

## Markit: デジタルスクラッピングシステムの設計と実装

川添瑞木† 米澤拓郎† 高汐一紀‡ 徳田英幸†,‡  
慶應義塾大学大学院 政策・メディア研究科†  
慶應義塾大学 環境情報学部‡

E-mail: [zoe,takuro,kaz,hxt]@ht.sfc.keio.ac.jp

近年、デジタルメディアの普及により、従来では紙メディアのみを対象としていたスクラッピングにおいても、デジタルメディアが用いられるようになった。紙メディアとデジタルメディア両者を対象とするスクラッピングの中でも、紙メディアからデジタルメディアへ変換し、蓄積する形態のスクラッピングは、その有用性に反して複雑で面倒な作業が人々の負担となっている。本稿では、紙からデジタルへのスクラッピングを、手間のかかる作業を伴わずに行う手法を紹介する。さらに、人が興味を持ったことの現れである印付けを用いたスクラッピングシステム **Markit** を構築する。本システムによって、ユーザは面倒な作業による負担を感じることなく、紙メディアからデジタルメディアへのスクラッピングができる。

### Markit : Designing and Implementation of Digital Sclapping System

Mizuki Kawazoe Takuro Yonezawa Kazunori Takashio Hideyuki Tokuda  
Graduate School of Media and Governance, Keio University

Spread of information technology has changed the form of information scrapping from only paper media to digital media. However, to scrap information from paper media to digital media is still burden task for users; they must capture information digitally by camera or OCR, and organize information manually. This research focuses to reduce the cost for scrapping physical media as digital media. For this purpose, we propose Markit - scrapping system using marking interaction. Users can capture physical information of their interest digitally only by marking them with sticky. We evaluate the usefulness of Markit interaction through usability test.

#### 1. はじめに

グーテンベルクによる活版印刷の発明以来、人々の周りには様々なメディアが登場し、情報が溢れている。そのような環境のなかで人々は、自らに必要な情報をメディアから収集し蓄積する行為であるスクラッピングを頻繁に行うようになった [1]。従来のスクラッピングは雑誌や新聞の切り抜きといった紙メディアに特化した行為であったが、近年のデジタルメディアの普及に伴い、Web サイトのブックマークやコンピュータ内へのファイルの保存など、デジタルメディアを対象とした新たな形態のスクラッピングが登場した。一方で、紙メディアも依然、情報取得の方法として盛んに利用されており、紙メディアを対象としたスクラッピングも広く行われている。

スクラッピングの対象メディアである紙メディアとデジタルメディアには、それぞれに長所と短所があることが論じられている [2], [3]。例えば、紙メディアには、閲覧性が高く扱いやすい一方で、多量の情報や動的に更新される情報は扱えないという特徴がある。またデジタルメディアには、検索や複製が容易であり多量の情報や動的な情報を扱えるが、電源が必要であり閲覧性が低く扱いにくいなど点もある。結果、人々は目的や嗜好によってメディアを使い分けてスクラッピングを行っているのが現状である。例えば、自身の子どもが活躍した記事を全て紙メディアでスクラップブックを作り保存する場合や、趣味の情報を Web で検索してデジタルメディアのままブラウザのブックマークとして保存する場合などがある。このように、人々は紙メディアとデジタルメディア

の両方を用いて、複数の形態でスクラッピングを行っている。

本研究では、スクラッピングの様々な形態のうち、紙メディアで得た情報をデジタルメディアに変換して蓄積する形態のスクラッピングに注目し、この形態を紙発見デジタル蓄積型と呼ぶ。さらに、紙発見デジタル蓄積型スクラッピングを支援することを目的とし、人が興味を持った際にとる動作である印付け行為を用いたスクラッピングシステム **Markit** を提案する。本システムによって、容易かつ自然な行為で、紙メディアからデジタルメディアへのスクラッピングが可能となり、両メディアの特性を生かした情報収集形態を促進することが期待される。

本稿は以下のように構成される。まず第 2 節では、本研究が対象とする紙発見デジタル蓄積型スクラッピングとその問題点について述べる。次に第 3 節では、**Markit** システムについて、目的と機能要件、アプローチを述べる。第 4 節では、本システムの設計と実装について述べる。第 5 節では本システムを用いた評価と考察を述べ、第 6 節では関連研究を挙げる。最後に第 7 節において、本稿をまとめる。

#### 2. 紙発見デジタル蓄積型スクラッピング

本項では、一般的なスクラッピング行為の流れを説明したのち、本研究が対象とする紙発見デジタル蓄積型スクラッピングについて述べる。最後に、紙発見デジタル蓄積型スクラッピングシステムを構築する際の論点をまとめる。

## 2.1 スクラッピングの流れ

スクラッピングは、以下の3つのフェーズによって行われる。

### ● 発見フェーズ

雑誌や新聞、webなどの様々なメディアから情報を取得し、興味のある情報を発見するフェーズである。新聞を読んでいる際に面白そうな本の書評記事を発見することや、好きなアーティストのコンサート情報をwebで発見することなどが例として挙げられる。

### ● 蓄積フェーズ

発見フェーズで発見した情報を、蓄積するフェーズである。例として、記事の切抜きをスクラップブックに貼り付けることや、ブラウザのブックマークにURLを追加する行為が挙げられる。

### ● 利用フェーズ

蓄積した情報を利用するフェーズである。例としては、関連情報の検索や情報の共有、蓄積された情報の見返しなどがある。

蓄積フェーズにおいて人は、利用フェーズでの利便性を考慮し、情報の分類とメタ情報付加というスクラッピング特有の作業を行う。

情報の分類とは、ブラウザのブックマークをカテゴリ別に整理するように、蓄積する情報を分類する行為である。蓄積された情報が多量になると、後に利用する際に目的の情報を発見することが困難になるため、分類によって情報の検索性を高めている。

メタ情報の付加とは、記事の掲載元やコメントなどの書き込みを行うなど、スクラップ対象記事のメタ情報を付加する行為である。メタ情報は、スクラッピング記事に関する説明だけでなく、記事の分類や検索にも用いられる。

また、利用フェーズの形態は、スクラッピングの目的によって変化する。好きな俳優に関するスクラッピングのように、情報の収集自体が目的である場合は、利用形態は蓄積した情報の見返しとなる。気になるニュースのスクラッピングのように興味深い情報を発見した際のスクラッピングでは、スクラッピングした情報の関連情報を検索することが考えられる。関連情報の検索を行う場合は、検索によって発見した情報もまたスクラッピングの対象となり、次々とスクラッピング情報が増えていくこととなる。

## 2.2 紙発見デジタル蓄積型

紙発見デジタル蓄積型スクラッピングとは、スクラッピングの発見フェーズを紙メディアで行い、蓄積と利用フェーズをデジタルメディアで行う形式のスクラッピングである。例えば、雑誌で見つけた記事を、スキャナで取り込み画像データとして保存する場合がこの形式に当てはまる。

現在、雑誌や新聞などの紙メディアからの情報取得は盛んに行われているが、紙メディアはかさばりやすく整理しにくいいため、取得した情報を紙のまま蓄積していると後の利用が困難である。一方で、デジタルメディアは多量の情報蓄積、容易な検索の実現という特性を持っているため、情報の蓄積に向い

ており、紙メディアで得た情報をデジタルメディアで保存したいというユーザの欲求は増加している。この欲求に応える型式が、紙発見デジタル蓄積型スクラッピングである。

## 2.3 論点

紙発見デジタル蓄積型スクラッピングを実現する際の論点を以下の3つにまとめる。

### 2.3.1 紙からデジタルへの変換

現在の紙発見デジタル蓄積型スクラッピングでは、非常に煩わしい作業が必要である。既存の紙発見デジタル蓄積型スクラッピングシステムの例として、メディアドライブ株式会社の「やさしく新聞スクラップ」を挙げる[4]。やさしく新聞スクラップでは、スキャナで読み取った紙メディアをOCRによって透明テキスト付きPDFへ変換する。このような、紙メディアからデジタルメディアへの変換をOCRとスキャナを用いて行う手法では、紙メディアを読む作業から、コンピュータを用いた作業に切り替える必要があり、作業が非常に面倒である。また、コンピュータを使う作業は不慣れなユーザにとって複雑な作業であり、負担となる。

このように、紙発見デジタル蓄積型のスクラッピングはユーザにとって有用な形態であるが、複雑で面倒な作業が必要であるために、気軽に行うことが出来ない。

### 2.3.2 スクラッピング特有の作業

前項で述べた、スクラッピング特有の作業である情報の分類とメタ情報の付加は、情報を蓄積する際に行うため、紙発見デジタル蓄積型スクラッピングではコンピュータを用いて作業することになる。ユーザは、紙メディアを読んでいる作業を大幅に中断しなければならない。また、スキャンした画像データへの書き込みなど、デジタルメディアへのメタ情報の付加は、自由な書き込みが難しいという問題がある。

### 2.3.3 スクラッピング情報の利用

ユーザがスクラッピング情報を利用するためには、閲覧性の高さや検索の容易性が必要となる。また、情報を蓄積する際に入力したメタ情報や分類情報を利用して、蓄積された情報が整理されている必要がある。

紙メディアのスクラッピング情報がデジタルメディアに変換されて蓄積される場合、紙メディアとデジタルメディアの差異を考慮した表示や、デジタルメディアの特性を生かした情報の整理と表示が行わなければならない。

## 3. Markit

本節では、本研究が提案するMarkitシステムについて述べる。まず目的を述べ、続いて機能要件をまとめる。最後に、本システムのアプローチである印付け行為を用いた紙発見デジタル蓄積型スクラッピングについて述べる。

### 3.1 目的

本研究の目的は、紙発見デジタル蓄積型スクラッピングを面倒な作業を伴わずに行うことである。紙発見デジタル蓄積型スクラッピングが容易に実現できることで、ユーザは自由な形態でスクラッピング作業を行うことが出来るようになる。

### 3.2 機能要件

前節で示した問題点を解決するために実現すべき機能を、以下のようにまとめる。

#### ● スクラッピング箇所の容易な伝達

紙メディアの情報をデジタルメディアへ変換するためには、ユーザが変換の対象となる箇所をシステムに対して容易に伝達できる必要がある。例えば、スクラッピング箇所を伝達するために、対象箇所を切り取ってスキャナで読み込むような作業は、ユーザに大きな作業負担を与える。

#### ● メタ情報の取得と蓄積

情報をスクラッピングする際には、スクラッピング対象の情報と共にメタ情報を保存する。そのため、保存すべきメタ情報を容易にコンピュータに伝達できる必要がある。また、メタ情報のなかでも、スクラッピングした日付や掲載元などの情報は、デジタルメディアの特性を生かし、システムが自動取得することで、ユーザの負担を軽減できる。

#### ● スクラッピング情報の分類

スクラッピング情報の分類方法はユーザの目的や嗜好によって異なっており、自動で分類を行うことは出来ないため、分類情報をユーザから取得する。ユーザが分類情報をシステムに指定する方法は、容易でなくてはならない。

#### ● スクラッピング情報の利用

ユーザが情報の見返しを容易に行えるように、スクラッピングされた情報の表現方法を工夫する必要がある。例えば、メタ情報はスクラッピング情報と共に表示すべきであり、ユーザが見返したいスクラッピング情報を容易に見つけられるように、スクラッピング情報はユーザが示した分類方法に基づいてきちんと分類されている必要がある。

### 3.3 アプローチ

本研究では、第 3.2 項で述べた機能要件を満たす解決手法として、印付け動作を用いたデジタルスクラッピングシステムを提案する。印付け動作とは、付箋の貼付やページの折り曲げのような、注目箇所に印をつける動作のことである。図 1 に例を示す。印付け動作の利用により、作業の流れを中断しないスクラッピングを実現する。

本研究のアプローチを、以下の 4 つにまとめる。

#### ● 付箋をトリガとしたスクラッピング箇所の伝達

本システムでは、印付け動作の一つである付箋の貼付によって、ユーザの注目箇所を検知し、スクラッピングを行う。ユーザは、スクラップし



図 1: 印付け動作の例

たい箇所に付箋を貼ることで、紙メディアの情報をデジタルメディアとして蓄積することができる。

#### ● ペンによる印付けを用いたメタ情報付加

本システムでは、ユーザは特に注目した単語や文節をペンで囲うことで、自らの注目箇所をスクラッピング情報のメタ情報として、保存することができる。また、掲載元や日付などの静的なメタデータは、自動的に取得する。

#### ● 付箋の色を用いた情報の分類

本システムでは付箋の色を用いて、情報の分類を行う。ユーザはあらかじめ、赤は仕事、青は趣味といったように、色ごとに分類情報を決めておき、対応した色の付箋を紙メディアに貼付することで、コンピュータを用いることなくスクラッピング情報を分類し蓄積することができる。

#### ● スクラッピングに特化したビューワの実装

本システムでは、蓄積した情報を見返すためのビューワを提供する。スクラップされた情報と共に保存されたメタ情報（雑誌名、発行年度及びユーザが独自に付加した情報等）を表示する機能を備え、ユーザから付箋を用いて指定された分類方法に基づいて、分類を行い表示する。これにより、ユーザの見返しを支援する。さらに、デジタルメディアの特性を生かし、スクラップされた情報の関連情報を、Web と他のユーザが蓄積した情報から検索し、自動的にスクラッピングする。これにより、ユーザが自ら関連情報を検索することなく、スクラップブックの内容を増強できる。

## 4. Markit システム

本節では、容易な紙発見デジタル蓄積型スクラッピングを実現する Markit システムの設計と実装について述べる。

### 4.1 想定環境

本システムは、雑誌や新聞などの紙メディア上の記事を対象とする。また、各記事には対応した 2 次元バーコードが印刷されており、記事と同一内容のデジタルメディアをデータベースから取得可能な環境を想定している。さらに、ユーザの上部には Web カメラが設置されており、ユーザはその下で読書を行うことを想定する。図 2 に、Markit システムの利用図を示す。



図 2: Markit 利用図

## 4.2 ハードウェア構成

ハードウェア構成を図 3 に示す。

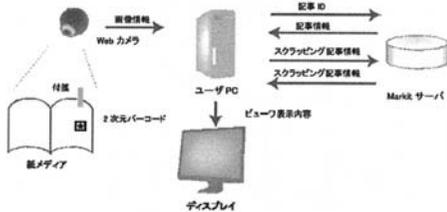


図 3: ハードウェア構成図

Markit サーバは、雑誌記事のメタ情報と全 Markit ユーザのスクラッピング記事情報を保持している。ユーザー PC は、ユーザーがスクラッピングした記事の情報を保持しており、印付け動作の取得や紙メディアからデジタルメディアへの変換、ビューワ管理を行う Markit クライアントシステムが動作している。

ユーザーの付箋貼付は、ユーザーの上部に設置された Web カメラを用いてユーザー PC が検知する。同時に、付箋が貼られた記事の 2 次元バーコードを読み取り、記事の情報を Markit サーバから取得する。さらに、スクラッピングされた記事の情報をディスプレイに表示する。

## 4.3 ソフトウェア構成

ソフトウェア構成を図 4 に示す。

Markit 設定リストは、付箋の色と分類の対応リストである。Markit ビューワを用いて、ユーザーによって生成される。記事情報データベースは、Markit サーバ上に設置され、雑誌に掲載されている記事の内容と掲載雑誌名、号数などのメタ情報が蓄積されている。スクラッピング記事情報データベースは、ユーザー PC 上に設置され、個々のユーザーがスクラッピングした記事の情報が蓄積されている。共有スクラッピング記事情報データベースは、Markit サーバ上に設置され、全ユーザーのスクラッピング記事情報が蓄積されている。

ユーザーが記事を発見し付箋を貼付すると、発見検知・蓄積モジュールがユーザーの印付け動作を検知し、付箋の色から蓄積の際に用いる記事の分類情報を取得

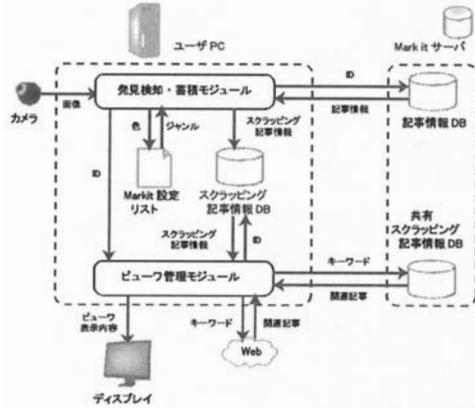


図 4: ソフトウェア構成図

する。さらに、2 次元バーコードを読み取り、Markit サーバ上の記事情報データベースから対象記事の情報を受け取る。取得した情報を、ユーザー PC 上のスクラッピング記事情報データベースへ蓄積する。

ユーザーがビューワを用いてスクラップした情報を閲覧する際は、ビューワ管理モジュールがユーザー PC 上のスクラッピング記事情報データベースから記事の情報を取得し、ビューワ上に表示する。さらにビューワ管理モジュールは、スクラッピングされた記事の関連情報を、Web と Markit サーバ上の共有スクラッピング記事情報データベースから自動に取得し、スクラッピングデータとしてスクラッピング記事情報データベースに保存する。

## 4.4 各モジュールの実装

各モジュールの実装について述べる。

### 4.4.1 データベース

データベースは全て MySQL を用いて実装した。以下に、スクラッピング記事情報データベースの ER 図を示す(図 5)。

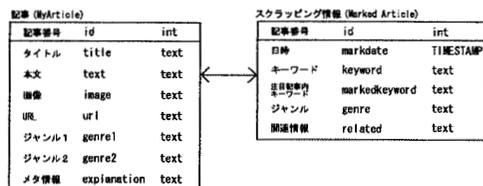


図 5: スクラッピング記事情報の ER 図

MyArticle テーブルは、ユーザーが所持するスクラッピング記事の情報を保持する。また、関連情報としてシステムが自動的に取得した情報も、MyArticle テーブルに入られる。id は、各記事につけられたユニークな ID である。title と text, image はそれぞれ記事

のタイトルと本文、記事内の画像の URL である。url は雑誌の URL であり、genre1,2 はそれぞれおおまかなジャンルと細かいジャンルを示す。explanation は、記事自体のメタ情報であり、掲載雑誌名と号数の 2 つがある。

MarkedArticle テーブルは、ユーザが注目し付箋を貼付した記事の情報である。id はスクラッピングした記事にユニークにつける ID である。markdate はスクラッピングした日時、keyword は記事のタイトルと本文を形態素解析して得られたキーワード、markd-keyword は、注目箇所内のキーワードである。genre はユーザが示した分類情報であり、related には関連情報の記事 ID がカンマ区切りで入っている。

#### 4.4.2 発見検知・蓄積モジュール

ユーザの注目記事の取得は、画像解析を用いて実装した。付箋の貼付は、画像の差分を取得しており、閾値を超える差分が生じた際に、差分発生部分の色と大きさを、付箋の色と大きさと比較することで取得する。また、検知する付箋の色は、ピンク、紫、黄緑、青、オレンジの 5 種類を実装した。ペンによる印付けには、ピンク色を用いた。

さらに、各記事に印刷された 2 次元バーコードには、JARToolKit [5] を用いた。JARToolKit は、ワシントン大学の HITL (Human Interface Technology Lab) によって開発された AR ライブラリである ARToolKit [6] の Java 版ライブラリである。ARToolKit で用いられる 2 次元バーコードの例を図 6 に示す。



図 6: ARToolKit における 2 次元バーコードの例

キーワードの取得には形態素解析を用い、形態素解析ライブラリは茶筌 [7] を用いた。また、形態素解析の結果から未知語と固有名詞をキーワードとした。

#### 4.4.3 ビューワ管理モジュール

本項では、ビューワ管理モジュールの実装について述べる。

ビューワ表示部分は以下のようにになっている。図 7 は、ユーザがスクラッピングした情報を中心に閲覧する際のスクリーンショットである。左のリストには、全てのスクラッピング記事が表示される。リストから一つを選択すると右側に詳細情報が表示される。スクラッピングした記事の下には雑誌名とスクラッピングした日時が、記事の周辺には関連情報が表示される。

図 8 に、システムが自動的にスクラッピングした関連記事を閲覧する際のスクリーンショットを示す。

ビューワには、利用の際の利便性のために、以下の 2 つのスクラッピング情報整理機能を実装している。第 1 に、ビューワはタブ切り替え式になってお



図 7: ビューワ スクラッピング記事表示時

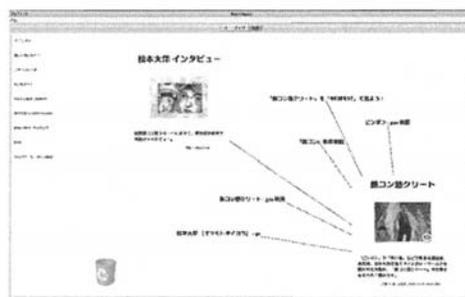


図 8: ビューワ 関連記事表示時

り、ジャンルごとに表示を切り替えられる。また、 unnecessary な記事は削除出来るように、ゴミ箱を用意している。

システムが Web から自動的に関連情報を取得する部分の実装には、GoogleAPI[8] を用いた。

## 5. 評価

本節では、Markit システムの評価を行う。評価は、基本性能評価とユーザビリティ評価の 2 種類を行った。

### 5.1 基本性能評価

本項では、Markit の基本性能評価として、付箋貼付の検知にかかる時間と正答率について実験し、考察を行う。

#### 5.1.1 検知所要時間

付箋の貼付を検知するためにかかる時間として、付箋を貼付してから検知された合図の音になるまでの時間を評価した。前節で述べた実装環境において、付箋の検知を 20 回行い、その平均所要時間を測定した。その結果、平均所要時間は 2.27(s) となった。測定結果のグラフを図 9 に示す。

平均所要時間の 2.27(s) は、ユーザが実際に利用する際にはほとんど負担にならない時間である。

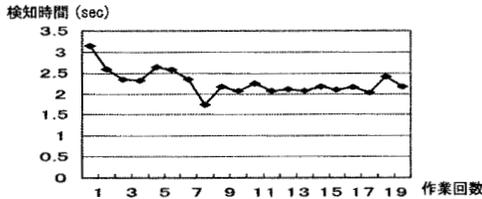


図 9: 検知所要時間

### 5.1.2 正答率

付箋の貼付の検知と、貼られた付箋の色の判断における正答率を評価した。2種類の雑誌と3種類の付箋を用いて、各々の組み合わせに対して付箋を真っ直ぐに貼った場合を10回、斜めに張った場合を10回、計120回のスクラッピングを行い、正答率を調査した。付箋は、ピンク、紫、黄緑の3種類を用い、雑誌は背景が白地のものとカラー写真のもの、2種類を用いた。貼付の検知における正答率を、表1に示す。

表 1: 貼付検知における正答率

背景	色	角度	貼付と色検知						正答率 (%)
白地	ピンク	垂直	○	○	○	○	○	○	100
		斜め	○	○	○	○	○	○	100
	紫	垂直	×	○	○	○	○	○	80
		斜め	○	○	○	○	×	○	80
	黄緑	垂直	○	○	○	○	○	○	100
		斜め	○	○	○	○	○	○	100
写真	ピンク	垂直	×	○	○	○	○	×	60
		斜め	×	○	×	×	×	○	40
	紫	垂直	○	○	○	○	○	×	80
		斜め	○	○	○	○	○	○	100
	黄緑	垂直	×	×	○	○	×	×	40
		斜め	×	×	×	×	×	○	20

調査の結果、色の判断における正答率は100%であった。記事の背景が白地の場合は、93%の高い確率で検知したが、背景がカラー写真の場合は、57%と低確率であった。特に、黄緑の付箋を用いた場合は、30%とほとんど正しく認識しなかった。この結果から、背景色がある場合は認識の精度が落ち、その中でも背景色と付箋の色の組み合わせによって認識率に差が出ることが分かった。これは、背景がカラー写真である場合は光の反射やコントラストの変化により画像解析を正確に行うことができないという、実装上の問題であると考えられる。また、付箋の傾きによる正答率の違いは見られなかった。

### 5.2 ユーザビリティ評価

本項では、ユーザビリティ評価について述べる。ユーザビリティ評価は、Jakob Nielsenによって定義[9]されたユーザビリティの指標である学習しやすさ (Learnability)、効率性 (Efficiency)、記憶しやすさ (Memorability)、間違えにくさ (Error Handling)、主観的満足度 (User Satisfaction) の5つに関して、実験を行い評価した。

## 5.3 実験

Markitシステムにより、紙メディアからデジタルメディアへの変換作業を負担なく行えることを評価するため、Markitを実際に用いて紙発見デジタル蓄積型のスクラッピングをする実験を行う。

### 5.3.1 実験環境

被験者は、初めて本システムを利用した10代から20代の学生10名である。机の上にMarkitクライアントが動作するPCとビューワが表示されているディスプレイ及びPCを置き、その前方にデスクトップライトとライトの上にカメラを設置する。試験者は、ライトの下に置かれた雑誌のスクラッピングを行う。スクラッピング対象となる記事は10種類、付箋の色は3種類用意した。

### 5.3.2 実験手順

実験の手順は、想定するシナリオとシステムについての簡単な口頭説明、実験、アンケートの3段階で行う。実験では、興味深い記事を発見し、付箋を選んで貼り付け、ビューワ上にスクラッピングした記事が表示されたのを確認するまでの作業を8回繰り返した。また、付箋を選んで貼り付け、スクラッピングされたことを確認するまでの時間を計測した。実験終了後に、10段階評価のアンケートを行った。

## 5.4 実験結果

実験結果として、一回の作業が終了するまでの平均所要時間を示したグラフを図10に、各被験者が作業に要した時間の平均値、最大値、最小値を示したグラフを図11に示す。また、アンケート結果の平均値を表2に示す。

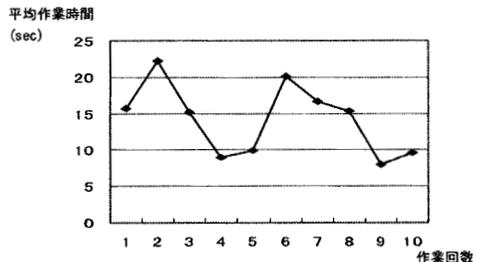


図 10: 作業時間平均

## 5.5 考察

本項では、実験結果からユーザビリティの5つの指標に基づいて考察を行う。

### 学習しやすさ

全ての被験者が簡単な説明だけでMarkitを利用できた。表2のアンケート結果においても、「使い方を習得するのは簡単だ」という質問に対する10人の平均値は8.7であり、学習は容易であるといえる。

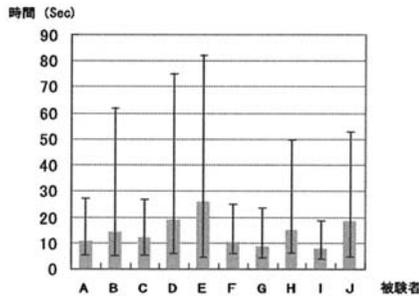


図 11: 被験者が作業に要した時間

表 2: アンケートの結果

付箋を用いた記事のスクラッピングについて (10段階)	
使い方を習得するのは簡単だ	8.7
作業に負担を感じた	4.5
必要だと思うことが全て出来る	6.9
やっていて楽しい	8.5
付箋の色による分類は分かりやすい	7.5
ビューアについて (10段階)	
見やすい	8.6
メタ情報の表示は役立った	8.8
関連・共有情報は役立った	7.8
関連・共有情報と元の記事との関連は分かりやすい	7

### 効率性

図 10 のグラフは、ゆるやかな右下がりとなっているが、基本的に作業回数によって作業時間が短くなる傾向は見られず、効率性が高いとは言えない。

### 記憶しやすさ

Markit システムは、付箋を貼付するという単純な動作を用いるため、被験者は簡単な説明で作業を理解し、実行出来た。よって、時間をかけても忘れにくいと考えられる。

### 間違いやすさ

図 11 の作業時間の最大値から、被験者は最長で 80 秒以上かけて作業を終えたことが分かる。これは、第 5.1.2 項の貼付検知における正答率で述べたように、背景色と付箋の色の組み合わせによってはほとんど検知されないため、成功するまで作業を行ったことが原因である。この実装上の問題は、表 1 のアンケート結果の、「作業に負担を感じた」という質問に対する回答の平均値が 4.6 と高くなっていることの原因であると考えられる。自由記述欄のコメントにおいても、貼付検知の成功率の低さに関するコメントが 6 件あり、ユーザに負担を与えたことが分かる。今後、付箋の検知部分の実装を見直し、精度を高めることが求められる。

### 主観満足度

前述のように、作業に対する負担が高いという結果

が得られたが、同時にアンケート結果の「やっていて楽しい」という質問に対する回答の平均値は、8.5 と非常に高い結果が得られた。自由記述欄のコメントにおいても、「付箋に貼るだけで情報を得ることが出来て簡単」「正しく認識したときは快感」というコメントを得た。このことから、今回の実験でユーザが感じた負担は実装面での問題であり、付箋の貼付を用いるという手法自体は好意的に感じられたことが分かる。

また、付箋の色による分類は、平均値 7.5 と良い評価を得られたが、自由記述欄において、「分類が増えるとわかりにくい」というコメントが 1 件得られた。ビューアに関しては、閲覧性、メタ情報の表示、関連情報の表示に関して、それぞれ高い評価を得た。しかし、「必要だと思うことが全て出来る」という質問に対する回答の平均値は 6.9 であり、自由記述欄のコメントにおいても、「記事のサムネイルが欲しい」「自分の閲覧履歴が見たい」など、ビューアに対する意見が寄せられた。これらの意見を踏まえ、今後の更なる実装が求められる。

## 6. 関連研究

### 6.1 やさしくデジタルファイリング

やさしく新聞スクラップは、メディアドライブ株式会社が開発したデジタル蓄積型スクラッピングシステムである [4]。様々な形式のファイルを一括で管理することが出来、紙発見デジタル蓄積型スクラッピングにも対応している。スキャナで読み取った紙メディアは、OCR によって透明テキスト付き PDF へ変換することができる。整理や後の利用における利便性を重視しており、様々な機能が実装されている。例えば、ファイルの分類、日付やファイル形式によるツリー構造での表示、重要な箇所にもマーカーを引いたりコメントの追加などができるアノテーション機能などである。

しかし、紙メディアからデジタルメディアへの変換を、OCR とスキャナを用いて行う手法では、蓄積フェーズでのユーザの負担が大きい。スキャナで読み込む作業だけでなく、分類とメタ情報付加もコンピュータを用いて行わなければならないため、紙メディアを読んでいる作業を大幅に中断するという問題点は解決されていない。

### 6.2 EnhancedDesk

小池らの提案する EnhancedDeck [10] は、紙とデジタル情報の統合を目的とした机型実世界志向システムである。机上の本や書類といった紙メディアに関連するデジタル情報を机上に投影することで、従来はユーザが行っている紙とデジタル情報の連携を自動的に行う。また、指によるデジタル情報の操作が可能であり、デジタル情報を直感的な動作で扱える。図 12 にシステムの利用手順を示す。紙メディアの認識には 2 次元バーコードを、手指の認識には赤外線カメラ、デジタル情報の投影にはプロジェクタを用いる。

EnhancedDesk は、指を用いたデジタル情報の操作によって、現実世界のものである紙とデジタル情報の統合という目的を達成している。本研究は、紙発見デジタル蓄積型スクラッピングを対象としており、スクラッピングには、記事の分類やメタ情報の付加、

スクラッピング記事の見返しといった、スクラッピング特有の作業がある。スクラッピングを支援するためには、それらのスクラッピング特有の作業に適したアプローチが必要となる。

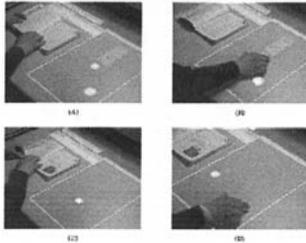


図 12: EnhancedDesk

## 7. 今後の課題と展望

本節では、本研究の課題と、今後の展望を述べる。

### 7.1 付箋の検知における精度の向上

評価の結果から、本システムの付箋検知部分の実装は、検知精度を向上する必要があることが分かった。検知精度が落ちる条件として、記事の背景がカラーである場合に付箋の色によって検知が困難であることが挙げられる。これは、付箋の貼付を検知するために画像解析を行っており、背景がカラー写真である場合は、光の反射やコントラストの変化により、正しく解析が行えないことが原因であると考えられる。今後は、実装の精度向上によって周囲の状況に左右されない解析を行い、ユーザに負担を与えないシステムにする必要がある。

### 7.2 紙メディアからの情報取得

本研究では、紙メディアから情報を取得するための想定環境として、取得する記事と同一内容のデジタルメディアが、取得可能な形で提供されていることを挙げている。そのため、デジタルメディアを提供できないような地方のローカル誌や個人的に作成された情報、過去に既に出版されている書籍などは扱うことが出来ないという問題がある。今後の展望として、OCR などの手法も取り入れ、記事と同一内容のデジタルメディアがない場合でも、紙発見デジタル蓄積型のスクラッピングを可能にしたい。

## 8. まとめ

本論文では、印付け動作を用いた紙発見デジタル蓄積型スクラッピングシステムの提案と実装を行った。

紙メディアとスキャナを利用したスクラッピングの場合では、紙メディアをスキャナで読み込んだ後、OCR による文字情報を取得し、メタデータの付加や分類を行う。この作業は煩雑であり、ユーザにとって大きな負担である。本論文では、このような問題点を解決するデジタルスクラッピングシステムとして **Markit** を提案した。Markit では、ユーザは付箋を貼ることで紙メディアの記事をデジタルメディアとして蓄積できる。さらに、スクラッピングに特化したビューワと関連情報の自動取得によって、デジタルな

スクラッピングブックを作成する。これにより、ユーザは簡単に気軽な作業によって、紙発見デジタル蓄積型スクラッピングを実現できる。

評価では、Markit は学習と記憶がしやすく、主観的満足度が高いという結果が得られた。これにより、印付け動作を用いた Markit のアプローチは、ユーザに負担をかけずに紙発見デジタル蓄積型スクラッピングを行わせることが可能であるといえる。

今後は、付箋貼付検知における画像解析の精度向上と、紙メディアからデジタルメディアへの変換において 2 次元バーコードを用いない手法を検討することで、本システムの改良・拡張をすすめる。

## 謝辞

本研究は、総務省「ユビキタスネットワーク制御・管理技術の研究開発 (ubila プロジェクト)」の一部として行われました。

## 参考文献

- [1] Catherine C. Marshall and Sara Bly. Sharing encountered information: digital libraries get a social life. In JCDL '04: Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, pp. 218-227, New York, NY, USA, 2004. ACM Press.
- [2] Jakob Nielsen, Differences Between Print Design and Web Design, <http://www.useit.com/alertbox/990124.html>
- [3] 面谷信. 紙への挑戦電子ペーパー 情報世界を変えるメディア
- [4] やさしく新聞スクラップ. <http://pac.mediadrive.jp/ynps/>
- [5] jARTToolKit. <http://jerry.c-lab.de/jartoolkit/>
- [6] ARTToolKit. <http://www.hitl.washington.edu/artoolkit/>
- [7] 茶筌. <http://chasen.naist.jp/hiki/ChaSen/>
- [8] Google API. <http://code.google.com/>
- [9] Jakob Nielsen. Usability Engineering. Morgan Kaufmann Pub; New Ed 版, 1994.
- [10] 小林貴訓. 赤外線画像を用いた指先実時間追跡による Enhanced Desk の実現. ヒューマンインタフェース学会 ヒューマンインタフェースシンポジウム HIS99, pp.417-422, 1999