An Adapted Approach for User Profiling in a Recommendation System: Application to Industrial Diagnosis

Fatima Zohra Benkaddour [1]*, Noria Taghezout [2], Fatima Zahra Kaddour-Ahmed [3], Ilyes-Ahmed Hammadi [3]

[1] Laboratoire d'Informatique Oran (LIO), École Normale Supérieure d'Oran (ENSO), Département des Sciences Exactes, filière Informatique, Oran (Algeria)
[2] Laboratoire d'Informatique Oran (LIO), Département d'Informatique, Université Oran1, Oran (Algeria)
[3] Département d'Informatique, Université Oran1, Oran (Algeria)

Received 15 January 2018 | Accepted 31 May 2018 | Published 12 June 2018

Abstract

In this paper, we propose a global architecture of a recommender tool, which represents a part of an existing collaborative platform. This tool provides diagnostic documents for industrial operators. The recommendation process considered here is composed of three steps: Collecting and filtering information; Prediction or recommendation step; evaluating and improvement. In this work, we focus on collecting and filtering step. We mainly use information result from collaborative sessions and documents describing solutions that are attributed to the complex diagnostic problems. The developed tool is based on collaborative filtering that operates on users' preferences and similar responses.

Keywords

Recommender Systems, Decision Support System, Industrial Diagnosis, Collaborative Filtering, Twitter.

DOI: 10.9781/ijimai.2018.06.003

I. Introduction

With the increasing volume of data that exist in information systems, the user can quickly be submerged by an informative mass. Moreover, these data are highly heterogeneous. As a result, targeting relevant information to the user becomes a major concern of a large number of research projects. Therefore, personalization is a suitable solution to this problem. Among the personalization tools, we find the recommendation systems.

Recommendations systems are defined by [1] as systems that can provide customized recommendations to guide the user to interesting and useful resources within an important data space. They play a major role in the information filtering systems dealing with how best it can recommend items or information which is relevant to the user. Recommendation systems can be applied to a variety of applications such as E-Commerce site.

The world is developing daily especially in the industrial diagnostics field. It becomes necessary to support the diverse diagnoses of the industry in order to achieve better performance and a sustainable and profitable industrial efficiency. In any industrial process, it is essential to establish the relative performance with the improvement and development of new technologies.

Generally, the industrial diagnosis helps to promote and develop expertise for a better result and an achievement of the objectives set by the company. Furthermore, industrial documents help to promote and develop diversity and knowledge.

The industrial domain is a field with different orientations and different sectors. It is preferable to move towards a recommendation system that gives to the user the way to acquire any documentation relating to his domain in a reduced time and without difficulties.

A. Problem Statement

The search for solutions to the breakdowns in the industrial environment is a difficult task that requires considerable research time. Finding these solutions in a short time will improve the company productivity. The work presented here is a part of collaborative decision support system [2]. This system is the first system that provides answers to non-woven operators for their diagnosis problems by using a domain ontology, which represents the knowledge source and case-based reasoning.

Due to the availability of industrial documents that describe most solutions to the complex problems of industrial operators arising from collaborative work; a real need for an information filtering tool was felt. The information gathered from collaborative sessions and Web 2.0 tools facilitate the development project of a recommendation system. In this context, the main objective of this work is to improve the search for diagnostic documents for an industrial operator by taking his preferences into account in order to provide better quality in the recommendations.

B. Contribution

The recommendation system in the industrial field has become the main focus for the development and efficiency of the company.

* Corresponding author.
E-mail address: benkaddourfatima@gmail.com

Our study consists of incorporating a collaborative support system with a documents recommender tool. The main goal of this tool is to provide relevant documents to industrial operators in a fast and simple way. For that, we consider a recommendation process of three steps which are: collecting and filtering information, prediction, and evaluating and improvement. In this work, we believe that identifying preferences of users allows more precision to the recommender tool. In this study, we give more importance to the first step.

The recommendation system that will plugin will have access to all the database including users, documents, and users' history data. The recommendation system will take this data as input and predict for example a new time-line for the logged in user.

We summarize our main contribution in the following:

- Implementing an extended architecture of collaborative system detailing the recommended component;
- Recognizing and collecting users 'preferences in two ways: explicitly and implicitly;
- Applying KNN (K Nearest Neighbors) method based on Cosine similarity to recommend documents after preparing and filtering users and documents data.

This paper is organized as follows: In section two, we give some works relative to recommendation systems. Besides, we also present our analysis. In section three, we detail our proposed approach by giving the global architecture of our recommendation tool and some details regarding the process of collecting and filtering information. In section four, we explain the experimental protocol. In section five, we discuss some obtained results. In section six, we present two scenarios of execution to describe in detailed the proposed approach. Finally, in section seven, we conclude with some future works.

II. Related Work

Today, as the Web is rapidly growing at a faster rate, finding relevant information has become extremely difficult. Information or content can be in any form such as music, video, images or text, which are the interest to the users. Therefore Recommendation systems come into picture. Recommendation systems are a sub-category of information filtering systems that help people to find products, correct information, and even other people as well.

There are a lot of works about filtering information and recommendation systems. We present some of them below.

In [1], authors proposed a social user profiling method to recommend online sites. These recommendations relied on some similar interest between users and their followers. The suggested approach was based on an extended matrix factorization model by incorporating both individual and shared users' interests, and multifaceted unsupervised similarities. Some experiences were conducted to show performance of this approach.

Authors presented in [3] a survey of collaborative filtering techniques. They sorted collaborative filtering into three categories: memory based, model-based, and hybrid Collaborative Filtering (CF) algorithms. These later were advanced with their advantages, limits, and solutions. Different metrics of evaluating recommender systems were given in this paper.

In [4], authors resolved the social recommendation (SR) problem by utilizing microblogging data via multi-view user preference learning. User preferences are presented as rating information, social relations, items information, and tagging information, which are the common representation of multi-view information. The items are recommended based on the learnt user preference.

Authors developed an alternating direction method of multipliers (ADMM) scheme to solve the proposed model. They evaluated their approach, by using two real world datasets.

In addition to the works cited above, authors proposed in [5] two recommendation models to solve the complete cold start (CCS) and incomplete cold start (ICS) problems for new items. These models are based on a framework of tightly coupled CF approach and deep learning neural network. They used a deep neural network SADE to extract the content features of the items, also a solution for cold start items (CSI) is provided. The problem of CSI exiting in CF model is solved by taking account of the content features and ratings into prediction.

The two proposed recommendation models are also evaluated and compared with ICS items, and a flexible scheme of model retraining and switching is proposed to deal with the transition of items from cold start to non-cold start status. Some experiments were conducted to support the proposed approach.

In [6], authors developed a multi-criteria approach for a recommender system. This later is developed in order to support decision makers in their activities by managing users 'profiles. An automated technique is used to ensure the evolution of the recommender system.

The paper cited in [7] presents a comparative study of different recommendation techniques. A Content-Based Recommendation is highlighted. This kind of recommendation makes easier to find relevant information to the user based on previous ratings and predictions. Authors gave a comparative study of different techniques; they concluded that hybrid approach will give better results.

Authors presented their work [8] as a comparison between explicit ratings methods. These methods represent users 'preferences. The most adequate method is used to rate web content, and will be utilized by any web recommendation systems.

As shown in [9], authors explored the effect of combining the implicit relationships of the items and user-item matrix on the accuracy of recommendations. They introduced Item Asymmetric Correlation (IAC), as a new method that generates the implicit item relationship based on the user-item matrix. In their work, they used relations as an additional dataset for the Matrix Factorization (MF) technique. This research considered the implicit relationship between items, the correlated items are extracted, and the new dataset is used in MF model as a regularization term.

After our analysis, we can say that a recommender system will be accurate if it takes in consideration users 'preferences. However, users 'preferences represent a significant amount of information. Using relevant information would considerably reduce research time, which would allow a good evolution of the recommender system.

We give a comparison between recommendation approaches in Table I.

Among the existing recommendation approaches, the hybrid approach remains the best.

A. Problems of Recommendation Systems

Recommendations Systems (RSs) have emerged with the evolution of information available on the Web. RSs come to solve the problem of information overload as well as its research. However, these systems encounter performance problems. We classify these problems into three categories: in the first category, we find the lack of information on users / items which is called the cold start problem (found in collaborative filtering). This problem is defined by [33] as the inability of the system to deal with new users or new items due to the lack of prior knowledge. The second category recurs the lack of information on an item, therefore, this item will never be a subject for commendation (sparsity problem). In other terms, inactive users, who have only expressed few ratings or interacted with few items, cause data sparsity, which makes it

TABLE I. Comparison Between Recommendation Approaches

Techniques	Advantages	Limits	References
Collaborative filtering	Does not require any knowledge about the content of the item or its semantics. The quality of the recommendation can be assessed. The higher number of users' accurate the recommendation.	Cold start problem. New items are recommended only if they are already rated by users. Problem of confidentiality. Complexity: in systems with a large number of items and users, the calculation grows linearly. The number of users is relative to the quality of the recommendation provided.	[3]
Content-based recommendation	Does not need for a large community of users to make recommendations. A list of recommendations can be generated even if there is only one user. Quality grows over time. Does not need for information about other users. Considers the unique tastes of users.	User Profile Requirement. Problem of recommendation of images and videos in the absence of Metadata. Content analysis is required to make a recommendation.	[10]
Hybrid approach	It always provides predictions to content of recommendation. It improves the user preferences for suggesting items to users.	Have increased complexity and expense for implementation. Need external information is not usually available.	[3], [7], [10]

hard to make an accurate recommendation, if the system relies only on users' ratings or their interaction records [32]. The third category presents the problem of information overload. This latter can be solved by collecting relevant users 'preferences. In the literature, there are two types of information acquisition explicit and implicit feedback. The explicit way is considered as the most representative indicator of the user's interest in an item [11]. The indirect way allows to observe the users 'behavior towards a given item. As a result, users are not required to rate the items directly. This act will not have any impact on the quality of the provided recommendation. The authors [11] find necessary to capture as much information as possible without the direct intervention of users, in order to [12] better determine their interests and needs. Some research has been conducted in this area, the authors in [13] have found a way to transform users 'behavior in the recommendation platform into explicit information through the "User Interactions Converter Algorithm (UICA)". This research helped to determine users' interest by analyzing and converting their behavior.

B. Real Recommender Systems

Currently, there are wide ranges of recommendation systems that are used in different areas. Table II includes the most popular recommendation systems.

TABLE II. The most Popular Recommendation Systems

Systems	item	References
Netflix	Films	[14]
YouTube	Videos	[15]
Facebook	Persons	[16]
TripAdvisor	Hotel, restaurant	[17]
Book crossing	Books	[18]
Google scholar	Scientific articles	[19]
Amazon	Objects	[21]

III. Proposed Approach

In this section, we describe in details our approach. Fig. 1 represents the general process of our approach. In the following, we give details of each step according to the chronological order of their appearance (shown in Fig. 2).

Fig. 1. General Process of recommendation.

1. Collecting users preferences: This step allows the user to enter his / her personal data by filling out forms that include questions such as: name, first name… and professional information such as years of experience, center of interest, etc. Some information is both static and dynamic since they vary over time according to the age or experience of the user.
2. Collecting preferences via Twitter: Through this step, the user can give his permission to retrieve his social information. Generally, users are more active on their social networks allowing us to recover their preferences indirectly (implicitly). In this study, we only operate on Twitter social network. Choosing Twitter is justified by its provision of public API developers, easy to use. We used Twitter API for collecting users preferences [34]. Furthermore, Twitter has several APIs to query its database, but also to build other services. These APIs are particularly rich by returning almost a hundred variables per query; the data concerning the tweets (date of publication, the text of the message, etc.), the author (creation date of the account, pseudo ...), the

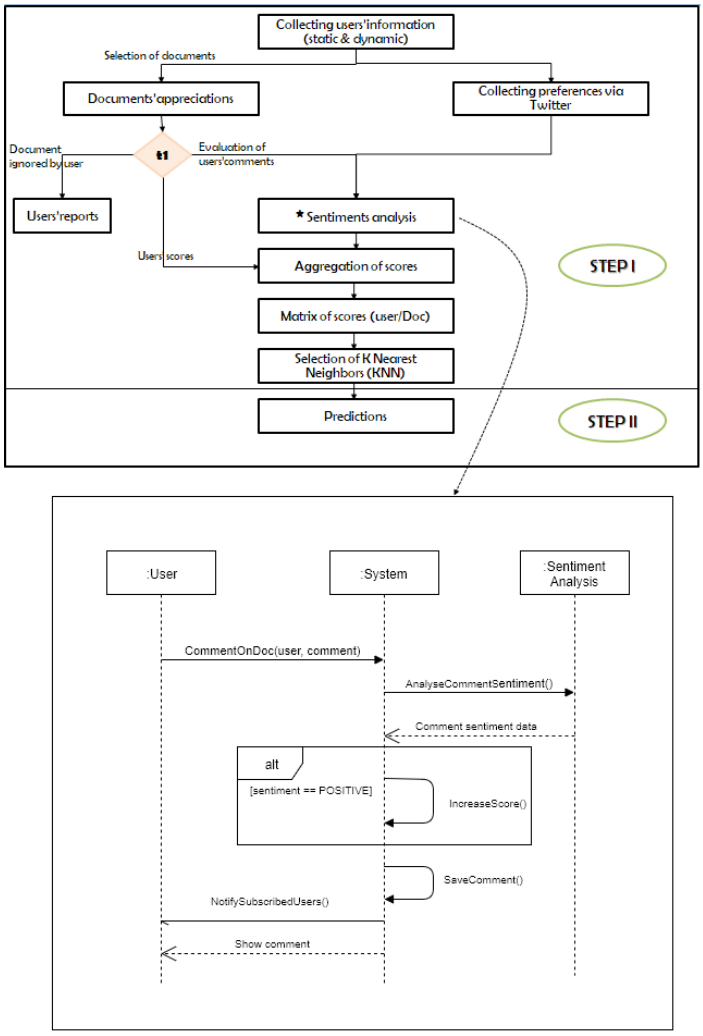

Fig. 2. Detailed process of collecting and filtering information step.

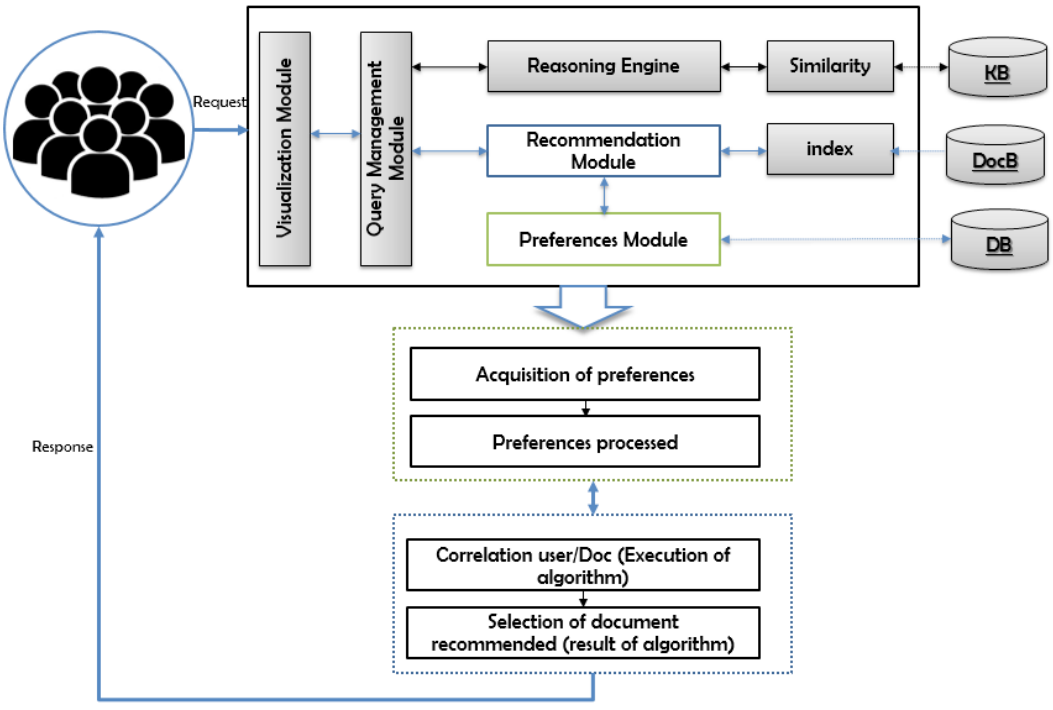

Fig. 3. A resumed overview of the collaborative system architecture.

entities contained in the messages (hashtags, mentions, urls ...) and information location (country, time zone, longitude / latitude).

Moreover, it is open source and allows analyzing users' behaviors through their activities and their posts [20]. This allows us to enrich the user profile with his behavior (We can be inspired by [20]). The user behavior will be considered as dynamic data (example: hashtag can be used as a center of interest) and consequently lead to an evolutionary recommender system over time. Behavioral study will also allow to build communities of users (cited in perspective) according to their behavior on social networks particularly on Twitter.

3. Document appreciations (document notation): the user is invited to evaluate some documents in order to better understand his preferences. The evaluation is done in two possible ways: by assigning a direct score varying from 1 to 10 and / or commenting on the proposed document.
4. Users' reports: If the document is not noted, the user is invited to answer some questions to better understand his expectations.
5. Sentiments analysis: user comments that appear on a given document are analyzed with API to detect feelings based on text processing [35] and categorized as Good where the obtained note vary between 8 to 10, Medium where it is between 5 to 7, and Null where it is between 1 to 4. These notes are utilized (average) with direct evaluation of a document, and constitute the final score of this document (as detailed in the scenario 2).The comment will be stored with the user profile (his history). The system will notify all the subscribed users (users that have also commented on this document), in when the document has a new comment. Posted comments for a given document provide feedback on the document itself. Users who give their opinions are thus bound in one way or another to the user who wants to have opinions on the document in question. Consequently, this link constitutes trusted network. In the case of industrial diagnosis, the trades people are considered to be persons of high confidence that is to say of first degree.
6. Aggregation of scores: here, the average of the notes of the documents is calculated. The matrix of scores (user / Doc) is established.
7. Selection of K Nearest Neighbors (KNN): This step allows to select the K nearest neighbors of the Active User in order to recommend relevant documents firstly. This selection is based on cosine similarity, which is a measure of similarity between users [22]. We opted for the KNN algorithm because we believe that is effective when the number of users is increasing in the platform. Our tool will only calculate the predictions between the k nearest neighbors instead of calculating them among all users. This allows a quick result. Moreover, this algorithm is easy to set up. Remember that this recommendation tool complements an existing CDSS system. The need for documents' recommendation is real.

$$\cos(u1,u2) = \frac{\sum_{i=1}^{d}(note(u1,i) * note(u2,i))}{\sqrt{\sum_{i=1}^{d}(note(u1,i))^2} * \sqrt{\sum_{i=1}^{d}(note(u2,i))^2}} \quad (1)$$

8. Calculating predictions: It represents the second phase of the recommendation process that is given in Fig. 1.

A. Global Architecture

We give below our tool architecture which is a part of a collaborative platform. Fig. 2 shows the detailed recommendation process shown in Fig. 1 (specifically Steps 1 and 2).

The first step in this process is to gather the information in mass of users in static type: name, first name ... and dynamic type such as the age, the experiment and the centers of interests.

Documents are presented for users' evaluation. This evaluation includes the score given to a document and / or the comment posted. The user can, if he wants, give access to his Twitter account.

If the document proposed to the user is not noted a questionnaire will be proposed to better understand the user's expectations (needs).

If the user posts a comment, this latter will be analyzed to explain his opinion and translated into a score through the feelings analysis step. This analysis is purely syntactic. It recognizes the terms of expressions such as good, bad, etc. We use an API for that [35].

The aggregation step groups the direct and indirect notes and normalizes them on a scale from 1 to 10. The User / Doc matrix is drawn at the end of this step.

To recommend a new document, we use the KNN to identify users with similar preferences to the active user. This method is based on Cosine measure. This latter represents the second step of Fig. 1.

Fig. 3 represents the summary architecture of the collaborative decision support system [2] for which we develop the recommendation module.

This figure zooms 2 essential modules the recommendation module and the preferences module.

Some information about the users (also preferences) of the recommendation tool developed in this paper come from this module shown in green. The blue part shows how the two modules influence each other.

All the documents of diagnostic that are stored in the documents base (DocB) have been evaluated by domain experts.

Proposed documents during a collaborative session are evaluated and corrected by the domain experts. These documents represent in the majority of times, detailed technical solutions to the problem faced by the industrial operator. We believe that the recommended documents in addition to the users' preferences influence the quality and the accuracy of the recommendation tool.

Table III summarizes some questions that allow experts to evaluate a diagnostic document.

TABLE III. Questions of Experts for Evaluating Diagnosis Documents

Id	Question
1	What is the machine part that needs to be changed?
2	What is the model of the new part?
3	What is the reference of the new part?
4	What is the machine to repair (reference)?
5	Please introduce the new solution.
6	What is the type of this problem? (Major or repetitive)
7	How is the proposed solution? (Simple to apply, complicated)
8	Do you find the proposed solution applicable?
9	Is the solution solving the problem in final way?

The diagram in Fig. 4 describes the main actions that a user can do on the platform. This latter can search by categories, tags, username or Doc name without being logged in. He can edit his profile to add more information, for example to add his Twitter social data to his profile to get more accurate recommendation based on his latest tweets. He can also view his notifications, and follow other users. The user can also access a book and like, share, Bookmark, comment, and read the document. All of these interactions are saved by the system for future uses by the recommendation system.

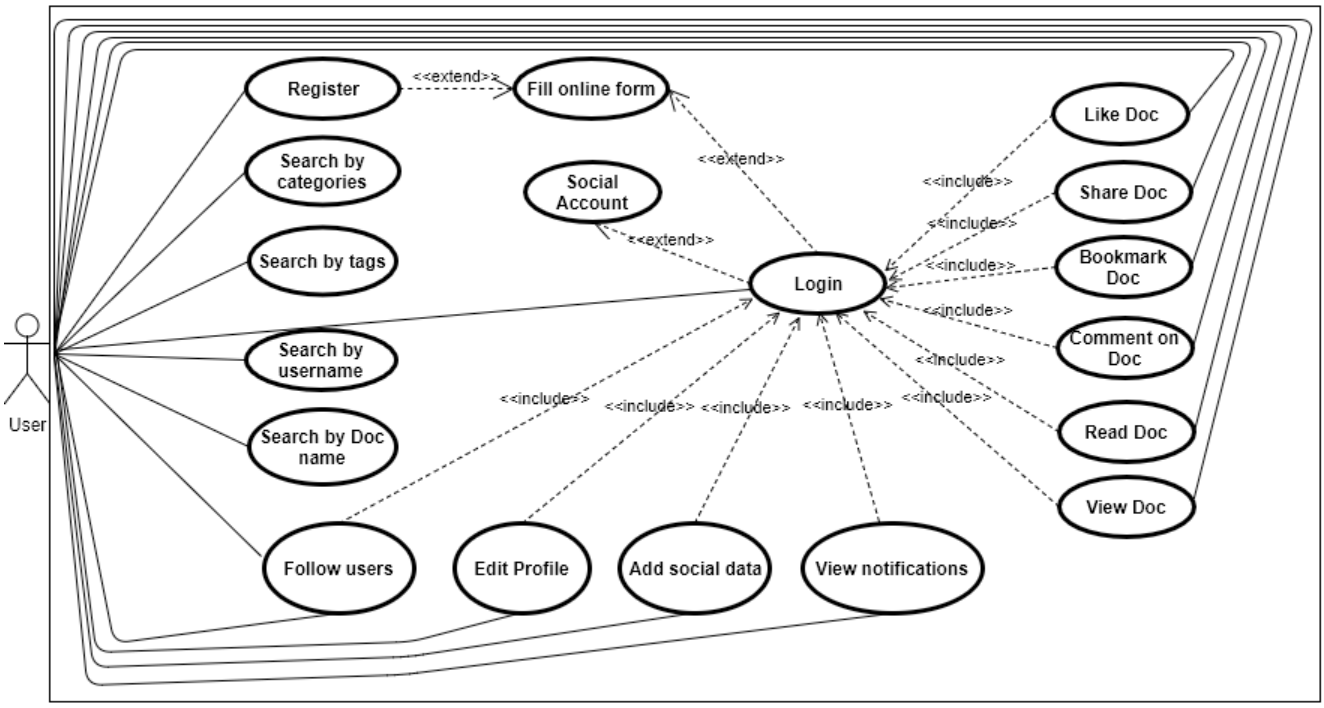

Fig. 4. User use-case diagram.

IV. Experimental Protocol

We conduct experiments "offline" (simulations) to evaluate our recommendation tool thus illustrating its efficiency, its performances and its scalability. In this section, we present the corpus of documents that will be used in the experiments of this work, as well as the evaluation measures.

A. Corpus

The experimental studies are carried out on a set of 40 industrial diagnostic documents and on the Book Crossing corpus. Book Crossing choice is justified by our motivation to analyse the behaviour of the tool developed facing a large number of documents. Book Crossing was collected in 2004 by Cao-Nicolas Ziegler from the Book Crossing community [18]. It is made up of 278,858 anonymous users who have provided 1,149,780 ratings on 271,379 books rated on a scale of 0 to 10. We have modified this corpus in such a way that it can be exploited by our tool. In book crossing the zero (0) is considered as null note. On the other hand, in our work 0 represents an unrated document. The grades assigned to the documents vary between [1, 10]. For that, we made a small modification to the notes in the book crossing corpus so that it can be manipulated by our tool.

B. Sample Used

To perform these experiments, we used a sample of 100 and 1600 votes for 40 diagnosis documents, we also used 10000 documents from Book Crossing to study the behavior of our approach. We make vary the k which is the number of the nearest neighbors according to the cosine similarity.

C. Experiences

In these experiments, we follow the example of calculation presented below. First, we calculate cosine similarity between a group of users. Second, we select K nearest neighbors and calculate predictions.

1) Example of Calculation

Table IV represents the ratings assigned by 6 users to 7 documents. Ux is a new user to whom we want to calculate the interest rate that D7 will bring.

TABLE IV. Matrix of Scores (User/Doc)

	D1	D2	D3	D4	D5	D6	D7
U1	1	2	7	1	8	10	3
U2	10	9	9	9	3	1	8
U3	7	7	8	9	10	1	1
U4	3	3	10	1	2	3	8
U5	8	9	1	8	9	1	1
U6	2	2	10	10	10	9	10
Ux	8	5	5	10	7	10	?

The principle is that users who shared the same preferences in the past are likely to share their preferences in the future.

Knowing that Ux noted D1, D2, D3, D4, D5 and D6 with respective notes 8, 5, 5, 10, 7 and 10. We use cosine measure to determine the similarity between these 7 users.

According to cosine (U, U) similarity:

Cosine (U1, Ux) = 0.777, Cosine (U2, Ux) = 0.774, Cosine (U3, Ux) =0.852, Cosine (U4, Ux) = 0.536, Cosine (U5, Ux) = 0.824, Cosine (U6, Ux) = 0.939.

The similarity threshold is set to 0.5, the similarity varies between [0, 1]. We keep the notes of the 3 nearest neighbors (in descending order) and calculate their average. The notes used in this example are the note of U6, U3 and U5 (10, 1 and 1). The average of these ratings is 3.3. As conclusion, document D7 will not be recommended to Ux.

Table V gives detailed notes according to the three nearest neighbors.

TABLE V.
Prediction Calculations According to the 3 Nearest Neighbors

	D1	D2	D3	D4	D5	D6	D7
U3	7	7	8	9	10	1	1
U5	8	9	1	8	9	1	1
U6	2	2	10	10	10	9	10
Ux	8	5	5	10	7	10	3.3

2) Evaluation with a Dataset

Below, we study the behavior of our approach with different mount of data knowing that an operator cannot evaluate the same documents many times. We use three samples: 1600 votes with 40 documents and 40 operators, 100 votes with 40 documents and 40 operators and 10000 votes of Book Crossing.

- *Sample of 1600 votes for diagnostic documents*

  Fig. 5 represents the Active user / Users similarity which varies between [0, 1]. The threshold of the similarity is set to 0.5. As shown in the figure, we notice that users with the respective id 4457, 5667, 832, 998, 593, and 8845 have similarity above the threshold, thus these users are similar to the active user. This result means that these users have shared the same preferences in the past. Hence, they are likely to share their preferences in the future.

  After the study of cosine similarity, we use KNN [22], [23] which determines the k users to calculate the prediction of the documents 'notes to be recommended to a given user.

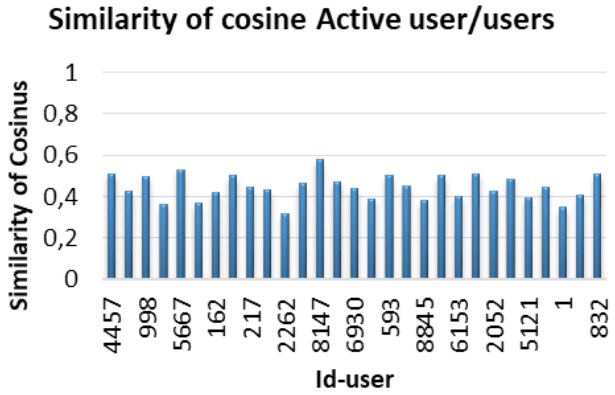

Fig. 5. Similarity of cosine calculated between the Active User and users of the sample.

We use the same set of users and documents. We give respectively the variable k the values 3, 20 and 30. The results obtained are presented in Fig. 6, 7 and 8.

K=3

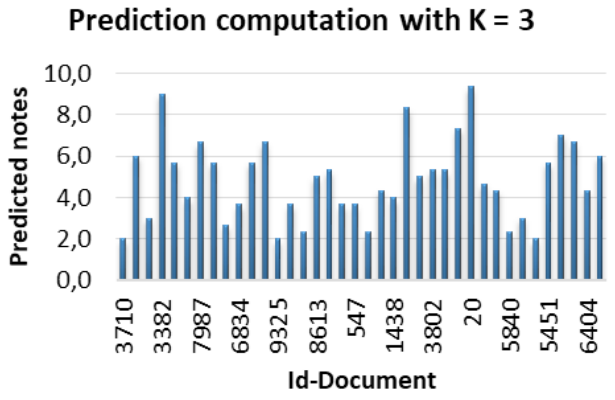

Fig. 6. Predicted notes with k=3.

K=20

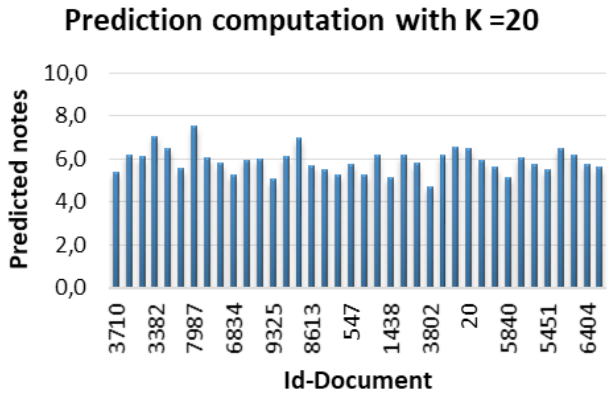

Fig. 7. Predicted notes with k=20.

K=30

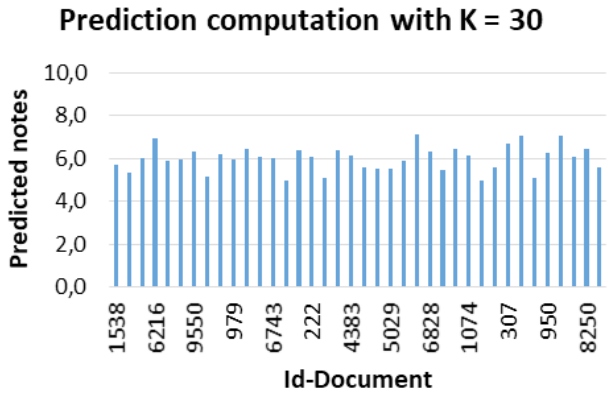

Fig. 8. Predicted notes with k=30.

Fig. 9 shows similarity between users and Active User. We see that the user with ID = 2 is more similar than others to the Active User. The sample presented here contains three (3) users and one hundred (100) votes.

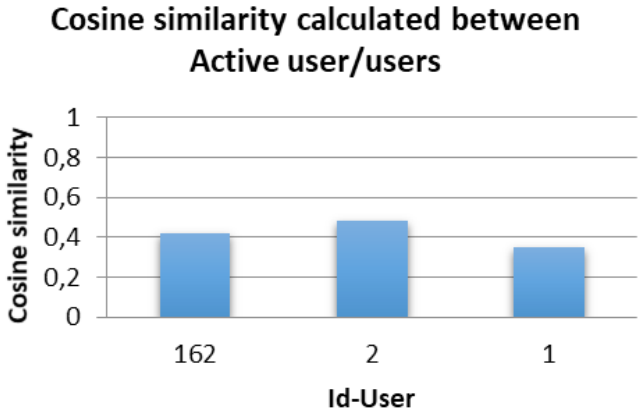

Fig. 9. Cosine similarity is calculated between the Active user and users of the sample.

Fig. 10 shows the calculated prediction for the diagnostic documents.

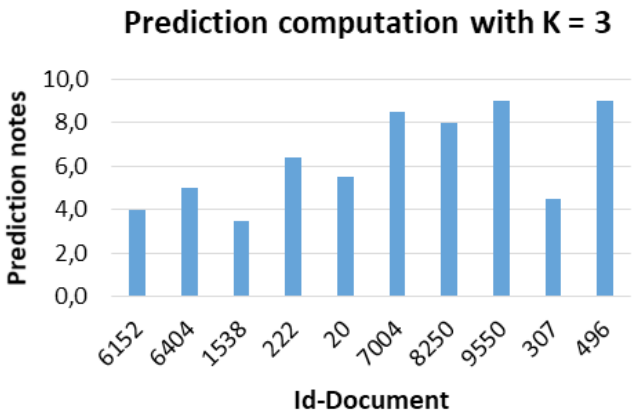

Fig. 10. Predicted notes with k = 3.

- *Sample of 10000 votes from Book Crossing*

  We conducted an experiment with 10000 votes taken from Book Crossing. The results obtained are shown in Fig. 11. We note that some values are null for some documents because Book Crossing interprets the zero as a non-given note.

  We notice that when the K increases, some documents appear as the best predicted documents. We also notice that the best documents predicted with k = 3 always remain the best predicted with k = 30 and k = 20 in the case of 1600 diagnostic documents.

D. Evaluation Measures

In this experimental protocol, we use 2 measurements of predictive evaluations MAE "Mean Absolute Error" and RSME "Root Mean Squared Error", which calculates the accuracy of the predictions against the actual assessment performed by the operator.

Let n be a set of test items, p (u; i) a note prediction of the user u for the item i and n (u; i) the actual score assigned by the user u for the item i [24].

The most commonly used measure is MAE, which is regularly used to evaluate the accuracy of a prediction. It corresponds to the mean absolute error between the actual evaluation and the prediction. The measure is calculated by the following formula:

$$MAE = \frac{1}{n}\sum_{k=0}^{n}\left|P(u,i) - n\,(u,i)\right| \tag{2}$$

The second RMSE measure raises the squared error before summing, which is useful when we want to give more criticality to the important errors [25]. The measure is calculated by the following formula:

$$RMSE = \frac{1}{n}\sum_{k=0}^{n}\sqrt{\left(P(u,i) - n\,(u,i)\right)^2} \tag{3}$$

Table VI gathers the notes assigned to the documents by a set of users as well as the predicted notes by the cosine measure while averaging the notes of the k nearest neighbors.

TABLE VI. PREDICTED AND ASSIGNED NOTES BY EACH USER

Id-user	ID-Document	Predicted Note	Attributed Note
6092	7335	7.1	7
	3382	7.5	7
	6216	7.1	7
	491	6.9	6
	3550	6.9	7
7887	7987	7.6	8
	3382	7	7
	7338	6.5	6
	8250	6.3	6
	6216	6.2	6
11111	7987	6.9	7
	3382	6.8	7
	7338	5.95	6
	8250	5.9	6
	6216	5.9	6
2222	7987	6.9	7
	3382	6.9	7
	7338	5.95	7
	8250	5.9	6
	6216	5.9	6
685	7987	6.4	7
	3382	6.35	6
	7338	6.3	6
	8250	6.2	6
	6216	6.2	6

After the calculation of MAE and RMSE presented above, we obtain the Table VII.

TABLE VII. RECOMMENDATION TOOL EVALUATION

Id-user	Real evaluation	MAE	RMSE
6092	6.8	1.1	2.13
7887	6.6	0.26	0.27
11111	6.4	0.11	0.21
2222	6.4	0.31	0.55
685	6.2	0.73	0.78

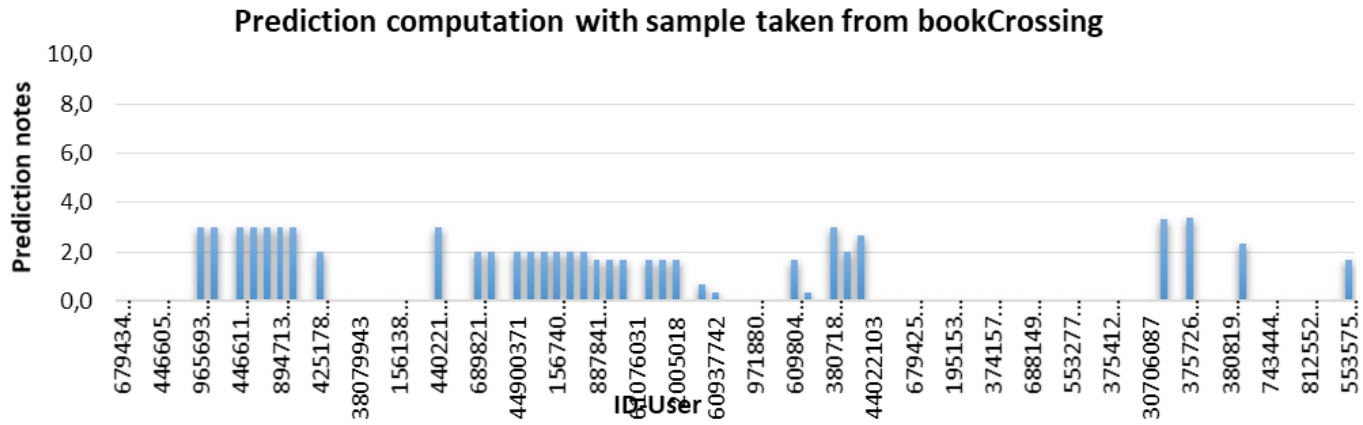

Fig. 11. Predicted notes using Book Crossing.

Fig. 12 shows the results shown in Table VII. We note that the error rate of our recommendation tool is much lower than the evaluation of the users.

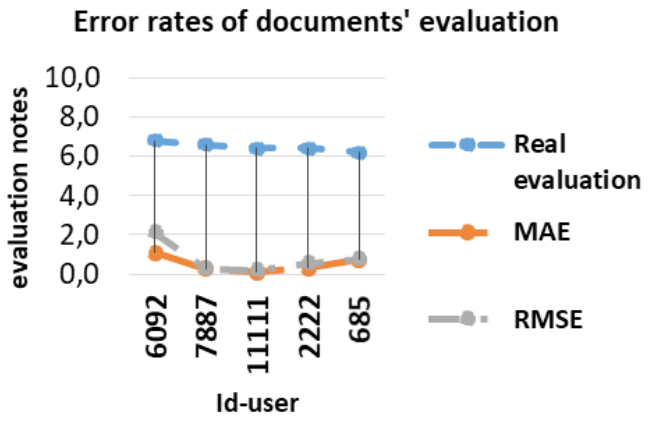

Fig. 12. Error rates of documents evaluation.

V. Results and Discussion

The data pre-processing phase that is presented in this article plays a very important role in the recommendation of the diagnostic documents proposed by the collaborative platform. Extracting relevant elements of recommendation is permitted by the pre-processing phase. This phase allows the acceptability of the provided recommendations to the users. In this phase, users preferences are discerned.

Our approach has some advantages which are summarized in the quality and speed of the recommendations that are provided to the users of the platform, and the use of users preferences especially industrial operators to avoid possible rejections. These recommendations orient to the solutions applied by users in the case of industrial diagnostics; a good recommendation in this case is translated into an effective resolution of the problem in order to allow a greater amount of production and a gain of money.

In order to demonstrate the effectiveness of the new tool and its influence in the collaborative platform, precisely in reasoning engine as shown in Fig. 3 (section 3) we use the same cases presented in [26].

We note that the reasoning engine performance is not affected only by the similarity measure but by the quality of the documents provided also. The quality for us is expressed in terms of relevance. This tool helps to increase the documents relevance by evaluating them through the domain experts as explained in section 3.We can justify this by increasing the number of relevant documents saved after integrating the recommendation tool. Besides, it will help to gain the confidence of operators and thus encourage them to reuse the tool to improve their profiles.

The objective of this evaluation is to study the impact of each prediction method on the performance of the recommendation tool. The methods studied are: FCS (Standard Collaborative Filtering) [24], D-BNCF-KNN Densified Behavioural Network based. Collaborative Filtering where only the direct neighbors [27] are involved in the calculation of the predictions, as well as our approach [28].

Table VIII shows MAE results of the compared methods. We can see that MAE is deteriorating in the case of D-BNCF-KNN. The application of the FCS brings an improvement compared to D-BNCF-KNN, we also notice that our approach is slightly than FCS.

TABLE VIII. MAE Results

Recommendation model	MAE
FCS	0.763
D-BNCF-KNN	1.074
Our work	**0.502**

VI. Scenario/Overview

In this section, we give an overview of the main Web features, mobile application for both the user and admin, and two scenarios of execution. The platform focuses on both Admin and User as the main users of the system. We use different documents which are not necessarily documents of industrial diagnosis. These documents are provided from different sources as Book Crossing.

- **Admin**: The admin logs in the system if he already has an account. He has the privilege to create, update and delete documents. He can also create, update, disable or delete other users from the platform. The admin has also the ability to see more features in the platform like the sentiment detection on users' comments. The admin has also all the features that a simple user has.
- **User**: The user logs in the system if he already has an account otherwise he creates a new one with his email that he needs to verify in order to interact with the system. After that, the user is logged in, he can navigate in the platform by reading, liking, bookmarking, sharing, and commenting on documents. He can follow other users and be followed by other users also.

A. Web Application

Fig. 13 shows an overview of sign in user. He can learn about all the platform features and choose to join or not. The user can login if he has already an Account. Otherwise, the user must provide a unique user-name that does not exists in the database with a valid email and password. After that, the user is registered. He will be redirect to a page that informs him about a verification link that was sent to his email inbox. Then, he will be redirected to a page to confirm his email.

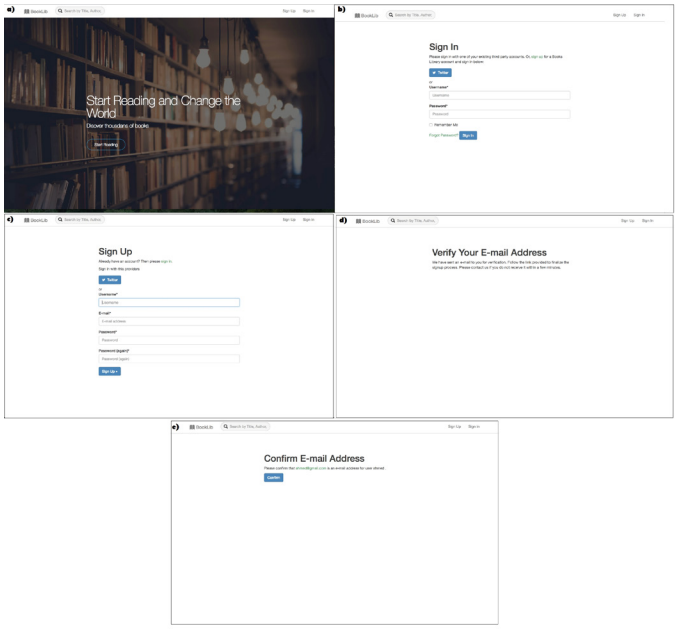

Fig. 13. An overview of sign in user process.

After the login the user is directly redirected to his time-line that contains all the recommended documents (books) according to his

profile as shown in Fig. 14. The recommendation shown here does not represent the final result, we recall that the aim of this work is to conceive and prepare the ground for the documents recommendation. The user can get more information about a document, and in this case the document is a book, by going to his details. The user can view his profile or the profile of another user. The user can view the documents that he started to read, his bookmarked documents, the users that he follows, and the users that are following him. The logged in admin can view the comments and sentiment as shown in Fig. 14 part "d".

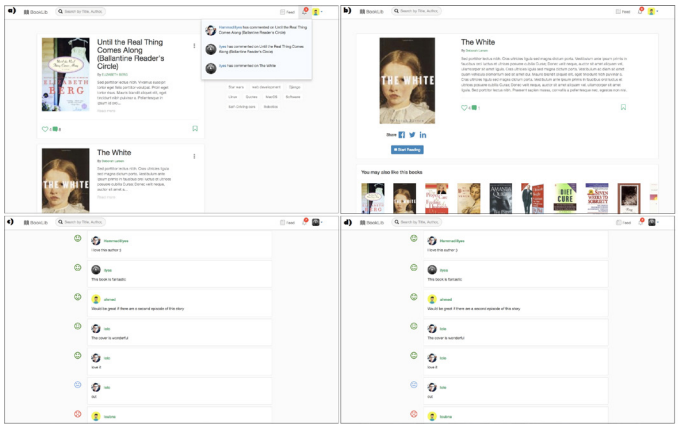

Fig. 14. User profile and documents details.

Moreover, it can access the admin panel by navigating to its url. The logged in admin can manage users and documents in the admin panel in the users and documents sections as shown in Fig. 15.

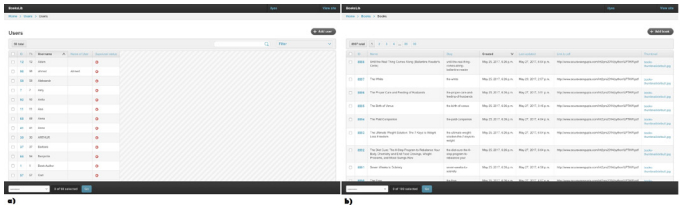

Fig. 15. Admin features.

B. Mobile Application

With the mobile app, the user can perform the same actions like in the Web application as shown in Fig. 16.

C. Scenarios

Here, we give two (02) scenarios of execution.

1) Scenario 01: Case of New User

The new user really represents a new operator who has just been recruited to the non-woven company "INOTIS", thing that remains unlikely. We have previously stated that this recommendation tool operates with a collaborative decision support system launched over the last 3 years on the INOTIS pilot company. We specify that operator data comes mainly from this system. The collection of user data lasted about one year. We mentioned as a future work the launch of this tool in the Web which opens the door to other operators from the world wide to join it. Each new operator goes through the registration step on the platform where he is invited to fill the requested information.

The user is not obliged to complete his profile; the registration form is designed in two categories "personal information of the user" which is required as: surname, first name, email address and password. The second category is "personal preferences" which is optional such as: favorite authors and centers of interest.

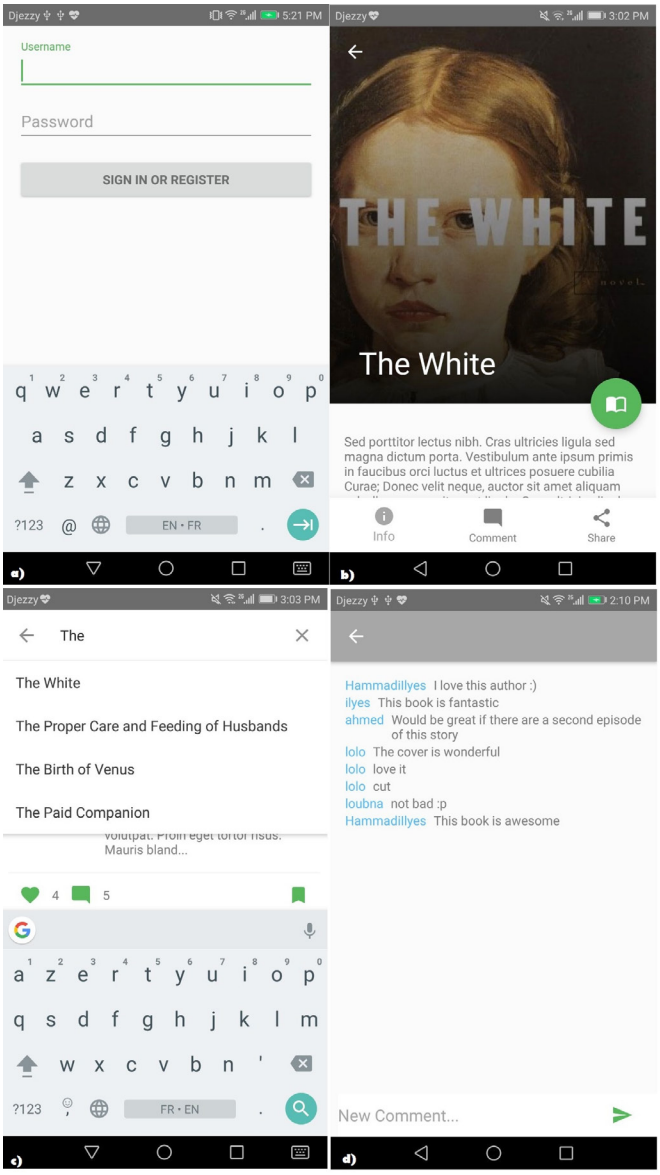

Fig. 16. An overview of the Web application.

Upon registration, he is redirected to his private area where he finds documents to note.

- If the new user doesn't complete his data profile (but he introduces his center of interest). The system uses this latter for example: the operator introduces HP machine, water pump. The system recovers the documents which contain these keywords in their description (knowing that a document has Keywords as a descriptor) in this example document whose id=38 is recovered, and some new documents.
- If this user doesn't insert his center of interest, the system proposes popular documents (i.e. popular documents are the top rated documents by other users of the platform) and new documents recently added to the documents base (the number of documents which are proposed to the evaluation is 10).

At any time, this operator can complete his profile by his personal preferences. As soon as it is done, he receives a recommendation according to the preferences expressed (as detailed in scenario 2).

2) Scenario 2: Case of an Old User (in 8 Steps)

In the case of an old user (registered user), the preferences are already

listed in the database. He can complete or update them by adding other areas of interest. The user has agreed to access his Twitter account.

The personal data of the user can be completed from his Twitter account such as: age ... for confidentiality, we cannot give a real example. It is enough to call the user "operator x".

The confidentiality of the data is ensured in the following 3 cases:

- In the case of an old operator who has already participated in the collaborative sessions by giving expertise (diagnosis). Data will not be used outside the recommendation platform.
- The notes attributed to the documents are visible only internally, that is to say only in the recommendation platform.
- When an operator comments on a document, the comment is only readable by the operators of the platform.

However, the operator is free to post a tip or an opinion on his social network for example: when he participates at a conference, he posts a tweet with hashtags. These will be exploited to enrich his list of centers of interest as explained below:

"#adhesive innovation enabling next generation hygiene products", adhesive is added as new center of interest for the operator who posts this tweet.

A set of 4 documents recently added to the database are proposed to the evaluation. The ratings assigned to D1, D2, D3 and D4 are 2, 4, 0, and 9 as a direct notation. The user has posted a comment to D4. The comment is evaluated as positive with the rating of 8/10, and then will be added to the user's profile. Users commenting on D4 will be notified.

The D3 has not been evaluated, a questionnaire is therefore proposed to the user. The answers to the questions are presented in Table IX.

TABLE IX. Users' Reports Questionnaire

Question	Answer
A part or the entire document is damaged?	No
The content of the document does not fall within your area of expertise?	No
Is the document too long?	Yes

After the assessment of the documents the score is calculated as shown Table X.

TABLE X. Documents Scores

Document	Score
D1	2
D2	4
D3	0
D4	(9+8)/2 = 8,5

In order to recommend a new document to this user, it is necessary to know its similar users having shared the same preferences in the past (document notation). Our tool is based on Cosine's measurement. We take an example of 5 documents noted by 4 users where the D5 is a new document not noted by operator x. We calculate the prediction for the operator x.

The similarity of Cosine is calculated in the following way, we detail the calculations for the user 1 and operator X.

Cos (U1, operator x) = 0.45 (according to the formula 4).

In the same way the other calculations are made. We get Table XI.

TABLE XI. Users / Documents Evaluation

	U1	U2	U3	U4
Operator x	0,45	**0,65**	**0,61**	**0,73**

Where k = 3 and the users closest to our user are U4, U2, and U3. The prediction is calculated according to the average of the notes of the similar neighbors: Operator x prediction (6 + 9 + 5)/3 = 6.66. The D5 will be appreciated by the operator X so it will be recommended (score is greater or equal to the threshold (5/10)).

As new documents coming from collaboration sessions of CDSS are offered to the different users (operators) of the platform according to their areas of interest this proposition is based on the descriptors "keywords" of the documents. The documents will be proposed according to the number of interests (keywords) that contain.

VII. Conclusion and Future Work

In this paper, we have presented a Web 2.0 platform that includes a document recommendation tool for industrial diagnosis. This tool is integrated in the collaborative decision support system in order to provide relevant solutions to industrial operators. The proposed approach is based on three essential steps for the recommendation namely: (i) data collection and filtering, (ii) prediction and (iii) evaluation. The input data for the implemented recommendation tool is varied like (i) considered data from collaborative system experiences such as operator preferences and relevant documents to recommend; (ii) information, preferences retrieved in a direct way through the forms, (iii) indirectly via the social network Twitter, or sentiments detection of comments posted. The prediction calculation is performed between the k-users using Cosine measure. It allowed us to consider similar users to the Active User and therefore obtain considerable reduction of the user / document matrix.

We discussed the results which are obtained from the recommendation tool and collaborative support system (without recommendation tool). This discussion presents the quality of documents that are recommended to industrial operators is not based only on the similarity measure of case-based reasoning [26] but on the number of relevant documents that are considered as inputs also. This comparative study allowed us to measure the number of relevant documents in the document base of collaborative support system.

We presented a comparison, which is conducted between our study and others from literature (FCS: Standard Collaborative Filtering, D-BNCF-KNN: Densified Behavioural Network based Collaborative Filtering where only the direct neighbors). The MAE results obtained from this comparison show that our approach has a low error rate compared to the other approaches.

The experiments are based on diagnostic and Book Crossing documents.

We can improve our work by:

- Using collaboration platform for identifying potential users networks,
- Evaluating the proposed tool with other data sets,

$$\cos(u1,u2) = \frac{\sum_{i=1}^{d}(note(u1,i) * note(u2,i))}{\sqrt{\sum_{i=1}^{d}(note(u1,i))^2} * \sqrt{\sum_{i=1}^{d}(note(u2,i))^2}} = \frac{((3*2)+(7*4)+(8*0)+(4*8*5)}{\left(\sqrt{(3^2+7^2+8^2+4^2+10^2)} * \left(\sqrt{(2^2+4^2+0^2+8*5^2)}\right)\right.} = 0{,}45 \quad (4)$$

- Testing other prediction measures to increase relevance,
- Widening of the recommendation tool to a variety of resource diagnostics (machines) in the field of non-woven, textile and other.
- Testing the satisfaction of the industrial operators;
- Testing the usability of the Web application according to the works presented in [29], [30] as a quantitative means for measuring the user's experience, and by using heuristics as a qualitative means such mentioned in [31].

References

[1] G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in Plastics, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.

[2] F.Z. Benkaddour, N. Taghezout, B. Ascar, "Towards a Novel Approach for Enterprise Knowledge Capitalization Utilizing an Ontology and Collaborative Decision-Making: Application to Inotis Enterprise". International Journal of Decision Support System Technology (IJDSST), 2016, vol. 8, no 1, p. 1-24.

[3] X. Su, and M. Khoshgoftaar. "A survey of collaborative filtering techniques". Advances in Artificial Intelligence, 2009, vol. 2009, p. 4.

[4] H. Lu, C. Chen, M. Kong, H. Zhang, and Z. Zhao, "Social recommendation via multi-view user preference learning". Neurocomputing, 2016, vol. 216, p. 61-71.

[5] J. Wei, J. He, K. Chen, Y. Zhou, and Z. Tang, "Collaborative filtering and deep learning based recommendation system for cold start items". Expert Systems with Applications, 2017, vol. 69, p. 29-39.

[6] A. Martin, P. Zarate, and G. Camillieri, "A Multi-Criteria Recommender System Based on Users' Profile Management". In: Multiple Criteria Decision Making. Springer International Publishing, 2017. p. 83-98.

[7] O., Sanjuan Martínez, C., Pelayo G-Bustelo, R., González Crespo, and E., Torres Franco, "Using recommendation system for e-learning environments at degree level". International Journal of Artificial Intelligence and Interactive Multimedia, 2009, vol. 1, no 2. p. 67–70.

[8] V. Satish Kumar, and P. Nymphia, "Survey on content based recommendation system". Int. J. Comput. Sci. Inf. Technol, 2016, vol. 7, no 1, p. 281-284.

[9] E. R, Nuñez-Valdez, J.M.C., Lovelle, O., Sanjuan-Martinez, C.E., Montenegro-Marin, and G. Infante-Hernandez. "Social voting techniques: a comparison of the methods used for explicit feedback in recommendation systems". International Journal of Interactive Multimedia and Artificial Intelligence, 2011, vol. 1, no 4.

[10] A.M., Dakhel, H.T., Malazi, and M., Mahdavi, "A social recommender system using item asymmetric correlation". Applied Intelligence, 2017, p. 1-14.

[11] J. Mathieu, Dyson, Mark, Callaway, and Duncan, "Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations". In: ACEEE Summer Study on Energy Efficiency in Buildings. 2012. p. 189-203.

[12] E. R, Núñez-Valdéz, J.M.C., Lovelle, O.S., Martínez, V., García-Díaz, P.O., De Pablos, and C.E.M., Marín, "Implicit feedback techniques on recommender systems applied to electronic books". Computers in Human Behavior, 2012, vol. 28, no 4. p. 1186-1193.

[13] E. R, Nuñez-Valdez, J.M.C., Lovelle, J.M., Cueva, G.I., Hernández, A.J., A.J., Fuente, and J.E., Labra-Gayo, "Creating recommendations on electronic books: A collaborative learning implicit approach". Computers in Human Behavior, 2015, vol. 51, p. 1320-1330.

[14] R., Hastings, and M., Randolph, from https://www.netflix.com/, created in 1997.

[15] S., Chen, C., Hurley, and J. Karim, from https://www.youtube.com/, created in 2005.

[16] M., Zuckerberg, and S. Sandberg, from https://www.facebook.com/, created in 2004.

[17] Mugly, and C. Petersen, from http://www.tripadvisor.fr, created in 2000.

[18] Book Crossing, from: http://www.BookCrossing.com, visited in 01/02/2017 (2001).

[19] Google, "Google scholar", from: https://scholar.google.com, created in 2004.

[20] F., Abel, Q., Gao, G.J., Houben, and K., Tao. Analyzing user modeling on twitter for personalized news recommendations. In International Conference on User Modeling, Adaptation, and Personalization (pp. 1-12). 2011, July. Springer, Berlin, Heidelberg.

[21] Bezos, J., "Amazon.com", from: https://www.amazon.com, created in 1994. Visited in 10/01/2017.

[22] Z. Wen, Recommendation Systems based on Collaborative filtering. CS229 Lecture Notes, 2008.

[23] W. Liang, G. Lu, X. Ji. J. Li, and D. Yuan, "Difference Factor' KNN Collaborative Filtering Recommendation Algorithm". In: International Conference on Advanced Data Mining and Applications. Springer, Cham, 2014. p. 175-184.

[24] U. Sharda and P. Maes, "Social information filtering: algorithms for automating "word of mouth"". In: Proceedings of the SIGCHI conference on Human factors in computing systems. ACM Press/Addison-Wesley Publishing Co., 1995. p. 210-217.

[25] S. Holmes, "RMS Error", from http://statWeb.stanford.edu/~susan/courses/s60/split/node60.html. Visited in 02/04/2017 (2000).

[26] F.Z. Benkaddour, N. Taghezout, B. Ascar, "Novel Agent Based-approach for Industrial Diagnosis: A Combined Use between Case-based Reasoning and Similarity Measures". International Journal of Interactive Multimedia & Artificial Intelligence, 2016, vol. 4, no 2.

[27] I. Esslimani, A. Brun, and A. Boyer, from social networks to behavioral networks in recommender systems. In: Social Network Analysis and Mining, 2009. ASONAM'09. International Conference on Advances in. IEEE, 2009. p. 143-148.

[28] I. Esslimani, Vers une approche comportementale de recommandation: apport de l'analyse des usages dans un processus de personnalisation. 2010. Thèse de doctorat. Université Nancy II.

[29] Schrepp, M., A. Hinderks, and J. Thomaschewski. Design and Evaluation of a Short Version of the User Experience Questionnaire (UEQ-S), International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 6, pp. 103-108.

[30] Schrepp, M., A. Hinderks, and J. Thomaschewski. Construction of a Benchmark for the User Experience Questionnaire (UEQ), International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 4, pp. 40-44.

[31] Bader, F., E. M. Schön, and J. Thomaschewski. Heuristics Considering UX and Quality Criteria for Heuristics, International Journal of Interactive Multimedia and Artificial Intelligence, 2017, vol. 4 no. 6, pp. 48-53.

[32] Q., Yuan, C., Li, S., Zhao, "Factorization vs. regularization: fusing heterogeneous social relationships in top-n recommendation". In: Proceedings of the 5th ACM conference on recommender systems. ACM, 2011, pp 245–252.

[33] J. Bobadilla, F., Ortega, A., Hernando, A., Gutierrez, "Recommender systems survey." Knowl-Based Syst vol. 46, 2013, pp.109–132.

[34] Nagard, E., L., "Use the Twitter API to collect tweets with Talend" from: http://www.erwanlenagard.com/general/tutoriel-utiliser-lapi-twitter-pour-collecter-des-tweets-sans-coder-avec-talend-1029. Accessed: 2017-05-20

[35] Sentiment analysis. http://text-processing.com/docs/sentiment.html. Accessed: 2017-05-29

F. Z. Benkaddour

F. Z. Benkaddour is an assistant professor at École Normale Supérieure d'Oran (ENSO). She holds her doctorate thesis in IRAD at university of ORAN1 Ahmed Ben Bella, Algeria in 2017. She received her Master degree on ID-IHM from the same university, in 2013. She is a member of the EWG-DSS (Euro Working Group on Decision Support Systems) since May 2015. She is also a member of the LIO (Laboratoire d'Informatique d'Oran) laboratory. Her research topics include Collaborative Decision Support System, Knowledge Management, Case Based Reasoning Systems, Multi Agents Systems, WEB technologies, and Recommender Systems.

N. Taghezout

N. Taghezout is an assistant professor at ORAN1 University. She holds her doctorate thesis in MITT at PAUL SABATIER UNIVERSITY in France in 2011. She also received another doctorate thesis in Distributed Artificial Intelligence from ORAN University in 2008. She holds a Master degree in Simulation and Computer aided-design. Dr Noria Taghezout conducts her research at the LIO laboratory as a chief of the research group in Modeling of enterprise process by using agents and WEB technologies. Since she studied in UPS Toulouse, she became a member of the EWG-DSS (Euro Working Group on Decision Support Systems). She is currently lecturing Collaborative decision making, Enterprise management and Interface human machine design. Her seminars, publications and regular involvement in Conferences, journals and industry projects highlight her main research interests in Artificial Intelligence.

F. Z. Kaddour-Ahmed

F. Z. Kaddour-Ahmed holds her Master degree on KEWT (Knowledge Engineering and Web Technology) at university of ORAN1 Ahmed Ben Bella, Algeria in 2017. She obtained her Bachelor's degree on CND (Computer Networks and Database) at university of Aïn Témouchent, Algeria in 2015. Her research topics include information systems, WEB technologies, and Recommender Systems.

I. A. Hammadit

I. A. Hammadi is a Full Stack Web Developer at Joboti since September 2017. He holds his Bachelor's degree in Computer Science in 2017 at university of ORAN1 Ahmed Ben Bella. His domains interest are Software Engineering and Development, Cloud Native Technologies, Machine Learning and AI.