The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1066-2243.htm



# Recommending quality book reviews from heterogeneous websites

Duen-Ren Liu, Wei-Hsiao Chen and Po-Huan Chiu
*Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan*

Received 1 September 2011
Revised 6 June 2012
Accepted 7 June 2012

## Abstract

**Purpose** – In recent years, readers have limited amounts of time to pick the right books to read from a market that is filled with similar types of books. Aiming to read only good books, readers tend to check book reviews written by others. However, it is very difficult to find good book reviews. The aim of this paper is to present a book review recommendation system that collects reviews from heterogeneous sources on the Internet and performs quality judgments automatically. Users can then read the top-ranked reviews suggested by this recommendation system.

**Design/methodology/approach** – In this paper, a book review recommendation system is constructed to collect, process, and judge the quality of book reviews from various heterogeneous sources. The quality measurement of book reviews uses review-evaluation techniques. The prediction results were validated with a ranking list produced by experts.

**Findings** – The proposed system is effective and suitable for recommending quality book reviews from heterogeneous sources. The proposed quality measurement method is more effective than other more commonly used methods.

**Originality/value** – This paper is one of the first to apply review evaluation techniques to the process of book review recommendation. The proposed system can collect and recognize book reviews from different websites with various forms of presentation. This evaluation shows that the quality measurement method produces better results than do other methods, such as ranking by rating score or by the date that the review was posted. Those methods are primarily used by commercial websites.

**Keywords** Book review recommendation, Heterogeneous data integration, Book review quality measurement, Readers, Books

**Paper type** Research paper

## 1. Introduction

In recent years, the market has been filled with very similar types of books, but readers have limited time to pick a good book to read. It is important for readers not to feel regret after purchasing a book. Because readers look to read only good books, they tend to check book reviews written by others. The world wide web has become thoroughly integrated into the life of modern people, and after the expansion of Web 2.0, a conceptual technology of communication and sharing, diversified book reviews have appeared across the internet. However, these book reviews are usually scattered around the internet and it is very difficult to acquire them in one system; moreover, it is difficult to integrate and present high-quality book reviews to users all at once. For example, when searching for book reviews of a certain book, the search results might be scattered among personal blogs and online bookstores. Reviews might even be found in forums or bullet board systems. Therefore, if the user wants to read all the

This research was partially supported by the National Science Council of Taiwan under grant NSC 100-2410-H-009-016 and NSC 100-2811-H-009-011.

Internet Research
Vol. 23 No. 1, 2013
pp. 27-46

1066-2243
DOI 10.1108/10662241311295764

book reviews for a certain book, the only way of getting data from heterogeneous web sites is by keyword searches; moreover, each web site would have to be accessed individually to determine whether the book review contained therein makes sense to the user. However, this method of using search engines is very time consuming and laborious.

Moreover, keyword searches will usually result in some irrelevant links because the search engine lacks semantic relationship analysis or accurate conditional search capabilities. Errors could easily be introduced if the judgment of the data was done merely through keywords. Even in similar web sites, the display format of data could be totally different. For example, the total number of pages of a book and the ages for which it is appropriate might not appear in all of the web sites of online bookstores. In this study, book reviews from diverse sources, for example, from a bookstore web site, the bibliophile social media site aNobii, a blog of a famous writer, or even the blog of an ordinary person, can be collected together. Promoting the book this way can stimulate an interested reader to progress from browsing to purchasing and reading the book.

There are a plethora of book reviews available, but their quality is difficult to guarantee. Readers do not usually have enough time to read all available reviews. Therefore, in addition to collecting book reviews from heterogeneous sources, we propose a method to evaluate the quality of each book review automatically. The order in which the book reviews are displayed depends on their quality to reduce the time users spend searching for high-quality reviews. The judgment or comparison of book reviews is required for an effective recommendation system. However, there is scant research for predicting the book review quality. The current recommendation mechanism is limited to individual web sites (Agichtein *et al.*, 2008; Denecke and Nejdl, 2009), without considering the recommendation across heterogeneous web sites. Although Amazon's online bookstore possesses many book reviews and book recommendation functions, it is limited to analyzing on its own internal data (Mizil *et al.*, 2009). The so-called top list (the most helpful customer reviews mechanism provided on the site) leads to a "rich get richer" effect (Lu *et al.*, 2010).

In this paper, we propose a novel recommendation system that can automatically collect heterogeneous book reviews, judge their quality and recommend quality book reviews to users. We propose a novel data structure to integrate heterogeneous book profiles and book reviews from different web sites. We also propose a review quality measurement approach to measure the quality scores of reviews. Several studies have been done to measure the quality of reviews (Figueiredo *et al.*, 2009; Jurca *et al.*, 2010; Nguyen *et al.*, 2007; Riggs and Wilensky, 2001), but none applies to book review quality evaluation. Our proposed book review recommendation system recommends book reviews according to the ranking of their quality scores. Users can read the top-ranked reviews suggested by the recommendation system. Finally, we conducted experiments to evaluate the effectiveness of our proposed system. The experimental results demonstrate that our review quality measurement approach can achieve better quality prediction results than other methods, such as ranking by rating score or the date of review posted. These last methods are used mostly by commercial web sites.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of related works. In Section 3, we describe our research objective, methods, system overview and the main modules of the proposed recommendation system. The details of review quality measurement are presented in Section 4. Section 5 details our

system implementation and evaluation. Then, in Section 6, we summarize our conclusions and consider future research directions.

## 2. Related works

Our proposed work focussed on sharing heterogeneous book reviews, judging their quality and recommending quality book reviews. Section 2.1 presents related works on opinion sharing in Web 2.0 applications that provide backgrounds and motivations for sharing book reviews on the internet. Section 2.2 illustrates related works on judging the quality of the information content or reviews. Moreover, user reputation is an important factor for judging the quality of opinions or reviews. Section 2.3 illustrates the related works on how to derive user reputations. Finally, related works on recommender systems are presented in Section 2.4.

### *2.1 Opinion sharing*

In Web 2.0 applications, user-generated content is extensive. Users share their experiences and feelings on the internet about items they have purchased. Consumers increasingly browse reviews about the products that they want to buy to avoid making a bad decision. The traditional face-to-face word-of-mouth recommendation has become an electronic word-of-mouth recommendation in the Web 2.0 environment (Cheung *et al.*, 2008; Karakaya and Barnes, 2010). Recommendations from a virtual community such as backpackers' forum could influence travelers' decisions (Ku, 2011). For the manufacturers, understanding consumer opinions about their own or competitors' products helps in developing a marketing plan or maintaining customer relationships (Morinaga *et al.*, 2002). The fact that there are various data sources makes users spend time collecting information. Decisions made after seeing a few opinions can very easily result in a bad choice. Integrating multiple information sources is an important issue for online consumers.

### *2.2 Information quality judgment*

The internet infrastructure supports simple and straightforward information acquisition, but it is less able to manage content in ways to reveal the value of information. Some evaluations are needed to judge the quality of the information. Information quality can be evaluated in several ways. Kim *et al.* (2007) provides an optimal answer selection metric, which can be divided into seven categories and 24 indices. Agichtein *et al.* (2008) uses three features (intrinsic, relation and usage features) to find high-quality content in social media. Wijnhoven *et al.* (2011) proposed to determine file retention policies by identifying casual relations between file parameters (e.g. frequency of access, file age, last modification time, file type and user grade) and the use values of files. They use factor analysis to identify important file retention parameters based on subjective use values of files measured by a sample of users. However, it is not clear how the identified parameters are used to generate a predictive use value measure. Poston and Speier (2005) investigated how content ratings and credibility indicators affect knowledge workers' content search and evaluation processes. Three credibility indicators, including "number of raters," "rater expertise" and "collaborative filtering (CF)," are examined to determine the effect of assisting knowledge workers in using content ratings.

Additionally, the quality of reviews has to be evaluated. Feature-based methods have been used to evaluate review quality; they can be classified into two categories, textual and non-textual. Textual feature-based approaches are mainly used to

summarize the content of reviews with a clear and brief review summary. Natural language processing and text mining techniques are employed in the summarization process (Morinaga *et al.*, 2002; Turney, 2002; Hu and Liu, 2004; Nguyen *et al.*, 2007; Ramkumar *et al.*, 2010). Non-textual feature-based approaches evaluate features such as time factor, review length, user reputation and social context. O'Mahony and Smyth (2010) classified the features into four categories: user reputation, social, sentiment and content features. The first three categories are non-textual, and the last category is textual. Jeon *et al.* (2006) proposed a framework to predict the quality of answers for Q&A (question and answer) applications with non-texture features. They suggested that high-quality answers are usually longer than low-quality answers although very long and low-quality answers also exist. Jurca *et al.* (2010) also found that lengthy comments are generally considered more useful in online review forums. The time attribute can help to determine the strength of the influence of a review. Riggs and Wilensky (2001) suggested that reviewers may be influenced by earlier reviews, so more credit may be given to earlier reviews. The reputation of the rater (reviewer) is also an important factor for quality judgment. Raters with high reputation usually provide high-quality reviews (Chen and Singh, 2001) and experts usually generate high-quality opinions within a specific domain (Kim *et al.*, 2009). Lu *et al.* (2010) used social context to predict review quality. However, there is no social information available between book reviewers from different book review sources. Additionally, there is no user behavior information in our collected book review dataset. It is difficult to analyze heterogeneous book reviews by using textual feature-based approaches since multi-lingual analysis of review content is required. Moreover, accurately understanding the contents of the text is very difficult with the current technology. Thus, we focus on non-textual features in this paper.

The above studies have investigated various methods and metrics to evaluate the quality of information such as Q&A answers, reviews, files or content for different applications. But these studies have not addressed the issue of judging the quality of book reviews. The method of applying these metrics to evaluate the quality of book reviews needs to be investigated. These evaluation metrics need to be appropriately adjusted for judging the quality of book reviews.

### *2.3 User reputation*

On community web sites, anyone can give reviews. However, there are high-quality and low-quality reviews. A high-quality review can be used as reference in decision making, whereas a low-quality review is not worth reading. In searching for high-quality reviews, user reputation provides an index. For evaluating review quality, user reputation can be used as a standard too. Among non-textual features, the answerer's acceptance ratio, i.e. the ratio of best answers to all the answers that the answerer answered previously, is highly correlated with high-quality answers and could be seen as a measure of user reputation (Jeon *et al.*, 2006). Although reputation is not clearly mentioned, user reputation can be derived by analyzing the reviews of the reviewer (Riggs and Wilensky, 2001). In general, a reviewer will have higher reputation if his or her ratings are more consistent with those of most reviewers (Chen and Singh, 2001; Riggs and Wilensky, 2001). Above researches provide methods for computing user reputation, but these methods are limited to single community web site. Data from multiple web sites cannot be integrated.

*2.4 Recommender systems*

The recommender system is widely used to provide suitable personalized information to users according to their needs and preferences (Adomavicius and Tuzhilin, 2005; Godoy *et al.*, 2010; Kazienko and Adamski, 2007; Liu and Shih, 2005; Sarwar *et al.*, 2000). There are two widely used filtering methods: CF (Konstan *et al.*, 1997; Schafer *et al.*, 2007), and content-based filtering (CBF) (Mooney and Roy, 2000; Pazzani and Billsus, 2007). The CF approach uses historical data related to preferences to find user neighbors or item neighbors, and makes recommendations based on similar users' opinions or similar items. CBF approach analyzes the customers' preferences on the attribute features of item to build up a personal feature profile, and then predict which items the customer will like. Most researches on recommender systems have been focussed on the technical aspects. Besides the technical aspects, a conceptual model with psychological constructs is proposed to analyze the consumer's adoption and use of a web site recommendation system from the perspective of consumer behavior (Martínez-López *et al.*, 2010). In this paper, we focus on the technical aspects of recommending quality book reviews from heterogeneous sources of book reviews for a specific book according to the ranking of their quality scores.

## 3. Book review recommendation system

*3.1 Research objective and approach*

The main objective of this research is to design an effective book review recommendation system that can help users find quality book reviews from heterogeneous sources on the internet. The goal is achieved by ranking book reviews based on the computation of their quality scores. A novel data integration structure is proposed to integrate heterogeneous book profiles and book reviews from different web sites into a single graphic user interface (web page) for convenience. The book profiles and book reviews from three online bookstores are collected and integrated based on the proposed data structure. The data structure provides a consistency rule so that data from different sources can be recognized by correct semantic context. Using the consistency property, we can apply a semantic query to these data to provide more useful user functions. Moreover, a novel book review quality measurement approach is proposed herein to evaluate the quality scores of reviews. According to the ranking of book reviews' quality scores, the system recommends the top-ranked reviews to the users. Note that we focus on recommending quality book reviews from heterogeneous sources of book reviews for a specific book without considering users' personal interests in the books. The recommendation of book reviews is conducted according to the ranking of the measured quality scores of book reviews rather than using personalized recommendation techniques such as CF or CBF methods.

Existing studies have investigated various textual and non-textual feature-based methods and metrics to evaluate the quality of information for different applications, as illustrated in Section 2.2. But these studies have not addressed the issue of judging the quality of book reviews. These evaluation metrics need to be appropriately adjusted for judging the quality of book reviews. In addition, Wijnhoven *et al.* (2011) proposed the use of factor analysis to identify important file retention features based on users' subjective use value of files. However, they did not explain how the identified features are used to generate a predictive use value measurement. In this work, we propose to derive the predictive quality score of a book review by a linear combination of the scores of the book review features with parameters to adjust their respective weightings. The parameters are systematically adjusted to the values that can result in

the best match with the human rankings of book reviews for the analytical dataset. We contribute by adopting appropriate features for judging the quality of a book review and proposing a novel systematic approach to derive the predictive quality scores of book reviews.

Section 3.2 gives an overview of our proposed book review recommendation system. The details of our methods for collecting and integrating book reviews from heterogeneous web sites are presented in Sections 3.3-3.7. The details of book review quality measurement approach are presented in Section 4.

*3.2 System overview and architecture*

Our proposed book review recommendation system is mainly developed to help users find quality book reviews from heterogeneous sources on the internet. The entire system uses book profiles to search for relevant book reviews from heterogeneous sources on the internet, and the output is a ranked list of multi-lingual book reviews that is sorted by book review quality score. When the user chooses to view a book review of a certain book, the top-ranked quality book reviews are displayed at the top of the page to the user.

To achieve this purpose, our proposed system is comprised of four functional modules which are required to generate the top-ranked quality book reviews from heterogeneous data sources on the internet, as shown in Figure 1. The book review

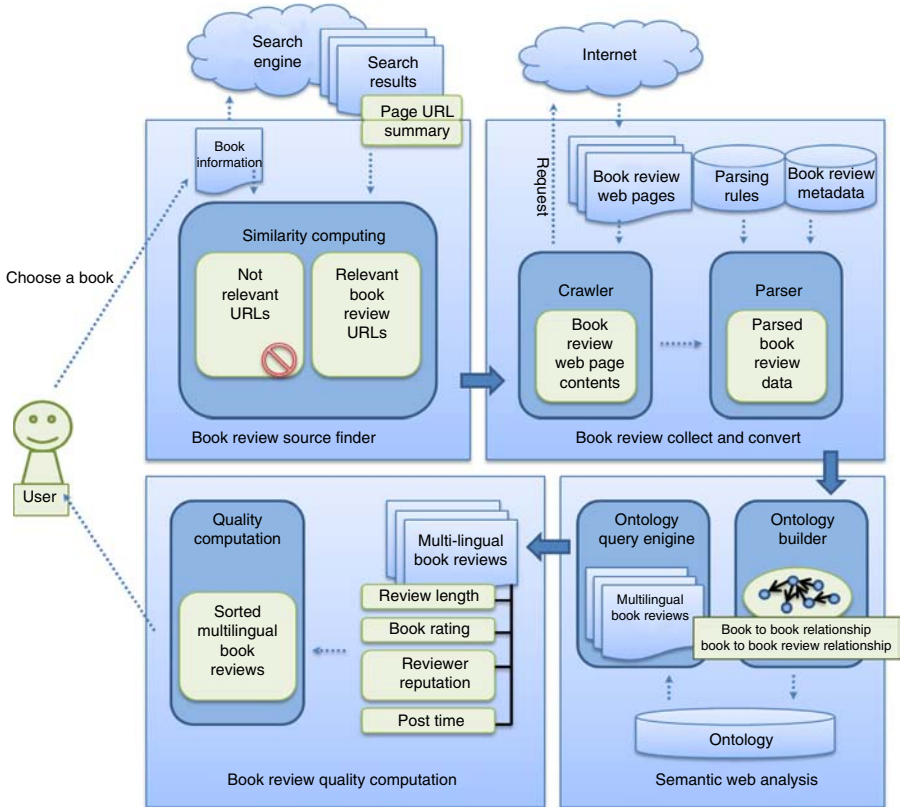


**Figure 1.**
System architecture

source finder (BRSF) locates relevant web pages of books and reviews from the internet by search engine. The BRSF module is described in Section 3.3. The book review collection and conversion (BRCC) process downloads web pages located by the BRSF module, and then uses pre-defined parsing rules to parse the information needed from the original web pages. The BRCC module is described in Section 3.4. The semantic web analysis (SWA) identifies semantic relations between books and book reviews based on ontology (semantic context), and integrates heterogeneous book profiles and book reviews by utilizing a heterogeneous data integration structure. The SWA module identifies the book reviews of a book with heterogeneous data formats and languages to obtain multi-lingual book reviews. The integration structure and SWA module are illustrated in Sections 3.5 and 3.6, respectively. The book review quality computation (BRQC) consists of a set of book review quality measurement functions; it utilizes information related to book reviews to derive the book review quality score. The BRQC module is described in Section 3.7. The more detailed descriptions of quality judgment are presented in Section 4. Finally, the book review quality scores are used for ranking the collected book reviews. All of the book reviews of the same book are ranked according to their quality score to be displayed to the user.

*3.3 BRSF*

Book reviews are scattered across the internet, so the system must search for them. The BRSF component, as shown in Figure 2, uses search engines, such as Google or Yahoo!, to find the URLs of book reviews by searching for strings of book information (e.g. book name, author, etc.). When the search engine returns raw results, judgments must be made about the relevant book information. Searches might return irrelevant web pages, for instance, the reviews of movies with scripts adopted from the book in question. The system finds relevant book information by using the content values of

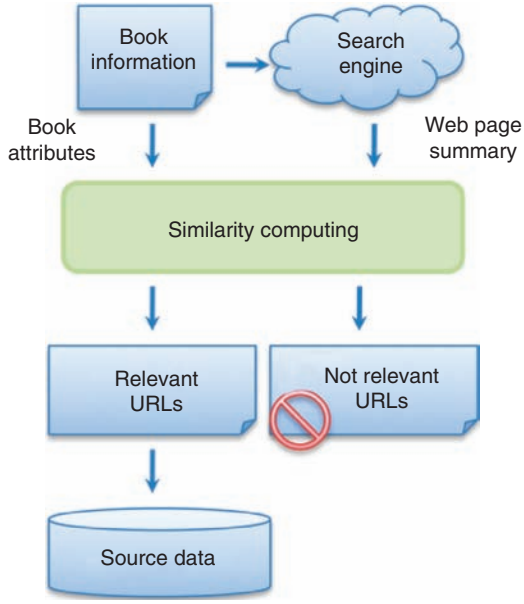


**Figure 2.**
Book review source finder

book attributes (e.g. book name, author) to compute the similarity in the relevance judgments. The URLs that are not relevant will be dropped; the relevant URLs are stored as the source data in the database.

*3.4 BRCC process*

We have located the information in the previous step. The reviews that were found must be downloaded from the internet. Because they are presented as web pages, the text of the review is mixed with HTML tags, advertisements and noisy text. The conversion process parses and fetches the desired information from the raw web page files. The core of the BRCC procedure is the item manager, which includes two major modules, the crawler and the parser, which are shown in Figure 3. The operation of the item manager includes three steps.

The first step judges the data source and generates a corresponding instance of the item manager. For example, if the input data source is a book, then a book item manager will be generated; if the input data source is a book review, then an instance of the book review manager will be generated. We use object-oriented programming to develop the item manager. The item manager is an abstract class; book/book review item manager is its derived class. Next, the crawler will follow the designated data sources to download the data and store them in the system to wait for the next processing step. In the last step, the parser will perform information parsing on the file obtained during the previous step. Each data source will have a corresponding information-parsing rule, and the parser will parse the file according to the rules applicable to the data type being processed. The parsing result is the related information of that object, for example, the reviewer, book rating, posted date and content of book review are parsed from the book review information, an example of which is shown in Figure 4.

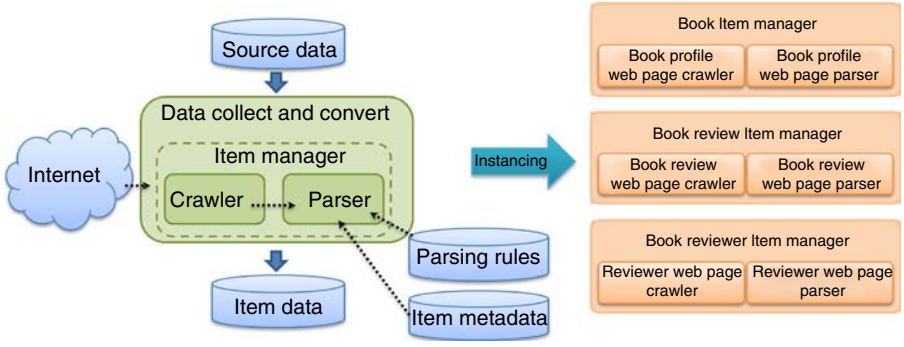


**Figure 3.**
Book review data collection and conversion process


| Review no. 1 | | Review no. 2 | |
|---|---|---|---|
| Field | Value | Field | Value |
| Reviewer | 蘇菲 | Reviewer | 盧豫榛 |
| Posted date | 11-06-2009 | Posted date | 11-11-2009 |
| Review | ..... | Review | ..... |
| Reviewer location | Taipei, taiwan | Reviewer location | Kaohsiung, taiwan |
| Book rating | 5.0 | Book rating | 5.0 |

**Figure 4.**
An example of parsed book reviews

*3.5 Heterogeneous data integration*

Book information and book reviews are collected from different data sources, for example, commercial bookstore web sites and personal blogs. Many of these sites have different formats for storing data. There are two primary problems encountered because of these differences. First, no two sources represent a book in exactly the same way. Thus, one web site might describe a book using five attributes, while another uses seven. The second problem is that even if two stores use the same attributes, the same book might have different values for the same attribute. A book written in English will be translated into a Chinese version, which appears on the Chinese bookstore web site. The same book has both English and Chinese titles and therefore different values for the title attribute, but, in fact, the two values have same semantic meanings. A mechanism is needed to store various attributes belonging to the same book. To solve the previously mentioned data integration issue, we propose a data storage structure as shown in Figure 5.

The data storage structure is comprised of a metadata part and a data part. The metadata part defines how to describe an item, that is, what attributes an item has. The data part stores the attribute values of an item. The metadata part solves the first problem, while the data part handles the second problem. The metadata part is twofold: class metadata and field metadata. The class metadata represents what type of thing the object is, for example, whether it is a book or book review. The field metadata represents the attributes owned by the object type it belongs to. For example, a book object includes the book title, author and publishing date; a book review object contains the author, posted date, location, review (content) and book rating. All the attributes of a book or book review, even if taken from multiple sources, are defined using this structure. When the system finds a new book attribute later, it is simply added to the field metadata. This addition will not affect existing data. The metadata part has the advantage of being able to integrate data from heterogeneous sources. By increasing the amount of class metadata information, more types of information can be provided by the system, so it would not be limited only to book reviews.

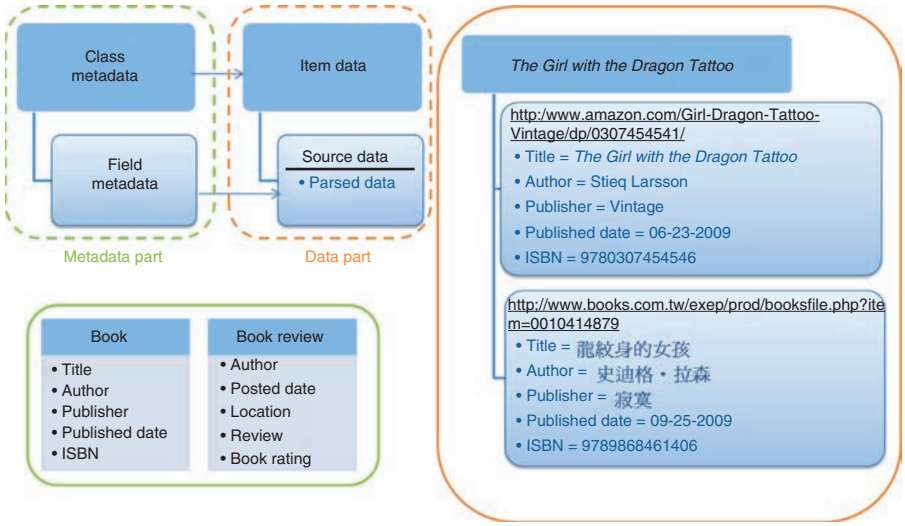


**Figure 5.**
Heterogeneous data integration data structure

The data part is split into three types: item data, source data and parsed data. The item data variable represents the object instance of a certain type, such as a book named *The Girl with the Dragon Tattoo*. The source data variable stores the address of the source of the data on the internet, which might be a web address such as the URL of a commercial bookstore web site. The parsed data variable stores the result of parsing the source data. At the right side of Figure 5, the two light blue squares represent two data sources for the book *The Girl with the Dragon Tattoo*, Amazon.com and Books.com.tw, respectively. The parsed data include the following fields: the book author, "Stieq Larsson," the publisher data, "Vintage," the ISBN, "9780307454546." So, if a book has various translated versions, they are stored as different source datum. It is possible to know what the book is, where the book data comes from, and what information the book has from the data part.

*3.6 SWA*

The most commonly used file format on the internet is HTML. However, HTML is mainly designed for humans to read and write. The computer cannot grasp the meaning information written in HTML format. To let computers recognize the meaning of data on the internet, it is necessary to mark up the data with metadata. This method is called the "semantic web." After enabling computers to comprehend content on the internet, some supervised tasks can be completed automatically by computers. One example is the integration of heterogeneous databases. Much of the data on the web have semantic relations. After the construction of the semantic web, computers will be able to judge the data more clearly by drawing inferences from the inference engine.

The data in our proposed system are collected from various web sites on the internet. Because of these variegated sources, it is necessary for our system to solve the synonym judgment problem (Figure 6) and establish the semantic relations between books and book reviews (Figure 7). These problems can be overcome using SWA. In Figure 6, different attribute names have the same meaning. We can use these synonyms to perform more complete retrievals of book information after building the synonym relationships. The relationships are represented in the semantic web. Some research projects use the WordNet to complete this work. However, it is possible to supervise the building of the synonym relationships in our system because the number of book attributes is low. We note that a word may have more than one meaning in different domains and may result in the problem of homonyms. In our research, the heterogeneous book data is collected for each target book, and the words contained in the book data are most likely in the same domain. Accordingly, our current system does not consider the problem of homonyms. Nevertheless, handling the problem of homonyms is important to improve our system, and is proposed as a future work.

The semantic relationships are established when a web page that contains a book review is crawled, data about the book, its author, content and book rating score is


**Figure 6.**
Synonyms

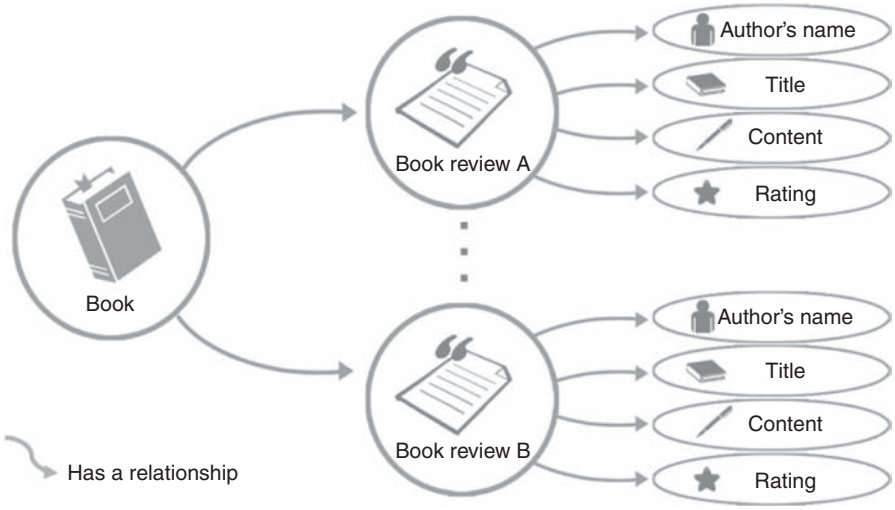


**Figure 7.**
Semantic relations between book and book reviews

discovered. The relations among the item data, source data and parsed data imply the "has-a" relationships. After the semantic web is constructed, we can use a database query system to retrieve the data that the user needs. Our system is based on the assumption that someone seeking book reviews wants to read or evaluate the book for purchasing. The system finds all book reviews of the book in the database. It returns a book review list that contains the book reviews of the target book from various sources, even if the reviews are in different languages. For example, book review A and book review B might come from different sources, as shown in Figure 7. Because of the connection made by the "has-a" relationship, we can understand that the two reviews are discussing the same book. Another example is the book *Harry Potter* in English is equivalent to the book "哈利波特" in Chinese.

### *3.7 Quality computation (QC) and recommendation*

To ensure the quality of the book reviews, they will be scored by several measures. The QC module is composed of a set of book review quality measurement equations. This module is such that it can be self-made. The book review quality measurement equation can be adjusted according to different needs. In this study, the book review quality is derived from three book review attributes and the book reviewer's reputation. The detailed method for deriving this content is illustrated in Section 4.

After the quality computation module has processed all of the related information, a book review score is obtained; the system then ranks the book reviews according to their scores. The book review with the highest score is displayed at the top, while the book review with the worst score is displayed at the bottom. This arrangement makes it easy for the user to start reading from the best book review until he or she has obtained sufficient information about whether to purchase or read the book.

## 4. Review quality measurement

Various textual and non-textual feature-based methods and metrics have been proposed to evaluate the quality of information for different applications, as illustrated in Section 2.2. It is difficult to analyze heterogeneous book reviews by using textual

feature-based approaches, since multi-lingual analysis of review content is required. Thus, we focus on non-textual features in this paper. There is no social context and user behavior information available in the collected book review dataset. Accordingly, we adopt four non-textual features of book review: book review length, time factor, book rating and reviewers' reputation, to measure the book review quality. The length attribute is easy to determine because there is no need for complicated analysis or computation (Jeon *et al.*, 2006). The review length represents the enrichment of the reviews, and generally longer reviews are more helpful for users (Jurca *et al.*, 2010). The time attribute represents the originality of the reviews, and can help to determine the strength of the influence of a book review. Earlier reviews are more influential than later reviews since a more recent review has the most antecedent reviews available for consultation (Riggs and Wilensky, 2001). Book rating is used to indicate the "book rating score," which is provided by book reviewers after reading the book. They can use 1-5 to judge how good is the book. For a specific book, when a reviewer's book rating comes closer to the average rating value of the book, the book review of that reviewer would generally be more reliable. Rating consistency is the closeness between the book rating given by the book reviewer for a specific book and the average value for all the reviews of that book; the higher the values of rating consistency, the higher the quality of the book reviews. Moreover, reviewers with high reputation usually provide high-quality reviews (Chen and Singh, 2001). The higher the reputation of a reviewer, the higher the quality of the book review written by that reviewer. In general, a rater will have higher reputation if his or her ratings are more consistent with those of most raters (Chen and Singh, 2001; Riggs and Wilensky, 2001). We adopt a similar idea to derive the reputations of reviewers. These attributes will be described in detail later.

The quality score of a book review is derived by a linear combination of the scores of the four book review features with parameters to adjust their respective weightings. In general, an earlier and longer book review with higher book rating consistency and user reputation will imply a higher quality score. The four book review features may have different importance (weightings) in deriving the quality score. The parameters of the book review quality score formula are adjusted to the values that result in the best match with the human rankings for the analytical data.

We scale all four attributes to be in the range from 0 to 1. Then they are aggregated with four factors, $\alpha$, $\beta$, $\gamma$ and $\lambda$. The computation method is as follows:

$$\text{Review quality score} = \alpha \times R_{u,b} + \beta \times L_{u,b} + \gamma \times T_{u,b} + \lambda \times UR_u$$

where $R_{u,b}$ is the rating consistency of the rating score given by a user $u$ to book $b$; $L_{u,b}$ the review length ratio of the review written by user $u$ of book $b$; $T_{u,b}$ the time variant of the date the review by user $u$ of book $b$ was posted; $UR_u$ the reputation of user $u$. $\alpha$, $\beta$, $\gamma$ and $\lambda$ are the weightings of the book review attributes. $\alpha$, $\beta$, $\gamma$, $\lambda \in [0, 1]$, $\alpha + \beta + \gamma + \lambda = 1$. The details of $L_{u,b}$, $T_{u,b}$, $R_{u,b}$ and $UR_u$ will be described in the following.

We try to determine the optimal combination of the parameter weights by testing all of their combinations. The weights are determined according to the best match with the human rankings that can be achieved under different combinations of weight assignments for the analytical data. To find the best weight combination, we tested various combinations of the $\alpha$, $\beta$, $\gamma$ and $\lambda$ parameters by enumerating their values systematically in increments of 0.1 ranging from 0 to 1.

*Review length ratio*: the book review length ratio is the ratio of the length of a book review to the longest book review for the same book. In addition, any difference in word count for book reviews near the top in length should not have a large effect. For example, if the largest word count is 1,000, then a book review with 950 words and book review of 980 words do not differ significantly in quality. Moreover, the review length is dependent on the style used and the language applied; hence, we take the square root of the previously mentioned length ratio to reduce such effect. A book review length ratio is calculated as follows:

$$L_{u,b} = \sqrt{\frac{I_{u,b}}{\mathrm{Max}_{u_j}(I_{u_j,b})}}$$

where $I_{u,b}$ is the length of the book review written by the book reviewer $u$ about book $b$. A higher value of the book review length ratio is more likely to indicate a higher quality book review. We note that this measure is applied to the reviews of the same book. The review length may vary for different kinds of book.

*Time factor*: according to the herd instinct (Avery *et al.*, 1999), the book reviewer may tend to follow the trend of the previous reviews made by most of the reviewers. Accordingly, earlier reviews are more influential and are more likely to provide constructive review comments than later reviews. Hence, the time factor $T_{u,b}$ is included in the analysis to accommodate such behavior:

$$T_{u,b} = 1 - \frac{Order(t_{u,b}) - 1}{n_b}$$

where $t_{u,b}$ is the posting time of the review written by book reviewer $u$ of book $b$; $n_b$ the total number of book reviews of book $b$; $Order(t_{u,b})$ the function that returns the time sequence of a book review. The review that is posted first has a time sequence of 1. Accordingly, the most recent book review has the smallest value of time factor, whereas the earliest book review has the largest one.

We note that later reviews are not necessarily of lower quality; however, it is difficult to ascertain the "true" quality of reviews. The time order of reviews does somewhat denote the originality of a review. Most viewpoints of later reviews are reactions to earlier reviews. Riggs and Wilensky (2001) also mentioned that since many reviewers may be influenced by earlier reviews, more credit may be given to earlier reviews. Thus, we give earlier reviews greater weight in regard to quality judgment.

*Rating consistency*: the closeness between the rating given by the book reviewer for a specific book and the average value for all the reviews of that book. When the reviewer's rating gets closer to the average value, the rating is more consistent. The rating consistency is computed as follows:

$$R_{u,b} = 1 - \frac{|r_{u,b} - \overline{r_b}|}{MaxRating - MinRating}$$

where $r_{u,b}$ is the rating given by book reviewer $u$ to book $b$; $\overline{r_b}$the average value of all the book ratings given by all the book reviewers to book $b$; *MaxRating* and *MinRating* the maximum and minimum values in the rating field, respectively.

Rating consistency $R_{u,b}$ can represent whether the opinion that book reviewer $u$ has about the book is the same as most of the people. When $\overline{r_b}$is close to *MaxRating*, it means that most of the people have given a high rating to book $b$. When $r_{u,b}$ gets closer to $\overline{r_b}$, book reviewer $u$ takes a view similar to most people. In contrast, when $r_{u,b}$ get farther away from $\overline{r_b}$, it means that the book reviewer $u$ has a view that differs from other people. The higher the values of rating consistency, the higher the quality of the book reviews.

*Reviewer's reputation*: the arithmetic mean of the rating consistencies of all books that the reviewer has previously reviewed. A user's reputation $UR_u$ consists of a user's general (overall) reputation $UR_u^o$ and category reputation, $UR_u^c$, i.e. specific reputation in certain book category $c$. A user's reputation $UR_u$ is computed as follows:

$$UR_u = \delta \times UR_u^o + (1 - \delta) \times UR_u^c$$

where parameter $\delta \in [0, 1]$ is used to adjust the relative the weightings of general reputation and category reputation of users. The computation of a reviewer's overall reputation $UR_u^o$ is as follows:

$$UR_u^o = (1 - \frac{1}{n_u + 1}) \times \frac{\sum_{b \in B_u} R_{u,b}}{n_u}$$

where $B_u$ is the set of all the books reviewed by the book reviewer $u$; $n_u$ the number of the books reviewed by $u$; and $R_{u,b}$ the rating consistency for book reviewer $u$ of book $b$.

Because different reviewers have different areas of expertise, book categories are introduced into the computations of reviewers' reputations. To represent the expertise of each book reviewer, the reputation of the reviewer in every book category is computed individually. The computation of the reviewer's categorical reputation $UR_u^c$ is as follows:

$$UR_u^c = (1 - \frac{1}{n_u^c + 1}) \times \frac{\sum_{b \in B_u^c} R_{u,b}}{n_u^c}$$

where $B_u^c$ is set of the books that are reviewed by $u$ and belong to a given category $c$; and $n_u^c$ the number of the books that belong to the category $c$ and are reviewed by $u$.

The higher the reputation of a book review's author, the more acceptable the book review. To prevent a reviewer with very few reviews being marked as having a good reputation and a review as high quality, a quantity factor $1 - \frac{1}{n_u^c + 1}$ will be added into the computation. When the number of book reviews is higher, the value will be higher; in contrast, if there is only one book review, then the value is only 0.5. If the value of the reputation of the book reviewer $u$ is higher, the book review quality of $u$ will be higher too.

## 5. System implementation and evaluation

We created a web-based prototype system based on the proposed system to demonstrate its feasibility and effectiveness. The web site is built on the Windows XP professional operating system. The web site architecture is model-view-controller. We use a complete Python web application framework called Django to construct this system. The MySQL database system was selected as a method to store and retrieve information.

We performed data abstraction techniques to model the data of book profiles, user reviews and reviewer profiles for manipulation in our system. These data models are presented in the Django framework as classes. The model layer of the Django framework initializes these classes, and then the data are accessed from the database to retrieve the attributes of the classes. Attributes are finally attached to instances of the classes. The user interfaces for our system are web pages that are composed of data models and template pages. The web pages (views) are rendered using the template layer of the Django framework. Controllers are the set of programming functions and are implemented to provide operations in our system. For example, book listing by category, list reviews of a book and show detailed information of a book. One controller performs one requested operation to manipulate the associated model objects to produce the information in the required corresponding view. Each view is rendered from a template with model objects manipulated by a controller.

After the prototype system was implemented, the system needed to collect the book information. It fetched book information in two steps. First, it retrieved books from a monthly list of books published. The list contains the book title and a hyperlink to the book's profile web page. Therefore, a book's information can be collected by parsing its web page profile. We collected data about books in Chinese that were published from August to November in 2009. About 4,400 books of various types were gathered. We also collected data about English books belonging to the "literature and fiction" category of Amazon.com web site published from May to July in 2011. A total of 19,926 book profiles were gathered.

We compared the ranking quality of our approach with three review ranking methods most frequently applied by online bookstores. The three baseline methods used are ordering reviews by book rating score, by earlier review date and by later review date. Since there is no standard result indicating the quality of book reviews, the evaluation is conducted by asking 13 graduate students to manually judge the quality of book reviews. The students were asked to read the target books. After reading the target books, they judge the quality of the book reviews by their own opinions and experiences. Each of them was requested to rank the book reviews according to the quality of the review. Some samples are shown in Table I. In the table, the order of columns represents the quality rank of the review; each row shows a user's judgment. From the table, we can see that the user named "Black" considered the 12th book review to be the best. The human ranking-order of a review is the average of 13 human rankings to the review.

We calculate the *MRRtopK* measures of various methods by comparing their ranking results with the human rankings as follows. For each book, we computed the mean reciprocal rank (*MRR*) for top-*K* ranked reviews, namely *MRRtopK*, to evaluate the ranking results. Our measure was modified from the *MRR* measure (Voorhees and Tice, 1999). For each ranking result derived from our proposed approach, or a compared method, we selected top-*K* reviews for evaluation using the *MRRtopK* measure. Let $R_i$ denote the $i$-th ranking-order review derived by a method. Let $rank(R_i)$ be the human ranking-order of $R_i$ derived from the average of human rankings. The higher the reciprocal rank of a review ($1/rank(R_i)$), the better the quality of the review. Thus, the higher the *MRRtopK* measure of a method, the better the performance of the method. The *MRRtopK* measure is listed as follows:

$$MRRtopK = \frac{1}{K}\sum_{i=1}^{K}\frac{1}{rank(R_i)}$$

In our experiments, $K$ is set as five; i.e. we selected the top-*5* reviews to evaluate the performance of various methods. To easily understand the *MRRtopK* measure, we also calculated the ratio of the *MRRtopK* measure to the perfect match $MRRtopK_{perfect}$ measure. The perfect match with human rankings is achieved when the top-$K$ ranked reviews of a method are also the top-$K$ ranked reviews derived by human rankings:

$$MRRtopK_{perfect} = \frac{1}{K}\sum_{i=1}^{K} 1/i$$

We chose 20 books as the experiment dataset for comparison; each of them had at least 20 book reviews. Our method and the three baseline methods were applied to rank the book reviews of the 20 books. For each method, we derived the *MRRtop5* measure of each book, and then derived the average *MRRtop5* measure of 20 books. The perfect match $MRRtop5_{perfect}$ value is:

$$MRRtop5_{perfect} = (\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5})/5 = 0.4567$$

Our proposed approach derives the predictive quality score of a book review by a linear combination of the scores of the book review features with parameters to adjust their respective weightings. The parameters are systematically adjusted to the values that can result in the highest *MRRtopK* measure, i.e. best match with the human rankings of book reviews for the analytical dataset. To find the best weight combination, which combines the scores of the four book review features, we tested various combinations of the $\alpha$, $\beta$, $\gamma$, $\lambda$ and $\delta$ parameters by enumerating their values systematically in increments of 0.1 ranging from 0 to 1. The best match is achieved when $(\alpha, \beta, \gamma, \lambda, \delta) = (0.2, 0.1, 0.2, 0.5, 0.3)$. Therefore, the values of $\alpha$, $\beta$, $\gamma$, $\lambda$, $\delta$ are set at 0.2, 0.1, 0.2, 0.5 and 0.3, respectively.

As we had obtained the proposed approach with the adjusted parameters, we then compared the ranking quality of our approach with three review ranking methods most frequently applied by online bookstores. The results are shown in Table II. In the

**Table I.**
A sample of the rankings of book reviews of the book, *The Girl with the Dragon Tattoo*

| User | Rank 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Divaka | r01 | r13 | r12 | r10 | r27 | r11 | r26 | r20 | r08 | r02 |
| Babyblue | r12 | r27 | r01 | r13 | r10 | r11 | r26 | r24 | r08 | r15 |
| Black | r12 | r27 | r13 | r11 | r01 | r18 | r02 | r10 | r20 | r04 |
| Vicky | r01 | r13 | r27 | r12 | r11 | r18 | r20 | r02 | r19 | r15 |

**Table II.**
Comparison between the review quality score algorithm and baseline methods

| Algorithm | *MRRtop5* | $MRRtop5/MRRtop5_{perfect} \times 100\%$ |
|---|---|---|
| Rank by rating | 0.1455 | 31.85 |
| Rank by earlier review date | 0.0747 | 16.35 |
| Rank by later review date | 0.0485 | 10.63 |
| Rank by review quality score | 0.2328 | 50.99 |

table, the *MRRtop5* and the percentage of perfect match ($MRRtop5_{perfect} = 0.4567$) are given. A larger value indicated a better result. As the table demonstrates, our proposed approach of ranking by review quality score has a better result than any of the baseline methods. It also shows that earlier reviews are better than later reviews.

## 6. Conclusion and future works

We have proposed an automatic book review recommendation system, which has solved the problem of searching for book reviews. Users usually have to use a search engine to find book reviews scattered across different web sites and must filter the results themselves. The user spends much time and effort searching for and filtering the information. This time and effort can be greatly reduced by a recommendation system. We also have proposed a heterogeneous data integration solution, which can solve many issues, such as increasing the size of a system in light of new information, and integrating data stored in different ways. Moreover, the issue of data attribute expansion has also been addressed by our system. Our system has the flexibility of using the semantic web to perform cross-language data integration, multi-language recommendations and providing results for complicated semantic queries. To predict the quality of book reviews, a novel book review quality measurement method is proposed. It considered three review attributes combined with the reviewers' reputations to predict the quality of book reviews. Our experiments show that the proposed review quality measurement method can obtain better prediction quality than other methods used widely by commercial web sites.

In this paper, we contribute to proposing a novel recommendation system that can automatically collect heterogeneous book reviews, judge their quality and recommend quality book reviews to users. There are two aspects to our contribution. First, while existing studies have investigated various textual and non-textual feature-based methods and metrics to evaluate the quality of information for different applications, these studies have not addressed the issue of judging the quality of book reviews. We contribute to adopting appropriate features for judging the quality of a book review and proposing a novel systematic approach to derive the predictive quality scores of book reviews. Second, the current studies did not consider the issue of recommending items from heterogeneous web sites. We contribute to proposing a novel data structure to integrate heterogeneous book profiles and book reviews from different web sites.

In the future, we will expand and evaluate this quality measurement equation and perform a weighting adjustment on the measurement indices so that the system can recommend book reviews to the user more accurately and effectively. We will also perform the evaluation with a larger number of books to examine the performance of our system. Moreover, our current work focusses on recommending quality book reviews from heterogeneous sources of book reviews for a specific book without considering users' personal interests on books. In future work, we will integrate the techniques of recommender systems such as CF and content-based approach to expand our system to recommend not only quality book reviews but also books that suit users' personal interests.

## References

Adomavicius, G. and Tuzhilin, A. (2005), "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17 No. 6, pp. 734-49.

Agichtein, E., Castillo, C., Donato, D., Gionis, A. and Mishne, G. (2008), "Finding high-quality content in social media", *Proceedings of the ACM International Conference on Web Search and Web Data Mining, Palo Alto, California, USA, February 11-12*, pp. 183-94.

Avery, C., Resnick, P. and Zeckhauser, R. (1999), "The market for evaluations", *American Economic Review*, Vol. 89 No. 3, pp. 564-84.

Chen, M. and Singh, J.P. (2001), "Computing and using reputations for internet ratings", *Proceedings of the 3rd ACM Conference on Electronic Commerce, Florida, USA, October 14-17*, pp. 154-62.

Cheung, C.M.K., Lee, M.K.O. and Rabjohn, N. (2008), "The impact of electronic word-of-mouth: the adoption of online opinions in online customer communities", *Internet Research*, Vol. 18 No. 3, pp. 229-47.

Denecke, K. and Nejdl, W. (2009), "How valuable is medical social media data? Content analysis of the medical web", *Information Sciences: An International Journal*, Vol. 179 No. 12, pp. 1870-80.

Figueiredo, F., Belém, F., Pinto, H., Almeida, J., Gonçalves, M., Fernandes, D., Moura, E. and Cristo, M. (2009), "Evidence of quality of textual features on the web 2.0", *Proceedings of the 18th ACM Conference on Information and Knowledge Management, Hong Kong, November 2-6*, pp. 909-18.

Godoy, D., Schiaffino, S. and Amandi, A. (2010), "Integrating user modeling approaches into a framework for recommender agents", *Internet Research*, Vol. 20 No. 1, pp. 29-54.

Hu, M. and Liu, B. (2004), "Mining and summarizing customer reviews", *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Seattle, WA, USA, August 22-25*, pp. 168-77.

Jeon, J., Croft, W.B., Lee, J.H. and Park, S. (2006), "A framework to predict the quality of answer with non-textual features", *Proceeding of the 29th ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, WA, USA, August 6-11*, pp. 228-35.

Jurca, R., Garcin, F., Talwar, A. and Faltings, B. (2010), "Reporting incentives and biases in online review forums", *ACM Transactions on the Web (TWEB)*, Vol. 4 No. 2, pp. 1-27.

Karakaya, F. and Barnes, N.G. (2010), "Impact of online reviews of customer care experience on brand or company selection", *Journal of Consumer Marketing*, Vol. 27 No. 5, pp. 447-57.

Kazienko, P. and Adamski, M. (2007), "AdROSA – adaptive personalization of web advertising", *Information Sciences*, Vol. 177 No. 11, pp. 2269-95.

Kim, S., Oh, J.S. and Oh, S. (2007), "Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective", *Proceedings of the American Society for Information Science and Technology*, Vol. 44 No. 1, pp. 1-15.

Kim, S.W., Chung, C.W. and Kim, D.E. (2009), "An opinion-based decision model for recommender systems", *Online Information Review*, Vol. 33 No. 3, pp. 584-602.

Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R. and Riedl, J. (1997), "GroupLens: applying collaborative filtering to Usenet news", *Communications of the ACM*, Vol. 40 No. 3, pp. 77-87.

Ku, E.C.S. (2011), "Recommendations from a virtual community as a catalytic agent of travel decisions", *Internet Research*, Vol. 21 No. 3, pp. 282-303.

Liu, D.-R. and Shih, Y.-Y. (2005), "Integrating AHP and data mining for product recommendation based on customer lifetime value", *Information & Management*, Vol. 42 No. 3, pp. 387-400.

Lu, Y., Tsaparas, P., Ntoulas, A. and Polanyi, L. (2010), "Exploiting social context for review quality prediction", *Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, April 26-30*, pp. 691-700.

Martínez-López, F.J., Rodríguez-Ardura, I., Gázquez-Abad, J.C., Sánchez-Franco, M.J. and Cabal, C.C. (2010), "Psychological elements explaining the consumer's adoption and use of a website recommendation system: a theoretical framework proposal", *Internet Research*, Vol. 20 No. 3, pp. 316-41.

Mizil, C.D.N., Kossinets, G., Kleinberg, J. and Lee, L. (2009), "How opinions are received by online communities: a case study on amazon.com helpfulness votes", *Proceedings of the 18th International Conference on World Wide Web, Madrid, Spain, April 20-24*, pp. 141-50.

Mooney, R.J. and Roy, L. (2000), "Content-based book recommending using learning for text categorization", Proceedings of the Fifth ACM Conference on Digital Libraries, San Antonio, TX, June 2-7.

Morinaga, S., Yamanishi, K., Tateishi, K. and Fukushima, T. (2002), "Mining product reputations on the web", *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, July 23-26*, pp. 341-9.

Nguyen, P., Mahajan, M. and Zweig, G. (2007), "Summarization of multiple user reviews in the restaurant domain", Microsoft Research Technical Report No. MSR-TR-2007-126, Microsoft, Redmond, WA, pp. 1-18.

O'Mahony, M.P. and Smyth, B. (2010), "A classification-based review recommender", *Knowledge-Based System*, Vol. 23 No. 4, pp. 323-9.

Pazzani, M. and Billsus, D. (2007), "Content-based recommendation systems", in Brusilovsky, P., Kobsa, A. and Nejdl, W. (Eds), *The Adaptive Web: Methods and Strategies of Web Personalization*, Springer, Berlin, pp. 325-41.

Poston, R.S. and Speier, C. (2005), "Effective use of knowledge management systems: a process model of content ratings and credibility indicators", *MIS Quarterly*, Vol. 29 No. 2, pp. 221-44.

Ramkumar, V., Rajasekar, S. and Swamynathan, S. (2010), "Scoring products from reviews through application of fuzzy techniques", *Expert Systems with Applications: An International Journal*, Vol. 37 No. 10, pp. 6862-7.

Riggs, T. and Wilensky, R. (2001), "An algorithm for automated rating of reviewers", *Proceeding of the 1st ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, Virginia, USA, June 24-28*, pp. 381-7.

Sarwar, B., Karypis, G., Konstan, J. and Riedl, J. (2000), "Analysis of recommendation algorithms for e-commerce", *Proceedings of the 2nd ACM Conference on Electronic Commerce, Minneapolis, MN, USA, October 17-20*, pp. 158-67.

Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S. (2007), "Collaborative filtering recommender systems", in Brusilovsky, P., Kobsa, A. and Nejdl, W. (Eds), *The Adaptive Web: Methods and Strategies of Web Personalization*, LNCS 4321, Springer, Berlin, pp. 291-324.

Turney, P.D. (2002), "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews", *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, Philadelphia, PA, USA, July 6-12*, pp. 417-24.

Voorhees, E.M. and Tice, D.M. (1999), "The TREC-8 question answering track evaluation", *Text Retrieval Conference TREC-8, Gaithersburg, Maryland, November 17-19*, pp. 83-105.

Wijnhoven, F., Amrit, C. and Bax, M. (2011), "The applicability of a use value-based file retention method", Proceedings of the 19th European Conference on Information Systems, Helsinki, June 9-11.

## Further reading

Guo, J., Xu, S., Bao, S. and Yu, Y. (2008), "Tapping on the potential of Q&A community by recommending answer providers", *Proceeding of the 17th ACM Conference on Information and Knowledge Management, Napa Valley, California, USA, October 26-30*, pp. 921-30.

Tsang, A.S.L. and Prendergast, G. (2009), "Is a 'star' worth a thousand words?: the interplay between product-review texts and rating valences", *European Journal of Marketing*, Vol. 43 Nos 11/12, pp. 1269-80.

Yang, J.M., Li, K.F. and Zhang, D.F. (2009), "Recommendation based on rational inferences in collaborative filtering", *Knowledge-Based Systems*, Vol. 22 No. 1, pp. 105-14.

**Corresponding author**
Duen-Ren Liu can be contacted at: dliu@mail.nctu.edu.tw