

個人での利用を指向した インターネット情報資源の収集・変換手法

中野伸吾^{†1} 終和佑^{†1}

現在、各市町村や自治体などで地域情報資源の開発と確保による、デジタルコンテンツの構築が行われている。また、計算機の低価格化に加え、様々な携帯型の端末が普及したことにより、ユーザはいつでも地域情報資源などにアクセスできるようになった。しかしその一方で、情報を得るためのコストにかかわる問題や、情報の埋没と信憑性に関わる問題、閲覧環境の相違による情報資源の互換性と不適応に関する問題が生じている現状である。本稿では、これらを解決するために、ユーザの利用環境とユーザが求めるオーソリティ・シチュエーションに合わせた情報の収集・変換・利用するシステムを提案・構築する。

Personal Policy Based Internet Content Harvesting and Converting System

SHINGO NAKANO^{†1} and WASUKE HIIRAGI^{†1}

Nowadays due to the popularity of various portable devices and decrease in prices of computers, users are able to access the information required at any time. As a result of which, problems with the diversification of information services, diversification of information provider's service and terminal performance differences are arisen. In this paper, the authors are proposing a system that can harvest the internet content which can also be modified and convert into different format that meets the user's policy, authority, environment and required information ranking in order to solve the arisen problems.

1. はじめに

1.1 地域情報資源の活用目的と背景

近年、様々な地域情報資源の開発と確保を行い、各市区町村でそれらを使ってデジタルコンテンツを構築する動きがある。加えて、観光に構築したデジタルコンテンツを利用す

る取り組みも多く見られる。このようなデジタルコンテンツは、各種観光ガイドや観光地に設置された案内板に利用されるだけでなく、Webを用いて発信されてきている。Webを用いることで、地域情報資源は発信者から一方的に送られてくるものではなく、利用者が自主的に収集、利用する情報になっている。さらに、昨今では、様々な携帯型の端末が普及し、ユーザはいつでも求める地域の情報を、観光中にいつでも取得できるようになっており、今後はより地域情報資源が利用されることが予想される。

しかし、Webを用いて発信される情報は各自治体や観光協会、その他関連企業や個人ごとに作成されており、その粒度や信憑性もまちまちである。自治体によってそれらをまとめたリンク集も作られているが、日々誕生・更新・消滅を繰り返すWebページに追いついていないのが現状である。さらに、ユーザが持ち歩く携帯端末の性能差も大きく、端末によっては閲覧の出来ないWebページや、情報も多く存在している。

本研究では、このようなデジタルコンテンツを発展させ、信頼性もつ情報基盤とするために、各地の地域情報資源と、地域情報資源を含むブログなどをリンクさせたデジタルアーカイブと、その閲覧システムを開発する必要があると考えている。現在、本研究では稚内市や港湾が所蔵する写真や印刷物のデジタル化を進めている。そして、稚内市について記述したWebページやブログを、それらデジタル化した資料を元に収集し、地域情報資源とWebアーカイブが有機的にリンクした総合的デジタルアーカイブの構築を進めている。

しかし、現在進めている総合的地域情報資源デジタルアーカイブには、旅行者にそのデータを利用させる際の手法が大きな問題として存在している。そこで本論文では、旅行者が旅行中に利用する端末、および各地に設置された案内板に、総合的地域情報資源デジタルアーカイブのコンテンツを表示させるための手法について検討する。また、それに併せてインターネット上に点在している様々な情報資源とデジタル化した地域情報資源を利用者の指向に合わせて変換し、表示するシステムを構築する。

1.2 情報資源の収集・提供に関する現状

現在、Webを利用して収集できる情報資源の量と種類は非常に多く、データ形式も様々なものである。そこで本論文では、地域情報に関するWebページを収集・提供するために以下の三つの問題が存在していると考えた。

- (1) 情報を得るためのコストに関わる問題
- (2) 情報の埋没と信憑性の保証に関わる問題
- (3) 閲覧環境の相違による情報資源の互換性に関わる問題

(1)の「情報を得るためのコストに関わる問題」は、インターネット上に、各種新聞社、ニュースサイト、電子掲示板、個人のブログ、SNSなど、ユーザに情報を提供するための様々

^{†1} 稚内北星学園大学
Wakkanai Hokusei Gakuen University

なサービスが存在することが原因である。このようなサービスが発展してきた背景としては、インターネットの商業的利用の発展と、それらサービスが提供する情報に対する検索ツールの高度化があげられる。そして、高度なツールの普及により、情報の中心的利用者が従来の専門家から、いわゆる一般ユーザーに移ってきた。

しかし、インターネット上にあるサービスは、サービスごとに操作方法や利用手順、ユーザインターフェース（以下 UI）が異なる場合や、サービスが指定するデータフォーマットを意識する必要がある場合もある。そのため、サービスの利用方法に習熟していないユーザーには敷居が高いサービスも増えてきている。結果として、ユーザーはサービスを使いこなすために、その習熟に多くのコストをかけることが求められるようになった。もし、習熟しないままそのサービスを利用すると、ユーザーにとって必要な情報がサービスに存在していても、収集することができない場合が想定される。

(2)の問題点である「情報の埋没と信憑性の保証に関わる問題」は、1.の問題点とも関連するが、情報資源を発信できる多様なサービスが登場・普及したことから起きている。従来、Webで情報を発信するためには、知識と技術が必要であったが、SNSやブログの登場により、一般のユーザーが簡単に情報を送受信することが可能になった。その結果、発信される情報は増加したが、同時にユーザーにとって有用な情報が埋もれてしまう可能性が増した。また、それに伴い情報資源の信憑性も保証しにくくなった。情報資源の信憑性が補償できなくなると、サービスやそれを紹介した自治体に対するユーザーの信頼が薄れる可能性がある。

また、地方自治体の持つ地域情報資源に目を向けると、作りっぱなしのものや、作成途中で放置され、配信しないまましまわれているものが散在している。このような地域情報資源も、観光者などの利用者にとっては有用な情報である可能性が高いが、陳腐化している可能性もあるため、これらの情報の再利用方法検討も急務である。

(3)の問題点である「閲覧環境の相違による情報資源の互換性に関わる問題」は、端末の対応しているデータフォーマットや性能、大きさなどの違いにより、情報資源の表示に差異が生じることから起きる。日本各地で作成された地域情報資源や、インターネット上に存在する情報資源は、様々な体裁・デザイン・フォーマットで作成されている。また、近年の電子書籍リーダーやタブレット機の普及(図1)[1]に伴い、近い将来には様々なタブレット型・スマートフォン型計算機が登場するだろうといわれている [2]。

このような高度な情報端末であっても、対応しているフォーマットが異なることがあるため、表示出来ない情報資源も存在する。この問題を解決するために、フォーマットの標準化プロジェクト [3] が複数進行しているが、従来の携帯電話のようにディスプレイのサイズ等の機械的な問題により表示が困難となる情報資源が存在したり、情報資源を発信する側が特定のデータフォーマットにしか対応していない場合も存在する [4]。

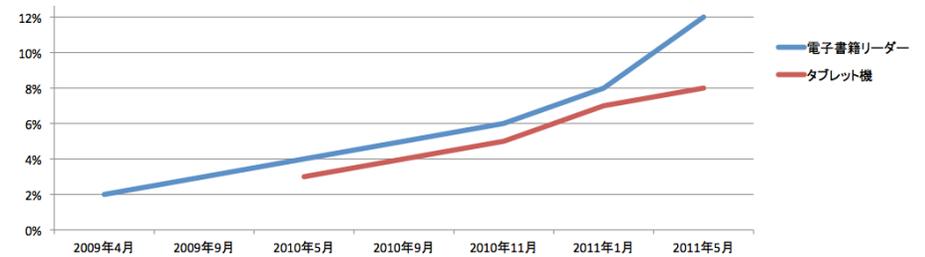


図1 米・成人男性が所有している電子書籍リーダーとタブレット機の普及率推移

2. 情報資源活用の現状

2.1 情報資源の収集・利用に関する現状とその問題点

前章で取り上げた問題を解決し、ユーザーにとって効率の良い情報収集を行うために、現在では以下のような方法がとられている。

- (1) 検索エンジンを使った情報資源の収集と利用
- (2) サービスが選別した情報資源の利用
- (3) SNSを利用した情報資源の探索と利用
- (4) ソフトウェアによる機械的な情報資源の探索と利用

(1)の方法は、Google[5]やYahoo[6]などによるWEBをベースとした検索エンジンによる情報収集である。福島[7]によれば「第二世代の検索エンジンを決定づける技術的なポイントは、インターネットを巡回して自動収集するクロウラを開発したことである。この技術によってウェブページの大規模収集が可能になった。」といわれている。しかし、質や分野を整理できないという欠点 [8,9] もあり、HTMLなどに記述された文字列をもとに検索を行うため、検索エンジンでは収集しにくいDeep Web[10]からの情報収集が困難である。また、各市町村や自治体によって管理されているために、一般ユーザーがアクセスしにくい地域情報資源からの情報収集も困難である。

(2)は、新聞やテレビのニュース番組など、情報閲覧者以外の第三者によって選別された情報を元に行う情報収集である。ユーザーが自ら情報収集を行う場合、長島ら [11]によれば「検索エンジンで検索は容易になるかもしれない。しかし、集めた情報は更に咀嚼、理解、検討といった情報処理の過程を経なければ役立つ情報にならない。そして情報処理にはやはり時間がかかる。つまり、時間の機会費用がサーチコストの大きな部分を構成しているかも

しれないのである。」とある。そのため、第三者によって選別された情報を利用することで、その手間を省くことができるが、新聞やテレビなどが扱う情報は時事性のあるものが多いため、過去の情報や、より専門的な情報を収集するには適していない。

(3)は、SNSを用いて関心のある情報、もしくはそれに関連する情報を収集するものである。日本でよく利用されているmixi[12]などのSNSは、自分の友人や知人、もしくは関心のあるトピックを指向する機能[13]も存在し、専門家を含む多様なユーザも情報発信を行っている。そのため、より膨大な情報群の中から専門的な情報を収集することが可能であるが、サービスの種類と膨大な情報群ゆえに、サービスを使いこなすことができず、必要な情報を収集することが困難な場合がある。

その情報は玉石混淆で多量なため、どのようなサービスを利用し、どんな方法で情報資源を選択すればよいのか、ユーザに迷いを与えてしまう。

これは佐藤[14]によると「情報過多」と呼ばれ、誰でもが手軽にあらゆる種類の情報を入手できるような環境が整い始めたことや、情報の蓄積や処理にかかる金銭的なコストが大幅に低下していることなどが要因とされる。

(4)は、よく利用するWEBサイトやブログなどを、RSSリーダーやブラウザのブックマーク機能などを用いて管理・収集する方法である。この方法は、ユーザが必要とする情報を提供しているサービスを自ら選択することができる。しかしこのような方法は、伊東ら[15]は「常に新鮮な情報を得るには複数のサイトを巡回しなければなりません。しかし、これは手間の掛かる作業です。」と述べている。そのため、ユーザ自身が登録したサービスをくまなく巡回する必要があり、そのコストは増加する。また、どのサービスを登録・利用すべきかの判断が困難な利用者や、短時間で情報資源を収集したいユーザにとっては、時間と手間をかける必要があるため、不向きな方法といえる。

ここに挙げたサービスは、近年のスマートフォン等の普及にともない、専用サイトや専用アプリケーションを使っても情報の提供を始めている。しかし、その対応状況は十分とはいえず、地方情報資源に関しては一般ユーザが用意にアクセスしにくい状態で保存されていたり、整理されていない情報資源も存在するため、利用できる情報資源の差がある。

2.2 アクセシビリティの向上に向けた活動

インターネット上に存在する情報資源や、各市町村や自治体によって作成された情報資源には、音声・映像・テキスト・代替テキスト等が混じった様々なコンテンツが遍在している。しかし、全てのユーザがあらゆる情報資源の言語を習得しているわけではない。また、そもそもユーザが利用する情報端末が情報資源に対応していない場合や、障がいがある人または高齢者への配慮が無い情報資源も存在する。それらの情報資源は、ユーザにとって知覚しにくい情報資源や、理解しにくい情報資源をということが出来るため、結果としてユーザの不満を生んでしまう。

この問題を解決し、インターネット上のウェブコンテンツをより広範な人に利用可能にするために、2008年12月にW3C Web Content Accessibility Guidelines 2.0(以下WCAG2.0)が定義されている。ここでは、WCAG2.0はWebページ制作者が守るべき原則をあげている[16,17]。この原則は、ウェブコンテンツがどんなユーザにとっても扱いやすく、管理者にとっても管理しやすいべきだと定めている。WCAG2.0はもともと、様々な障がいのある人に対して、コンテンツをよりアクセシブルにするために規定されたが、障がいがない人にとっても非常に有用である。なぜならば、上記の原則が整っていない情報資源は、閲覧の際に余計な手間がかかる可能性があるからだ。その結果、場合によってはユーザにとって情報閲覧にコストが生じ、情報資源の取得・閲覧時に不利益になる場合がある。

したがって、WCAG2.0のような統一された制約やガイドラインを設けることは、情報過多とサービスの乱立における混乱を軽減することにも繋がる。

2.3 個人での利用を指向した情報収集に残された課題

インターネット上に存在する情報資源や、各市町村や自治体によって作成された情報資源を収集する方法として、様々な方法が行われているが、前述した方法やサービスではいくつもの課題が残されている。本研究では、課題を大きく4つにわけ、その解決手法を考察する。

- (1) ユーザのポリシーに応じた情報収集・提供手法の開発
- (2) ユーザから独立した組織による情報資源の選別手法の開発
- (3) 有料情報資源の収集手法の開発

(1)の課題は、2.1で述べた「検索エンジンを使った情報資源の収集と利用」と「サービスが選別した情報資源の利用」に関するものである。ユーザが自らWebページを閲覧する場合と異なり、自動的に情報を収集する場合は、前もってユーザにどのような情報を収集するか記述してもらう必要がある。それによってユーザの求める情報資源がより鮮明になり、情報資源を収集する際の条件となる。

また、閲覧環境によって情報資源が閲覧出来ないことを防ぐため、キーワードだけでなく、ユーザが利用する情報端末に関する情報も記述してもらうことが理想である。そうすることにより、システム側がユーザが利用する情報端末に合わせて情報資源の再変換が行いやすくなり、従来閲覧不可能であった情報資源の閲覧が可能となる。

(2)の課題は、2.1で述べた「SNSを利用した情報資源の探索と利用」と「ソフトウェアによる機械的な情報資源の探索と利用」に関するものである。ユーザが最低限の手間で情報資源を収集し、無駄なコストはかけないことが望ましい。また、サービスの種類や利用方法に習熟していない人であっても目的の情報資源を集められることが望ましい。そのため、情報資源の選別は、ユーザのポリシーを考慮した上で、自動的に行われることが理想である。

しかし、このような横断的なデータ収集はユーザや各種商業的/非商業的サービスから独立した組織が行うことになる。そのため、ユーザのポリシーをどのように反映するか考察が必要である。

(3)の課題は、有料の情報資源についての扱いに関するものである。インターネット上には無料の情報資源の他に、有料の情報資源を取り扱うサービスも存在する。そのため、ユーザが求める情報資源が有料情報だった場合、サービスの種類や扱い方に詳しくない人は、その情報資源を収集・利用できない可能性がある。そのため、ユーザが求める情報資源が有料であっても、ユーザのポリシーによっては収集することが望ましい。

3. 利用者環境に合わせた情報の収集・提供

本研究では、前章までの議論を踏まえ、個人で利用することを目的として、情報を収集・蓄積し、利用者の環境に合わせて提供を行う手法を提案する。

この手法では、ユーザはまず、システムが情報資源の収集に必要なデータと、システムが再変換を行う際に必要なデータをポリシーとして入力する。(図2(1))

次に(1)で入力された「収集時に必要なデータ」を元に、クローラを用いて、インターネット上から必要な情報資源の収集を行う(図2(2))。ここで集めた情報資源は、後述するEPUB型のデータフォーマットに変換が行われる(図2(3))。

最後に、(1)で入力された「変換時に必要なデータ」を元に、各端末とユーザ環境に合わせた再変換を行う(図2(4))。最後に、システムはユーザが利用している情報端末に、再変換した情報資源を出力する(図2(5)(6))。

この手法の要求は以下の通りである。

- (1) 有料情報資源を含めた情報資源収集
- (2) 情報資源をユーザのポリシー順に取捨選択・並べ替え
- (3) 情報資源収集時のコスト削減と情報収集サービス選択の簡易化
- (4) 収集された情報の変換・提供と変換箇所の明示

これにより、無駄なコストをかけずに、ユーザの環境や求めるポリシーに見合った情報資源の収集が行うことができるようになる。各要求の解説を行う。

3.1 有料情報資源を含めた情報資源収集

無料情報資源と有料情報資源を含めた情報資源の収集を同時に行うことにより、ユーザが負担するコストを軽減することにつながり、ユーザが求める情報資源の収集が容易になる。それは現在、ソフトウェアによる有料と無料をあわせた情報資源の収集は難しく、手動による収集方法が主流だからである。

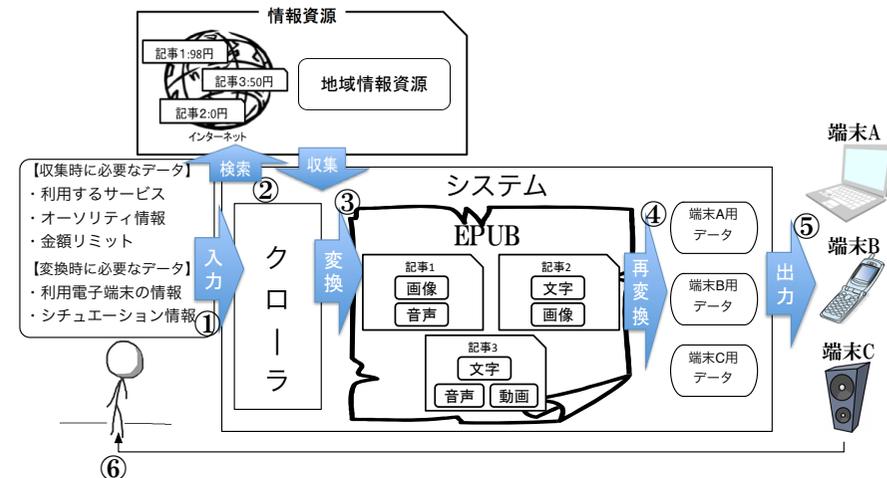


図2 有料と無料の情報資源を合わせた収集・管理手法

最近では Twitter や Facebook にある情報資源を電子新聞にまとめて発行するサービス [18] も現れてきてはいるが、限られた情報提供サービスのみの対応となっている。その理由は、情報資源を提供するサービスのシステム・UI などの違いが壁となっているからである。そのため、現時点では異なる情報提供サービスや、有料情報資源を組み合わせる情報資源を収集・管理するサービスは存在しない。

しかし以前より、iTunes[19]による音楽のマイクロコンテンツ化 [20] に始まり、書籍の電子化 [21] が叫ばれている。そのため有料情報資源は今後もますます市場に出回り、より普及していくことが予想される。

3.2 情報資源をユーザのポリシー順に取捨選択・並べ替え

ユーザが情報資源を提供するサービスを自ら選択し、そこから収集された情報資源に対してオーソリティを考慮したルールを与えることで、収集した情報資源群は、利用する際に目的の情報資源を探す手間を必要としない、理想の情報資源群となる。第三者によって取捨選択された情報資源を利用するのではなく、ユーザ自身がサービスを選択することによって、ユーザが求める情報資源を収集しやすくすることができる。また、オーソリティを考慮したルールを与えることによって、収集した情報の中から必要な情報資源を、無駄なコストをかけずに閲覧することができる。

ユーザが求める情報資源を収集しやすくし、コストを減らすことは、有料情報資源の普及に伴って予想される情報過多の拡大によって予想される、情報資源の埋没を防ぐことにつながる。

3.3 情報資源収集時のコスト削減と情報収集サービス選択の簡易化

ユーザが情報資源を収集するためのサービス選択を簡易化し、情報資源の収集自体は自動的に行うことによって、より専門的な情報資源の収集に対応しつつ、手動管理における情報資源収集のコストを減らすことにつながる。ユーザが情報資源収集のために情報提供サービスを巡回することは、サービスごとにシステム・UIなどが異なるため、場合によってはユーザが無駄なコスト負担してしまう。

ユーザに無駄なコストを負担させないためにも、利用したいサービスの選択はよりシンプルで扱いやすいことが望ましい。また、インターネット上にある情報資源は、老若男女、年齢を問わない幅広いユーザが利用する。そのため、サービスの種類や扱い方に詳しくない人でも、手軽に目的の情報資源を収集するためのサービスを選択でき、情報資源の収集は自動的に行われることが望ましい。

3.4 収集された情報の変換・提供と変換箇所の明示

収集した情報資源をユーザの環境に合わせて変換・提供することにより、1章で述べた「閲覧環境の相違による情報資源の互換性に関わる問題」の解決につながる。これにより、収集した情報資源のデータフォーマットが、ユーザが利用している情報端末で情報資源が利用できない場合でも、格納されている情報を文字や画像などのレベルで再変換することにより、利用が可能になる場合がある。

また、画像や動画コンテンツなどのリッチコンテンツが、利用ユーザが所持している情報端末に対応していなかったり、ユーザが表示を望まない場合、リッチコンテンツをは再変換された情報資源には存在しないことが望ましい。しかし、リッチコンテンツが抜き取られたことによって情報資源の意味が通じなくなってしまう事態を防ぐために、再変換された情報資源には、該当箇所に変換処理を施したことを示すための印を残す。

このように格納されている情報を分解し、再変換することで、ユーザが利用している情報端末の、画面サイズやスペックなどに適した情報資源の利用が期待できる。

4. MED システムの構築

4.1 インターネット上での情報収集における問題解決のために

本研究では、1章で取り上げた「情報を得るためのコスト問題」と「情報の埋没と信憑性の保証問題」を解決するために、オーソリティを考慮して情報資源を自動的に収集し、「閲覧環境の相違による情報資源の互換性と不適応問題」を解決するためにユーザの登録した状況(以下シチュエーション)別に利用したいサービスと検索用メタデータを管理できる仕組みを提案する。さらに、提案した手法を用いて Making My EPUB Document System (以下 MED システム) を構築した。(図4)

それにより、ユーザが求める情報をユーザが求めた順番で、自動的に情報資源を収集する

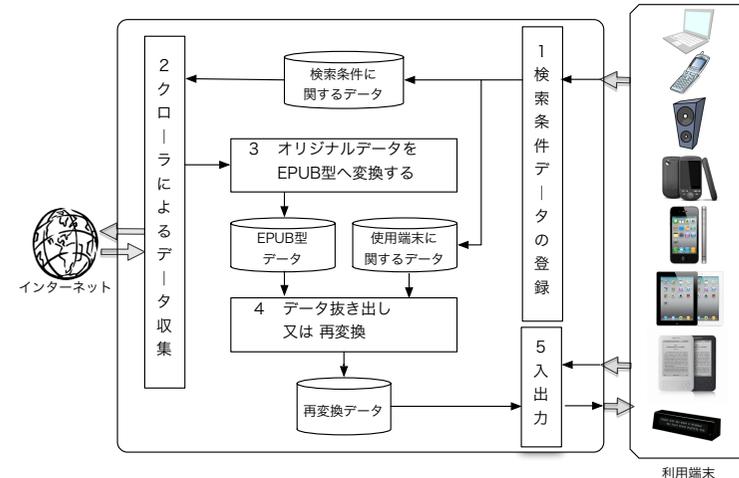


図3 MED システム全体図

ことができるようになる。それと同時に、ユーザは手間や時間というコストを払う必要がなく、情報端末や環境が原因で情報資源が閲覧できない事態が起きにくくなる。

それを実現させるため、MED システムでは、システム全体を4つの機能に分けている。その機能は以下の通りである。

- (1) Policy データの登録
- (2) クローラによるデータ収集
- (3) EPUB への変換と蓄積
- (4) データの変換と提供

4.2 Policy データの登録

MED システムには、検索条件に関するデータとして、利用ユーザに以下の Policy データを入力してもらう。

[情報資源収集の際に使用するデータ]

- 利用するサービス
- サービスごとの検索用キーワード
- サービスのオーソリティ
- サービスごとの金額リミット

[利用端末にあわせて再変換する際に使用するデータ]

- 利用情報端末に関するデータ
- シチュエーションに関するデータ

4.3 クローラによるデータ収集

前述の Policy データの中の「情報資源収集の際に使用するデータ」を用い、指定されたサービスごとに、インターネット上からの情報資源の収集を行う。システムに登録されているサービスは、サービスごとに利用方法が異なるため、それぞれのサービスごとに収集用の記述を用意し、それをシステムに設置することで、情報資源の収集を行う。

4.4 EPUB への変換と蓄積

本研究では、情報資源の中に格納されている文字や画像などの情報を分解し、再変換するための基本データフォーマットとして EPUB[22, 23] を用いる。それは EPUB がシンプルな内部構造を持ち、かつ構成要素全般にインターネット標準技術 (HTML・CSS・SVG 等) が用いられており、多くのソフトウェアやブラウザが EPUB に対応しやすく、また内部構造も分かりやすい分類になっているため、非常に扱いやすいためである。また、一つの EPUB ファイルの中には、複数の章を配置することができるため、出版者側は望んだ順序で情報資源を配置することができる。また、現在市場に出回っている主要な情報端末は、ほぼ EPUB のデータフォーマットに対応しており、新しく策定された EPUB3.0 は文字ポイントの可変は勿論、書字方向・和欧文混植・縦中横・行頭行末禁則・ルビ処理などに対応しており、様々な環境や与えられた条件に対して、柔軟に対応することができる。

4.5 データの変換と提供

前述の Policy データの中の「利用端末にあわせて再変換する際に使用するデータ」を用い、ユーザの利用環境と、指定されたシチュエーションに合わせた、データフォーマットの再変換を行う。この際、ユーザの使用端末が EPUB に対応している場合、ユーザが指定したシチュエーションに体裁を直した状態で、EPUB のデータフォーマットをそのまま利用する場合もある。

5. 実装内容

5.1 ユーザ環境とシチュエーションを入力するフォーム

本システムは、まず最初に、ユーザが情報資源を閲覧する際のシチュエーションに関するデータと、ユーザの利用情報端末に関するデータを入力する。入力するデータ項目は、以下の通りである。

[利用情報端末に関するデータ]

- 画面サイズ

- 画像データ、音声データ、動画データなどが利用が可能かどうか
- 色彩表示が可能かどうか

[シチュエーションに関するデータ]

- 情報資源を閲覧する際のパターン情報

5.2 有料情報資源のリミット金額とオーソリティを考慮した入力フォーム

次に、ユーザにとって理想の情報資源を、できるだけコストをかけずに収集するため、以下の入力フォームをシステムに実装した。

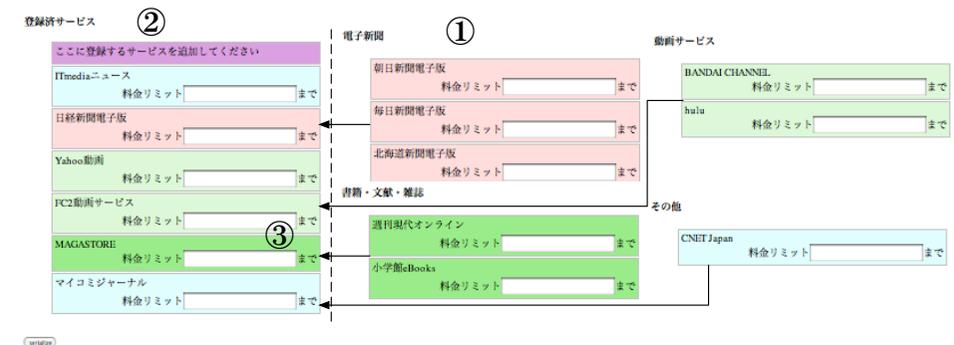


図 4 入力フォーム

ユーザは、未登録のサービスグループの欄 (図 4 (1)) から、登録済みサービスの欄 (図 4 (2)) に、利用したいサービスを移動させる。登録済サービスの配置場所は、動的に項目が移動可能であり、上にある項目ほどオーソリティが高いと判断される。一覧に目的の提供サービスが無い場合は、別フォームの登録申請フォームからシステムに登録して欲しいサービスの名前や URL を記述し、通知を受け取ったシステム管理者が手動で MED システムへ登録作業を行う。利用したいサービスが DB に登録されていない場合、設定フォームからシステム管理者にサービス登録の申請を行うことができる。

また、このとき同時に、サービスごとの有料情報資源収集の金額リミットも設定する。

5.3 実装環境と DB テーブル

現段階では、本システムは Web ブラウザ上で動作する。実装環境として、OS は Ubuntu(10.10)、システムは PHP(5.3.3-1)、DB 環境は MySQL(5.1.49-1)(表 1) で構築している。動作確認は、MacOS X(10.6.8) 上の、Google Chrome(13.0.782.215) と Safari(5.1.(6534.50)) のブラウザで行った。

テーブル名	説明
default_service	システムに予め登録してある情報提供サービスの保存用テーブル
pinfo	MEDシステムにログインするためのログインidとpassword保存用テーブル
situation_info	ユーザのシチュエーション保存用テーブル
selected_service	ユーザが選択したサービス、優先順位、金額リミット保存用テーブル
search_keywords	検索用キーワード保存用テーブル
unregistered_service	システムが未登録の情報提供サービスへの対応用テーブル

表 1 MED システム DB テーブル一覧

5.4 出力結果

今回、インターネット上に存在する情報資源を、想定されるユーザの Policy に乗っ取って、変換し、出力を行った。その結果を下記に示す。

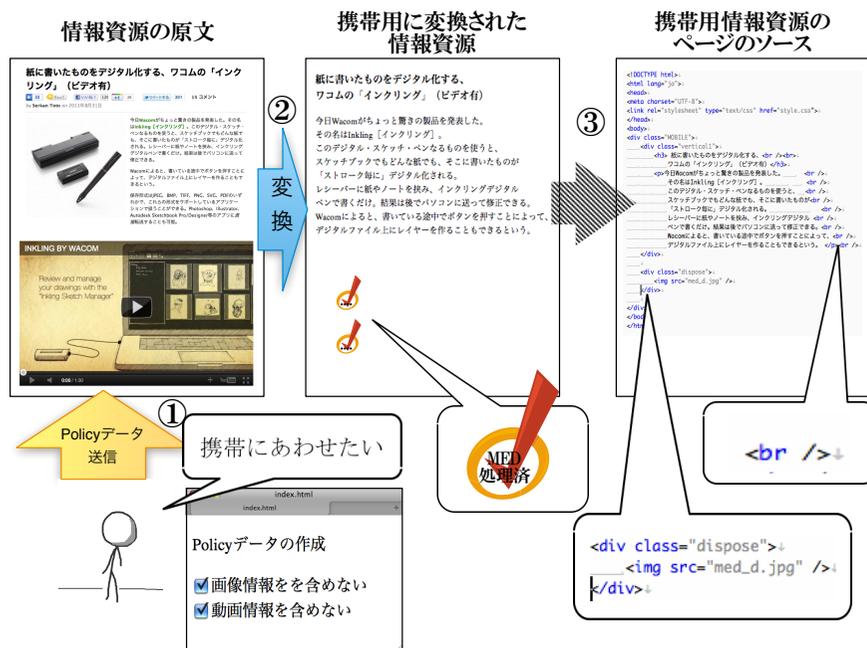


図 5 再変換・出力された情報資源

システムは、入力された Policy データ (1) をもとに、原文の情報資源の変換を行う (2)。

図5のように、ユーザが望む Policy が「テキストデータのみを端末のサイズに合わせて読みたい」場合には、Web 上でその旨を入力する。すると Policy データに、NoPic と wide='n' が追加される。変換されたデータには情報資源にあった画像や動画のデータが消去されると、変換後の情報には、添削した情報に関する印が残る。それによってユーザは情報の欠落を知ることができ、必要な場合は、より高性能な計算機を使ってさらにその情報を確認することができる。

6. おわりに

本研究では、各市町村や自治体などが作成した地域情報資源と、インターネット上に存在する情報資源を対象とした「情報を得るためのコストに関わる問題」「情報の埋没と信憑性の保証に関わる問題」「閲覧環境の相違による情報資源の互換性に関わる問題」を解決するため、インターネット上に存在する様々な地域情報資源と情報資源の中から、有料の情報資源と無料の情報資源を考慮し、利用ユーザのポリシーに対応する情報資源の自動的収集・管理・利用方法を提案した。これにより、多大なサービスの種類と情報量によって生まれるとコストや信憑性の問題に加え、1章に記した「閲覧環境の相違による情報資源の互換性に関わる問題」を解決するための手法を確認することができた。

今後、実装したシステムの検証を行いつつ、利用ユーザがどのサービスや情報源を利用して良いかわからない場合に対応するため、予め推奨サービスを提示する仕組みを加えたいと考えている。

また、ユーザが目的の情報資源を閲覧するためのコストを最小限に抑え、効率よく閲覧するために、ユーザの利用環境と登録されたポリシーに合わせて収集した有料と無料の情報資源を EPUB 型から別のデータフォーマットに変換する手法を考察し、さらに端末の性能差に左右されない情報資源を閲覧を試みる。

謝辞 本研究にご協力頂いた、筑波大学大学院図書館情報メディア研究科 杉本重雄教授に感謝の意を表す。本研究は科研費 (23700286) の助成を受けたものである。

参考文献

- 1) Kristen Purcell. "E-reader Ownership Doubles in Six Months - Pew Research Center". PewResearchCenter Publications. <http://pewresearch.org/pubs/2039/e-reader-ownership-doubles-tablet-adoption-grows-more-slowly>, (参照 2011-08-28)
- 2) 高橋正和, 三柳英樹. "【電子書籍最前線〜キーパーソンに聞く】 「電子書籍元年」の次に来る, EPUB 3 とタブレット端末の普及元年 -INTERNET Watch". INTERNET Watch. http://internet.watch.impress.co.jp/docs/column/ebookint/20110207_425462.html, (参照 2011-08-28)
- 3) 一般社団法人日本電子書籍出版社協会. "電子書籍交換フォーマット". 電子書籍交換

- フォーマット標準化会議 (参照 2011-07-04).
- 4) 山口真弘.”イーストに聞いてみた：電子書籍フォーマット「EPUB 3」ってぶっちゃけどうよ？ (1/4) - 電子書籍情報が満載! eBook USER”. eBook USER. <http://ebook.itmedia.co.jp/ebook/articles/1108/10/news005.html>, (参照 2011-08-16)
 - 5) Google. “Google”. Google. <http://www.google.co.jp/>, (参照 2011-08-30)
 - 6) Yahoo Japan Corporation. “Yahoo! JAPAN”. <http://www.yahoo.co.jp/>, (参照 2011-08-30)
 - 7) 福島俊一. インターネット検索エンジン：検索エンジンの仕組みと技術の発展. 情報の科学と技術. 2004, vol.54, no.02, p.66-71.
 - 8) 上田修一. “サーチエンジンにおける検索手法”. サーチエンジンにおける検索手法. <http://www.slis.keio.ac.jp/~ueda/semi/99search.html>, (参照 2011-07-28)
 - 9) 大森敬介, 中藤哲也, 山田, 泰寛, 原由加里, 廣川佐千男. “複雑な検索機能を持つ検索サイトの動向調査”. 電子情報通信学会第 15 回データ工学ワークショップ. 伊勢志摩, 2004-03-04, 電子情報通信学会データ工学専門委員会, 日本データベース学会. 2004,(ページ番号なし). <http://hdl.handle.net/2324/2956>, (参照 2011-08-23)
 - 10) Bergman, Michael K. White paper: the Deep Web: Surfacing Hidden Value. The Journal of Electronic Publishing. Vol.7, No.1, 2001. <http://www.press.umich.edu/jep/07-01/bergman.html>, (参照 2011-08-27).
 - 11) 長島直樹, 新堂精土. 情報サーチと消費者行動：消費者はネット情報をどのように使っているか. 経営情報学会誌. 2002, 11(3),p.17-36, (参照 2011-08-29).
 - 12) mixi. “ソーシャル・ネットワークワーキング サービス [mixi(ミクシイ)]”. mixi. <http://mixi.jp/>, (参照 2011-08-30)
 - 13) 大向一輝. SNS の現在と展望：コミュニケーションツールから情報流通の基盤へ. 情報処理. 2006, 47(9), p.993-1000, (参照 2011-08-28)
 - 14) 佐藤研司. “いかに情報を使いこなすのか：情報過多時代における情報の活用 (ナレッジ・マネジメントとレコード・マネジメント ii：学問と知識)”. レコード・マネジメント：記録管理学会誌, 2000, pp. 1?7,(参照 2011-07-03).
 - 15) 伊東栄典, 池田嗣徳, 亀岡謙, 高木早智子. 九大ポータル My.Kyushu-U の試作. 情報基盤センター広報 学内共同利用版. 2005, 6(1), p.1-6, (参照 2011-08-28)
 - 16) Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden. “Web Content Accessibility Guidelines (WCAG) 2.0”. W3C Recommendation. <http://www.w3.org/TR/2008/REC-WCAG20-20081211/>, (参照 2011-03-23)
 - 17) 財団法人日本規格協会情報技術標準化研究センター情報アクセシビリティ国際標準化に関する調査研究開発委員会ウェブアクセシビリティ国際規格調査研究部会. “WCAG 2.0 (W3C 勧告) 日本語訳 [原題：Web Content Accessibility Guidelines (WCAG) 2.0]”. 財団法人日本規格協会. <http://www.jsa.or.jp/stdz/instac/committee-acc/W3C-WCAG/WCAG20/index.html>, (参照 2011-03-23)
 - 18) Apple. “アップル - iTunes - あなたを楽ませるすべてのエンターテインメントが、ここに.”. iTunes. <http://www.apple.com/jp/itunes/>, (参照 2011-08-30)
 - 19) 佐々木俊尚. “電子書籍の衝撃”. デイスクヴァー・トゥエンティワン, 2010, 303p.
 - 20) 長尾真, 遠藤薫, 吉見俊哉. “書物と映像の未来 - - グーグル化する世界の知の課題とは”. 岩波書店, 2010, 208p.
 - 21) SmallRivers.” paper.li ? Twitter や Facebook を新聞として読む”. Paper.li, <http://paper.li/>, (参照 2011-08-17)
 - 22) IDPF. “EPUB 3 — International Digital Publishing Forum”. idpf, International Digital Publishing Forum Trade and Standards Organization for the Digital Publishing Industry. <http://idpf.org/epub/30>, (参照 2011-08-26)
 - 23) 秋元良仁. “EPUB の多国語対応に向けた取組と事例報告”. 情報知識学会誌, 2010, pp. 366?382,(参照 2011-07-05)