Automatically Generate EPUB eBook from Wiki and Linked Data

Lee-Jang Yang Der-Chen Huang

Department of Computer Science and Engineering, National Chung-Hsing University
Taichung 402, Taiwan
huangdc@nchu.edu.tw

Received 12 November 2014; Revised 21 February 2015; Accepted 12 May 2015

Abstract Wiki is a collaborative authoring platform for collective information to quickly gaining popularity in content publication. However, Wiki lacks of the reusability and interoperability to support a wider range of publication requirements. On the other hand, Linked Data approach, which may be considered as a new wave of Semantic Web, is turning from the conventional HTML-based Web to RDF-based Web, and creating added value from data. Linked Data also makes the World Wide Web into a global database that we call the Web of Linked Data. Clients (human and agent) can query Linked Data from multiple sources at once and combine it on the fly. EPUB specification is the most widely adopted and deployed standard for electronic publishing. A significant challenge of Web content (Both wiki and RDF) rendering is how to automatically convert Web content into the EPUB eBook that can be accepted by EPUB eBook reader. To address Web content rendering issues, this study develops two EPUB Generators, including Wiki-to-EPUB Generator (WEGen) and RDF-to-EPUB Generator (REGen), to automatically generate EPUB eBook from Wiki content and RDF triples. The EPUB Generator contains three main modules: XML Converter, Digital Content Repository and EPUB eBook Composer. To demonstrate the feasibility of WEGen and REGen, two EPUB eBooks are automatically generated from wiki contents and DBpedia endpoint to reveal the reusability and interoperability for EPUB publication.

Keywords: EPUB, eBook, Wiki, DBpedia, WEGen, XML, REGen, RDF

1 Introduction

Within the last two to three years, the Internet has greatly changed our way of sharing resources and information. As well known, Web 2.0 is recognized as the next generation of web applications proposed by T. O'Reilly [1]. The main feature of Web 2.0 applications is that they provide a medium for the sharing and exchange of resources. These resources, such as Web 2.0 documents allow web developers to take advantage of these resources to enrich their own applications or produce new integrated solutions by integrating resources, which they could not have provided on their own.

Wiki is one of the most famous Web 2.0 applications. A large number of participants can create new pages or modify existing pages through their web browser easily in wiki platform. It is a collaborative authoring system for collective information which is quickly gaining popularity in content publication. In this work, we adopt MediaWiki [2] as the collaborative authoring platform. This study uses the term Wiki throughout to mean MediaWiki unless clearly specified otherwise. MediaWiki is an extensible open-source Wiki engine written in PHP. It provides a complete wiki pages management engine, including authorization control, version control and database storage, as well as database design and development capabilities.

The term Linked Data is first introduced by Tim Berners-Lee [21], which is referred to a series of best practices to publishing and connecting structured data on the Web. Technically speaking, Linked Data is based on Semantic Web technologies [15, 16], and uses RDF [22] data model to describe statements that link arbitrary data resources on the Web and it can facilitate to infer new data resources at runtime through the RDF links, and then provide more complete answers as new data resources appear on the Web. More specifically, Linked Data which may be considered as a new wave of RDF, is turning from the conventional HTML-based Web to RDF-based Web, and provides the means to reach the goal of Semantic Web, also known as Web of data, machine-readable Web, or Web 3.0.

Besides the above, Linked Data also makes the World Wide Web into a global database that we call the Web of Linked Data. Clients (human and agent) can query Linked Data from multiple sources at once and combine it on the fly. In practice, to make sure the concept of Linked Data is correctly followed; there are four principles proposed in Tim Berners-Lee [21]:

1. Use URIs as names for things.
2. Use HTTP URIs so that a client (machine or human reader) can look up these things.
3. When someone looks up a URI, provide useful RDF information through SPARQL [23] query.

4. Include RDF links to other URIs, so that a client (ma chine or human reader) can discover more things.

However, the Web of Linked Data is a big collection of RDF triples, where the subject of one triple is a URI reference in the namespace of one dataset, and the object of the triple is a URI reference in the namespace of another. In addition, by using HTTP URIs to identify resources, HTTP protocol as retrieval mechanism and RDF data model to represent resource descriptions, Linked Data is directly built upon the general architecture of the Traditional Web.

DBpedia [24], plays a key role on the Web of Linked Data. In fact, DBpedia is the machine-readable version of Wikipedia, which is the most successful and popular wiki site of current Web; DBpedia is automatically generated by processing the pre-existing structured information on each wiki page in Wikipedia. More specifically, it is therefore possible that DBpedia has already created a corresponding URI for the resource one intend to describe. And as we known, it is essentially a huge collection of RDF graphs. There are three different methods one can use to interact with DBpedia:

1. **Using SPARQL to Query DBpedia:** DBpedia provides a public SPARQL endpoint service; http://dbpedia.org/sparql. In practice, this endpoint is generally used directly by remote agents.
2. **Direct Download of DBpedia Datasets:** DBpedia provides a public download URL; http://wiki.dbpedia.org/Downloads2014, to directly download its RDF dumps.
3. **Access DBpedia as Linked Data:** DBpedia provides a lookup service; http://lookup.dbpedia.org/api/search.asmx, that returns DBpedia URIs for a given set of keywords.

EPUB [4] is the most widely adopted and deployed standard for electronic publishing, including books, magazines, manuals and educational, professional and scientific publications. EPUB significantly increase the accessibility, reusability, and interoperability of eBooks to support a wider range of publication requirements. This work focuses on converting different Web content into EPUB eBook. The Web content conversion is a type of transcoding, which is a technology used to adapt computer application displays and content so that they can be viewed on any of the increasing number of diverse devices on the market.

This study develops two EPUB Generators, including Wiki-to-EPUB Generator (WEGen) [20] and RDF-to-EPUB Generator (REGen), to automatically generate EPUB eBook from Wiki content and RDF triples. The remainder of paper is organized as follows. The next section presents some related works. Section 3 is an introduction to the EPUB publication. A flow-oriented aspect to describe the EPUB Generator architecture is presented in Section 4. In Section 5, we demonstrate that the WEGen and REGen can automatically generate EPUB eBook for various EPUB eBook readers to display the same Wiki content with different renderings. Finally, summary and concluding remarks are included.

2 Related Work

EPUB specification is the most widely adopted and interchange format standard for digital publications, such as digital books and documents. The current EPUB 3 specification consists of a set of four specifications, including EPUB Publications 3.0 [6], EPUB Content Documents 3.0 [7], EPUB Open Container Format (OCF) 3.0 [8], and EPUB Media Overlays 3.0 [9]. The EPUB Publications 3.0 specification defines overarching conformance requirements for EPUB publications and EPUB readers at the publication level.

The EPUB Content Documents 3.0 specification institutes profiles of HTML5, CSS, and SVG for creating content documents of EPUB publications. The EPUB Open Container Format (OCF) 3.0 specification defines a container structure for encapsulating the sets of related digital contents that comprise one or more EPUBR publications into a ZIP archive. The EPUB Media Overlays 3.0 specification relies on a subset of SMIL to define a format and a processing model for synchronization of text and audio.

In recent years, several research studies [10-12] have focused on the conversion of digital content in EPUB eBook format, such as PDF, Facebook. In [10], the authors present a mobile Facebook application on Android which turns Facebook contents into an EPUB eBook format for offline-reading purpose. Two main eBook formats are supported by most eBook reader: PDF and EPUB formats. The PDF format is widely used to share electronic documents allowing cross-platform readability. However, it is not ideal for the reusability to support a wider range of publication requirements. On the opposite, the EPUB eBook format is reusable and well suited for eBook readers. In [11], the authors describe a system designed for the conversion of PDF books in the EPUB format. But, they do not address the issue of how a collaborative authoring platform, such as Wiki, can be automatically converted Wiki content into the EPUB eBook which can be accepted by EPUB eBook reader. The study develops the WEGen to address the issue. In addition, the study also develops the REGen to automatically convert DBpedia RDF triples into the EPUB eBook.

3 EPUB Publication

The EPUB 3 specification is a distribution and interchange format standard for digital publications and documents [4]. It defines a means of representing, packaging and encoding structured Web content, such as HTML5, CSS, SVG, images.

3.1 EPUB Open Container Format

The EPUB Open Container Format (OCF) 3.0 defines a file format and processing model for encapsulating the sets of related resources that comprise one or more EPUB Publications into a single-file container [8]. It also defines the rules for the representation of this abstract container within a ZIP physical container. An EPUB OCF is an abstract container that defines a root directory for the contents of the container. A physical container of EPUB OCF is composed of a mimetype file, a META-INF directory, and a content directory, and depicted in Figure 1. The mimetype file in the root directory is reserved for using by EPUB OCF ZIP Containers.

The root directory of an EPUB OCF includes a mandatory directory named META-INF that is used to store the following special files. Each file is written in XML standard.

Required file: container.xml

It is used to describe the document that is the point of entry for each embedded Publication. An example of container.xml is shown in Figure 2.

Optional file: signatures.xml

It contains digital signatures for various assets and license agreements.

Optional file: encryption.xml

It contains information regarding the encryption of the EPUB resources.

Optional file: metadata.xml

It is used to describe metadata about the EPUB OCF container.

Optional file: rights.xml

It is used to store information about digital rights.

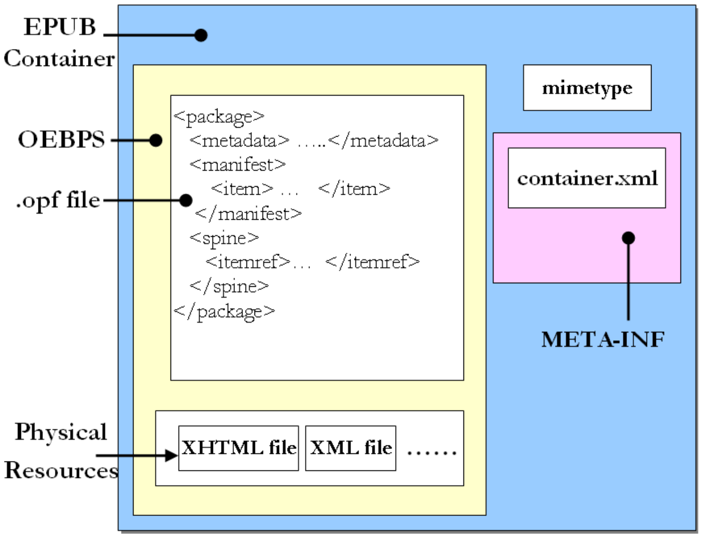


Fig. 1. EPUB eBook Container

```
<?xml version="1.0" encoding="UTF-8"?>
<container version="1.0" xmlns="urn:oasis:names:tc:opendocument:xmlns:container">
<rootfiles>
  <rootfile full-path="OEBPS/content.opf" media-type="application/oebps-package+xml"/>
</rootfiles>
</container>
```

Fig. 2. The content of container.xml

The contents of each of the individual publications can be stored within a specific directory under the EPUB eBook container's root. In Figure 2, the subdirectory OEBPS is relative to the root directory for the EPUB OCF Container. The OEBPS directory contains an OFP file and a set of physical resources. The OPF file is an XML file that describes an EPUB Publication. It identifies all other files in the EPUB publication and offers metadata information about them. The physical resources, such as XML files, HTML files, XHTML files, images files, are included in the OEBPS directory.

3.2 Open Packaging Format

The Open Packaging Format (OPF) defines the mechanism to provide additional structure and semantics to the EPUB publication. The OPF mainly aims to enhance the semantic representation of EPUB publication for computer-interpretable effects. An OPF file is written in XML to describe metadata of EPUB publications. The basic schema of OPF is shown in Figure 3. The OPF provides four main functions [13]. (1) OPF describes and references all components of the EPUB eBook, such as markup files, navigation structures, and images. (2) OPF defines the schema of XML to describe the publication-level metadata. (3) OPF assigns the linear reading-order of the EPUB publication. (4) OPF provides a mechanism to assign a declarative global navigation composition.

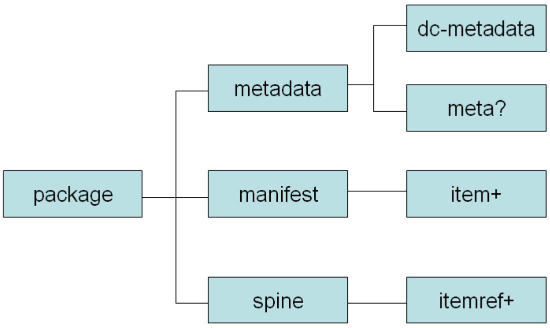


Fig. 3. The EPUB eBook schema

These basic elements in the OPF are described as follows:

package element
It is the root element of the OPF document and encapsulates EPUB publication metadata and resource information. The package element must contain metadata element, manifest element, and spine element.

metadata element
It encapsulates meta-information of EPUB publications. The metadata element consists of dc-metadata element and meta element.

dc-metadata element
It represents the elements in the Dublin Core Metadata Element Set [14]. The minimal required metadata that EPUB publications must contain three Dublin Core elements, including title, identifier, and language elements.

meta element
The meta element is optional. It is used to extend the description capability of dc-metadata.

manifest element
It provides a complete list of the resources that constitute the EPUB publication, each represented by an item element.

item element
It represents an EPUB publication resource. Each item element in the manifest element identifies a EPUB publication resource by the IRI provided in its href attribute.

spine element
It is used to provide the default linear reading order of the EPUB publication content by defining an ordered list of item elements.

itemref element
It is used to refer to a specific EPUB publication resource. The child itemref elements of the spine element represent a sequential list of EPUB publication resources. The order of the itemref elements defines the default reading order of the EPUB publication. Each itemref element must reference an item element in the manifest element via its idref attribute.

4 EPUB Generator

This section develops two EPUB Generators, including Wiki-to-EPUB Generator and RDF-to-EPUB Generator. These Generators contain three main modules: XML Converter, Digital Content Repository and

EPUB eBook Composer. Due to the different content sources, two XML Converters, including Wiki-XML and RDF-XML are constructed separately, yet the other two modules are the same.

Digital Content Repository contains digital publication resources based on XML and Web standards. Examples of publication resources include the package container, content documents, style sheets, audio, video, images, embedded fonts and scripts.

EPUB eBook Composer searches and integrates related content documents for encapsulating a set of related resources into a single-file EPUB Container. It supports a ZIP-based packaging and distribution format for EPUB eBooks.

4.1 Wiki-to-EPUB Generator

Wiki-to-EPUB Generator (WEGen) automatically generates EPUB eBook from Wiki content. Figure 4 shows the WEGen architecture with the flow control sequence.

Wiki-to-XML Converter is developed based on flat-based transcoding approach to convert Wiki content into XML documents, including container.xml, OPF document, and XHTML documents. It enables an efficient execution of the transcoding, which conforms to the EPUB specifications. The Wiki-to-XML Converter can be started via a request URL from Wiki platform. The markup mapping between Wiki and XHTML is presented in Table 1.

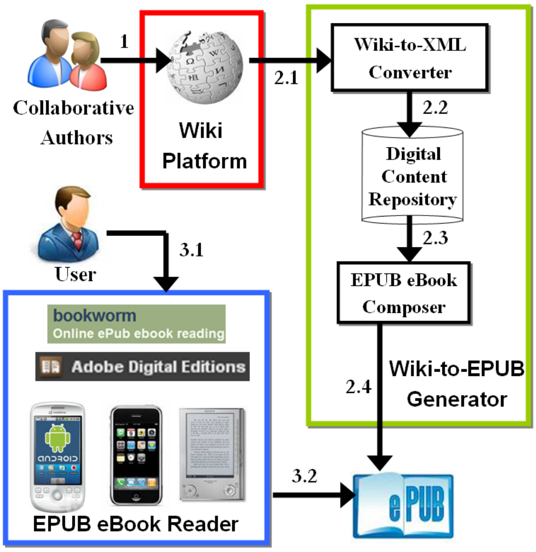


Fig. 4. The flow-oriented WEGen architecture

Table 1. Markup mapping between Wiki and XHTML

description	Wiki markup	XHTML
Heading	=	<h1><hr />
bold	'''	<b>
italicize	''	<i>
space		&nbsp;
line breaks		<br />
blank line		<p />
block		<div>
Retaining new lines and spaces	<poem>	<pre>
line	----	<hr />
lists	*	<ul><li>
lists	#	<ol><li>
Internal hyperlink	[[id | topic]]	<a href="id"> topic</a>
external hyperlink	[URL topic]	<a href="URL"> topic</a>
table	{| border="1"…. |}	<table> ..</table>
head field	!	<th>
data field	||	<td>
comment	<!-- -->	<!-- -->

The information flow of the WEGen as following shows:

1. This step is a community of collaborative authors to edit content on Wiki platform.
2. This step describes how the WEGen automatically generate the EPUB eBook.
   2.1 The Wiki-to-XML Converter retrieves a Wiki page, and then depends on the predefined flat-based transcoding, converting the Wiki page to a specific XML-based document, such as container.xml, OPF document, and XHTML documents.
   2.2 The specific XML-based document is saved into the Digital Content Repository,
   2.3 The EPUB eBook Composer depends on the OPF document to retrieve the Digital Content Repository to get the related digital contents, such XHTML documents, CSS documents, images.
   2.4 The EPUB eBook Composer packages the related digital contents into an EPUB eBook.
3. This step shows the user how to read the EPUB eBook with various readers.
   3.1 User uses an EPUB eBook reader, such as smartphone, eBook reader, Web-based eBook reader, and Adobe Digital Editions.
   3.2 EPUB eBook reader downloads EPUB eBooks via the Internet.

4.2 RDF-to-EPUB Generator

RDF-to-EPUB Generator (REGen) automatically generates EPUB eBook from RDF content. Figure 5 shows the REGen architecture together with the flow control sequence. In general, the most popular representation of RDF data model is in XML serialization format (RDF/XML). Hence,

RDF-to-XML Converter is developed based on direct-based transcoding approach to convert RDF content into XML documents, including container.xml, OPF document, and XHTML documents. The RDF-to-XML Converter can be started via a query request to DBpedia public SPARQL endpoint (http://dbpedia.org/sparql). Figure 6 shows the opening page of the SPARQL explorer.

The information flow of the REGen occurs as follows.

1. This step is a client (human or agent) to request DBpedia public SPARQL endpoint. More specifically, for human, this endpoint could be a Web page. In addition, this endpoint also provides a set of Web APIs that can be invoked by the calling agent.
2. This step describes how the REGen automatically generates the EPUB eBook.
   2.1 The RDF-to-XML Converter retrieves RDF triples, and then converts the RDF statements to a specific XML-based document, such as container.xml, OPF document, and XHTML documents. In general, These RDF statements are mainly a long list of property-value pairs, and are often presented in a XHTML table format.
   2.2 The specific XML-based document is saved into the Digital Content Repository,
   2.3 The EPUB eBook Composer depends on the OPF document to retrieve the Digital Content Repository to get the related digital contents, such XHTML documents, CSS documents, images.
   2.4 The EPUB eBook Composer packages the related digital contents into an EPUB eBook.
3. This step shows the user how to read the EPUB eBook with various readers.
   3.1 User uses an EPUB eBook reader, such as smartphone, eBook reader, Web-based eBook reader, and Adobe Digital Editions.
   3.2 EPUB eBook reader downloads EPUB eBooks via the Internet.

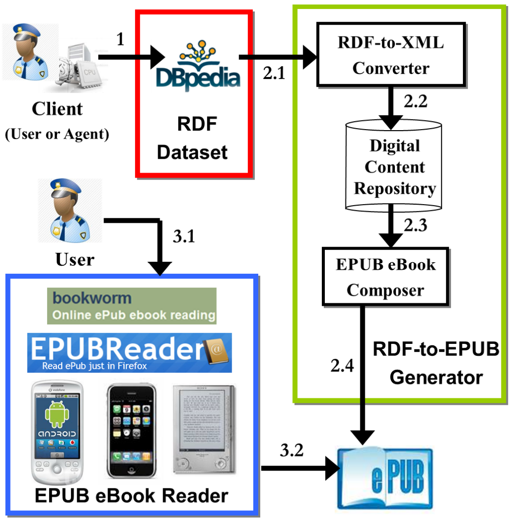
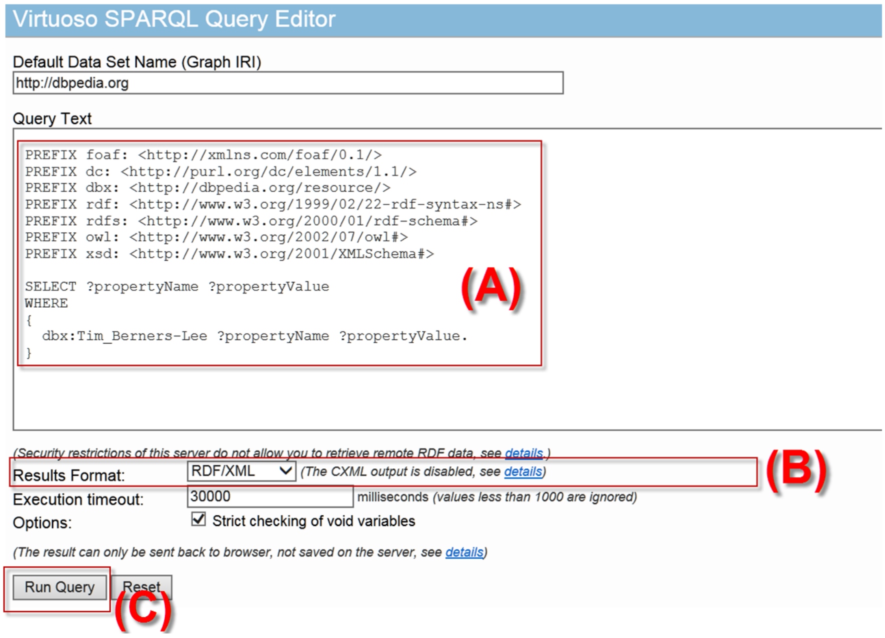


Fig. 5. The flow-oriented REGen architecture



Default Data Set Name (Graph IRI)

http://dbpedia.org

Query Text

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX dbx: <http://dbpedia.org/resource/>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?propertyName ?propertyValue
WHERE
{
  dbx:Tim_Berners-Lee ?propertyName ?propertyValue.
}
```

(Security restrictions of this server do not allow you to retrieve remote RDF data, see details.)

Results Format: RDF/XML *(The CXML output is disabled, see details)*

Execution timeout: 30000 milliseconds *(values less than 1000 are ignored)*

Options: ☑ Strict checking of void variables

(The result can only be sent back to browser, not saved on the server, see details)

Fig. 6. DBpedia public SPARQL endpoint

4.3 WEGen vs REGen

In previous sections, we have separately argued the flow-oriented architecture of WEGen and REGen. As a summary, the major features of comparison between them are summarized in Table 2.

Table 2. Major features comparison between the WEGen and REGen

Feature	WEGen	REGen
Client Access	human	human or agent
Content Source	wiki page	RDF triples
XML Transcoding	flat-based	direct-based
XHTML Markup	yes	no
Applicability	Web 2.0	Web 3.0

5 EPUB Generator Applications

5.1 An application with the WEGen

Section 4.1 shows the components of WEGen along with flow-oriented architecture. In this section, an XML Q&A EPUB eBook is automatically generated from XML Q&A Wiki contents to demonstrate the feasibility of WEGen. The list of XML Q&A items (A) and the list's wiki markup (B) are shown in Figure 7. The list's wiki markup is transformed by the Wiki-to-XML Converter into an OPF document that describes metadata of the EPUB eBook, as shown in Figure 4. A Wiki page (A) is written in Wiki markup (B). The Wiki markup document is transformed by the Wiki-to-XML Converter into an XHTML homepage (C) that provides dynamic hyperlinks, as shown in Figure 8.

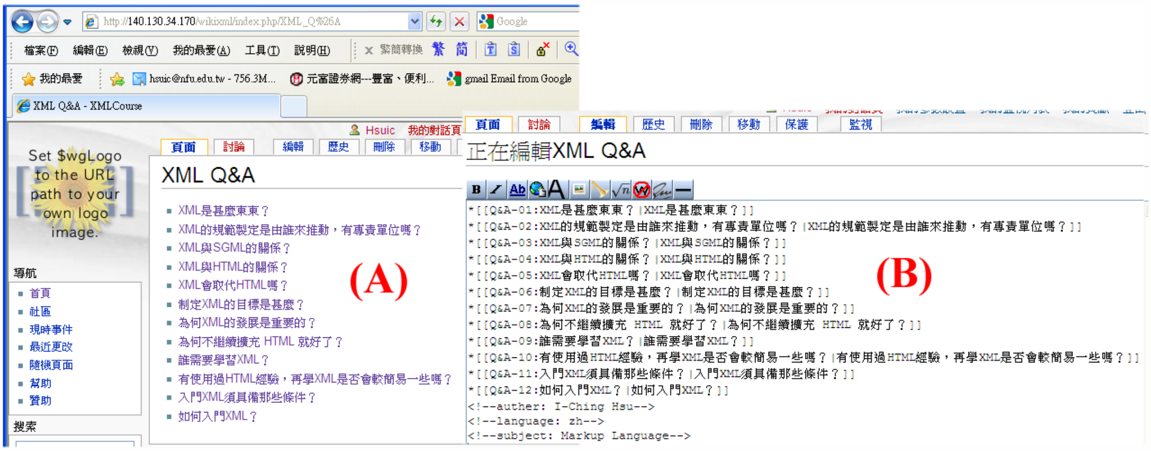
The EPUB eBook Composer parses the OPF document (see Figure 4) to extract all item elements, and then depends on the predefined flat-based transcoding, converting these Wiki pages to XHTML documents. Finally, the EPUB eBook Composer package these XHTML documents to an XML Q&A EPUB eBook. The XML Q&A EPUB eBook can be displayed in online Web-based EPUB reader bookworm, as shown in Figure 9.

5.2 An application with the REGen

Section 4.2 shows the components of REGen along with flow-oriented architecture. In this section, an RDF based EPUB eBook is automatically generated from DBpedia contents to demonstrate the feasibility of REGen.

First, let us use the WWW inventor Tim Berners-Lee as an example, the corresponding Wikipedia description page is shown in Figure 10, and this page has the following URL:

http://en.wikipedia.org/wiki/Tim_Berners-Lee


http://140.130.34.170/wikixml/index.php/XML_Q%26A
Google
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
繁簡轉換 繁 簡
我的最愛
hsuic@nfu.edu.tw - 756.3M...
元富證券網---豐富、便利...
gmail Email from Google
XML Q&A - XMLCourse
Hsuic 我的對話頁
頁面 討論 編輯 歷史 刪除 移動 保護 監視
Set $wgLogo to the URL path to your own logo image.
導航
首頁
社區
現時事件
最近更改
隨機頁面
幫助
贊助
搜索
(A)
(B)

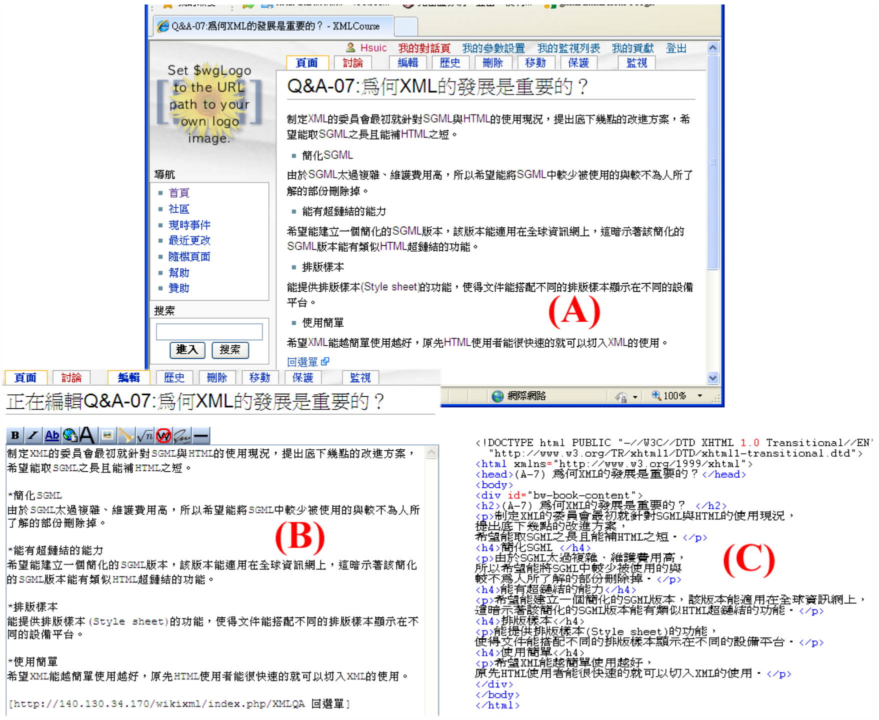
XML Q&A

- XML是甚麼東東？
- XML的規範製定是由誰來推動，有專責單位嗎？
- XML與SGML的關係？
- XML與HTML的關係？
- XML會取代HTML嗎？
- 制定XML的目標是甚麼？
- 為何XML的發展是重要的？
- 為何不繼續擴充 HTML 就好了？
- 誰需要學習XML？
- 有使用過HTML經驗，再學XML是否會較簡易一些嗎？
- 入門XML須具備那些條件？
- 如何入門XML？

正在編輯XML Q&A

```
*[[Q&A-01:XML是甚麼東東？|XML是甚麼東東？]]
*[[Q&A-02:XML的規範製定是由誰來推動，有專責單位嗎？|XML的規範製定是由誰來推動，有專責單位嗎？]]
*[[Q&A-03:XML與SGML的關係？|XML與SGML的關係？]]
*[[Q&A-04:XML與HTML的關係？|XML與HTML的關係？]]
*[[Q&A-05:XML會取代HTML嗎？|XML會取代HTML嗎？]]
*[[Q&A-06:制定XML的目標是甚麼？|制定XML的目標是甚麼？]]
*[[Q&A-07:為何XML的發展是重要的？|為何XML的發展是重要的？]]
*[[Q&A-08:為何不繼續擴充 HTML 就好了？|為何不繼續擴充 HTML 就好了？]]
*[[Q&A-09:誰需要學習XML？|誰需要學習XML？]]
*[[Q&A-10:有使用過HTML經驗，再學XML是否會較簡易一些嗎？|有使用過HTML經驗，再學XML是否會較簡易一些嗎？]]
*[[Q&A-11:入門XML須具備那些條件？|入門XML須具備那些條件？]]
*[[Q&A-12:如何入門XML？|如何入門XML？]]
<!--auther: I-Ching Hsu-->
<!--language: zh-->
<!--subject: Markup Language-->
```

Fig. 7. The list of XML Q&A items


Q&A-07:爲何XML的發展是重要的？ - XMLCourse
Hsuic 我的對話頁 我的參數設置 我的監視列表 我的貢獻 登出
頁面 討論 編輯 歷史 刪除 移動 保護 監視
Set $wgLogo to the URL path to your own logo image.
導航
首頁
社區
現時事件
最近更改
隨機頁面
幫助
贊助
搜索
進入 搜索
網際網路
100%
(A)
(B)
(C)

Q&A-07:爲何XML的發展是重要的？

制定XML的委員會最初就針對SGML與HTML的使用現況，提出底下幾點的改進方案，希望能取SGML之長且能補HTML之短。

- 簡化SGML

由於SGML太過複雜、維護費用高，所以希望能將SGML中較少被使用的與較不為人所了解的部份刪除掉。

- 能有超鏈結的能力

希望能建立一個簡化的SGML版本，該版本能適用在全球資訊網上，這暗示著該簡化的SGML版本能有類似HTML超鏈結的功能。

- 排版樣本

能提供排版樣本(Style sheet)的功能，使得文件能搭配不同的排版樣本顯示在不同的設備平台。

- 使用簡單

希望XML能越簡單使用越好，原先HTML使用者能很快速的就可以切入XML的使用。

回選單

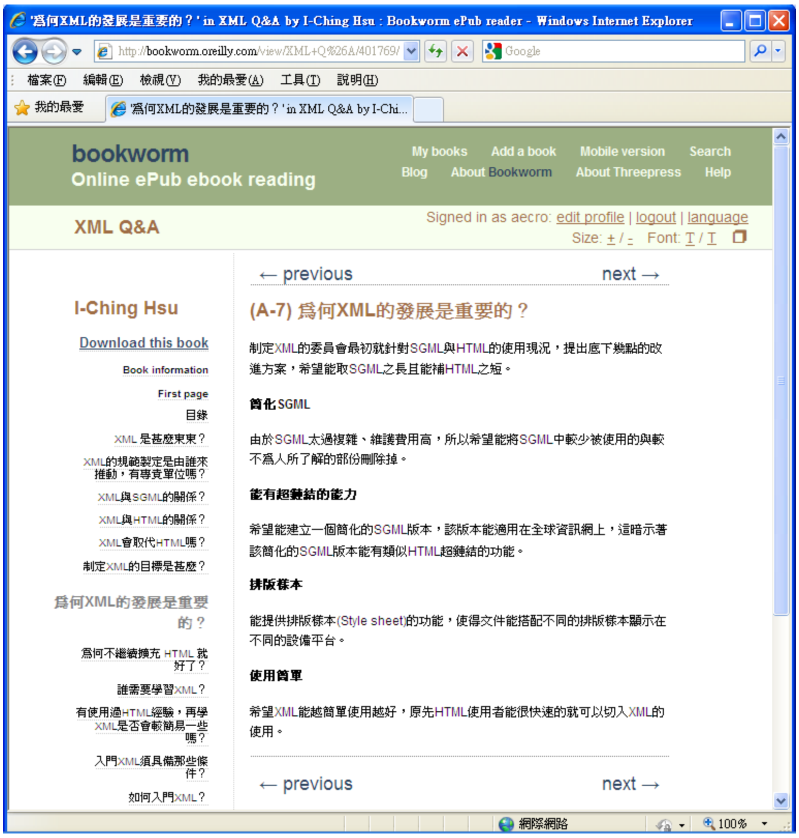
正在編輯Q&A-07:爲何XML的發展是重要的？

```
制定XML的委員會最初就針對SGML與HTML的使用現況，提出底下幾點的改進方案，
希望能取SGML之長且能補HTML之短。

*簡化SGML
由於SGML太過複雜、維護費用高，所以希望能將SGML中較少被使用的與較不為人所
了解的部份刪除掉。

*能有超鏈結的能力
希望能建立一個簡化的SGML版本，該版本能適用在全球資訊網上，這暗示著該簡化
的SGML版本能有類似HTML超鏈結的功能。

*排版樣本
能提供排版樣本(Style sheet)的功能，使得文件能搭配不同的排版樣本顯示在不
同的設備平台。

*使用簡單
希望XML能越簡單使用越好，原先HTML使用者能很快速的就可以切入XML的使用。

[http://140.130.34.170/wikixml/index.php/XMLQA 回選單]
```

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Transitional//EN'
  "http://www.w3.org/TR/xhtml1/DTD/xhtml1-transitional.dtd">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>(A-7) 爲何XML的發展是重要的？</head>
<body>
<div id="bw-book-content">
<h2>(A-7) 爲何XML的發展是重要的？ </h2>
<p>制定XML的委員會最初就針對SGML與HTML的使用現況，
提出底下幾點的改進方案，
希望能取SGML之長且能補HTML之短。</p>
<h4>簡化SGML </h4>
<p>由於SGML太過複雜，維護費用高，
所以希望能將SGML中較少被使用的與
較不爲人所了解的部份刪除掉。</p>
<h4>能有超鏈結的能力</h4>
<p>希望能建立一個簡化的SGML版本，該版本能適用在全球資訊網上，
這暗示著該簡化的SGML版本能有類似HTML超鏈結的功能。</p>
<h4>排版樣本</h4>
<p>能提供排版樣本(Style sheet)的功能，
使得文件能搭配不同的排版樣本顯示在不同的設備平台。</p>
<h4>使用簡單</h4>
<p>希望XML能越簡單使用越好，
原先HTML使用者能很快速的就可以切入XML的使用。</p>
</div>
</body>
</html>
```

Fig. 8. The Wiki markup converts to XHTML

More specifically, the URL is composed from the prefix, “http://en.wikipedia.org/wiki”, together with the subject, “Tim_Berners-Lee”. Next, we start using SPARQL to query DBpedia endpoint. As shown in Figure 6, (B) depicts the results format RDF/XML, which represents an RDF graph as an XML document. And (A) shows a query representation that is used to find what properties the DBpedia has used to describe the subject’s corresponding resource name, “dbx: Tim_Berners-Lee”, where dbx: stands for prefix namespace; “http://dbpedia.org/resource”. Next, we run query (C) and the DBpedia return the RDF document, as shown in Figure 11. The RDF document is transformed by the RDF-to-XML Converter into an XHTML homepage. Lastly, in similar fashion, the EPUB eBook Composer automatically generates the corresponding EPUB eBook, and can be displayed in EPUBReader that is an addition of Firefox browser, as shown in Figure 12.


'爲何XML的發展是重要的？' in XML Q&A by I-Ching Hsu : Bookworm ePub reader - Windows Internet Explorer
http://bookworm.oreilly.com/view/XML+Q%26A/401769/
Google
檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)
我的最愛
'爲何XML的發展是重要的？' in XML Q&A by I-Chi...
bookworm
Online ePub ebook reading
My books Add a book Mobile version Search
Blog About Bookworm About Threepress Help
Signed in as aecro: edit profile | logout | language
Size: + / - Font: T / T
網際網路
100%

XML Q&A

I-Ching Hsu

Download this book

Book information

First page

目錄

XML 是甚麼東東？

XML的規範製定是由誰來推動，有專責單位嗎？

XML與SGML的關係？

XML與HTML的關係？

XML會取代HTML嗎？

制定XML的目標是甚麼？

爲何XML的發展是重要的？

爲何不繼續擴充 HTML 就好了？

誰需要學習XML？

有使用過HTML經驗，再學XML是否會較簡易一些嗎？

入門XML須具備那些條件？

如何入門XML？

← previous next →

(A-7) 爲何XML的發展是重要的？

制定XML的委員會最初就針對SGML與HTML的使用現況，提出底下幾點的改進方案，希望能取SGML之長且能補HTML之短。

簡化SGML

由於SGML太過複雜、維護費用高，所以希望能將SGML中較少被使用的與較不爲人所了解的部份刪除掉。

能有超鏈結的能力

希望能建立一個簡化的SGML版本，該版本能適用在全球資訊網上，這暗示著該簡化的SGML版本能有類似HTML超鏈結的功能。

排版樣本

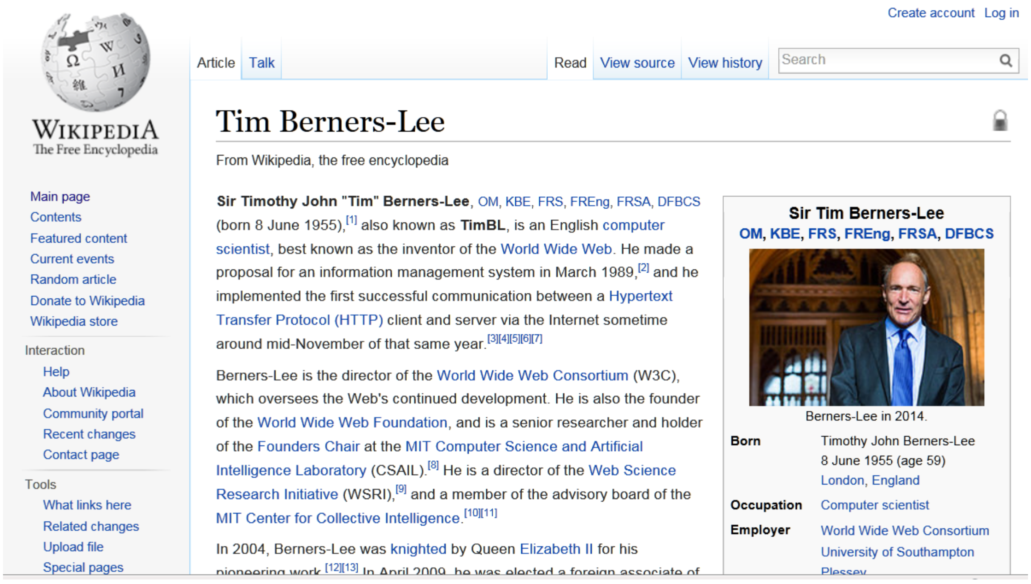
能提供排版樣本(Style sheet)的功能，使得文件能搭配不同的排版樣本顯示在不同的設備平台。

使用簡單

希望XML能越簡單使用越好，原先HTML使用者能很快速的就可以切入XML的使用。

← previous next →

Fig. 9. The XML Q&A EPUB eBook displays in bookworm reader



Tim Berners-Lee

From Wikipedia, the free encyclopedia

Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FREng, FRSA, DFBCS (born 8 June 1955),[1] also known as **TimBL**, is an English computer scientist, best known as the inventor of the World Wide Web. He made a proposal for an information management system in March 1989,[2] and he implemented the first successful communication between a Hypertext Transfer Protocol (HTTP) client and server via the Internet sometime around mid-November of that same year.[3][4][5][6][7]

Berners-Lee is the director of the World Wide Web Consortium (W3C), which oversees the Web's continued development. He is also the founder of the World Wide Web Foundation, and is a senior researcher and holder of the Founders Chair at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL).[8] He is a director of the Web Science Research Initiative (WSRI),[9] and a member of the advisory board of the MIT Center for Collective Intelligence.[10][11]

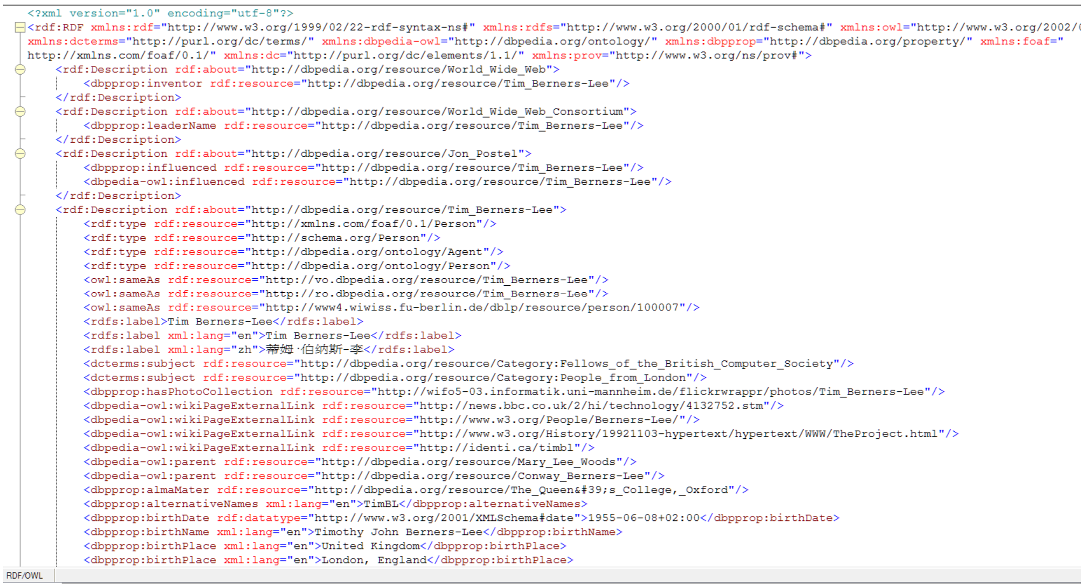
In 2004, Berners-Lee was knighted by Queen Elizabeth II for his pioneering work.[12][13] In April 2009, he was elected a foreign associate of

Sir Tim Berners-Lee
OM, KBE, FRS, FREng, FRSA, DFBCS

Berners-Lee in 2014.

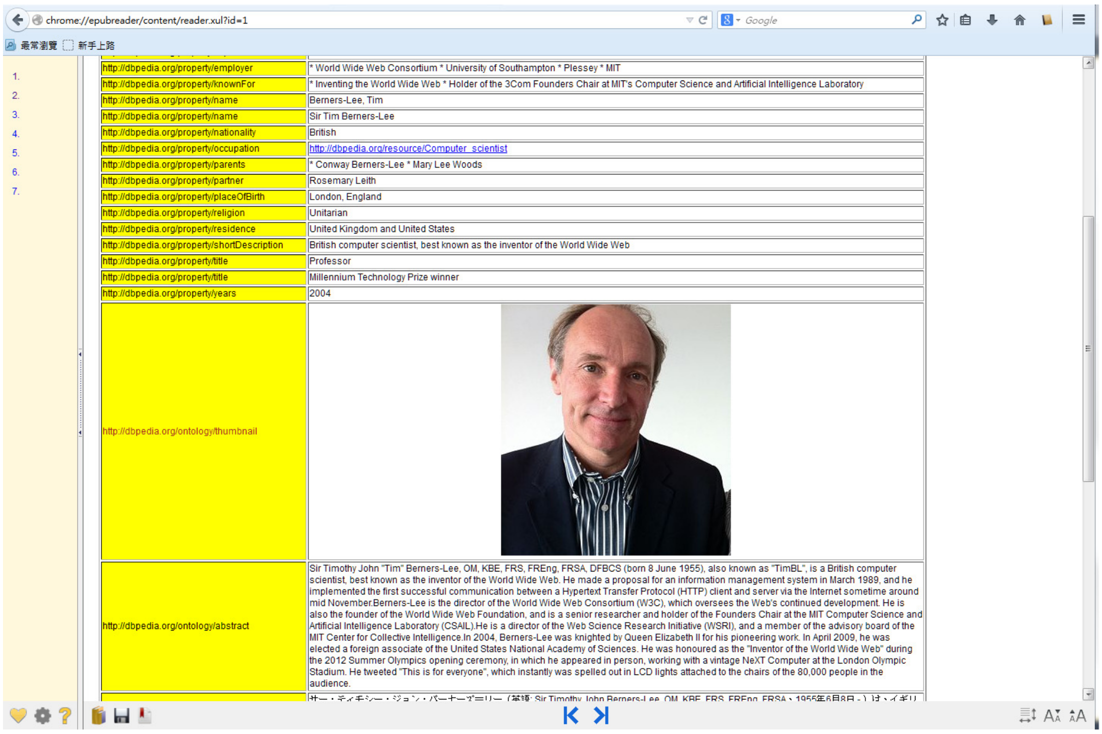
Born	Timothy John Berners-Lee 8 June 1955 (age 59) London, England
Occupation	Computer scientist
Employer	World Wide Web Consortium University of Southampton Plessey

Fig. 10. Tim Berners-Lee's wiki page in Wikipedia



```
<?xml version="1.0" encoding="utf-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:owl="http://www.w3.org/2002/(
xmlns:dcterms="http://purl.org/dc/terms/" xmlns:dbpedia-owl="http://dbpedia.org/ontology/" xmlns:dbpprop="http://dbpedia.org/property/" xmlns:foaf="
http://xmlns.com/foaf/0.1/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:prov="http://www.w3.org/ns/prov#">
    <rdf:Description rdf:about="http://dbpedia.org/resource/World_Wide_Web">
        <dbpprop:inventor rdf:resource="http://dbpedia.org/resource/Tim_Berners-Lee"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://dbpedia.org/resource/World_Wide_Web_Consortium">
        <dbpprop:leaderName rdf:resource="http://dbpedia.org/resource/Tim_Berners-Lee"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://dbpedia.org/resource/Jon_Postel">
        <dbpprop:influenced rdf:resource="http://dbpedia.org/resource/Tim_Berners-Lee"/>
        <dbpedia-owl:influenced rdf:resource="http://dbpedia.org/resource/Tim_Berners-Lee"/>
    </rdf:Description>
    <rdf:Description rdf:about="http://dbpedia.org/resource/Tim_Berners-Lee">
        <rdf:type rdf:resource="http://xmlns.com/foaf/0.1/Person"/>
        <rdf:type rdf:resource="http://schema.org/Person"/>
        <rdf:type rdf:resource="http://dbpedia.org/ontology/Agent"/>
        <rdf:type rdf:resource="http://dbpedia.org/ontology/Person"/>
        <owl:sameAs rdf:resource="http://vo.dbpedia.org/resource/Tim_Berners-Lee"/>
        <owl:sameAs rdf:resource="http://ro.dbpedia.org/resource/Tim_Berners-Lee"/>
        <owl:sameAs rdf:resource="http://www4.wiwiss.fu-berlin.de/dblp/resource/person/100007"/>
        <rdfs:label>Tim Berners-Lee</rdfs:label>
        <rdfs:label xml:lang="en">Tim Berners-Lee</rdfs:label>
        <rdfs:label xml:lang="zh">蒂姆·伯纳斯-李</rdfs:label>
        <dcterms:subject rdf:resource="http://dbpedia.org/resource/Category:Fellows_of_the_British_Computer_Society"/>
        <dcterms:subject rdf:resource="http://dbpedia.org/resource/Category:People_from_London"/>
        <dbpprop:hasPhotoCollection rdf:resource="http://wifo5-03.informatik.uni-mannheim.de/flickrwrappr/photos/Tim_Berners-Lee"/>
        <dbpedia-owl:wikiPageExternalLink rdf:resource="http://news.bbc.co.uk/2/hi/technology/4132752.stm"/>
        <dbpedia-owl:wikiPageExternalLink rdf:resource="http://www.w3.org/People/Berners-Lee/"/>
        <dbpedia-owl:wikiPageExternalLink rdf:resource="http://www.w3.org/History/19921103-hypertext/hypertext/WWW/TheProject.html"/>
        <dbpedia-owl:wikiPageExternalLink rdf:resource="http://identi.ca/timbl"/>
        <dbpedia-owl:parent rdf:resource="http://dbpedia.org/resource/Mary_Lee_Woods"/>
        <dbpedia-owl:parent rdf:resource="http://dbpedia.org/resource/Conway_Berners-Lee"/>
        <dbpprop:almaMater rdf:resource="http://dbpedia.org/resource/The_Queen&#39;s_College,_Oxford"/>
        <dbpprop:alternativeNames xml:lang="en">TimBL</dbpprop:alternativeNames>
        <dbpprop:birthDate rdf:datatype="http://www.w3.org/2001/XMLSchema#date">1955-06-08+02:00</dbpprop:birthDate>
        <dbpprop:birthName xml:lang="en">Timothy John Berners-Lee</dbpprop:birthName>
        <dbpprop:birthPlace xml:lang="en">United Kingdom</dbpprop:birthPlace>
        <dbpprop:birthPlace xml:lang="en">London, England</dbpprop:birthPlace>
```

Fig. 11. Tim Berners-Lee's RDF document return from DBpedia



http://dbpedia.org/property/employer	* World Wide Web Consortium * University of Southampton * Plessey * MIT
http://dbpedia.org/property/knownFor	* Inventing the World Wide Web * Holder of the 3Com Founders Chair at MIT's Computer Science and Artificial Intelligence Laboratory
http://dbpedia.org/property/name	Berners-Lee, Tim
http://dbpedia.org/property/name	Sir Tim Berners-Lee
http://dbpedia.org/property/nationality	British
http://dbpedia.org/property/occupation	http://dbpedia.org/resource/Computer_scientist
http://dbpedia.org/property/parents	* Conway Berners-Lee * Mary Lee Woods
http://dbpedia.org/property/partner	Rosemary Leith
http://dbpedia.org/property/placeOfBirth	London, England
http://dbpedia.org/property/religion	Unitarian
http://dbpedia.org/property/residence	United Kingdom and United States
http://dbpedia.org/property/shortDescription	British computer scientist, best known as the inventor of the World Wide Web
http://dbpedia.org/property/title	Professor
http://dbpedia.org/property/title	Millennium Technology Prize winner
http://dbpedia.org/property/years	2004
http://dbpedia.org/ontology/thumbnail	
http://dbpedia.org/ontology/abstract	Sir Timothy John "Tim" Berners-Lee, OM, KBE, FRS, FREng, FRSA, DFBCS (born 8 June 1955), also known as "TimBL", is a British computer scientist, best known as the inventor of the World Wide Web. He made a proposal for an information management system in March 1989, and he implemented the first successful communication between a Hypertext Transfer Protocol (HTTP) client and server via the Internet sometime around mid November.Berners-Lee is the director of the World Wide Web Consortium (W3C), which oversees the Web's continued development. He is also the founder of the World Wide Web Foundation, and is a senior researcher and holder of the Founders Chair at the MIT Computer Science and Artificial Intelligence Laboratory (CSAIL).He is a director of the Web Science Research Initiative (WSRI), and a member of the advisory board of the MIT Center for Collective Intelligence.In 2004, Berners-Lee was knighted by Queen Elizabeth II for his pioneering work. In April 2009, he was elected a foreign associate of the United States National Academy of Sciences. He was honoured as the "Inventor of the World Wide Web" during the 2012 Summer Olympics opening ceremony, in which he appeared in person, working with a vintage NeXT Computer at the London Olympic Stadium. He tweeted "This is for everyone", which instantly was spelled out in LCD lights attached to the chairs of the 80,000 people in the audience.

Fig. 12. Tim Berners-Lee's EPUB eBook display in EPUBReader of Firefox

6 Conclusion

Wiki is a popular and widely collaborative authoring platform for collective information which is quickly gaining popularity in digital content publication. However, Wiki is limited in the reusability, interoperability, and offline reading. EPUB significantly increase the accessibility, reusability, and interoperability of digital publications to support a wider range of publication requirements, including various digital content, navigation order, flexible layouts, and global typography features. Therefore, EPUB can improve the above issues of Wiki. This work presents two EPUB Generators, including Wiki-to-EPUB Generator (WEGen) and RDF-to-EPUB Generator (REGen), to automatically generate EPUB eBook from Wiki content and RDF triples.

At least two research topics are of interest for future study. Firstly, further research will be to extend the EPUB Generator with Semantic Web technologies [15, 16] to support additional intelligence in the developed

EPUB eBooks by deducing new adaptation rules. By integrating logic rules into EPUB Generator, this approach can describe additional semantics of navigation order and personal applications. Semantic Web Rule language (SWRL) [17] seems to be the most appropriate language to further study, because it currently is the main language for representing logic rules in the Semantic Web. Secondly, future direction of development is to investigate how to integrate context-aware technologies into personal needs to facilitate the implementation of context-aware personalized search applications [18, 19]. We will attempt to evaluate our proposed method by applying context-aware factors, including time, location, task, and preference for dynamically tailor the EPUB eBook to meet personal needs.

References

[1] T. O'Reilly, What Is Web 2.0. 2005. Available from: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[2] C. Nentwich, L. Capra, W. Emmerich, "xlinkit: A consistency checking and smart link generation service," *ACM Transactions on Internet Technology*, Vol. 2, No. 2, pp. 151-185, 2002.

[3] Lightweight Markup Language. Available from: http://en.wikipedia.org/wiki/Lightweight_markup.

[4] G. Conboy, M. Garrish, M. Gylling, W. McCoy, M. Makoto, D. Weck. EPUB 3 Overview. 2011. Available from: http://idpf.org/epub/30/spec/epub30-overview.html.

[5] I.-C. Hsu, L.-P. Chi, S.-S. Bor, "A platform for transcoding heterogeneous markup documents using ontology-based metadata," *Journal of Network and Computer Applications*, Vol. 32, No. 3, pp. 616-629, 2009.

[6] M. Gylling, W. McCoy, M. Garrish, EPUB Publications 3.0, 2011. Available from: http://idpf.org/epub/30/spec/epub30-publications.html.

[7] M. Gylling, W. McCoy, E.J. Etemad, M. Garrish, EPUB Content Documents 3.0, 2011. Available from: http://idpf.org/epub/30/spec/epub30-contentdocs.html.

[8] J. Pritchett, M. Gylling, EPUB Open Container Format (OCF) 3.0. 2011. Available from: http://idpf.org/epub/30/spec/epub30-ocf.html.

[9] G. Conboy, M. Gylling, W. McCoy, M. DeMeglio. EPUB Media Overlays 3.0. Available from: http://epub-revision.googlecode.com/svn-history/r1952/trunk/build/EPUB_MediaOverlays30.html.

[10] C.-Y. Chuang, Y.-B. Lin, Z.J. Ren, Y.-T. Yeh, "User-generated E-books from Facebook contents," in P*roceedings of 2011 7th International Wireless Communications and Mobile Computing Conference (IWCMC)*, IEEE Computer Society, pp. 1918 - 1922, 2011.

[11] S. Marinai, E. Marino, G. Soda, "Conversion of PDF books in ePub format," in *Proceedings of 2011 International Conference on Document Analysis and Recognition*, IEEE Computer Society Press, pp. 478-482, 2011.

[12] M. Bernius, R. Langille, G. Paddock, S. Battle, "Using EPUB as a framework for the automated collection, tagging, and cross-media transformation of web content for re-publication," in *Proceedings of IS&T/SPIE Electronic Imaging 2010, Imaging and Printing in a Web 2.0 World*, 2010.

[13] Open Packaging Format (OPF), 2010. Available from: http://idpf.org/epub/20/spec/OPF_2.0.1_draft.htm.

[14] Dublin Core Metadata Element Set, 2010. Available from: http://dublincore.org/documents/dces/.

[15] I.-C. Hsu, "Personalized web feeds based on ontology technologies," *Information Systems Frontiers*, Vol. 15, No. 3, pp. 465-479, 2011.

[16] I.-C. Hsu, "SXRS: An XLink-based recommender system using semantic web technologies," *Expert Systems with Applications*, Vol. 36, No. 2, pp. 3795-3804, 2009.

[17] I. Horrocks, P.F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, and M. Dean, SWRL: A Semantic Web Rule Language Combining OWL and RuleML, 2003. Available from: http://www.daml.org/2003/11/swrl/.

[18] G. Solskinnsbakk, J.A. Gulla, "Combining ontological profiles with context in information retrieval," *Data & Knowledge Engineering*, Vol. 69, No. 3, pp. 251-260, 2010.

[19] I. Cantador, A. Bellogín, I. Fernández-Tobías, S.L. Hernández, "Semantic Contextualisation of Social Tag-based Profiles and Item Recommendations," in *Proceedings of 12th International Conference on E-Commerce and Web Technologies (EC-Web 2011)*, pp. 101-113, 2010.

[20] L.-J. Yang, D.-C. Huang, I-C. Hsu, "Digital library 2.0: Automatically generate EPUB eBook from wiki platform," in *Proceedings of International Conference on Business and Social Sciences (ICBASS)*, 2014, pp. 501-507, 2014.

[21] C. Bizer, T. Heath, T. Berners-Lee, "Linked data-the story so far," *International Journal on Semantic Web and Information Systems*, Vol. 5, No. 3, pp. 1-22, 2009.

[22] G. Klyne, J. J. Carroll, Resource Description Framework (RDF): Concepts and Abstract Syntax, 20 Aug. 2012. Available: http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/

[23] E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF, 15 Jan. 2008. Available: http://www.w3.org/TR/rdf-sparql-query/.

[24] S. Auer, C. Bizer, G. Kobilarov, I. Lehmann, R. Cyganiak, Z. Ives, "Dbpedia: A Nucleus for a Web of Open Data" in *Proceedings of 6th International Semantic Web Conference, 2nd Asian Semantic Web Conference (ISWC 2007 + ASWC 2007)*, pp. 722-735, Springer Berlin Heidelberg, 2007.