

電子書籍にメタデータ付きアノテーションを付けることで 想起の起点，到達点またはハブとして利用する システムの提案

山口琢^{†1} 大場みち子^{†2} 高橋修^{†2}

読書では，ある本で読んだ一節を別の状況で思い出したり，あるいはその本の一節を読んだときに別の情報を思い出したりする経験を通して，我々はその本の一節，ひいては本全体の内容への理解を深める．Webの世界ではすでにこのような情報の連鎖が実現されており，電子化された書籍も，この連鎖の中に位置づけることが可能なはずである．我々は，読書におけるこのような想起体験を，電子化してインターネットと接続することで，より豊かなものにした．その可能性を示すことが本稿の目的である．我々は，状況に応じてコンテンツを提示する情報想起サービス(reminder)を想定して，関連するシステムを研究している．本稿では，そのようなサービスにおいて想起を起こす状況としての起点，想起されるコンテンツとしての到達点，あるいは想起の連想を中継するハブとして，電子書籍を利用するシステムを提案する．このようなシステムでは，時に数千，数百ページになる書籍の一部分を，このような処理の単位として扱う必要がある．我々は，電子書籍の一節にアノテーションを設定し，かつそのアノテーションにオープンなメタデータを付けることで，そのような想起処理が可能となることを示す．また，従来の推薦システムでは，ハブとして普遍性を持った辞書的なデータを使用するが，そのようなオントロジーの構築には高い専門性が求められるので，誰が構築するのかという困難な問題があった．しかし，本の一節の想起を目的とするならば，普遍性を気にすることなくサービスが成立する可能性を示す．

Proposal for a reminder system which uses E-books as triggers, targets and hubs by metadata annotation

TAKU YAMAGUCHI^{†1} MICHIKO OBA^{†2}
OSAMU TAKAHASHI^{†2}

Reading books, we improve comprehension of a passage or the whole book, by remembering that passage in some other context or by being reminded some other things by that passage. That is familiar in the Web system and we want to let E-books to participate in the information chain. In this paper we will show a information reminder system which reminds registered information in some appropriate context. The system use E-books as reminding triggers, targets and hubs by metadata annotations.

1. はじめに

今後，電子書籍への主な関心は，冊子レイアウトの継承からは外れていくだろう．読書では，ある本で読んだ一節を別の状況で思い出したり，あるいはその本の一節を読んだときに別の情報を思い出したりする経験を通して，我々はその本の一節，ひいては本全体の内容への理解を深める．Webの世界ではすでにこのような情報の連鎖が実現されており，電子化された書籍も，この連鎖の中に位置づけることが可能なはずである．

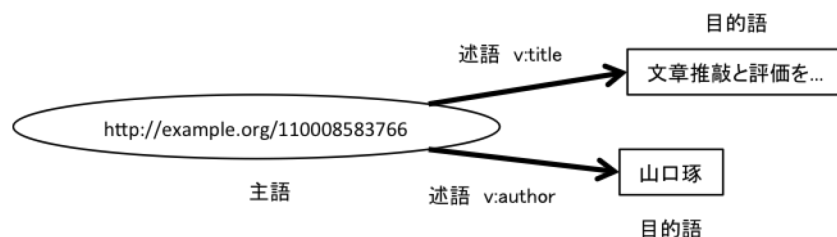
我々は，読書におけるこのような想起体験を，電子化してインターネットと接続することで，より豊かなものにした．その可能性を示すことが本稿の目的である．我々は，状況に応じてコンテンツを提示する情報想起サービス

(reminder)を想定して，それを支えるシステムを研究している．本稿では，そのようなサービスにおいて想起を起こす状況としての起点，想起されるコンテンツとしての到達点，あるいは想起の連想を中継するハブとして，電子書籍を利用するシステムを提案する．

このようなシステムでは，時に数千，数百ページになる書籍の一部分を，このような処理の単位として扱う必要がある．我々は，電子書籍の一節にアノテーションを設定し，かつそのアノテーションにオープンなメタデータを付けることで，そのような想起処理が可能となることを示す．また，従来の推薦システムでは，ハブとして普遍性を持った辞書的なデータを使用するが，そのようなオントロジーの構築には高い専門性が求められるので，誰が構築するのかという困難な問題があった．しかし，本の一節の想起を目的とするならば，普遍性を気にすることなくサービスが成立する可能性を示す．

^{†1} 公立はこだて未来大学 大学院
Future University Hakodate, Graduate School of Systems Information Science

^{†2} 公立はこだて未来大学
Future University Hakodate



```
<span
  xmlns:v="http://rdf.data-vocabulary.org/#"
  about="http://example.org/article/110008583766">研究報告
「<cite property="v:title">文章推敲と評価を支援するシステムの開発に向けての考察</cite>」の著者は
  <span property="v:author">山口琢</span>です.
</span>
```

図 1 RDFa による記述例
Figure 1 description by RDFa

1.1 電子書籍の今後

2011 年 10 月, International Digital Publishing Forum(IDPF)が, 電子書籍のフォーマット EPUB のバージョン 3.0 を確定した.

EPUB 3 において重要な点の 1 つは, 縦書きやルビなどを取り入れることで, 日本語対応のみならず EPUB の「国際化は大きく前進した」(村田 2012)ことである. また, EPUB は HTML5 や CSS など W3C の標準を採用しており, EPUB が「Web と電子書籍のシナジーを追求する」(村田 2012)ものである点も重要だ. すなわち, リフローや固定といったレイアウト機能は, 紙の出版物からの移行や既存出版サービスとの連続性において重要ではあるが, 書物の未来を考える上では機能の一部にすぎない. 掲示板のコメントやタイムラインなどソーシャル・コンテンツでは, リンクされたり部分的に埋め込まれたりすることで総体として動的に変化・成長するプロセスや, その過程で多言語が混在してコンテンツが国際化していくさまを, 顕著に見ることができる. このような Web コンテンツの現在の姿に, 電子書籍の近未来が垣間見える.

1.2 Linked data

Linked data とは, インターネット上の文書が HTML で記述されて互いにリンクすることで利用が促進されたように, データもインターネットに置いて互いにリンクすることによって豊に活用しようとする考え方である. すなわち次の原則に基づいてデータを用意する: (1) 物事の識別に URI を使い, (2) それらを参照してたどるのに HTTP を使い, (3) たどられたら有益な情報を RDF/XML などの標準形式で提供し, (4) その情報には他の物事へのリンクを含めることで他の物事が発見されやすくする (Berners-Lee 2006).

Resource Description Framework (RDF)ではデータを主語(subject), 述語(predicate), 目的語(object)の 3 つの要素で記

述する. 3 点セットで表現するので, これをトリプルと呼ぶ. この主語は URI であり, 目的語にも積極的に URI を使うことでデータ間を積極的にリンクし, 文書の Web のようにデータの Web を実現しようとするのが linked data の考え方である.

RDF は抽象的な記法であり, 具体的なデータは RDF/XML や RDFa で記述(serialize)する.

1.3 RDFa

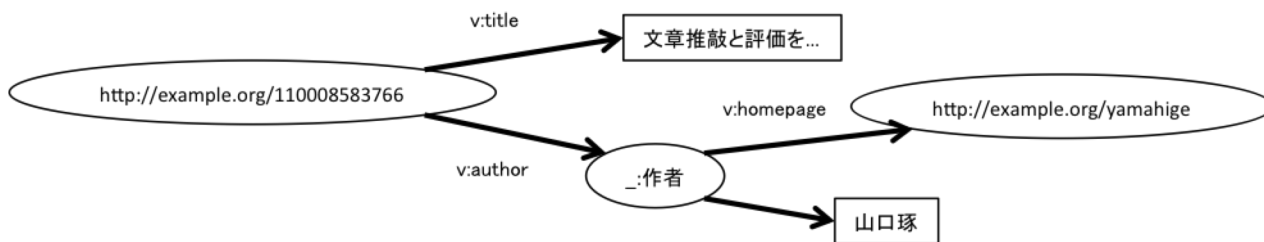
RDFa は, RDF をインターネットの Web 文書, すなわち XHTML 文書に埋め込むことで記述(serialize)する方法である. RDFa による記述例を示しつつ, linked data の趣旨も併せて説明する.

図 1 では, 研究報告のタイトルと著者を表すトリプルが上のダイアグラムのようなとき, これを下のコードのように XHTML に埋め込む.

このとき, 目的語が URI になるようにデータを作成することで, このデータがさらに他のデータにつながるようにしよう, というのが linked data の原則である. この原則に則って図 1 のトリプルを改善したのが図 2 である.

図 1 では, 研究報告に関心を持ったユーザーが, そのタイトルと著者の文字列(literal)にたどりついて, そこで終わりである. 図 2 では, 研究報告に関心を持ったユーザーを, その著者のホームページに連れて行ける可能性がある. このように, データをリンクすることで, データ全体の利用を促進する.

データを記述する記法は RDFa の他にも提案・採用されている. 本稿は特に RDFa を推すものではない. データが XHTML 文書に埋め込まれることをコードで具体的に示す例として採用している.



```

<span
  xmlns:v="http://rdf.data-vocabulary.org/#"
  about="http://example.org/article/110008583766">研究報告
「<cite property="v:title">文章推敲と評価を支援するシステムの開発に向けての考察</cite>」の著者は
  <span rel="v:author">
    <a rel="v:homepage" href="http://www.example.org/yamahige" property="v:name">山口琢</a>
  </span>です.
</span>
    
```

図 2 RDFa の記述例，目的語が URI になっている
 Figure 2 description by RDFa: contains out going link

1.4 アノテーション

RDFa で図 1 や図 2 のように記述した場合，トリプルの主語は，この記述を含む文書そのものではないことに留意されたい。主語は about 属性で示される。これは，XHTML の head 要素における meta 要素でメタ情報を記述場合と異なる。meta 要素で記述されるデータの主語は，それを含む XHTML 文書自身である。例えば「著者」といった場合，それは XHTML 文書の著者を指す。

図 1，図 2 のような記述が可能であることは，データの主語が，そのデータを含む文書でなくてもよいことを意味する。つまり，XHTML 文書の中に，他の主語に対する述語-目的語を，つまり他のデータに関するデータ，すなわちメタデータを記述することができる。

では，勝手な主語-述語-目的語のトリプルを文書に挿入して，いわばスパム化してよいかというと，そうではなからう。いざんとして，それが埋め込まれた文章の内容とマッチすることが求められる。あるいは，マッチするかをチェックすることで，スパム化に対して一定の歯止めとなる。これが埋め込み型メタデータのメリットである。このようなスパム・フィルタリングについては別稿に譲り，本稿ではこれ以上とりあげない。

2. 関連研究

冒頭で述べた読書における想起体験を，状況駆動の情報推薦とみなして検討する。結論から言えば，想起体験は ToDo リストのような想起システム(reminder)によって実現可能ではないかと考えている。

2.1 推薦技術

情報推薦システム研究の歴史は長い。情報推薦システムでは，プロファイリングによってユーザーの特徴情報を獲得して利用する。コンテンツに基づく推薦では，コンテンツの特徴とユーザー・プロフィールを比較する。協調フィルタリングでは，ユーザー同士のプロフィールを比較する。コンテンツに基づく推薦では，コンテンツの特徴とユーザーの特徴を直接比較するため，セレンディピティの高い推薦が難しい。協調フィルタリングでは，まったく新しいコンテンツやまったく新しいユーザーについて推薦が難しい。すなわち，一定の使用期間を経てコンテンツの評価やユーザー・プロフィールを獲得する必要がある。

また，コンテンツとプロフィールの比較に普遍性を持つ辞書を介在させることで，推薦結果の正しさを高める研究がある。そこでは，そのような辞書の構築や維持が課題となる。

我々は，まず，システムを想起させる仕組みと位置づける。ユーザーが予め「適切な状況で思い出したい」と指定した情報を，後に状況を検知して想起させる。その上で，このシステムを前提に，そこに推薦の要素を加味することが可能だろう。まず，想起システムとして成立させることが目標である。

ポイントは，ユーザーには種類があり，これ全体を想起サービスとして構築するところにある。以下で説明する。

3. アプローチ

我々のシステムは，従来の推薦技術と照らし合わせると，

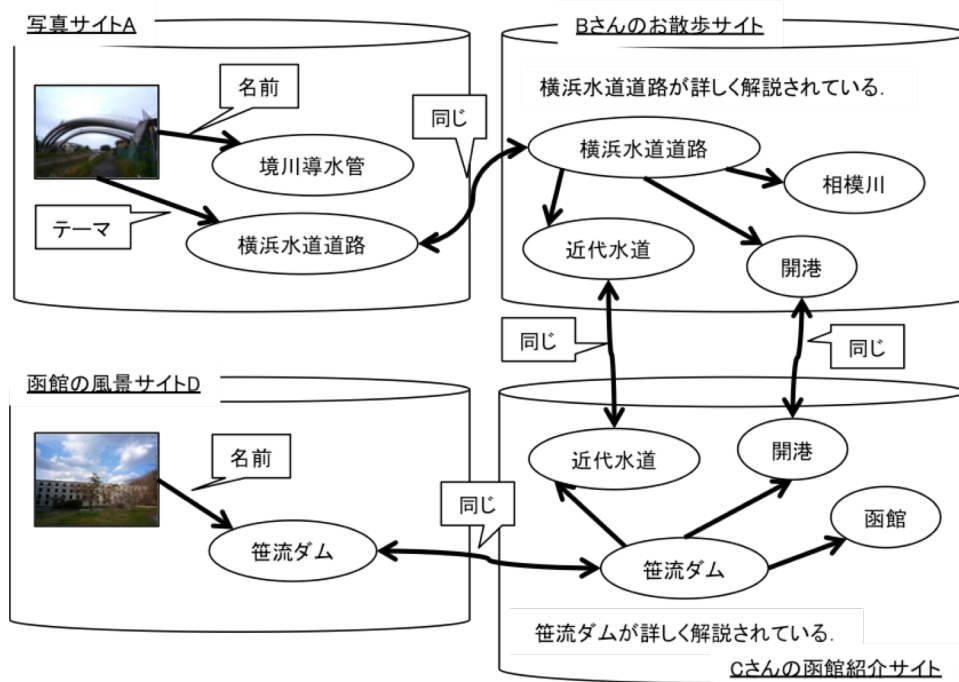


図3 新規観光客誘致のシナリオ概念図

Figure 3 Inviting new customer

推薦システムというよりむしろ想起サービス (reminder)と呼ぶのが近い。

ユーザーに種類があるということを示すため、新規観光客誘致のケースで説明する。

3.1 ハブによる想起サービス(reminder)

このシステムでは、想起のきっかけとなる情報、想起させられる情報、そして想起の連想ルールとなるハブ情報は、本であれ写真であれ、すべて RDF で表現されている。情報を想起させられるユーザーのプロファイルは、ユーザーがシステムに対して予め設定した：(1) 後で想起させて欲しいと設定した情報、(2) 連想ルールとして利用して欲しいと指定したハブ情報、(3) 想起のきっかけとなる情報に「今、さらされた」ことである。

しかし、サービスのユーザーは、情報を想起させられるユーザーだけではない。以下、新規観光客誘致を例に説明する。

3.2 例：新規観光客誘致

観光振興施策において、観光客に新しい旅行先・訪問先を見つけてもらう問題を考える。観光客が伸び悩む、あるいは減少傾向にある一方で、観光客の多くがリピーターである場合には、新しい観光客を誘致する必要があるだろう。旅行情報の入手先としてインターネットが上昇しており、そのような仕組みがインターネット上にあることが望ましい。

地域情報を扱うシステムのユーザーには、情報発信者を

含めて設計する必要がある。地域の、まだ観光客に広く知られていない情報を、身近で豊かなものにするには、住民や業者などが自ら情報発信することが望ましい。

それらの地域情報は、地域外や分野外の情報とも接続可能である方が、観光客に新しい情報を提示するために有利であろう。前述の linked data をオープン・データに適用する Linking Open Data (LOD, または Linked Open Data)が進められており、行政が情報を処理可能なデータとして公開する動きが欧米で盛んである。この処理可能なデータというのが RDF/XML などである。これに対し、PDF や画像で公開する、従来のいわゆる「情報公開」では、公開された情報から再計算などする場合に、再入力が必要となり、活用が阻害されていた。

3.2.1 目的

新規観光客誘致に即して述べると、われわれの目的は、次のような枠組みを提案することである：直接は探していない地域の情報を、システムのユーザーに推薦することができる。推薦される情報は、地域の住民や事業者が発信したものである。推薦される情報は、情報間の関連付けに基づいて選択される。情報は LOD の仕組みで関連づけられ、必ずしも直接的な関連ではない、多段階の関連をたどることもスコープに入れている。推薦された情報はなぜ推薦されたのか、ユーザーに関連が分かる。推薦される情報や関連付けに利用される情報は、提案システムだけでなく他のシステムからも自由にアクセス可能なデータであるとする。

ここで、ユーザーが 2 種類あることに着目されたい。1 つは、新たな観光先を推薦される観光客であり、もう 1 つは観光客を呼びたい地域である。さらにもう 1 つ、システムのユーザーとしてハブ制作者がいる。

水道道路」となっている。

情報発信者は函館市に住んでいて、花見や紅葉狩りに訪れる笹流ダムの写真をネットに投稿した。ダムの部分には「笹流ダム」と名前がついている。

想定するシステムは、ネット上の情報を利用して、「境川

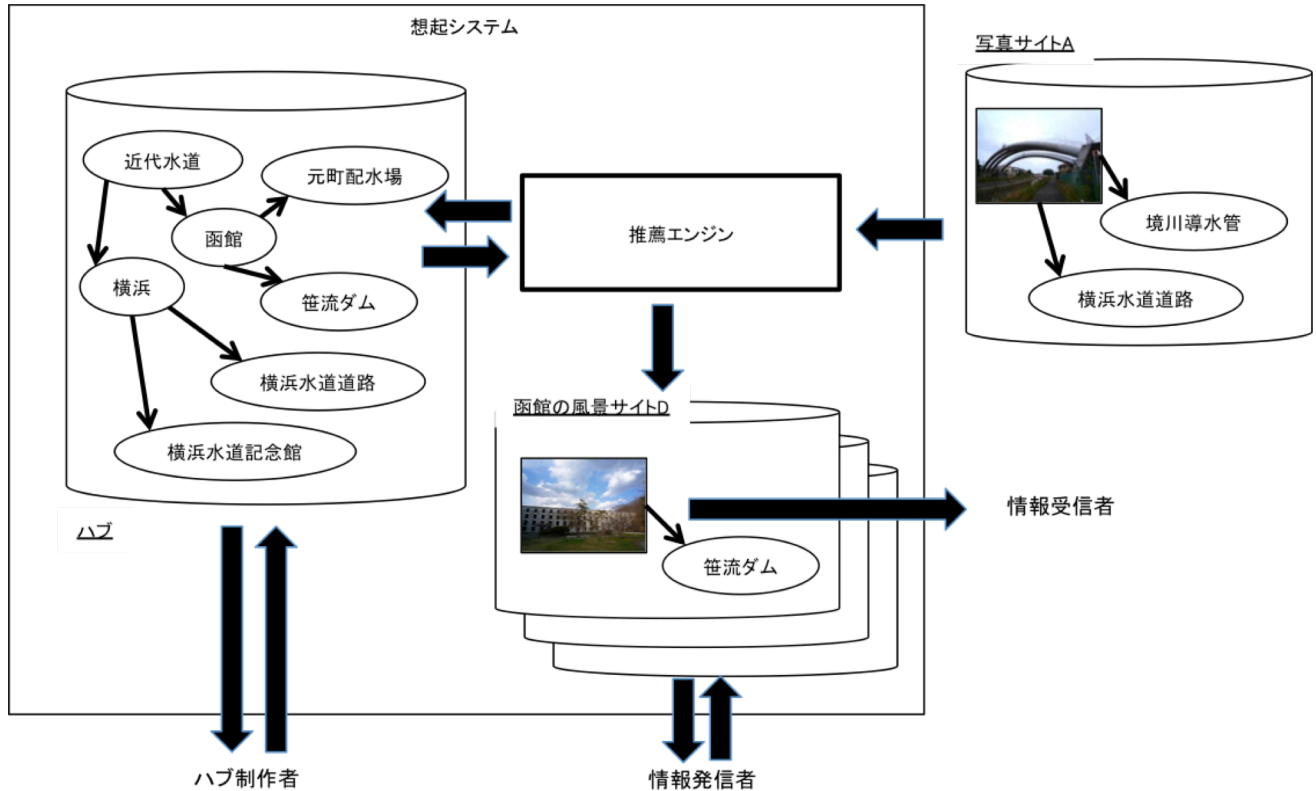


図 4 システム概要

Figure 4 System overview

3.2.2 課題

相互に関連づけられたデータのネットワークをたどって、あるデータ A を元に推薦データを得ようとする、情報が豊であればあるほど候補は膨大となる。適切な枝刈りが必要である。

また、情報は地域の住民などが発信するもので、IT の専門家とは仮定しない。特に情報の関連付けにあたって、理解しやすい UI が必要となる。例えば、ブログの記事や SNS への投稿にタグをつけるように容易な UI である。

3.2.3 シナリオ

横浜の、水道道に興味がある観光客に、函館市の笹流ダムを薦めるケースを想定する (図 3 参照)。

情報受信者は横浜市に住んでいて、近々、函館を入口に北海道旅行を予定している。幕末・明治維新の歴史に興味があり、関連する横浜の史跡にも興味がある。近所には相模川から引いた導水管が境川を渡る導水管があり、ネットの写真サイトでその写真を見つけた。写真中の導水管には「境川導水管」と名前がついており、写真の主題は「横浜

導水管」に興味を持った北海道を旅行する予定の横浜市に人に、いくつかの候補の一つとして「笹流ダム」を提示する。

例えばネットに、これらとは別に、横浜の近代水道の歴史を紹介したサイトや、函館の歴史を紹介したサイトがあると。横浜の近代水道の歴史を紹介したサイトでは、「近代水道」の説明として「横浜水道道路」が「開港」と関連づけて説明されている。函館の歴史を紹介するサイトでは、「笹流ダム」が、やはり「近代水道」と「開港」と関連づけて紹介されている。ここで「横浜水道道路」、「近代水道」、「開港」、「笹流ダム」が、どれも同じコトを指していれば、4 つのデータやサイトをまたがって関連をたどることで、「境川導水管」の写真と「笹流ダム」の写真を関連づけるコンピューター処理が可能だろう。

3.2.4 アプローチ

アプローチのポイントは、目標とするシステムを、図 3 よりもシンプルに、図 4 のようにすることである。

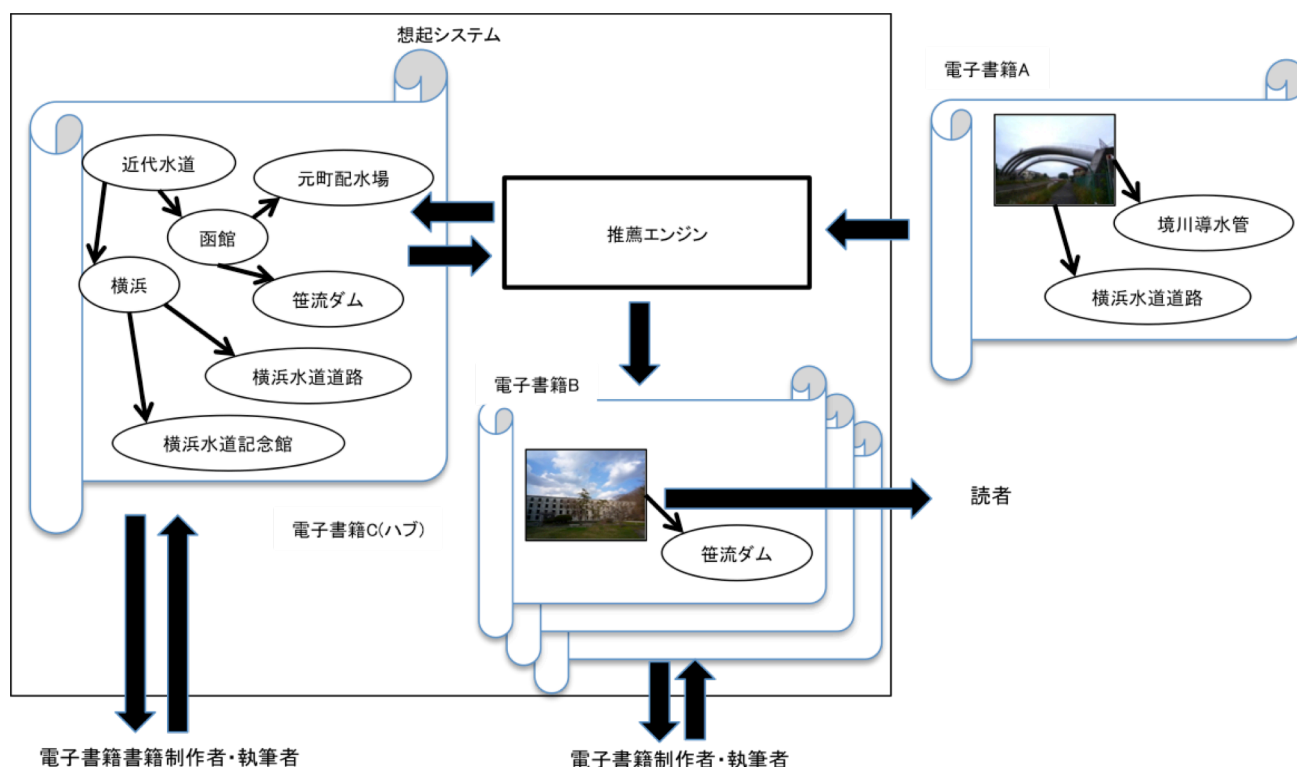


図5 電子書籍による想起システム

Figure 5 Reminder by E-Book

ハブデータ

図3とは異なり、システムは、ある特定の1つのサイト/データ群を特別扱いして、探索の中間経路として利用する。これは、いぜんとして情報発信者が作成したデータであり、特別な辞書のようなものではない。例えば、各自治体のホームページで見られる地域の年表や、TVの歴史ドラマ紹介サイト、あるいは教科書・教材(歴史、地理、音楽、etc.)などである。

推薦エンジン

推薦エンジンは、ユーザーが着目した「境川導水管」とハブサイトと推薦先のデータが存在するサイトを入力として、推薦先データ・サイトから「笹流ダム」などいくつかを選択する。

推薦エンジンも汎用的なものでもなく、適応分野に応じて個性があってよい。このときに、ハブサイトで使われている、特定の関係を重要視するかもしれない。それらは、例えばハブが年表であれば、年代、場所、主題、関連する出来事などである。

推薦エンジンは、ハブを含む3つのサイトと、仕組みとしては別ものであるが、ユーザーには、最初に着目した写真サイトとセットで、あるいは推薦される風景サイトとセットで提供されるかもしれない。

情報受信者

情報受信者は、出発点の写真サイトAで、ある写真に着目することで、関連する情報の推薦を受ける。推薦された中から、気になるサイトを参照するかもしれないし、しないかもしれない。

Eコマース・サイトなどと異なるのは、写真サイトAは、システムの外側にある点である。受信者にとって、推薦される情報はオマケである。情報ショッピングをしているわけではなく、何の推薦もなくとも構わない。Eコマース・サイトでは、何かを調べたり何かを購入したりすれば、その都度他の商品が推薦される。

情報発信者

情報発信者は、自分が発信した情報の利用動向を見ることができる。それ見て発信の仕方を変えることもあるだろう。例えば、「笹流ダム」と名前をつけるだけでなく、「近代水道」という主題を追加することで、ページビューを増やしたいと思うかもしれない。

情報発信者にとって、このシステムは、受信者に「想起させる」システムである。

ハブ制作者

ハブとされた情報の制作者は、自分が制作した情報が推薦に与える影響を見ることができる。それを見て、ハブを書き換えるかもしれない。

以上のように、情報受信者、情報発信者、ハブ制作者の3者の振る舞いによって、その後のシステムの挙動が変わる。とくに、情報発信者やハブ制作者は、作る・発信する情報によって、システムの振る舞いを変えることができる。

4. 電子書籍による想起システム

図4における3種類の情報として、それぞれ独立に電子書籍を利用しようというのが本稿の提案である。

RDFaのようなアノテーション型のメタデータであれば、必ずしも書籍自体を主語とせず **linked data** を表現できるので、図4と同じことが伝書籍でも可能である(図5参照)。

読者

ある読者は水道関連のことがらに興味があり、水道に関連する想起システムに登録している。

読者が横浜史の電子書籍Aを、今まさに読書中だとしよう。読者がその書籍のあるページ/部分にさしかかったときに、電子書籍リーダーから呼び出された想起システムが、その部分に関連する別の水道関連の情報を、読者に提示する。

読者が電子書籍Bを1週間かけて読書中だとしよう。ある写真サイトAを開いたときに、Webブラウザーにインストールした想起システム拡張機能が、想起システムを呼び出して、関連する別の情報を読者に提示する。

電子書籍制作者・執筆者

電子書籍Bの制作者・執筆者は、単に書籍の売り上げが分かるだけでなく、どの部分がどの状況で想起され、さらに読者に参照されたかを知ることができる。これを、改訂の参考にしたたり、別の書籍企画の参考にしたたりするかもしれない。

電子書籍からどの部分を選んでどのようなメタデータを付けるかは、制作者・執筆者の裁量による。

このようなことは、現在、Webサイトの企画・運営では普通に行われていることである。

ハブとしての電子書籍

事典のような書籍は、連想を媒介するハブとして使えるかもしれない。

5. まとめ

電子書籍の内容にメタデータ付きアノテーションを付けることで、情報想起のきっかけとしての起点、想起する内容としての到達点、または連想を中継するハブとして利用

するシステムを提案した。

紙の書籍に対する電子書籍のメリットは、軽量化、検索性、アクセシビリティなどだけではない。電子書籍はシステムが利用可能な電子化知識であり、様々な立場のユーザーが、それぞれの目的に応じて内容を工夫することで、結果的にシステムの振る舞いに影響を与え、システムの質を向上させる可能性もある。

今回は、コンピュータ・システム・レベルで動作することをプロトタイプで確認した。

今後は、ハブなどデータの規模の影響を調査する。また、それぞれのユーザーの立場での評価と、システム全体、およびサービス全体を評価する手法を検討する。

参考文献

- 1) 山口琢, 大場みち子, 高橋: 修地域情報システムにおける **Linked Open Data(LOD)**による住民参加型データ構築手法の提案, 電気学会研究会資料. IS, 情報システム研究会, 2012
Taku Yamaguchi, Michiko Oba, Osamu Takahashi, Consideration for a participatory data construction method with **Linked Open Data(LOD)** for regional information system, 2012
- 2) 中島秀之, 橋田浩一, サービス工学としてのサイバーアシスト-10年早すぎた? プロジェクター, 産業技術総合研究所, *Synthesiology* 3(2), 96-111, 2010
- 3) 電子書籍フォーマットEPUBと日本語組版: 日本でメインストリームにいる人間は国際標準化の舞台ではまず勝てない, 科学技術振興機構, 情報管理 55(1), 13-20, 2012
- 4) 深見嘉明, 小林巖生, 嘉村哲郎, 加藤文彦, 大向一輝, 武田英明, 高橋徹, 上田洋, **Linked Open Data**によるボトムアップ型オープンガバメントの試み, 情報処理学会研究報告 Vol. 2011-DD-79(1), 1-8, 2011
- 5) Tim Berners-Lee, Design Issues: **Linked Data**, <http://www.w3.org/DesignIssues/LinkedData.html>, 2006
- 6) 神崎正英: 「セマンティック HTML/XHTML」, 毎日コミュニケーションズ, 2009
- 7) 土方嘉徳, 嗜好抽出と情報推薦技術(嗜好抽出・情報推薦の基礎理論), 情報処理学会, 情報処理 48(9), 957-965, 2007