 構築する Web サイトの概要

本論で構築する Web サイトの利用対象者は、本学の学生とする。また、筆者自身がキュレータとなり、複数の手段を用いて情報を収集し、再構成したものを Web ページとして構築する。

3.1 各手段の役割

Web スクレイピング、データベース、MTA のそれぞれが本論の Web サイトのコンテンツを構成するにあたり、情報収集の手段としてもつ長所に応じて役割を分担する。以下の表 3.1 は本論で想定する各手段の役割を示す。

表 3.1: 各手段の役割

Web スクレイピング	Web ページとしてデータが記述されている場合
データベース	Web 上に公開されているが、Web ページとして記述されていない場合
MTA	Web 上にデータが公開されていない場合

3.2 コンテンツの内容の決定

掲載するコンテンツの内容を提示する。コンテンツ決定に際し、他キュレーションサイトが提供している情報を収集した。以下は、収集したコンテンツの概要をまとめたものである。

他サイトが提供している情報の概観

- 芸能に関する情報
- 映画に関する情報
- 音楽に関する情報
- テレビ情報
- アニメ情報
- ゲームに関する情報
- 書籍情報
- スポーツに関する情報
- ペットに関する情報
- 経済に関する情報
- 政治に関する情報
- 海外の情勢に関する情報
- 恋愛に関する情報
- 美容に関する情報
- 衣服に関する情報
- 科学技術に関する情報
- IT 技術に関する情報
- 新商品の情報
- 新店舗に関する情報
- 今週のイベント情報
- 料理のレシピ情報
- グルメ情報
- 外食店に関する情報
- 地域ごとの食材や食品に関する情報
- 旅行に関する情報
- インテリアや雑貨に関する情報
- 健康に関する情報
- ライフスタイルに関する情報
- ビジネススキルに関する情報

また、本学の学生 10 人と意見を交わし、在学中に必要と感じた情報を挙げてもらった。以下は提案の一覧である。

在学生の提案

- 学内の時間割情報
- 学内の補講情報
- 学内のイベントに関する情報
- 学外のイベントに関する情報
- 学生や講師へのインタビュー記事
- 市内のバスの運行情報
- 交通情報
- 酒田市の天気
- 求人情報
- 就職活動に関する情報
- 資格試験に関する情報
- 不動産情報
- スーパーマーケットの割引情報

本論では対象とする地域を酒田市域及び庄内地方に限定するため、以上の結果のうち、対象とする地域が庄内地域の外であるものを除外した。また、研究を地域の活性化に繋がたいと考えているため、酒田市内の企業や店舗に関する情報を優先して取り扱う。加えて、学生の学内での活動及び地域連携活動に有用であると感じた提案を採用する。これらをまとめた内容は以下の通りになる。

選択したコンテンツ

- 書籍情報
- スポーツに関する情報
- 衣服に関する情報
- 外食店に関する情報
- 新商品の情報
- 新店舗に関する情報
- 健康に関する情報
- ライフスタイルに関する情報
- 学内の時間割情報
- 学内の補講情報
- 学内のイベントに関する情報
- 学外のイベントに関する情報
- 学生や講師へのインタビュー記事
- 市内のバスの運行情報
- 交通情報
- 酒田市の天気
- 求人情報
- 就職活動に関する情報
- 資格試験に関する情報
- 不動産情報
- スーパーマーケットの割引情報

以上を踏まえ、コンテンツとして提供する内容を決定した。以下は、本論の Web サイトにおいて実際に組み込んだコンテンツの内容の一覧である。

本論で作成するコンテンツ

- 酒田市の天気予報（当日、翌日、週間）
- 庄内地方の道路情報
- 酒田市内巡回バスの時刻表（かんぼの宿発、酒田駅正面口発）
- 本学自動時間割表
- 学内補講情報
- 学内のイベント情報
- 学外のイベント情報
- 酒田市内の開店中のラーメン店情報
- 賃貸物件情報

3.2.1 サイトレイアウト

Web サイトのレイアウトは「左袖ナビゲーション型」[16]を参考にする。このレイアウトパターンは画面左にナビゲーションが常に表示されるため操作性が高い。また、ナビゲーションが増加しても縦に伸びるため、多数の項目を第 1 階層に表示する際に便利な構造となっている。レイアウトの詳細は図 3.1 の通りである。

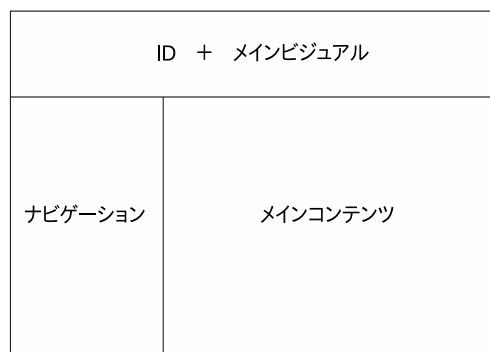


図 3.1: Web サイトレイアウト

第4章 システム解説

今回構築する Web サイトのシステムについて記述する。

4.1 HTTP 通信に基づく Web ページの表示

ブラウザから Web サーバにアクセスをすることをリクエストといい、Web サーバからブラウザへの応答をレスポンスという。リクエストの内容は図 4.1 に、レスポンスの内容は図 4.2 に示す。ユーザがリクエストを Web サーバに送り、もしリクエストが成功した場合、Web サーバは処理結果を返す。

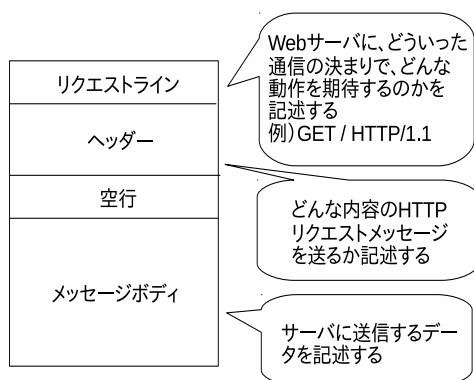


図 4.1: HTTP リクエストの内容

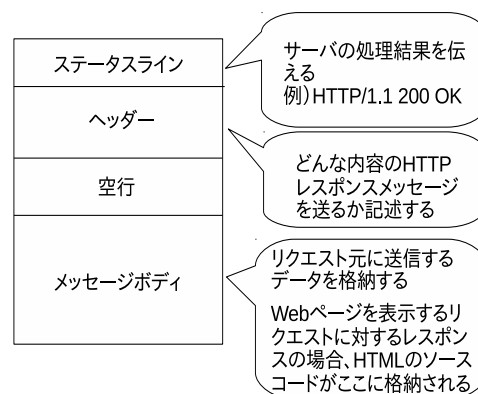


図 4.2: HTTP レスポンスの内容

なお、Web サーバは HTTP を用いてブラウザと通信することはできるが、プログラムを実行する機能は持っていないため、実行する機能を追加する必要がある。プログラムを動かす機能を持つ Web サーバのことを Web アプリケーションサーバという。Web アプリケーションサーバにプログラムを配置することで Web アプリケーションが実行可能となる。

4.2 サブレットを用いた Web アプリケーション

サブレットとは、サーバ上で Web ページなどを動的に生成したり、データ処理を行ったりするために、Java で作成されたプログラム及びその仕様である。サブレットの文法に従って作成したサブレットクラスは、サブレットコンテナを持つ Web アプリケーションサーバに配置することで実行できるようになる。サブレットは以下の 3 つのルールに従ってクラス定義をする。

- javax.servlet.http.HttpServlet クラスを継承する¹
- doGet(doPost) メソッドをオーバーライドする²
- サブレット関係のクラスをインポートする³

上記のルールに則って作成したプログラムは、図 4.3 の流れで処理を実行し、Web ページを表示する。

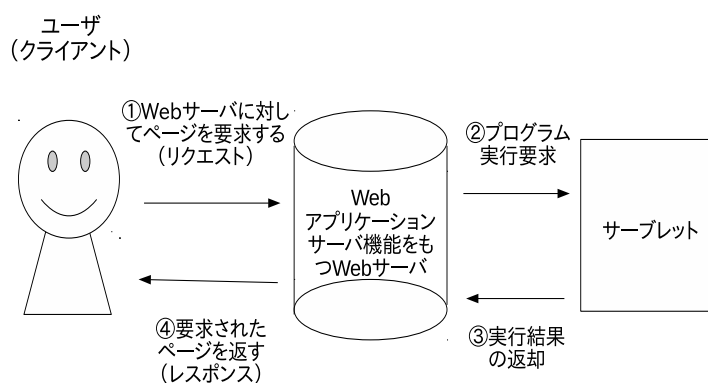


図 4.3: Web アプリケーションの実行の流れ

¹HttpServlet クラスは、サブレットクラスの元となるクラスで、これをサブクラスに引き継ぐことで、引き継いだサブクラスはサーバ上で動的な Web ページやデータ処理を行うことができる。

²HttpServlet クラスは抽象クラスであるため、このクラスを継承したサブレットは、doGet() メソッドや doPost() メソッドなどの HTTP メソッドで、HTTP 通信においてどんな処理をしたいかを上書きし、処理を実行できる。

³javax.servlet クラスと javax.servlet.http クラスが該当し、これらのクラスはサブレットプログラムの作成に関わるものがまとめられている。これらのクラスをインポートすることにより、サブレットプログラムを容易に作成できる。

4.3 システムの概観

本論のキュレーションサイトの利用の流れは以下の通りである。

1. ユーザはメインページにアクセスする
2. アクセスを受けた後、Web アプリケーションサーバは各プログラムに実行要求を出す
3. 要求を受けたプログラムは各々処理を実行し、結果として Web ページを Web アプリケーションサーバに返す
4. 処理結果を受け取った Web アプリケーションサーバはコンテンツ部分に処理結果を表示する
5. ユーザは、処理結果を受けたメインページで情報を確認することができる

今回作成する Web サイトの概観は図 4.4 のようになる。

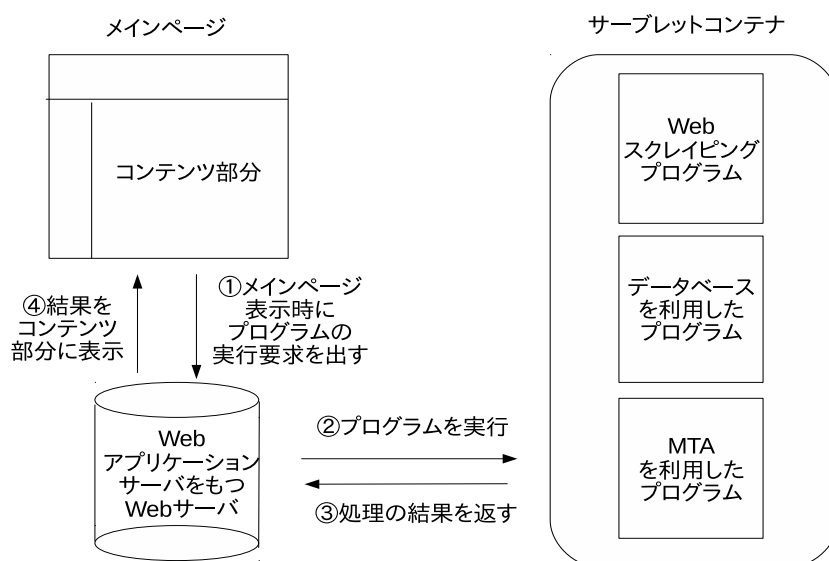


図 4.4: システムの実行の流れ

4.4 各プログラムの実行の流れ

コンテンツ部分を構成するプログラムの実行の流れを記載する。

4.4.1 Web スクレイピングプログラム

本論の Web スクレイピングプログラムの要点は以下の通りである。

1. 対象とする Web サーバにリクエストを発行する
2. 発行したリクエストに基づいたレスポンスを受け取る
3. 受け取ったレスポンスのメッセージボディ部分から HTML ソースを取り出す
4. 取り出した HTML ソースを基にパターンマッチングを行う
5. パターンマッチングで取り出した要素を基に Web ページを生成する

Web スクレイピングプログラムの実行の流れは以下の図 4.5 のようになる。

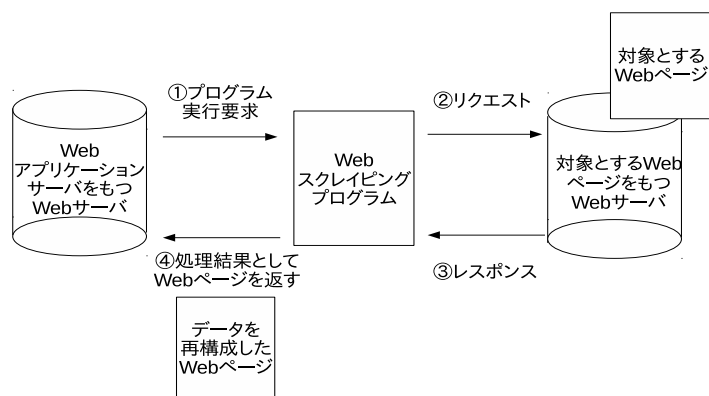


図 4.5: Web スクレイピングプログラムの実行の流れ

4.4.2 データベースを利用したプログラム

本論のデータベースを利用したプログラムの要点は以下の通りである。

1. Web ページに公開されているデータを集める
2. 集めたデータをデータベースの中に保存する
3. 保存したデータに対して操作を行う
4. 操作の結果によるデータを基に Web ページとして編成する

データベースを利用したプログラムの実行の流れは以下の図 4.6 のようになる。

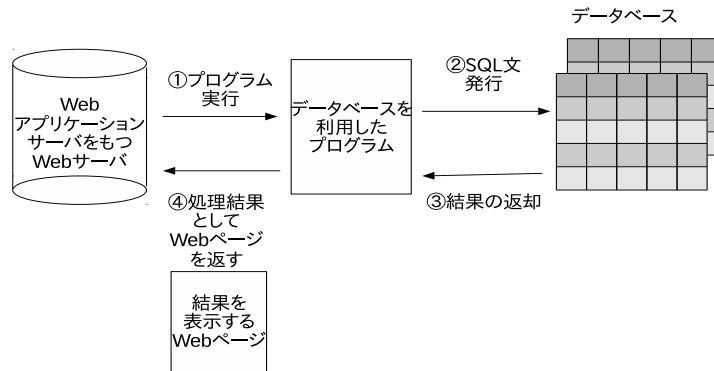


図 4.6: データベースを利用したプログラムの実行の流れ

4.4.3 MTA を利用したプログラム

本論の MTA を利用したプログラムの要点は、以下の通りである。

1. Web 上に公開されていないデータを画像ファイル収集する
2. 収集した画像ファイルをメールに添付し、特定のメールアドレスに送信する
3. 送信された画像ファイルを公開するディレクトリに保存する
4. 保存された画像ファイルを基に Web ページを生成する

MTA を利用したプログラムの実行の流れは以下の図 4.7 のようになる。

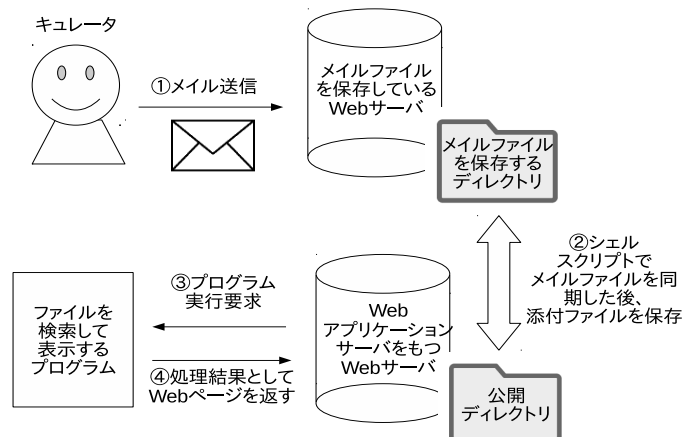


図 4.7: MTA を利用したプログラムの実行の流れ

第5章 構築したWebサイト

上記のプログラム群をまとめ、レイアウトデザインに則って構築した Web ページは以下の図 5.1 のようになった。レイアウトの枠組みは HTML と CSS によって記述した。また、ナビゲーション部分にはボタン型のリンクを配置し、メインコンテンツ内の指定の部分を表示できるようにした。



図 5.1: 構築したキュレーションサイト

現時点でメインコンテンツとして提示しているプログラムを以下に提示する。

賃貸物件情報 本学には、大学周辺にある賃貸物件を利用して大学に通っている学生がいる。この学生たちを対象として、より効率的に物件を探索できる情報サイトを作成することを試みた。本論では、酒田市で賃貸物件情報を取り扱う不動産会社のうち、東洋開発の提供する Web ページのデータを基に Web ページの再構築を試みた。実行結果は上記の通りである。今後、他社との比較を行うことができる Web ページとすることを考慮している。

時間割情報 本学では、学期の初めに時間割表が紙媒体で配布される。また、PDF ファイルとして本学の Web ページにも公開されている。しかし、その内容は非常に細かく、目的の講義を検索するのに時間がかかる。また、紙媒体であるため取り扱いに手間がかかる。そこで、時間に合わせて自動的に現在の講義だけを表示することで、目的の内容にすばやく辿り着くことができるようなプログラムを作成した。実行結果は上記の通りである。電子媒体であるため資源の節約にもなり、学生は Web サイトを確認するだけで、目的の講義に辿り着くことができる。

補講情報 本学の講義の補講情報は、学生課前の掲示板に紙媒体で掲載される。学生はその内容を確認するために学生課に向かう必要があり、時間と労力の大幅なロスになっている。そこで、紙媒体を電子媒体に変換することで、簡単に情報を獲得でき、学生は時間を有効に使うことができると考えた。プログラムの実行結果は上記の通りである。

天気予報 酒田市の天気予報を、Yahoo!天気予報の Web ページに Web スクレイピングをかけ、新規に Web ページを生成する。本論のプログラムの実行時は 2016 年 12 月 7 日 18 時であり、対象とする Web ページの URL は「<http://weather.yahoo.co.jp/weather/jp/6/3530.html>」である。このプログラムは、当日と翌日、翌々日から始まる一週間分の天気予報を提示する部分だけを取り出して表示するプログラムである。酒田市は、季節による天候の変化が激しい地域であるため、Web ページとして提示した。実行結果は図 5.2 のようになる。



図 5.2: 酒田市の天気予報を表示する Web ページ

道路情報 酒田国道交通事務所は、庄内地域の道路に設置された Web カメラの情報を提供している。この情報は画像として提示されており、5分おきに更新される。また、更新された画像は、更新以前の URL に上書きする形で提示されている。プログラムでは、Web スクレイピングを用い、この URL を基に新規に Web ページを生成する。本論でのプログラムの実行時は 2016 年 12 月 9 日 18 時 30 分であり、対象とする URL は「<http://www.thr.mlit.go.jp/sakata/road/view/index.html>」である。図 5.3 は、実行結果の一部である。



図 5.3: 庄内地域の道路交通情報を提供する Web ページ

バスの運行情報 酒田市内では市が運営する巡回バスのるんるんバスが運行している。また、酒田市はこのバスの時刻表を PDF ファイルとして提供している。このプログラムでは、乗車場と降車場を選択して検索を実行することで、プログラムの実行時間にもっとも近い乗車時刻と、そのバスの到着時刻を提示するものである。今回はかんぼの宿から出発するるんるんバスの時刻表を基にしたプログラムを提示する。検索ページは図 5.4 に、実行結果は図 5.5 に記載する。実行時は 2016 年 12 月 7 日 19 時であり、対象とする Web ページの URL は「http://www.city.sakata.lg.jp/sangyo/kotsu/runrunbas/fukushibus-jikokuhyo.files/Univ.TimeTable_2810.pdf」である。

●かんぼの宿発

乗車場: ▼
 降車場: ▼

出発:大学前=>20:17:00

到着:酒田駅正面口=>20:32:00

[戻る](#)

図 5.4: 検索ページ

図 5.5: 実行結果

学内のイベント情報 本学の Web ページでは、大学が管轄するイベントの情報を一覧表として提供している。このプログラムでは、本学のイベント情報のうち、プログラムの実行時期に最も近い順に 10 件のイベント情報を提示するものである。以下の図 5.6 は、2016 年 12 月 7 日 20 時にプログラムを実行した結果であり、対象とする Web ページの URL は「<http://www.koeki-u.ac.jp/event/>」である。

東北公益文科大学 最新イベント10件

2016.12.2	第2回グローバルセミナー「詩を通して世界の文化に触れるタベ」を開催しました
2016.12.2	[12月17日(土)]観光・まちづくりフォーラム『山形県、庄内地方におけるアジア・インバウンド戦略について考える』を開催します
2016.11.29	[12月6日(火)]大学院履修交流協定締結記念シンポジウム「大学院教育と地域貢献 一関学と公益大の取組みとこれから」を開催します
2016.11.22	[12月17日(土)]冬のオープンキャンパスについて
2016.11.15	[11月25日(金)]グローバルセミナー「詩を通して世界の文化に触れるタベ」を開催します。
2016.11.9	イングリッシュカフェを開催しました
2016.10.31	2016年度東北公益文科大学大学院公開講座「政策研究特論b(地域政策に関する学際的研究)」を開講します
2016.10.26	[11月3日(木・祝)]東北公益文科大学ニュージーランド研究所15周年記念「小松隆二先生記念講演会」を開催します
2016.10.26	[11月3日(木・祝)]ローカル・イノベーション・セミナーVol.6「人と地域の「らしさ」を生かす働き方とは？」を開催します

図 5.6: 本学のイベント情報を提供する Web ページ

酒田市のイベント情報 酒田市の Web ページでは、酒田市内で行われる今後のイベント情報が掲載されている。学内のイベント情報と同様に、プログラムの実行時期に最も近い順に情報を収集するものであるが、提示するイベント情報は最も近い月を基準としている。プログラムの実行結果は、図 5.6 のようになる。実行時は 2016 年 12 月 7 日 20 時であり、対象とする Web ページの URL は「<http://www.city.sakata.lg.jp/event/index.html>」である。

酒田市 今月のイベント一覧

エコキャンドル作り参加者募集
今月のイベント・講座情報
成人式のお知らせ
森山直太郎15thアニバーサリーツアー「絶対、大丈夫」チケットについて
ファミリー健康体力測定会を開催します
酒田市立資料館
元気みらいワークショップ～思いを形に～
使用済み小型家電の無料回収のお知らせ
庄内空港利用振興協議会
平成29年度 認可保育所・認定こども園の入所・入園申込みについて
琢成・宮野浦・亀ヶ崎・飛鳥地区合同津波避難訓練および緊急速報メール訓練配信(庄内全域)を実施します

図 5.7: 酒田市のイベント情報を提供する Web ページ

酒田市内のラーメン店の開店情報 酒田市内には複数のラーメン店があり、本学の学生も利用している。学生が望む時間帯に開店している店舗の情報を得ることで円滑に訪問を行うことができる

と考える。このプログラムでは、Google が提供している店舗情報の中から時間に関わるデータをデータベースに保存し、プログラムの実行時に合わせて開店しているラーメン店を一覧表示するプログラムである。本論作成時には GoogleAPI¹を利用してプログラムを作成する予定であったが、筆者の環境では実現できなかったため、代替案としてデータベースを利用するに至った。実行時は2016年12月7日20時であり、実行結果は図5.8のようになる。

開店中のラーメン店

店名	開始	終了
酒田ラーメン花鳥風月	11:00:00	20:00:00
新月	11:00:00	19:00:00
三日月軒東中の口店	11:00:00	19:00:00
ニンニクげんこつラーメン花月酒田みずほ通り店	11:00:00	00:00:00
麺工房さらしな	17:30:00	19:30:00
新旬屋麺中町店	18:00:00	00:00:00
つけ麺道癒庵	11:00:00	21:00:00
三日月軒駅東支店	10:30:00	19:00:00
福の家	17:00:00	20:00:00
三日月軒中町店	11:00:00	18:30:00
味好本店	17:00:00	20:30:00
ラーメン尾浦	11:00:00	19:00:00

図 5.8: 酒田市のラーメン店の開店状況を提示する Web ページ

なお、本論のプログラムに利用したものと提供元を表 5.1 に記載する。

表 5.1: 本論で利用したものおよび提供元一覧

提供元	利用したもの
東洋開発	http://www.toyokaihatsu.com/cgi/toyo2/toyo.cgi
東北公益文科大学	2016 年度時間割【2013 年度以前入学】の PDF ファイル
Yahoo!Japan	http://weather.yahoo.co.jp/weather/jp/6/3530.html
酒田河川国道事務所	http://www.thr.mlit.go.jp/sakata/road/view/index.html
酒田市役所	酒田駅大学線の PDF ファイル
東北公益文科大学	http://www.koeki-u.ac.jp/event/
酒田市役所	http://www.city.sakata.lg.jp/event/index.html
Google	https://www.google.co.jp/maps/search/%E9%85%92%E7%94%B0%E5%B8%82+%E3%83%A9%E3%83%BC%E3%83%A1%E3%83%B3%E5%B1%8B/@38.9018624,139.8222046,14z/data=!3m1!4b1%

¹Google が提供するプラットフォームやサービスのリソースへアクセスするための API である。

第6章 結論

今回構築した Web サイトについての結論と、課題として見つかった点や今後の予定を、各利用手段と全体を俯瞰したものを分けて考える。

6.1 結論

本システムは、紙媒体、電子媒体問わず収集した情報を集約してユーザに提供する方法として有効である。また、即時性の求められる情報を提供することに対しても効果的なシステムである。

6.2 今後の課題

6.2.1 Web スクレイピングプログラムについて

今回の手法を用いて Web スクレイピングを行う際には、対象となる Web ページの構造に大きく依存するという欠点があることが分かった。元となる Web サイトの構造が変化する度に、プログラムを更新していく必要がある。また、JavaScript で記述された内容は、HttpClient がサーバサイドでアクセスを行うため、Web スクレイピングをかけることができなかった。Selenium WebDriver ツールによってブラウザに対して Web スクレイピングをかけることができるため、今後取り組んでいきたい。

6.2.2 データベースを利用したプログラムについて

今回提示したプログラムのデータは簡素なものであったが、今後はデータベースの構造を駆使したプログラムの作成にも着手したい。また、Web スクレイピングプログラムと連携させ、Web スクレイピングが扱えないときでも過去の内容のうち最も最新の情報をデータベースを用いて表示できるようにすると、より有用性の高い情報提供の手段にできると思われる。

6.2.3 MTA を利用したプログラムについて

本論では、Web サイトの作成者が撮った写真を表示する手法で行ったため、表示される写真にはばらつきがあったり、見えづらかったりといった人的手法による欠陥が見受けられた。また、文字認識を行うことができるツールの中で、tesseract OCR¹を用いて文字認識を行い、HTML 内で文字を表示させる予定であったが、ツールの精度が不正確であり、別の文字に化けることがあるため、実現に至らなかった。今後は機械学習により文字認識の精度を上げることも取り組んでいきたい。

¹オープンソースの光学文字認識ソフトウェアである。日本語にも対応している。

6.2.4 全体を通して

現段階では Web サイトの公開を行っていないが、もし公開を行う場合は、利用元となった Web ページの所有者に認可をもらう必要がある。今後はキュレーションの考え方をビジネス利用すること、ひいてはキュレーションの考え方自体に企業や組織が難色を示すことが想定されるため、承諾を得るためには、正しい理解の共有と信頼関係を構築しなくてはならない。また、今回作成した内容が学生の需要に則したものであるのか本論作成時点では判別できないため、アンケートや他の学生との更なる交流を以て運用していきたい。加えて、情報の信頼性の確保のためにも、データの発信元が正しい情報を提供しているのを見極める必要がある。UI についても、データの増加や端末の変化に対応していきたい。

今後の展望として、まずは現行コンテンツの機能を増やす方針で進めることを考えている。例えば、るんるんバスの自動時刻表示プログラムを改良し、指定した降車場に指定した時間で着くために乗る必要のある乗車便を確認できるようにしたい。他にも、ファッションのコーディネートを表示し、そのコーディネートに必要な物品が購入できる場所を提示するプログラムを作成することを視野に入れている。また、Android や iOS に対応したアプリケーションとし、SEO 対策に頼らず、必要な人だけが利用できるようにすること、広告などを使わず非営利な活動の下進めていくべきであること、元媒体の隆盛に貢献できるマーケティングを学生や地域住民と連携して考えていくことで、キュレーションサイトの弊害を抑えることを考える必要がある。本論で挙げた RSS の利用法としては、元サイトの更新に応じて、元サイトに影響を与えすぎないように発信していくこともできると考える。加えて、初期の段階では Web スクレイピングのみで Web ページの構築をすることを考えていたが、必要なデータの収集ができない場合なども出てくるのが分かった。そのため、今後は API やオープンデータを利用して学生向けの料理のレシピを提供したり、商店街の割引情報を提供したりしていく。

上記の結果から、今後も酒田市内の住民と学生の交流に、キュレーションサイトが貢献できるように工夫を重ねる必要がある。そして、キュレータの意志が公益の精神を追究するものであることを望む。

参考文献

- [1] 東北公益文科大学. "地(知)の拠点整備事業(大学COC事業)の概要 — 東北公益文科大学 地(知)の拠点整備事業(大学COC事業) 公式ホームページ" .
<http://coc.koeki-u.ac.jp/tinokyoten>,(参照 2016-12-06).
- [2] 内閣府. "第1節 高齢化の状況 | 平成28年版高齢社会白書(概要版) - 内閣府" .
<http://www8.cao.go.jp/kourei/whitepaper/w-2016/html/gaiyou/s1.1.1.html> ,(参照 2016-11-16).
- [3] 総務省. "平成27年通信利用動向調査の結果" .
<http://www.soumu.go.jp/johotsusintokei/statistics/data/160722.1.pdf>,(参照 2016-11-16).
- [4] RSS Listing. "RSS リスティング - ブログの更新一覧をホームページに表示" .
<http://berss.com/>,(参照 2016-11-16).
- [5] 株式会社 Gunosy. "グノシー 雑談力がつくニュースアプリ" .
<https://gunosy.com/>,(参照 2016-10-30).
- [6] 株式会社 Gunosy. "経営成績 - IR 情報 | 株式会社 Gunosy (グノシー) | 情報を世界中の人々に最適な形で届ける" .
<http://gunosy.co.jp/ir/highlight/>,(参照 2016-12-05).
- [7] 株式会社グライダーアソシエイツ. "キュレーションマガジン antenna [アンテナ]" .
<https://antenna.jp/>,(参照 2016-10-30).
- [8] 時事通信社. "D e N A、医療情報サイト非公開に = 記事信ぴょう性に疑い: 時事ドットコム" .
<http://www.jiji.com/jc/article?k=2016112900977&g=soc>,(参照 2016-11-29).
- [9] 川崎 克巳. Java スパイダーツールサンプル&クックブック. 秀和システム,2005,377p.
- [10] Oracle Corporation and/or its affiliates. "Java テクノロジーについて" .
https://java.com/ja/about/whatis_java.jsp ,(参照 2016-11-16).
- [11] The Eclipse Foundation. "Eclipse - The Eclipse Foundation open source community website." .
<https://eclipse.org/> ,(参照 2016-11-16).
- [12] The Apache Software Foundation. "Apache Tomcat®; - Welcome!" .
<http://tomcat.apache.org/index.html> ,(参照 2016-11-16).
- [13] Oracle Corporation and/or its affiliates. "MySQL" .
<https://www-jp.mysql.com/> ,(参照 2016-11-16).
- [14] D. J. Bernstein . "qmail: the Internet's MTA of choice" .
<http://cr.yip.to/qmail.html> ,(参照 2017-02-10).
- [15] Russell Nelson. "netqmail" .
<http://www.qmail.org/netqmail/> ,(参照 2017-02-10).
- [16] 新田光敏. Web デザインの基本ルール.SB クリエイティブ,2009,21p.