

個人の知識を整理するための デジタルドキュメントネットワーク構築システムの提案

松永賢次^{†1} 吉越将紘^{†1} 吉田享子^{†1}

電子書籍の普及によって個人が多くの本をデジタルドキュメントとして所有するようになってきている。また、その他の様々な形式のデジタルドキュメントの所有量も増加の一途をたどっている。これらの大量のデジタルドキュメントが、個人の知識を整理するための有用な知識源となるように、関連するものをリンク付けてネットワークを構築するシステムを提案する。このシステムにより利用者は、複数のデジタルドキュメント間にメモやリンクを作成でき、それらの情報を利用して検索することができる。

A PERSONAL DOCUMENT NETWORK BUILDING SYSTEM FOR ORGANIZING KNOWLEDGE

KENJI MATSUNAGA^{†1} MASAHIRO YOSHIKOSHI^{†1}
KYOKO YOSHIDA^{†1}

The popularity of eBooks has spread in recent years with the launch of eBook readers and more contents. The volume of digital documents as well as eBooks owned by individuals is also increasing, necessitating the development of a method to manage them. This study led to the proposal of a personal network building system for organizing knowledge related to books and documents. This system allows users to mark important sentences of the documents, link between them and attach a memo to a mark and a link. User can search documents in various ways by using memos on marks and links.

1. はじめに

近年、電子書籍の普及によって個人が多くの本をデジタルドキュメントとして所有するようになってきている。また、すでに多くの資料などが様々な形式のデジタルドキュメントとして個人のストレージに保有されており、今後もデジタル化された文書の量は増加の一途をたどると予想される。従来から紙媒体の書籍や資料は、個人の知識を整理するための有用な知識源であり、人類は紙ベースの書籍や資料を利用して知的作業を行ってきた。今後、紙の文書に代わる大量のデジタルドキュメントをもとに、知識を整理したり新しい意味を加えたりする知的作業をすることが求められる。そのためには、従来の作業方法をもとにデジタルの利点を活用し支援するシステムが必要となると考えられる。

電子書籍に関する研究としては、電子書籍への情報を付加するためにアノテーションを利用した知識共有型電子書籍などの研究がある[1]。これは、コメントやブックマークなどの書籍に関連したアノテーションを他のユーザと共有することで、書籍の特定の箇所に関わる議論や疑問の解決や要点抽出など、電子書籍ならではの新しい書籍の読み方を提案しているものである。

我々は、すでに電子書籍のためのツールとして、複数の電子書籍をユーザの用途に合わせて関連付け、並行して読んだり、複数の本から必要なページを抽出して新しい電子

書籍を作成したりするアプリケーションを提案した[2]。これは、電子書籍を利用した学習を想定して、机上に複数の本を並べて関連のあるページを見比べながら作業するという、紙の書籍での作業方法を電子書籍で実現させるためのシステムであった。また、電子書籍だけでなく pdf (Portable Document Format) 形式のファイルを含めた複数のデジタルドキュメント間を横断して検索し、検索結果に基づきリンク機能を用いて新しい電子書籍としてまとめ、その本を閲覧することができるシステムも提案した[3]。しかしこれらのシステムは、紙媒体の書籍のフレキシブルな使用法を電子書籍で実現することを目的に作成されたものであり、デジタルドキュメントの特性や利点を活かした上で、知的作業を支援するための利用法を検討したものではなかった。

本稿では、多くの書籍や資料の文書を扱う研究者にインタビューを行い、知識の整理の方法や新しい知見を加えるために役に立つ方法を検討したうえで、デジタルドキュメントのネットワークを構築するシステムを提案する。

2. 背景

人類は長年にわたって、紙媒体の書籍や資料などを用いて知識の整理や習得を行ってきた。そのため、デジタル化が普及した現在でも、デジタルドキュメントを印刷してからそこにメモを書くなど、紙ベースの作業の習性から抜け出すことが難しい。しかし、すでに Amazon では電子書籍の売り上げが紙の本の売り上げを抜いており、電子書籍の普及が遅れているといわれる日本においても、今後計画さ

^{†1} 専修大学
Senshu University

れている学校の教科書の電子書籍化などが進められればさらに電子書籍の利用の拡大が進むとが予想される[4][5]。また一方、個人がすでに所有している紙の本や資料もスキャナーなどでデジタル化して保存する人が増えている。ストレージ価格の低下により大容量のファイルをローカルに保有することや、クラウド上のストレージも利用可能となっており、大量のデジタルドキュメントを保有する環境が整っている。これらの個人所有のデジタルドキュメントについては今後、どのように持つかという管理方法と、持ったものをどう使うかという利用方法が問題となると考えられる。

デジタルドキュメントを整理・閲覧するためのソフトウェアとしてはすでにいくつかの電子書籍や pdf のリーダーや関連するアプリケーションが提供されている。それぞれのリーダーには、単に文書を読むだけではなく、いくつかの編集用の機能が付いている。電子書籍では、しおりをつける、デバイス間でページを同期する、辞書を使う、メモを貼り付ける、マーカーを引く、SNS で共有する等が可能である。pdf ファイルの編集方法としては、ノート注釈、テキストのハイライト表示、しおり（ブックマーク）、文書校正、図形描画、塗りつぶし、画像挿入、図形描画、リンクの作成などが可能となっている。また管理ソフトでは、ワープロや表計算ファイルへの逆変換、ページの入替え、複数の PDF ファイルを目次形式でグループごとにまとめ管理することなどができる。また、文書にメタデータとしてタグをつけて検索や管理に利用している。しかし、これらのものは、主にひとつの書籍や文書を読んでコメントをつけるなどの作業に重点を置いて作られており、複数の書籍を読んでその関連を考察しながら知識を組み立てるような読み方には対応していない。本稿では、書籍や資料を扱う研究者を対象として、多くの書籍や資料をデジタルドキュメントとして扱う研究を支援するシステムを提案する。

3. 多くの文書を文献として用いる研究の方法

今回は、広い研究分野の中から特に多くの書籍や資料を扱って研究している文系の研究方法について調査した。そのために、日本文学や日本古代史を研究している大学院生 2 名に、研究の方法や書籍の利用法についてインタビューを行った。インタビューの内容としては、情報の収集・管理の方法、資料を用いての作業内容、デジタル化された資料で使用方法について質問し以下の回答を得た。

(1)情報の収集・管理方法

- ・現在、古典の書籍の出版についてはまだデジタル化が進んでいるとはいえないが、研究者個々で pdf にするなどの作業はかなり行われている。
- ・本を優先として集めているが、電子化されている論文やデータベース化された資料も増えてきているのでその情報も利用している。集める本の数は、研究のテーマによって

異なり、数冊から数千冊まで様々である。

- ・紙の本の必要な部分をコピーしてファイリングして使っている。古書をスキャナーで pdf ファイルにして保存している人も多い。

- ・集めた資料を、自分のいくつかのテーマにそって、分類してファイルを分けて保存している。テーマが A,B,C とある場合、A,B,C の 3 つに関連するものは、3 つコピーして保存している。ただ、それを後で探すのが大変である。

(2)資料を用いての作業内容

- ・本から自分の研究に必要な部分を探し、アンダーラインを引き付箋を貼る。メモなどを書き込むこともある。

- ・他の本や資料と関連する場合も、同様にメモを書き込む。

- ・関連する資料は、書籍だけでなく写真などの画像の場合もある。

- ・メモとして書く内容の属性は分野によって重要視されているものが異なる。

- ・時代や日付で並び替えたいときは、表計算ソフトなどを使用している。そのための入力作業や並べ替え作業がかなりの負担となっている。

(3)デジタル化された資料で使用方法

- ・文書の重要なところにメモを書き込みたい。それを見て、簡単に思い出しの作業ができ、早く整理ができるようにしたい。

- ・1 つの資料に対して関連ある論文を抽出し、それらの解釈の違いを表示できるとよい。

- ・本と本の関連ある部分がリンクづけされ、そこに情報が付加できるとよい。

- ・検索された場所からもとの本や資料の前後の文章も読めるとよい。

- ・付加された情報を年代別、関連人物別、関連ワード別など、色々な視点から検索できるとよい。

インタビューの結果から、文系の研究では「書籍や資料などの情報を集め、それらの中から必要なものを選択し、それらの中にある関係を見つめ、そこに新しい解釈や知見を加える」ことが重要と考えられた。そのためには、デジタル化した書籍や資料に研究者がメモを書き、そのメモ同士を関連づけて新しいネットワークを形作る、またはすでに得られているネットワークの形に新しい知見を与えるなどの作業を支援できるシステムが必要と考えられる。

4. システム概要

本稿では、個人の知識を整理する作業を支援するデジタルドキュメントネットワーク構築システムを提案する。本システムを用いて、デジタルドキュメントから抽出した文章（以下、マークと記述）にメモをつけ、マークとマークをリンクで結びつけることで、ネットワークを作成できる。リンクにもメモをつけることができる。また、それぞれのメモは、「キー：値」の形式で保存され、異なる研究分野に

あわせたキーを設定することができる。値としては、数値やテキスト以外に、URL でファイルや Web ページを設定すれば、画像ファイルやインターネットの資料と関連付けることも可能である。作成されたネットワークを利用して、マークのメモやリンクのメモを使って検索することが可能となる (図 1)。

4.1 システムの機能

本システムの機能には、「ネットワークの作成」と「ネットワークの検索」の 2 つがある。

ネットワークは、「マークをつける」ことと「マーク間にリンクの関係をつける」ことの 2 つを行うことで作成される。まず、個人のストレージにある電子書籍や資料から、検索キーワードなどを用いてマークをつけるデジタルドキュメントを抽出する。その中からマークしたいページの文章を選択し、そこにマーク (マーカーでいくつかの文章をマーキングする、または範囲指定する) をつける。このマークにはメモをつけることができる。メモをつけた日付、関連事項、関連人物、コメント文など「キー: 値」形式で入力できる。マーク間にリンクの関係をつけるためには、マークの中からリンクをつけたいものを 2 つ以上選択し、その間にリンクを設定する。リンクにもメモをつけることができ、メモの内容はマークにつけたメモと同様のものを設定できる (図 2)。

電子書籍や資料に対して作成されたネットワークはメモの情報をもとに検索することができる。まず作成された複数のメモにつけられた内容から、検索キーワードなどを用いてメモを選ぶ。マークに付いたメモの場合は、メモがつけられたマークされている文章を表示する。リンクについているメモの場合は、リンク先の 2 つ以上の文章を表示する。表示されたものは、マークされている文章の箇所だけでなく、その前後の文章やページを見ることが可能である。また、選んだメモに関連付けられたすべての書籍と資料についてネットワークの全体像を概観することも可能である (図 3)。

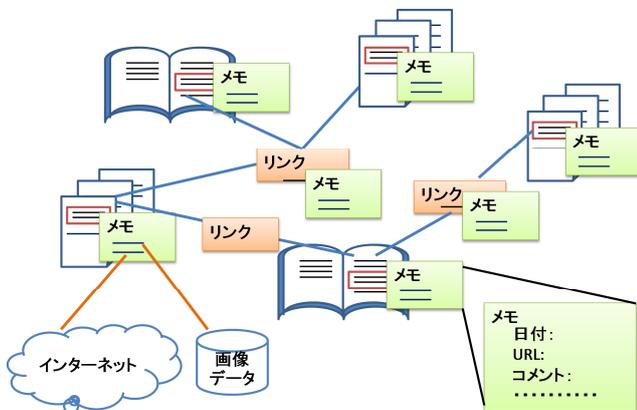


図 1 デジタルドキュメントネットワーク

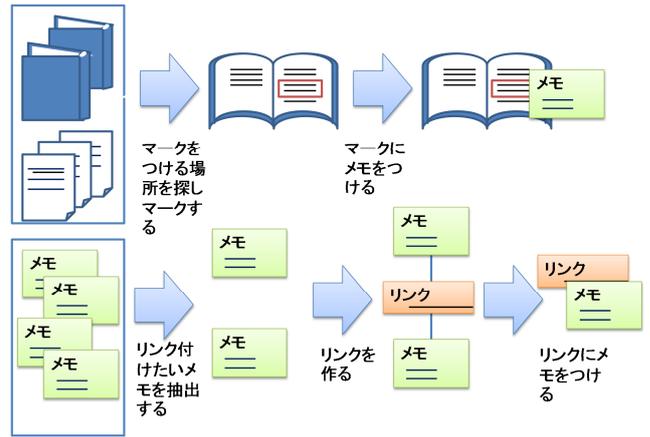


図 2 ネットワークの作成

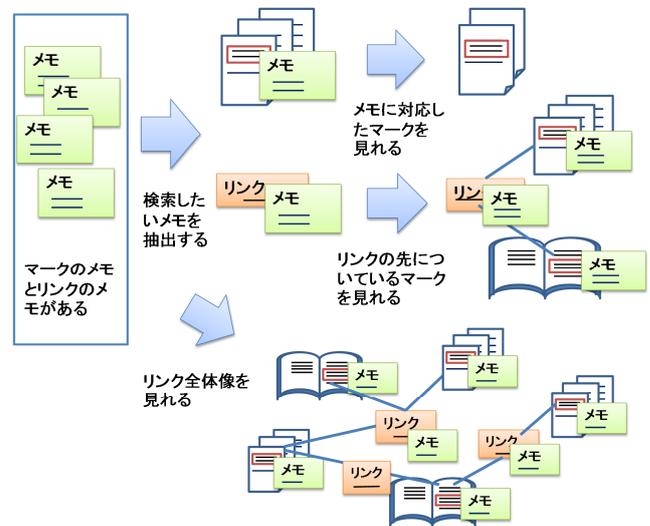


図 3 ネットワークの検索

4.2 システムの画面

システムの機能を利用者が容易に利用できるよう、画面案を作成した (図 4)。

一番左には、文書を整理するタブと、マークとリンクを整理するタブを用意している。これにより、文書を起点にするルートと、マーク及びリンクを起点にするルートの両方から情報にアクセスできるようにしている。

図 4(a)は、ある文書を選び、その文書を読み進め、マークしたい範囲を選び、それに対するメモを記述する利用画面を示している。

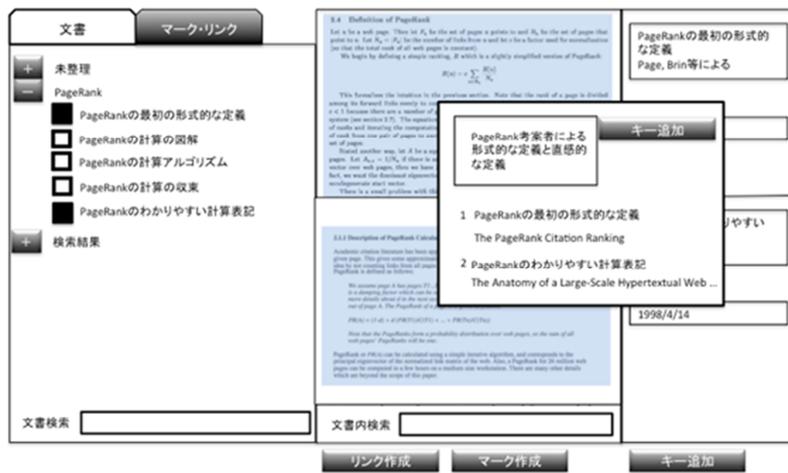
図 4(b)は、いくつかのマークが既にある状況で、マークを 2 つ選び、それらに対するリンクを作成し、メモを記述する利用画面を示している。

ネットワークを作成する (リンクを付ける) 作業と、ネットワークを検索する (リンク対象のマークを探す) 作業を、利用者が交互に繰り返していくので、その際に、利用者が混乱しないようなインターフェースになるようにした。



文書を選択 → マークする箇所を選択 → マークのメモを作成

(a) 文書の中からマークを作成し、マークに対するメモを記述



マークを複数を選択 → リンク作成をポップアップ → リンクのメモを作成

(b) リンク付けしたいマークを選択し、リンクを作成しメモを記述

図 4 システムの画面案

4.3 システムの構成

システムの構成は、図 5 となる。電子書籍の形式としては、プロプライエタリなものとして Amazon の Kindle が採用する「Topaz」、ボイジャーの「book」などや、オープンなものとして IDPF が推進する「ePUB」が挙げられる。本システムでは、扱う電子書籍の形式は特に限定していないが、デジタルドキュメントとしては pdf 形式を想定している。それらの文書を入力として書式変換処理を行い、本システム内で用いる同じ形式のファイルを出力する。本システム内の統一したファイルとしては、ePUB 形式を共通のフォーマットとしている。また書式変換時に、文書に関するメタ情報のファイルも作成する。マーク作成・編集処理では、これら 2つのファイルを参照し、結果をデータベースに格納する。リンク作成・編集、メモ作成・編集処理では、データベース内のマーク・リンク・メモの情報を利用し結果をデータベースに保存する。また同時に、マーク・リンク・

メモに関するメタ情報も作成する。マーク・リンク・メモのデータベースを参照しながらリンク作成・編集と、メモ作成・編集処理を行う。検索処理では、ePUB 形式の共通フォーマットファイル、文書のメタ情報、マーク・リンク・メモや、マーク・リンク・メモのメタ情報を参照しながら、検索結果を表示する。

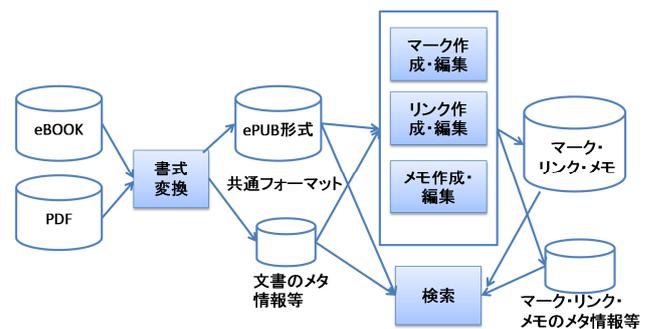


図 5 システム構成

5. 考察

今回は、文系の研究利用の視点から本システムを提案したが、ここでは、本システムとWebを比較して今後の可能性を検討する。

現在、インターネットの発達により、Webの中には多くの知識や情報が集積されている。複数のページによって構成されるWebサイトは、1冊の本として捉えることもできる。Webサイトと電子書籍を比較してみると、使用されている技術は基本的にどちらもHTML/CSS/JavaScriptといったWebの技術であり大きな違いはない。異なる点としては、Webサイトは常時内容の変更が可能であり最新の情報を得ることができるが、電子書籍は完成された出版物であり、内容を変更することを前提とはしていない。電子書籍は、編集者による校正を経た本として、あるテーマに沿って体系的にまとめられている書作物であり、一般的には著作内容について筆者や責任の所在が明確で信頼性はWeb情報よりも高い。特に、初版から長く読み継がれたものは読者による評価を受けており、専門書などは知識がまとまって体系化されており価値が高いと考えられる。

また、Webの特性であるリンク機能を見ると、Webサイトはサイト内の別のページへ自由に移動でき、別のサイトのページへリンクすることもできる。しかし、Webページのハイパーリンクは一つのWebページから他のWebページへと1対1で1方向へ結ぶものである。また、リンクの参照先がページ全体ないしはページ内の1箇所しか指定できない。ページ内の一部の文章をマークして、その部分全体にリンクの関係をつけることはできない。本システムの場合は、リンクは1つの本の抽出文章から関係する複数の本の抽出文章へ関連付けることができ1対多で双方向のリンクを設定することや、参照したい文章の範囲をマークとして後から定義することもできる。それによって、リンクやメモの情報を用いながら、1冊の本だけにとどまらず複数の書籍の間を相互に行き来することが可能となる。Webの情報がリンクで結ばれることで新しい価値が付加された様に、電子書籍の情報もリンクで結ばれることで新しい知見を得たり、新しい価値が付加できる。

また、Webは多くの人によって作成された情報であるため、個々に表現の仕方が異なり、文書の意味を統一するのが難しいという問題がある。本システムで作成されたネットワークは、個人の書籍や資料に限定されたクローズされた空間である。コメントとして記述されるメモの内容も個人が書くもので内容や言葉の意味は統一されている。自分が選んだ書籍・資料類、選んだ文章、自分の書いたメモでWebとは違うコンテンツの一貫性があると考えられる。

マーク・リンク・メモの情報は、個人が所有している書籍につけられたものであるが、これらの情報自体を利用することも可能である。たとえば、友人や同じ研究室内でマ

ーク・リンク・メモの情報を共有することによって、先輩から後輩への研究の継承に利用することが可能となる。また、本システムを小グループで使ったり研究者間での議論に利用したりすることもできる。

作成された書籍のネットワークについては、研究のための書籍で作成されたものだけでなく、趣味や他の仕事に関する書籍を対象として別の書籍ネットワークをいくつでも構築することができる。紙媒体の書籍は、本棚に分野ごとに分類・整理されて保存されることが多く書籍間の関連までを把握することは難しい。しかし、本システムを用いれば、書籍の文章と他の書籍の文章を関連付けることが容易になり、紙媒体では難しい新しい分類・整理の方法が容易となる。

6. おわりに

今回は、多くの書籍を集め、それらの中から知識を体系だてさらに新しい知見を見つけ出そうとしようとしている文系の研究に焦点をあて、書籍の利用法や整理の方法を調べた。そのうえで、デジタルドキュメント化された書籍や資料を使用した研究に役立つシステムについて提案した。今後は、この内容をもとにさらに文系の研究者の意見を聞き、試作システムを作成した上で、システム評価のための実験を行う予定である。

また今回は文系の研究者を対象として1つのフレームワークとして提案したが、他の分野についても研究の方法を学んでシステムに取り入れることも予定している。

参考文献

- 1) 秋山博紀, 安村通晃: アノテーション付加による知識共有型電子書籍の提案, 情報処理学会研究報告, HCI, ヒューマンコンピュータインタラクション研究会報告, 2011-HCI-142(13), 1-8, 2011-03-10.
- 2) 吉越将紘, 久住あも, 天野ほのか, 吉田享子: 複数の本を関連付けて読むための電子書籍アプリケーションの提案, 情報処理学会研究報告, 情報システムと社会環境研究会報告, 2012-IS-119(12), 1-6, 2012-04-15.
- 3) Masahiro Yoshikoshi, Kenji Matsunaga, Kyoko Yoshida: A Personal Document Network Building System for Digital Document Searches, C. Stephanidis (Ed.): Posters, Part II, HCII 2013, CCIS 374, pp. 458-461, 2013.
- 4) Amazon Press Releases Jan 27, 2011; Amazon.com Announces Fourth Quarter Sales up 36% to \$12.95 Billion <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=1521090&highlight=>
- 5) 全小学校に「デジタル教科書」(日本経済新聞 2012年2月10日) http://www.nikkei.com/article/DGXNASFB1003Y_Q2A210C1L83000/