# 1

# General Introduction to Recommender Systems

## 1.1. Putting it into perspective

Before the emergence of modern information systems, individuals developed the habit of recommending products or services through "word of mouth", sharing certain social or cultural affinities [OBR 77, SHA 95]. This approach, which can be qualified as social, pursued the principle of sharing an individual experience with others, in areas, at first, as wide as culture or handicraft and then industry. Beyond the reputation tied to the intrinsic quality of a product, there were assessments that emerged through the prism of sociocultural mediums which also improved products and services.

Today, offers – whether information or products – are increasing day-by-day, proposed on the Internet. Beyond a certain threshold, too much information can lead to a deterioration of the quality of the message, which we refer to as information overload [LEV 98, CHE 09]. For the end user in search of information, it is of interest for the system to carry out preprocessing in order to filter the least important elements, in line with their expectations. The development of automated recommender systems (RecSys) is therefore a foreseeable phenomenon for contributing toward resolving the problem of information overload, valuing content and focusing attention on the user in such a context of overabundance.

Chapter written by Ghislaine CHARTRON and Gérald KEMBELLEC.

The first recommender systems, using “collaborative filtering”, had the aim of using the volume of community evaluations in order to propose personalized cultural advice, based on evaluation statistics and the correlation of user profiles [RES 94].

As early as 2000, Burke remarked that many commercial websites such as Amazon or even eBay had understood the purpose of contextualizing peripheral hyperlink offers consulted by the user [BUR 00]. Commercial search engines have even created related products such as “Google AdSense” in order to optimize advertising profits by taking advantage of recommendations based on the contents of queries, or even e-mails[1]. The principle is simply to propose private advertisers to provide hyperlinks directed toward their website in the margin of content selected by the user. This second method is called the content-based method.

With the arrival of social networks, be they in the public or professional spheres, sharing and the evaluation of content have become a mass worldwide phenomenon. As a result of this unprecedented generation of data, mercantile diversions are common and have led AFNOR[2] to propose standards for controlling the phenomenon [AFN 13].

## 1.2. An interdisciplinary subject

The first notable papers confirming recommender systems as a dedicated area of study and research involved computer science specialists as well as economists invested in the emerging development of e-commerce. The issue of information systems unified them; it has become a decisive factor in the decision-making of organizations. Thus, the precursory paper by Paul Resnick (AT&T) and Hal R. Varian (Berkeley School of Information Management) in 1997 focused on the functional analysis of five precursory recommender systems by mostly concentrating on the business model and risks of corruption of such systems [RES 97]. In 2000, Robin Burke, a researcher in computer science, prioritized mentioning the emergence of large catalogs and the required assistance for the consumer in making their choices; his articles focused on the design of algorithms and their

[1] Google Mail FAQ, regarding advertising advice. Accessed online on 29 August 2014 at https://support.google.com/mail/answer/6603.

[2] AFNOR is the French standardization organization and is a member of the ISO.

performance [BUR 00]. E-commerce and recommendation algorithms were originally linked.

The data in Table 1.1, collected by consulting the digital library of publications of the Association for Computing Machinery (ACM) about the thematic area of “Recommender System” in the titles of articles, show the increase in interest in this subject over the last 5 years. This count remains partial compared to the set of articles published by other publishers on this subject over the same period. The growth of information as well as the major development of online commercial platforms explains for the most part the stakes associated with the issue; its development goes hand in hand with the optimization of information systems and the needs of e-marketing.

| | |
|---|---|
| 1999–2003 | 64 articles |
| 2004–2008 | 318 |
| 2008–2013 | 740 |

**Table 1.1.** *Increase in the number of articles dedicated to recommender systems in the library of the ACM (http://dl.acm.org/)*

The international conference on recommender systems (RecSys) was held in 2007 by the ACM and gathered many RecSys specialists. The 8th meeting of the conference will be held in Silicon Valley at the end of 2014[3].

The literature shows that the computer approach is focused on the performance of algorithms, their robustness, the design and comparison of systems based on semantic, social as well as hybrid data. The proposed evaluation is often centered around the interaction with the technical system, but does not take into account the more qualitative approaches centered around the user. The computer approach also takes into consideration questions related to the transparency, clarification, trust and measurement of recommendation diversity. The ongoing renewal mainly includes combinations with other technologies: notably the Web of data, Big Data and automated sentiment analysis.

E-commerce approaches are mostly focused on new techniques which can direct potential clients to targeted products and services. The combinations of different types of recommendation have been tested in

[3] http://recsys.acm.org/recsys14.

fields such as tourism and cultural industries (selling of books, music, on-demand video). Recommender systems are considered to be marketing tools and technologies specific to "business intelligence", a set of methods and technologies which transform data into useful information for decision-making in industry.

From the point of view of information science, identified works are more recent; they highlight the use of such systems for developing discovery functions in digital libraries and library catalogs [WAK 12]. Qualitative evaluations of recommendations, the perspective of users and psychological factors are all perspectives of analysis which are specific to recommender systems and which open up new areas of research in this field with the help of abundant literature on techniques and algorithms. Several conferences are focused, however, on the user experience with these recommender systems by assessing their acceptance or rejection placed in this context. It is notably the aspects of visualization, clarification, transparency, trust, and help in decision-making which are the objects of investigations by researchers from various subject areas[4].

## 1.3. The fundamentals of algorithms

Here, we introduce the foundations of recommendation systems, models and methods to provide a better context for the later chapters. This conceptual appropriation is intended to be neutral and factual; it will pave the way for the presentation of more involved points of view in the rest of this book.

### 1.3.1. *Collaborative filtering*

Historically, the first system proposed was based on collaborative filtering. This method assumes an authentication of users on the content management platform and, of course, personal input. Once a document has been proposed to the user by the system on the basis of criteria researched during the creation of the profile and/or the use of an additional internal search engine by the user, the latter will propose the possibility of attributing a rating to it. This rating can be an intrinsic assessment of the document, or

[4] See http://www.di.uniba.it/~swap/DM/programme.html, 1st Workshop on Decision Making and Recommendation Acceptance Issues in Recommender Systems (DEMRA 2011).

an assessment of the relevance to the context of the search and its main intentions.

This rating will be preserved within the system to be reused. According to the "memory-based" or heuristic collaborative filtering, ratings can help predict the assessment of a user $\alpha$ of an item based on that of another user $\beta$, having regularly rated in a similar way. In order to determine which user $\beta$ is most similar to user $\alpha$, the Pearson correlation is often used [RES 94]. This method is also referred to as "Word of Mouth" [SHA 95] or "People-to-People Correlation" [SCH 99].

Let $r$ be the Pearson correlation coefficient which in our case compares ratings, from 0 to 10, of 2 users for a collection of items. We note that this function is integrated into modern spreadsheets[5]. The correlation will be weak if the coefficient is less than 0.5 and strong if it tends toward 1.

Pearson correlation:

$$r = \left(\frac{\sum_{i=1}^{N}(\alpha_i-\bar{\alpha})\,.(\beta_i-\bar{\beta})}{\sqrt{\sum_{i=1}^{N}(\alpha_i-\bar{\alpha})^2}\,.\sqrt{\sum_{i=1}^{N}(\beta_i-\bar{\beta})^2}}\right) \qquad [1.1]$$

Example of the computation of the similarity between users having rated a set of items. Table 1.2 displays a collection of user assessments for certain items.

| Votes | User A | User B | User C | User D | User E |
|---|---|---|---|---|---|
| Item 1 | 9 | 10 | 7 | 10 | 9 |
| Item 2 | 7 | 6 | 2 | 1 | 1 |
| Item 3 | 5 | 1 | 5 | 5 | 4 |
| Item 4 | 3 | 5 | 3 | 2 | 2 |
| Item 5 | 1 | 3 | 5 | 7 | 6 |

**Table 1.2.** *Example of a sample of ratings*

Table 1.3 displays the correlation coefficients computed two by two for the collection. The values in bold show strongly correlated users.

[5] See: http://office.microsoft.com/en-us/excel-help/pearson-function-HP010342758.aspx for MS Excel And https://wiki.openoffice.org/wiki/Documentation/How_Tos/Calc:_ PEARSON_ function for Open Office Calc accessed online on 17 October 2014.

| Correlation | User A | User B | User C | User D | User E |
|---|---|---|---|---|---|
| User A | X | **0.699** | 0.243 | 0.215 | 0.246 |
| User B | **0.699** | X | 0.265 | 0.341 | 0.413 |
| User C | 0.243 | 0.265 | X | **0.977** | **0.669** |
| User D | 0.215 | 0.341 | **0.977** | X | **0.996** |
| User E | 0.246 | 0.413 | **0.669** | **0.996** | X |

**Table 1.3.** *Similarity of users based on their Pearson correlation*

In the example, for the values presented in Table 1.2, the results displayed in Table 1.3 show that each user can benefit from the assessments of at least one other user with a similar profile to theirs (correlation close to 1).

Once the number of user ratings has reached the maximum value, it can be used for offering a more precise prediction method referred to as "model-based" prediction which uses user profiles [BRE 98]. In this second method, the profile types are established by grouping those which have given similar ratings. These are the profile types or models which will be used to give out recommendations.

### 1.3.1.1. *Advantages and drawbacks of collaborative filtering*

The first advantage of recommendations based on collaborative filtering is that familiarity with the area of knowledge is not required for searching for information [BUR 02]. This system also facilitates the recommendation to be extended to genres which are correlated to the area of knowledge by using the other interests of similar profiles. This elicited serendipity is referred to by Burke as "cross-genre niches" [BUR 02]. According to Poirier *et al.*, because of its independence from the representation of data, this technique can be applied to contexts where analysis of the content is difficult to automate [POI 10]. We also add that for image, audio and video documents, metadata is rarely available. In this context, outside of collaborative filtering (or a preliminary significant descriptive crowdsourcing effort), there would not be an alternative recommendation method. The last positive aspect is that the quality of the recommendation proposed through collaborative filtering increases with the use of the system.

Claypool *et al.* have highlighted a certain number of problems in initial recommendation methods [CLA 99]. For example, in the initial state, a recommender system based on collaborative filtering is unusable due to a "coldstart". This coldstart problem manifests itself in the following way: without ratings no recommendation is possible. This difficulty is reproduced every time an item or user is added. With an overly low number of evaluations for a vast corpus, the data will be too sparse to establish enough correlations. This phenomenon is referred to as "sparsity" [CLA 99].

It is also shown that the principle of popularity will be favored by collaborative filtering. The more an item is favorably rated, the more it will be recommended and therefore rated again. This principle of self-generated notoriety therefore seems to be a result of age rather than the actual quality as perceived by users. This problem can be made up for, or on the contrary intensified by, a downfall of social recommendation systems, namely rating fraud through multiple identities. It can be tempting to modify recommendations from a marketing perspective by leaving ratings under multiple identities. This technique is referred to as "shilling" and is the object of many studies [LAM 04, BUR 06].

### 1.3.2. *Content filtering*

The other classic filtering method is based on the description and analysis of the content proposed by the system. This process is mainly based on text analysis techniques, but can be extended to various forms of content containing metadata. Digital text documents which are already well equipped with a wealth of metadata and linked to catalog records illustrate this point.

The content-based recommendation technique is based on the relationship between the user and metadata associated with the items stored in the knowledge base [BOU 04, LEE 06].

The user can voluntarily enter their preferences during their signup to the service: they are "provided". The other possibility is to compute preferences through the observation of their behavior [ADO 05]. In this case, they are "calculated" and put into vectors.

User preferences are represented in the form of a vector containing the most representative preferences of the user. These key terms can have a statistically determined value depending on their frequency in documents visited and/or rated by the user within the corpus [BAL 97]. For example, it is possible to use the *tf* algorithm to weight key terms from texts [SAL 88].

Frequency of a term in a document:

$$tf_{(m,d)} = \frac{n}{card(d)} \qquad [1.2]$$

EXAMPLE 1.1.– Let us consider a document *d* containing 100 words in which the term *m* appears *n* times with *n* = 3. The frequency of the term (*tf*) for *m* in document *d* is therefore the quotient between the number of occurrences *n* of the word *m* in document *d* and the total number of words in *d*. In this example, this gives 3/100.

The inverse of the frequency of documents [JON 72] is therefore computed with the logarithm of the quotient between the cardinal number of the whole of the corpus *C* and the cardinal number of the sub-corpus $C'$ of documents of *C* containing term *m*. The number 1 is added to the denominator in order to generalize the function in the case of the absence of terms in the corpus.

Inverse of the frequency of a word in the corpus:

$$idf = \log\left(\frac{card(C)}{1+ C'_{m,C}}\right) \qquad [1.3]$$

EXAMPLE 1.2.– Suppose that we have 10 million documents in the corpus *C* and that the term *m* appears in one thousand of these. If we apply this to our example, the *idf* is log (10 000 000/1 000), thus 4. The value of *tf.idf* in our example is therefore 0.03 × 4 = 0.12. Thus, the term *m* will statistically be weighted with a coefficient of 0.12 in document *d* of corpus *C*.

This basic algorithm is rarely used on its own, and has been replaced by more recent and sophisticated combinations, such as Terrier [OUN 05], notable with *okapi* BM25, but remains the basis for the weighting of the representative terms of documents in text corpuses.

Methods based on the vectorization of queries show promising results. Berry *et al.* have suggested the recovery of the query in matrix form through the popular latent semantic indexation (LSI) algorithm. The algorithm creates a vector space of reduced dimensions which offers a representation in *n* dimensions of a set of documents [DUM 88]. When a request is submitted, its numerical representation is compared with the cosine of other documents in the database, and the algorithm returns the documents with the smallest distance. This method can be adapted to recommending documents according to the needs of users.

#### 1.3.2.1. *Advantages and drawbacks of content filtering*

The advantages of content filtering are similar to those observed in collaborative filtering [BUR 00]. Thus, knowledge of the area is not required by the user, since recommendations are based on corpus data. The accuracy of the system recommendations will also evolve with the size of the corpus. However, a system based solely on corpus data will not be able to propose "serendipity" in the absence of user correlations. Furthermore, as pointed out by Poirier, each user is absolutely independent of others. Thus, a user who would have appropriately filled their profile with their interests will receive recommendations even if they are the only one to be registered [POI 10].

The main drawback of a content-based recommender system is first, as for collaborative types, the case posed by new users who do not have established profiles and therefore no "observed" reference data. Moreover, it is also very difficult to index non-text-based data. The users will be typecast into a particular search context, the one which has already been set as their area of interest. This problem is referred to as "overspecialization", which eliminates any possibility of serendipity through the proposal of related subjects.

### **1.3.3.** ***Hybrid methods***

Trivially, the hybridization of recommender systems is the result of the combination of collaborative filtering and content-based methods. This vision for hybridization was refined by Burke and then by Adomavicius and Tuzhilin [BUR 02, ADO 05].

Burke made a list of the following seven hybridization techniques [BUR 02]:

– weighted: the recommendation value of an item is based on the sum of available methods. For example, P-Tango [CLA 99] gives an equal value to both collaborative filtering and content-based filtering. This value is then weighted by a confirmation of the users;

– switching: the system chooses to apply either a data-based method or social filtering depending on the search context of the user;

– mixed: this technology facilitates the proposal of recommendations from traditional methods with the aim of limiting the drawbacks of each classic method;

– features combination: this method offers the possibility of enriching data which has been integrated *a priori* into the system with the ratings of users, which enriches the database *a posteriori*. The computation of the recommendation is carried out over all of the data;

– cascade: this process consists of a double analysis of user profiles. The first is used to highlight potential candidates, the second to refine the selection of users;

– features augmentation: this is a technique which is similar to the previous one for the first pass-through. If the number of candidates is too high on the first pass-through, then a second will carry out a secondary discrimination by integrating the data of recommended items;

– meta level: as for the first two methods, it involves filtering users twice in order to determine similarities. The difference is that the first pass-through makes possible the generation of a model or profile type of the user.

Adomavicius and Tuzhilin have proposed a classification of hybrid recommendation methods based on three points of focus [ADO 05]:

– combining separate recommenders: the collaborative method and the content-based method are applied separately, then their predictions are combined;

– adding content-based characteristics to collaborative models: this system uses the classic collaborative “People-to-People Correlation” approach, to which it adds recommendations based on the classification of the content and the interests indicated by users;

– adding collaborative characteristics to content-based models: the principle of this model is not to reverse the previous one, but to incorporate characteristics of the "model-based" group profile collaborative method into the content-based approach;

– single unifying recommendation model: construction of a general model which incorporates the characteristics of two models within a same algorithm.

### 1.3.4. *Conclusion on historical recommendation models*

The timelines of the first two types of recommendation model overlapped in the 1990s.

Collaborative filtering recommender systems are based on the statistical processing of opinions expressed by users. It was found that data-based methods are adapted to automatic language processing rules, namely automatic indexing and the weighting of representative terms. In order to mitigate the weaknesses inherent to these initial models, hybrid methods have emerged since the end of the 1990s. We will examine the ways in which these different algorithms have been implemented in online applications.

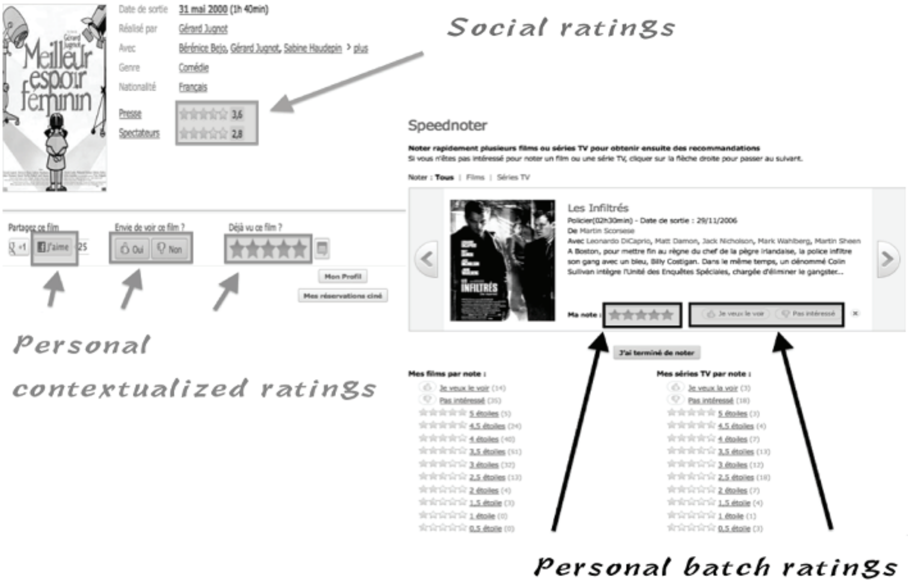
## 1.4. Content offers and recommender systems

### 1.4.1. *Culture and recommender systems*

#### 1.4.1.1. *Recommendation and cinema*

Historically, researchers (GroupLens) have mostly been interested in the application of recommender systems to the cultural domain with cinema and film ratings [ALS 97]. Film database interfaces are available to users in return for a rating. This method, used in MovieLens, is exactly that presented in section 1.3.1 [SCH 07]. Based on the ratings of each user, it is possible to provide recommendations.

The French cinema listings website Allociné contextually proposes an offer with similar ratings for each presented film. The improvement of this recommender system is based on the introduction of stars to the Internet user, which represent an evaluation, as well as the popular Facebook "Like" mechanism or even "Would you like to watch this film yes/no" (see Figure 1.1, top left). This website also offers the possibility of rating films in batches, on a scale of 1 to 10 if one has seen the film, or indicating whether the user is interested or not (see Figure 1.1, bottom right). The principle is to consecutively assess a large number of cinematographic works and therefore facilitate the system to create the most accurate profile of our preferences in this department. Additional propositions will be more accurate as the number of rated films increases.



Date de sortie 31 mai 2000 (1h 40min)
Réalisé par Gérard Jugnot
Avec Bérénice Bejo, Gérard Jugnot, Sabine Haudepin > plus
Genre Comédie
Nationalité Français
Presse 3,6
Spectateurs 2,8

**Noter rapidement plusieurs films ou séries TV pour obtenir ensuite des recommandations**
Si vous n'êtes pas intéressé pour noter un film ou une série TV, cliquer sur la flèche droite pour passer au suivant.

Les Infiltrés
Policier(02h30min) - Date de sortie : 29/11/2006
De Martin Scorsese
Avec Leonardo DiCaprio, Matt Damon, Jack Nicholson, Mark Wahlberg, Martin Sheen
A Boston, pour mettre fin au règne du chef de la pègre irlandaise, la police infiltre son gang avec un bleu, Billy Costigan. Dans le même temps, un dénommé Colin Sullivan intègre l'Unité des Enquêtes Spéciales, chargée d'éliminer le gangster...

**Mes films par note :**
Je veux le voir (14)
Pas intéressé (35)
5 étoiles (5)
4,5 étoiles (24)
4 étoiles (40)
3,5 étoiles (51)
3 étoiles (32)
2,5 étoiles (13)
2 étoiles (4)
1,5 étoile (3)
1 étoile (0)
0,5 étoile (0)

**Mes séries TV par note :**
Je veux la voir (3)
Pas intéressé (18)
5 étoiles (3)
4,5 étoiles (4)
4 étoiles (7)
3,5 étoiles (13)
3 étoiles (12)
2,5 étoiles (18)
2 étoiles (7)
1,5 étoile (4)
1 étoile (1)
0,5 étoile (3)

**Figure 1.1.** *Allociné's rating context*

1.4.1.2. *Recommendation and literature*

For the recommendation of literary works, we mention the social network for readers Goodreads and the French network Babelio[6]. Goodreads initially modeled its recommendation system on metadata sourced from Amazon. Filtering was therefore based on this data. The partnership which linked the social network with the online selling giant then ended, with Goodreads employing Discovereads, a social algorithm developed at Stanford which

[6] Also see Chapters 9 and 10 that are dedicated, respectively, to the offer of French literary suggestions in the first case, and more specifically Babelio in the second case.

only uses data from users on a corpus of metadata pertaining to the contents of books[7].

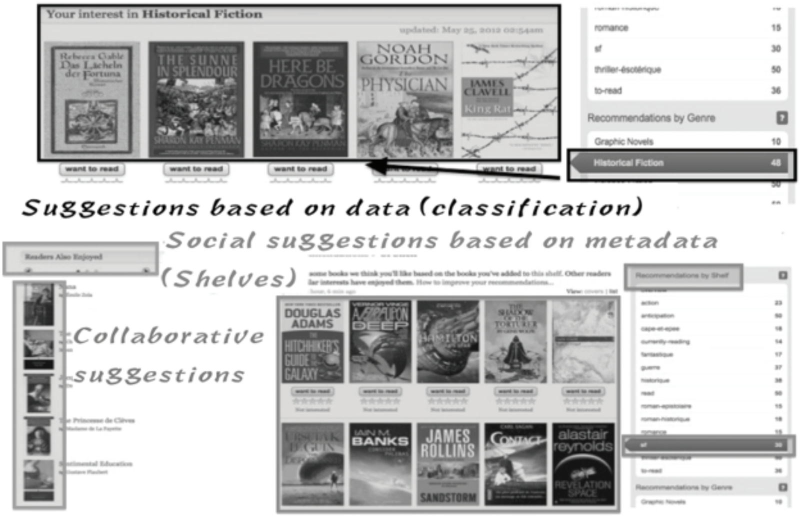
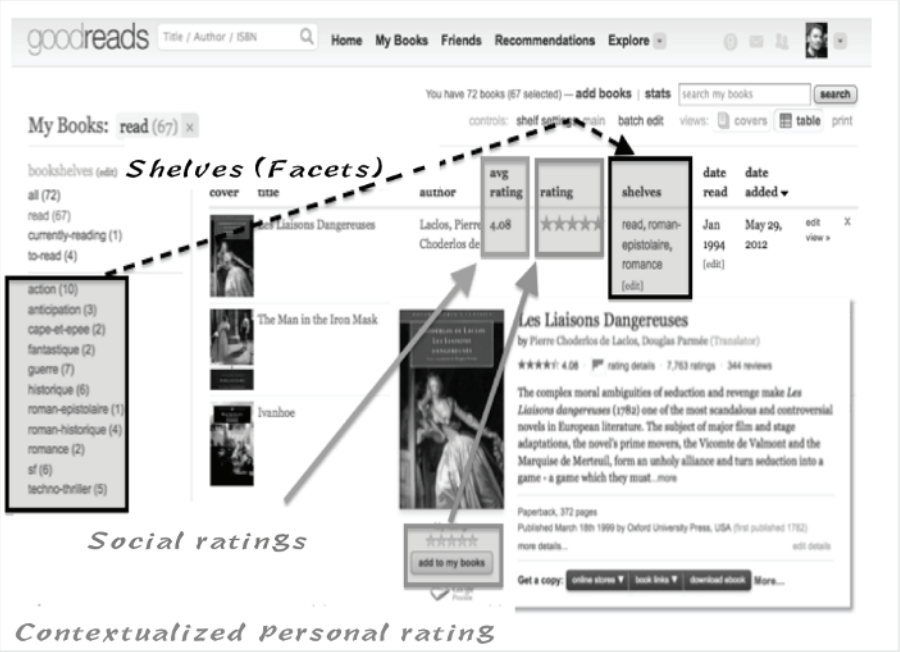


**Figure 1.2.** *Goodreads suggestions*

Thus, armed with its 8 million users, its database of 300 million rated books and a correlation algorithm, Goodreads can individually offer reliable recommendations based on 20 rated books. A layer of hybridization of data recommendation intervenes due to a typology of the books. In fact, the books are classified according to a taxonomy of literary genres (Graphic Novels, Historical Fiction, Science Fiction, Thriller, etc.) in which our preferred genres will be defined. The upper section of the illustration shown in Figure 1.2 proposes to show recommendations based on the contents of bibliographic records, which are sourced from summaries or metadata. The left side of Figure 1.2 illustrates the offer of a social suggestion based uniquely on user ratings. Furthermore, as illustrated in the bottom right section of Figures 1.2 and 1.3, it is possible to use book "covers" in order to organize our books into "virtual bookshelves" [MAN 99, HUD 11, DES 12]. These shelves will be reused by the system in order to propose relevant reads to other users.

[7] See http://www.libraryjournal.com/lj/newsletters/newsletterbucketljxpress/892038-441/goodreads_launches_book_recommendation_feature.html.csp accessed online on 29 August 2014.



| cover | title | author | avg rating | rating | shelves | date read | date added ▾ | |
|---|---|---|---|---|---|---|---|---|
| | Les Liaisons Dangereuses | Laclos, Pierre Choderlos de | 4.08 | ★★★★★ | read, roman-epistolaire, romance<br>[edit] | Jan 1994<br>[edit] | May 29, 2012 | edit<br>view » |
| | The Man in the Iron Mask | | | | | | | |
| | Ivanhoe | | | | | | | |

Les Liaisons Dangereuses

by Pierre Choderlos de Laclos, Douglas Parmée (Translator)

★★★★½ 4.08 · rating details · 7,763 ratings · 344 reviews

The complex moral ambiguities of seduction and revenge make *Les Liaisons dangereuses* (1782) one of the most scandalous and controversial novels in European literature. The subject of major film and stage adaptations, the novel's prime movers, the Vicomte de Valmont and the Marquise de Merteuil, form an unholy alliance and turn seduction into a game - a game which they must...more

Paperback, 372 pages

Published March 18th 1999 by Oxford University Press, USA (first published 1782)

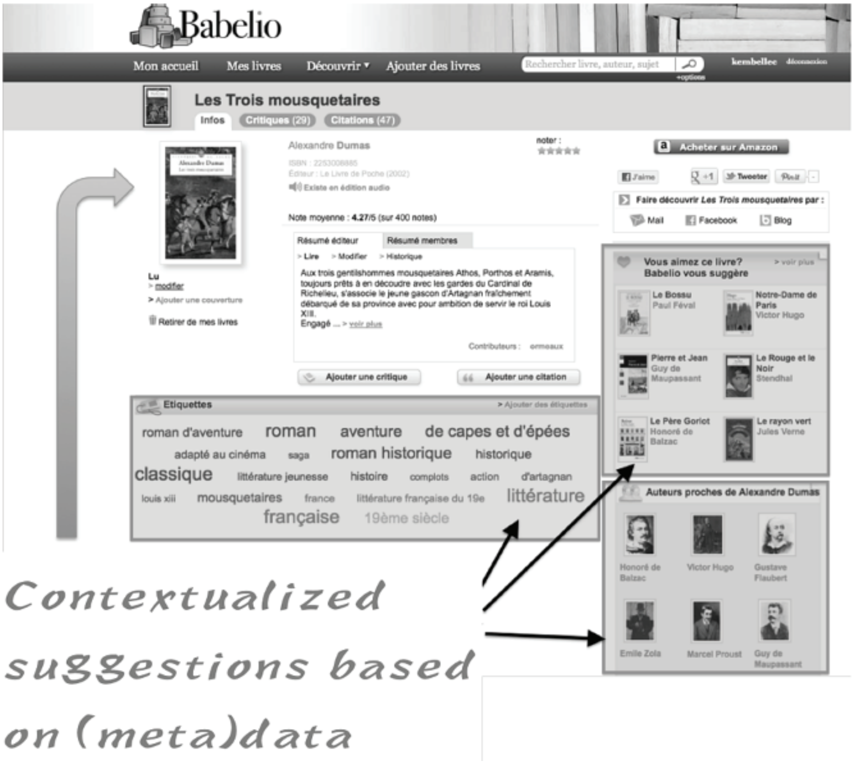
more details... edit details

Get a copy: online stores ▼ book links ▼ download ebook More...

**Figure 1.3.** *Goodreads virtual bookshelves*

The Babelio system is quite similar to Goodreads with a webcam scanner system for integrating books using barcodes or ISBN codes. Babelio offers the rating and annotation of books with classifications as well as book "labels". The recommender system of this system is based more on data than on collaborative filtering. Indeed, Babelio offers contextualized recommendations on the same page as the listing of a book (see in Figure 1.4 the "word cloud" of the book). The suggestion is in the form of "Do you like this book? Babelio suggests (... similar books)" (see Figure 1.4, top right). The system also offers to display other books by the same writer or those authors considered to be "similar", without specifying in what way (see Figure 1.4, bottom right).

Social recommendation is also present, to a lesser extent, with the possibility of accessing the library of users who have liked a particular book. This social offering has the initial assumption that the books of "friends" are also those that will interest me. This concept is referred to as "homophilia" or proximity of readers' social networks [GUI 11, AIE 10]. Once a user has been identified as "similar" to our profile and has been accepted as such, we can rely on their ratings and preferences. It seems as if Babelio prioritizes this system, with a focus on user comments and summaries of books rather than the usual correlation algorithm.



Aux trois gentilshommes mousquetaires Athos, Porthos et Aramis, toujours prêts à en découdre avec les gardes du Cardinal de Richelieu, s'associe le jeune gascon d'Artagnan fraîchement débarqué de sa province avec pour ambition de servir le roi Louis XIII.
Engagé ... > voir plus

Contextualized suggestions based on (meta)data

**Figure 1.4.** *Recommendations and data with Babelio*

A port of Babelio has been made for use within a traditional OPAC system[8]. The municipal library of Toulouse is equipped with this platform. The social suggestions network is described as "very rich and very active" [KRA 11]. However, this institution does not integrate, in terms of users, the

[8] See http://www.babeltheque.com, accessed online on 29 August 2014.

sufficient "critical mass" of data for the autonomous use of a social recommendation network [WAK 12]. The condition for offering a useable service from the start is therefore to rely on platforms already containing content (in order to avoid a coldstart). The verdict made from this offering is positive with "coherent" suggestions estimated at 80% and an improvement of Babelio's user base.

#### 1.4.1.3. *Recommendation and general culture*

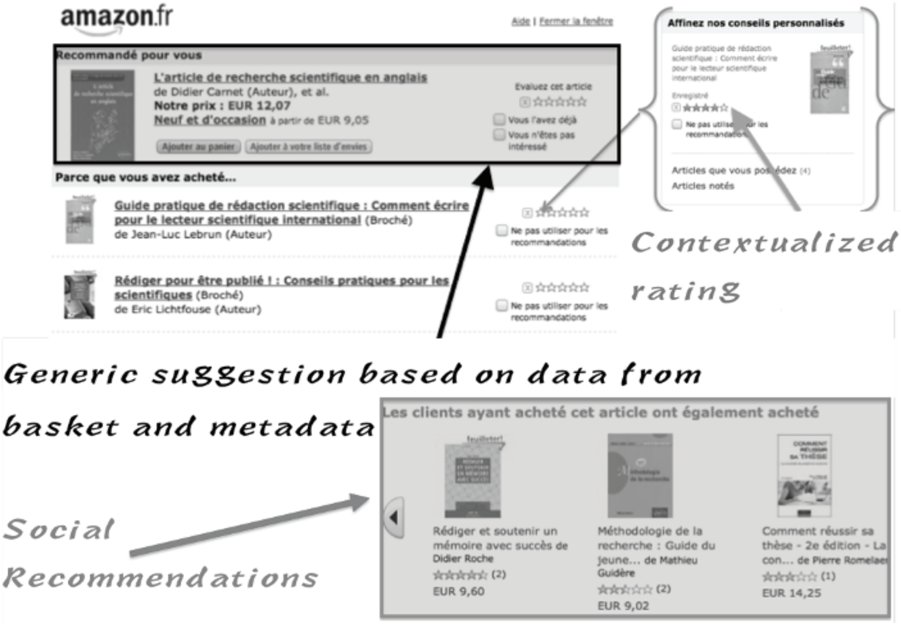
Hunch's personalized recommender system proposes general cultural recommendations. From the user's point of view, initialization is very controlled since one must first answer closed-ended or semi-open questions on preferences and interests. The user must then rate an initial selection of cultural items as a sequence of videos, books, images or even info graphics related to the selected topics. This data-based step helps in the creation of a profile and the ability to class the Internet user into a user group in order to propose more tailored selections of cultural items. Each element is classed into a thematic with a compatibility percentage associated with the user profile. After a few dozen validations – or invalidations – of propositions, correlations with users with similar interests are offered. It is important to note that the collection of topics is not necessarily identical, but the evaluations within these common topics will correspond quite accurately. The aim of these correlations is to establish experimental relations between users in order to begin proposing hyperlinks to topics that would be considered of interest, the content of which is properly indexed. It would then be possible to have a new content because of others sharing the same preferences on common topics of interest. This system is very effective for discovering new content adapted to individual preferences on targeted themes. This efficiency has not escaped the notice of the commercial players since the commercial giant eBay has bought back Hunch in order to adapt it to its online sales platform. The use of recommender systems for commercial means is not a unique case; many major online commercial players turn recommendation into very effective personalized sales systems[9].

### **1.4.2.** ***Recommender systems and the e-commerce of content***

More broadly, the use of recommender systems is growing rapidly within the framework of e-commerce. It has become rare to find an e-commerce

[9] For more detail on this, see Chapter 3.

service which does not provide purchasing recommendations. The aim of these systems is to contextually provide, with varying levels of success, buying advice on “interesting” products in relation to the “needs” of the client.



amazon.fr

Aide | Fermer la fenêtre

Recommandé pour vous

L'article de recherche scientifique en anglais
de Didier Carnet (Auteur), et al.
**Notre prix : EUR 12,07**
Neuf et d'occasion à partir de EUR 9,05

Evaluez cet article
☒ ☆☆☆☆☆
□ Vous l'avez déjà
□ Vous n'êtes pas intéressé

Parce que vous avez acheté...

Guide pratique de rédaction scientifique : Comment écrire pour le lecteur scientifique international (Broché)
de Jean-Luc Lebrun (Auteur)
☒ ☆☆☆☆☆
□ Ne pas utiliser pour les recommandations

Rédiger pour être publié ! : Conseils pratiques pour les scientifiques (Broché)
de Eric Lichtfouse (Auteur)
☒ ☆☆☆☆☆
□ Ne pas utiliser pour les recommandations

Affinez nos conseils personnalisés

Guide pratique de rédaction scientifique : Comment écrire pour le lecteur scientifique international
Enregistré
☒ ★★★★☆
□ Ne pas utilise… pour les recommandation…

Articles que vous pos…édez (4)
Articles notés

Les clients ayant acheté cet article ont également acheté

Rédiger et soutenir un mémoire avec succès de Didier Roche
☆☆☆☆☆ (2)
EUR 9,60

Méthodologie de la recherche : Guide du jeune... de Mathieu Guidère
☆☆☆☆☆ (2)
EUR 9,02

Comment réussir sa thèse - 2e édition - La con... de Pierre Romelae…
☆☆☆☆☆ (1)
EUR 14,25

Ajouter au panier
Ajouter à votre liste d'envies
feuilleter!
Contextualized rating
Generic suggestion based on data from basket and metadata
Social Recommendations

**Figure 1.5.** *Recommendations on Amazon*

The most well-known example is the recommender system used by Amazon’s online library. As illustrated at the bottom of Figure 1.5, a link is proposed next to every title: “Clients who purchased this book also purchased ...”. These propositions come from the analysis and evaluation of the buying habits of Amazon clients. The upper left section of the same illustration proposes recommendations based on content, in other words the metadata from the listings of products. On the right side, a hybrid offer helps create a personalization of a recommendation on the best sales in relation to previous purchases and the currently visited listing.

### 1.4.3. *The behavior of users*

Online personalization proposes recommendations for products and services based on the online purchases of clients or their browsing habits. Personalization applications reduce information overload and provide services of added value.

However, their adoption could be held back by concerns from clients regarding the confidentiality of information. A study has thus been dedicated to determining whether confidentiality *versus* the quality of a recommender service would have a significant impact on its adoption by customers [LI 12]. Slightly unexpectedly, investigations have shown that users are ready to divulge personal information when using services which they deem to be of high quality. The results even go so far as to show that clients who are susceptible to using online personalization are also susceptible to paying for such a service [LI 12].

Without elaborating on the monetization of the system, an analysis of forums related to French recommendation services indicates an immediate subscription and a high involvement from users. This acceptance can even lead to a feeling of frustration in the case of an interruption of the service[10].

On the contrary, it is possible to attribute the high subscription rate of the online bibliographical management service Mendeley in French-speaking scientific environments, in part, to its recommendation service [KEM 12, 13].

These partial results would therefore suggest that the perceived quality of personalization could be decisive regarding concerns over confidentiality. Providers of these services could therefore pursue the objective of improving the quality of their offered personalization services in order to compensate for preoccupations over privacy. Despite worries over privacy, there would however be an opportunity for companies to monetize recommendation. However, one could wager that this return on investment would be accompanied by increasingly strong constraints on the ethically sensible use of personal data and its safety.

[10] See reactions to service interruptions of recommender systems of both Allociné and Goodreads: http://allocine.uservoice.com/forums/25482-allociné-v6-evolutions-et-idé/suggestions/347269-recommandations; http://www.sobookonline.fr/non-classe/goodreads-victime-de-la-licence-damazon/.

## 1.5. Current issues

Although specialists in algorithms and e-commerce have up until now produced the majority of works on recommender systems since the start of the 2000s, other questions arise regarding the perception of users: What prevents users from adopting this type of service? How do users rate recommendations which are proposed to them? Do these lead to new practices, new purchases? What are the possible determining factors regarding the trustworthiness associated with these tools?

As the issue of personal data protection becomes a growing priority, accompanying a current reform of the European directive on this matter, how can these types of tools create a context of trust while preserving the efficiency of its algorithms which rely on the tracking of users? Will we see the emergence of actors taking on the role of building trust? A strong need for legal and normative action dealing with recommender systems is felt in France[11].

Another worrying issue regarding specialists in information searching is as follows: Will algorithmic recommendation open up or funnel the practices of users? Will we assist in the reproduction of the "blockbuster" phenomenon to the detriment of a proper sustainability of diversity? In the cultural domain, what will be the impact on the evolution of reading, listening and watching practices of cultural content? These are questions which solicit many investigations.

## 1.6. Bibliography

[ADO 05] ADOMAVICIUS G., TUZHILIN A., "Towards 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. 734–749, 2005.

[11] A revision of the rules on data protection so that citizens can control their personal data and so that companies can move more easily within the EU was adopted at the European Parliament's committee on Civil Liberties on 22 October 2013. As a response to situations of mass surveillance, deputies have introduced backups for transfers of data to third countries, the requirement of having explicit consent, the right to be forgotten and higher penalties for companies breaching these rules.

[AFN 13] AFNOR, NF Z74-501, Avis en ligne de consommateurs – Principes et exigences portant sur les processus de collecte, modération et restitution des avis en ligne de consommateurs, 2013, available online: http://www.boutique.afnor.org/norme/ nf-z74-501/avis-en-ligne-de-consommateurs-principes-et-exigences-portant-sur-les-processus-de-collecte-moderation-et-restitution-des-avis/article/808897/fa178349#info.

[AIE 10] AIELLO, L.M., BARRAT A., CATTUTO C., *et al.*, "Link creation and profile alignment in the aNobii social network", *IEEE 2nd International Conference on Social Computing*, Minneapolis, MN, pp. 249–256, 2010.

[ALS 97] ALSPECTOR J., KOICZ A., KARUNANITHI N., "Feature-based and clique-based user models for movie selection: a comparative study", *User Modeling and User-Adapted Interaction*, vol. 7, no. 4, pp. 279–304, 1997.

[BAL 97] BALABANOVIĆ M., SHOHAM Y., "Fab: content-based, collaborative recommendation", *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, 1 March 1997.

[BER 99] BERRY M.W., DRMAC Z., JESSUP, E.R., "Matrices, vector spaces, and information retrieval", *SIAM Review*, vol. 41, no. 2, pp. 335–362, 1999.

[BOU 04] BOUTELL M.R., LUO J., "Bayesian fusion of camera metadata cues in semantic scene classification", *Computer Vision and Pattern Recognition, Proceedings of the 2004 IEEE Computer Society Conference on IEEE,* Washingtion DC, USA, vol. 2, pp. 623–630, 2004.

[BRE 98] BREESE J.S., HECKERMAN D., KADIE C., "Empirical analysis of predictive algorithms for collaborative filtering", *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence,* Morgan Kaufmann, San Francisco, CA, pp. 43–52, 1998.

[BUR 00] BURKE R., "Knowledge-based recommender systems", *Encyclopedia of Library and Information Science*, vol. 69, no. 32, pp. 175–186, 2000.

[BUR 02] BURKE R., "Hybrid recommender systems: survey and experiments", *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, 2002.

[BUR 06] BURKE R., MOBASHER B., WILLIAMS C., *et al.*, "Classification features for attack detection in collaborative recommender systems", *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, *ACM Press*, New York, p. 542, 2006.

[CHE 09] CHEN Y.-C., SHANG R.-A., KAO C.-Y., "The effects of information overload on consumers subjective state towards buying decision in the internet shopping environment", *Electron. Commer. Rec. Appl.*, vol. 8, no. 1, pp. 48–58, 2009.

[CLA 99] CLAYPOOL M., MIRANDA T., GOKHALE A., *et al.*, "Combining content-based and collaborative filters in an online newspaper", *Proceedings of Recommender Systems Workshop ACM SIGIR,* ACM Press, New York, pp. 40–48, 1999.

[DES 12] DESFRICHES-DORIA O., "La classification à facettes pour la gestion des connaissances dans les organisations", *Etudes de communication,* vol. 39, pp. 173–198, 2012.

[DUM 88] DUMAIS S.T., FURNAS G.W., LANDAUER T.K., *et al.*, "Using latent semantic analysis to improve access to textual information", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems – CHI'88*, ACM Press, New York, pp. 281–285, 1 May 1988.

[GUI 11] GUILLAUD H., "Les livres de mes amis sont miens: homophilie et proximité dans les réseaux sociaux de lecteurs", *La Feuille l'édition à l'heure de l'innovation*, 2011, available online: http://lafeuille.blog.lemonde.fr/2011/12/29/homophilie-et-proximite-dans-les-reseaux-sociaux-de-lecteurs/.

[HUD 11] HUDON M., MUSTAFA EL HADI W., "Organisation des connaissances et des ressources documentaires", *Les Cahiers du numérique*, vol. 6, no. 3, pp. 9–38, 28 February 2011.

[JON 72] JONES K.S., "A statistical interpretation of term specificity and its application in retrieval", *Journal of Documentation*, vol. 28, no.1, pp. 11–21, 1972.

[KEM 12] KEMBELLEC G., Technologie et pratiques bibliographiques associées à l'écriture scientifique en milieu universitaire, Paris, 2012.

[KEM 13] KEMBELLEC G., "La médiation technologique autour des pratiques rédactionnelles et bibliographiques en milieu universitaire français", *Documentaliste-Sciences de l'Information,* vol. 50, no. 1, pp. 62–69, 2013.

[KRA 11] KRAJEWSKI P., "Quelques retours sur Babelthèque", *Des Bibliothèques 2.0*, 27 October 2011. Available at http://bibliotheque20.wordpress.com/2011/10/27/quelques-retours-sur-babeltheque/.

[LAM 04] LAM S.K., RIELD J., "Shilling recommender systems for fun and profit", *Proceedings of the 13th Conference on World Wide Web,* pp. 393–402, April 2004.

[LEE 06] LEE B., CHEN W.-Y., CHANG E.Y., "A scalable service for photo annotation, sharing, and search", *Proceedings of the 14th Annual ACM International Conference on Multimedian,* ACM Press, New York, p. 699, 2006.

[LEV 98] LÉVY P., *Qu'est-ce que le virtuel ?* La Découverte, Paris, 1998.

[LI 12] LI T., UNGER T., "Willing to pay for quality personalization? Trade-off between quality and privacy", *European Journal of Information Systems*, vol. 21, no. 6, pp. 621–642, March 2012.

[MAN 99] MANIEZ J., " Du bon usage des facettes", *Documentaliste-Sciences de l'Information*, vol. 36, no. 4–5, pp. 249–260, 1999.

[OBR 77] O'BRIEN K., TERRENCE V., "Information handling in consumer decisions", *Journal of the Academy of Marketing Science*, vol. 5, no. 3, pp. 229–232, 1977.

[OUN 05] OUNIS I., AMATI G, PLACHOURAS V., *et al.*, "Terrier information retrieval platform", in LOSADA D.E., FERNÁNDEZ-LUNA J.M., (eds.), *Information Retrieval,* Springer, Heidelberg, Germany, vol. 3408, pp. 517–519, 21 March 2005.

[POI 10] POIRIER D., FESSANT F., TELLIER I., "De la Classification d'Opinion à la Recommandation: l'Apport des Textes Communautaires", *TAL : traitement automatique des langues*, revue semestrielle de l'ATALA, vol. 51, no. 3, pp. 19–46, 2010.

[RES 94] RESNICK P., IACOVOU N., SUCHAK M., *et al.*, "GroupLens", *Proceedings of the ACM Conference on Computer-Supported Cooperative Work,* ACM, New York, pp. 175–186. 1994.

[RES 97] RESNICK P., VARIAN HAL R., "Recommender Systems mmende tems", *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, March 1997.

[SAL 88] SALTON G., BUCKLEY C., "Term-weighting approaches in automatic text retrieval", *Information Processing & Management,* vol. 24, no. 5, pp. 513–523, 1988.

[SCH 07] SCHAFER J.B., FRANKOWSKI D., HERLOCKER J., *et al.*, "Collaborative filtering recommender systems", in BRUSILOVSKY P., KOBSA A., NEJDL W., (eds.), *International Journal of Electronic Business*, vol. 2, no. 1, p. 77, 2007.

[SCH 99] SCHAFER J.B., KONSTAN J., RIEDI J., "Recommender systems in e-commerce", *Proceedings of the 1st ACM Conference on Electronic Commerce,* ACM, New York, pp. 158–166, 1 November 1999.

[SHA 95] SHARDANAND U., MAES P., “Social information filtering: algorithms for automating “Word of Mouth””, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, New York, pp. 210–217, 1 May 1995.

[WAK 12] WAKELING S., CLOUGH P., SEN B., *et al.*, “Readers who borrowed this also borrowed…, recommender systems in UK libraries”, *Library Hi Tech.*, vol. 30, no. 1, pp. 134–150, 2012.