

EPUBCFIを用いた読書関連情報の管理と検索

片岡 えり[†] 天笠 俊之^{‡§} 北川 博之[‡]

[†] 筑波大学情報学群情報科学類 [‡] 筑波大学システム情報系情報工学科

[§] 宇宙航空研究開発機構宇宙科学研究所学際科学研究系

1 はじめに

ここ数年、ネットワーク上で様々な情報を公開、共有するソーシャルネットワーキングサービスが数多く登場し、その一環としてソーシャルリーディングが行われている。ソーシャルリーディングサービスは、インターネット上で読書情報を多人数で共有する機能を提供する。既存のサービスとしてはブックログ¹、booklook²などがあり、いつどこで誰がどのような本を読んだのか、読書の進捗状況、読んだ本に関する評価やレビューなどの情報が共有されている。

その一方、スマートフォンやタブレット端末、電子書籍リーダーの普及に伴って電子書籍が現在急速に浸透しつつある。電子書籍のフォーマットのひとつにEPUB[1]がある。EPUBの現在のバージョンであるEPUB3.0[2]ではコンテンツの識別子としてEPUBCFI[3]が策定されている。EPUBCFIを用いることで、電子書籍内の任意の一点を唯一に識別することが可能となる。このため、CFI同士を比較することでコンテンツ内における順序関係を特定することができる。また、テキストだけでなく画像、音声、動画に対しても有効であるといった特徴があり、さまざまな応用が考えられる。

今後、電子出版物の普及に伴い、電子書籍を対象にしたソーシャルリーディングの取り組みが高度化することが考えられる。そこで本研究では、電子書籍リーダーを利用した読書情報の共有を目的としたシステムを提案し、読書関連情報の管理と検索の手法について議論する。

2 基本事項

2.1 EPUB

EPUB[1]はIDPF[4]が定めた電子出版物のフォーマットである。EPUBはWebで用いられている標準規格、例えばXML³、HTML5⁴、CSS⁵などを利用しているため、誰でも自由にEPUB形式の電子出版物を作

ることが可能となっている。EPUBの現在のバージョンはEPUB3.0[2]であり、2011年10月にIDPFによって提唱された。

EPUBの特徴の一つとしてリフロー機能が挙げられる。リフローとは文書が画面サイズに応じて自動的に最適化される方式のことである。ページが可変であるという想定のもとでコンテンツが作成されるため、例えば文字サイズの変更や端末の向きの変更などがあった場合や画面サイズの異なる端末で表示した場合に、最適なレイアウトに自動的に変更される。これにより、読み手の最適な環境で読むことが可能となる。

2.2 EPUBCFI

EPUBCFI[3]はCanonical Fragment Identifierと呼ばれる部分文書識別子のことで、IDPF[4]によって2011年10月に策定された。これはEPUB出版物のある部分を指し示すことができ、出版物の中のある一点に対するCFIは一つしかないことが保証されている。また、CFIのみを比較することで文書中における順序が判別できる。さらにテキストだけでなく画像や音声、動画についても詳細な場所、時間について参照することができるため、さまざまな応用が考えられる。EPUBCFIは

```
#epubcfi(/6/4!/4/8[chap4]/24/12/15:3)
#epubcfi(/6/8!/4/2/11:30~23.5@5.75:97.6)
```

のような文字列で表現される。

3 読書情報管理システム

本研究では、EPUBCFIを利用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提案し、読書関連情報の管理・検索を行う。EPUBCFIを用いることでEPUBコンテンツ内の任意の場所を一意に識別することができるため、例えば現実世界における付箋を貼ったり書き込みをするような行為を、電子書籍上で実現できる。さらにそれらを多人数で共有したり、検索・マイニングを行うことにより、現在では不可能なリッチな読書体験の共有を実現することが期待される。

読書情報管理システムの前提として、EPUB出版物はユーザによって購読され、EPUBクライアントを利

Management and retrieval of reading related information using EPUBCFI

Eri Kataoka[†](erikata@kde.cs.tsukuba.ac.jp),

Toshiyuki Amagasa[‡](amagasa@cs.tsukuba.ac.jp) and

Hirokyu Kitagawa[‡](kitagawa@cs.tsukuba.ac.jp)

[†]College of Information Science, University of Tsukuba

[‡]Division of Information Engineering, Faculty of Engineering, Information and Systems, University of Tsukuba

[§]Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency

¹<http://booklog.jp/>

²<http://booklook.jp/>

³<http://www.w3.org/TR/xml/>

⁴<http://www.w3.org/TR/html5/>

⁵<http://www.w3.org/Style/CSS/>

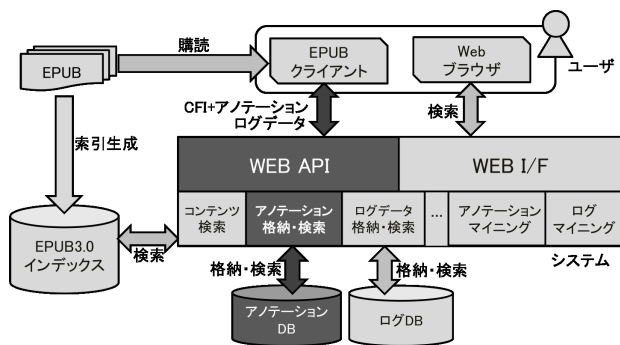


図 1: 読書情報管理システムの概要。

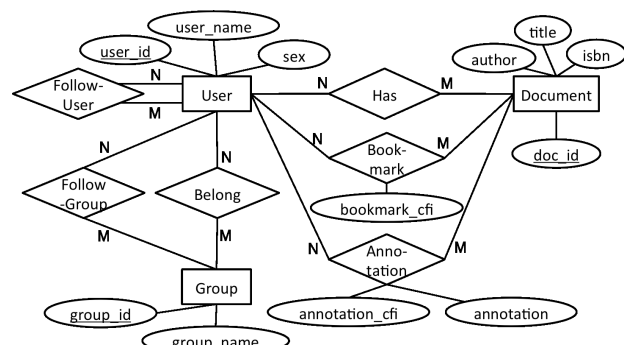


図 2: データベーススキーマ。

用して読まれることを仮定する。また、EPUB クライアントを用いてアノテーションの作成や検索を行うだけでなく、Web ブラウザを用い Web インタフェースを通じて各種アノテーション検索や他ユーザのフォローといったソーシャルなアクティビティを行うこともできる。システムの概要を図 1 に示す。

このシステムでは、EPUB3.0 出版物のファイルが利用できる場合はそれをデータベースに登録するとともに、内容に関する索引を生成する。これはコンテンツに対する検索を行なうために利用されるが、ここでは詳しく述べない。EPUB クライアントからは、EPUBCFI によって識別される読書関連情報、すなわちアノテーションやブックマークが送信される。さらに、EPUB クライアントの操作状況をログデータとして取得する。取得したアノテーションとログは、それぞれ Web API を通じてアノテーションデータベース、ログデータベースに格納される。データベースに格納されたデータは、必要に応じて Web API から利用される。これにより EPUB クライアントからの読書情報の共有が可能となる。収集した各種データについては、各種マイニングを行なうことによって、有益な情報を発見することも考えられる。

本研究では図 1 で示した読書情報管理システムのうち 1) EPUB クライアントからのデータ取得、2) Web API 設計、3) 取得したデータのデータベースへの格納、4) 取得したデータの検索について扱う。

4 実装

本研究では、図 1 で示した読書情報管理システムのうち黒色で示した箇所について実装を行った。本研究におけるソーシャルサービスの実装には Radium[5] を用いた。Radium は IDPF[4] と支援企業によるプロジェクトで、EPUB 出版のためのシステムとレンダリングエンジンのオープンソースリファレンスである。本研究では Radium を拡張し機能を追加する形でソーシャルサービスを開発した。また、Radium には EPUB3.0 の CFI ライブラリ⁶が組み込まれているため、EPUBCFI

の解釈と解析には同ライブラリを用いた。EPUBCFI の取得位置は Radium 上の文書をクリックした時の一点とし、取得した EPUBCFI はアノテーションと共にローカルストレージに保存する。

ローカルストレージに保存された読書関連情報は適切なタイミングでアノテーションデータベースに格納される。アノテーションデータベースのデータベーススキーマを図 2 に示す。システムには、EPUB クライアントや Web クライアントからアクセスするための Web API を用意した。ローカルストレージに蓄積された情報は、Web API を用いて格納される。

表示の際には、レンダリングされる対象コンテンツおよび周辺に含まれるアノテーションを、EPUBCFI をキーとして検索し、コンテンツとともに重畳表示する。

5 まとめ

本稿では、EPUB に対する読書関連情報の管理・検索を目的とした読書情報管理システムについて述べた。読書情報管理システムによって、電子書籍上での新しい読書情報の共有、さらには EPUB 上でのソーシャルリーディングが可能となる。また、読書の高度化だけでなく、教育などへの応用も考えられる。

今後の課題としては、図 1 で示したシステムの実装を進めるとともに、EPUBCFI をキーとした効率的な検索を可能にする索引法の検討、EPUB クライアントから取得する情報の検討などを行う。

参考文献

- [1] EPUB : <http://idpf.org/epub/>.
- [2] EPUB3 : <http://idpf.org/epub30/>.
- [3] EPUB Canonical Fragment Identifier : <http://idpf.org/epub/linking/cfi/epub-cfi.html>.
- [4] IDPF : <http://idpf.org/>.
- [5] Radium : <http://readium.org/>.
- [6] Lichan Hong, Ed H. Chi, Raluca Budi, Peter Piroli, Les Nelson. SparTag.us: A Low Cost Tagging System for Foraging of Web Content. AVI 2008.

⁶<https://github.com/readium/EPUBCFI>