

**科学研究費助成事業 研究成果報告書**

平成 28 年 6 月 20 日現在

機関番号：12102

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25330124

研究課題名(和文) EPUB3.0を核とした知識集積型ソーシャルリーディング基盤に関する研究

研究課題名(英文) A study on knowledge-accumulating social reading system based on EPUB3.0

研究代表者

天笠 俊之 (AMAGASA, Toshiyuki)

筑波大学・システム情報系・准教授

研究者番号：70314531

交付決定額(研究期間全体)：(直接経費) 3,600,000円

研究成果の概要(和文)：電子書籍の普及とともに、電子書籍リーダを利用して、同一の書籍に関する読書体験の共有やアノテーションによる知識の集積が期待されている。そこで本研究では、電子書籍のオープンな規格であるEPUB3.0を対象に、知識集積型ソーシャルリーディング基盤に関する研究を行った。具体的には、1) EPUB3.0関連技術を活用したソーシャルリーディング基盤およびクライアントの開発、2) EPUBCFIを用いたEPUBに対するアノテーション方式の開発、3) EPUBリーダを対象としたログ取得方式の開発およびログを用いたユーザの読書行動マイニング、4) 半構造データに対する効率的な処理手法の開発を行った。

研究成果の概要(英文)：Due to the recent diffusion of electronic books (e-books), it is expected that readers can share reading experiences, annotations, and/or knowledge regarding different publications with the help of e-book readers. For this reason, in this research, we have developed a social reading system for EPUB3.0, which is an open standard for e-books, that allows knowledge accumulation. More precisely, we have performed the following researches: 1) development of social reading system and client exploiting EPUB3.0 and its related standards, 2) an annotation scheme for EPUB publications using EPUBCFI, 3) development of user log extraction method and data mining of user activities using the logs, and 4) development of efficient processing techniques for semi-structured data.

研究分野：データ工学

キーワード：電子書籍 EPUB ソーシャルリーディング EPUBCFI アノテーション

### 1. 研究開始当初の背景

(1) Amazon を始め、楽天やその他の企業が電子書籍を提供している。利用者は専用端末やスマートフォン、タブレット端末等を用いて、電子フォーマットで提供される書籍を読むことができる。電子書籍の普及とともに、一般の書籍に加えて、教科書や電子マニュアル、技術文書等の展開が期待されている。さらに、XHTML 等のオープンな規格のみから構成された電子書籍フォーマットである EPUB3.0 がリリースされ、今後急速な普及が見込まれる。

(2) 電子書籍は、単に読書を電子化するだけでなく多くのことが可能になる。例えば、電子書籍リーダーを利用して、同一の書籍に関する読書体験の共有やアノテーションによる知識の集積等が可能になる。

(3) さらに電子書籍リーダーによって利用者の利用行動や時間、場所等のコンテキスト情報が取得できる。これにより、利用者の読書行動を利用者のコンテキスト情報と組み合わせることによって、読書行動のより詳細な分析が可能になる。

### 2. 研究の目的

(1) 以上の背景を受け、本研究では、電子書籍のオープンな規格である EPUB3.0 を対象に、知識集積型ソーシャルリーディング基盤に関する研究を行う。

### 3. 研究の方法

(1) 以上の目的を達成するために、以下の項目について研究を行う。

(2) EPUB3.0 および関連技術を活用することによって、ソーシャルリーディング基盤およびクライアントを開発する。

(3) ソーシャルリーディング基盤において、EPUB 書籍に対するアノテーションを共有することによる知識集積を可能にするため、アノテーション化に r 方式に関する検討を行う。特に、EPUB 関連規格である EPUBCFI を活用する。

(4) 開発するソーシャルリーディング基盤および EPUB クライアントを対象とした、利用者の読書行動ログ取得方式を開発するとともに、ログを用いたユーザの読書行動マイニング手法を開発する。

(5) 電子書籍を含む半構造データに対する各種処理・マイニング手法に関する研究開発を行う。

### 4. 研究成果

(1) 本研究では、EPUB を対象とした電子書籍向けの新たなソーシャルリーディングシス

テムを開発した。概要を図 1 に示す。

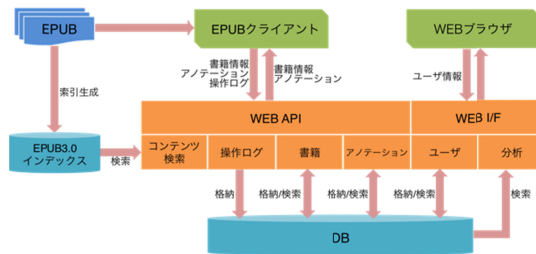


図 1 システム概要

ユーザは電子書籍リーダー (EPUB クライアント) を用いて読書を行なう。また Web ブラウザを使う場合があるが、これはユーザ自身の情報を格納・検索、もしくは、Web とのインタフェースのためのものである。EPUB クライアントは EPUB 出版物を読むだけでなく、ブックマークやアノテーションなどの読書情報の書き込みも行なう。読書情報は Web API を通じてデータベースへ格納される。さらに読書情報はユーザ間で共有される。他のユーザの読書情報の取得にも Web API を用いる。サーバサイドでは、ブックマークやアノテーションなどの収集した読書情報を保持するだけでなくさまざまな分析を行うことができる。また EPUB 出版物からインデックスを生成する。

(2) クライアントではユーザ、書籍、ブックマーク、アノテーション、操作ログについてそれぞれ操作が発生する。操作情報はクライアント側で保持する。これはオフラインでも読書を行なうことができるようにするためである。保持した情報はユーザの操作により送受信が行われる。本システムでは 1) ユーザ情報の送信、2) 書籍情報の送信・受信、3) ブックマークの作成・編集・削除・送信・受信、4) アノテーションの作成・編集・削除・送信・受信、5) 操作ログの生成・送信の機能を持つこととした。図 2 に開発したプロトタイプのスクリンショットを示す。

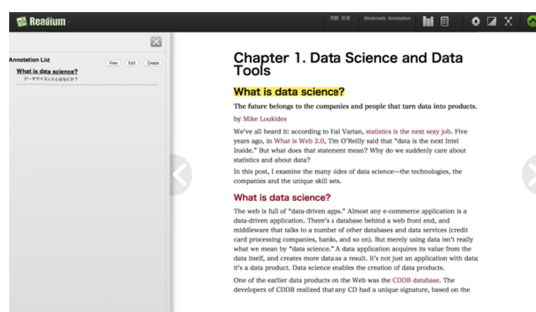


図 2 クライアントのスクリンショット

(3) 開発したプロトタイプシステムを用いて実験を行った。実験では、O'Reilly Radar Team 著 Big Data Now を用い、グループによ

る輪講を実施し、その間の利用者の読書行動を取得した。取得したデータを元に、利用者の読書行動に関する分析を行った。具体的には、1)ブックマークおよびアノテーション位置による利用者の興味を引くコンテンツの同定、2)操作ログを分析することによる利用者ごとの進捗分析、3)利用時間の分析。以下ではいくつかの結果を示す。

(4) まず、ブックマークおよびアノテーションの位置を集計することにより、利用者が興味を持つコンテンツが分かるかどうかを検証した。結果を図2に示す。図ではブックマークされた箇所をコンテンツの区間毎に示している。結果から、ブックマークが特定の位置に集中しており、利用者の興味を引くコンテンツを示している可能性があることが示された。

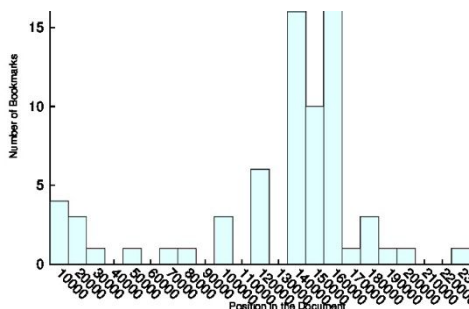


図2 ブックマーク位置の分析

(5) 次に利用者の読書進捗の結果を示す(図3)。横軸は時間、縦軸は書籍内の位置を表している。図から、時間に伴う読書の進捗をみる事ができる。

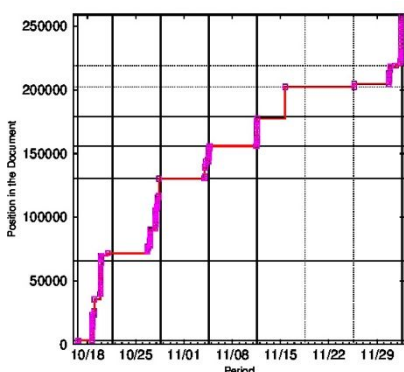


図3 読書進捗分析の結果

(6) 本研究では、ソーシャルリーディングシステムやクライアントの開発、ユーザ行動マイニング手法に加えて、ソーシャルリーディングに関係する種々の研究開発を行った。

EPUB 書籍を対象とした EPUBCFI に基づく索引手法。

木構造データに対する類似部分木検索の高速化。

グラフにおける PageRank/ObjectRank の高速な推定手法。

半構造データストリームに対するパース式を考慮したキーワード検索手法。

文献データベースにおけるトピック変遷の検出手法。

## 5. 主な発表論文等

〔雑誌論文〕(計2件)

伊藤寛祥, 天笠俊之, 北川博之, “引用情報を利用した論文データベースにおけるトピックの変遷の検出”, 日本データベース学会和文論文誌 Vol. 14, No. 13, 2016年3月(査読あり)。

坂倉悠太, 山口祐人, 天笠俊之, 北川博之, “部分グラフに基づく効率的なPageRank 推定”, 日本データベース学会論文誌, Vol. 13-J, No. 1, pp. 32-39, 2014年10月(査読あり)。

〔学会発表〕(計14件)

丸山純, 天笠俊之, 北川博之, “EPUB 出版物のための EPUBCFI に基づく索引の提案”, 情報処理学会第78回全国大会, 6ZD-07, 神奈川県横浜市, 2016年3月10日~3月12日。

小柳涼介, 天笠俊之, 北川博之, “木構造データに対する類似部分木検索の並列化”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2016), D5-4, 福岡県福岡市, 2016年2月29日~3月2日。

Eri Kataoka, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A Social Reading System for EPUB Publications," Proc. of the 17th International Conference on Information Integration and Web-based Applications & Services (iiWAS 2015), Article No. 6, Brussels, Belgium, December 11-13, 2015.

Yuta Sakakura, Yuto Yamaguchi, Toshiyuki Amagasa, Hiroyuki Kitagawa, "An Improved Method for Efficient PageRank Estimation," 25th Int'l Conf. on Database and Expert Systems Applications (DEXA 2014) (2), pp. 208-222, Munich, Germany, Sept. 1-5, 2014.

伊藤寛祥, 天笠俊之, 北川博之, “非負値行列分解を用いた論文データベースにおけるトピック変遷の検出”, 情報処理学会第77回全国大会, 5N-03, 京都府左京区吉田本町 2015年3月17日~3月19日。

伊藤寛祥, 天笠俊之, 北川博之, “論文データベースにおけるトピックの変遷の検出”, F3-6, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2015), 福島県郡山市, 2015年3月2日~3月4日。

片岡えり, 天笠俊之, Franck Gass, 北川博之, “ EPUB を対象としたソーシャルリーディングシステムにおけるユーザ分析”, D2-2, 第7回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2015), 福島県郡山市, 2015年3月2日~3月4日.

小柳涼介, 天笠俊之, 北川博之, “ 構造情報の再利用による XML データに対する類似検索の高速化”, 第13回情報科学技術フォーラム (FIT2014), D-001, 茨城県つくば市, 2014年9月3日~5日.

小柳涼介, 天笠俊之, 北川博之, “ 大規模 XML データにおける効率的な重複データ検出”, 情報処理学会第76回全国大会, 3N-8, 東京都足立区, 2014年3月11日~3月13日.

小柳涼介, 天笠俊之, 北川博之, “ テキストおよび構造の類似度に基づいた XML データに対する効率的な類似検索”, D7-4, 第6回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2014), 兵庫県淡路市, 2014年3月2日~4日.

坂倉悠太, 山口祐人, 天笠俊之, 北川博之, “ 部分グラフに基づく効率的な PageRank 推定”, D6-4, 第6回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2014), 兵庫県淡路市, 2014年3月2日~4日.

Savong Bou, Toshiyuki Amagasa, Hiroyuki Kitagawa, “ Path-based Keyword Search over XML Streams”, 第6回データ工学と情報マネジメントに関するフォーラム (DEIM フォーラム 2014), 兵庫県淡路市, A1-5, 2014年3月2日~4日.

Eri Kataoka, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A System for Social Reading based on EPUB3" (short), The 15th Int'l Conf. on Information Integration and Web-based Applications & Services (iiWAS2013), pp. 72-76, Vienna, Austria, Dec. 2-4, 2013.

Yuta Sakakura, Yuto Yamaguchi, Toshiyuki Amagasa, and Hiroyuki Kitagawa, "A Local Method for ObjectRank Estimation," The 15th Int'l Conf. on Information Integration and Web-based Applications & Services (iiWAS2013), pp. 92-101, Vienna, Austria, Dec. 2-4, 2013.

## 6. 研究組織

### (1) 研究代表者

天笠 俊之 (AMAGASA TOSHIYUKI)

筑波大学・システム情報系・准教授