REVISTA CUBANA
DE TRANSFORMACIÓN DIGITAL

RNPS 2487
Vol. 1 • Nro. 1 • enero-abril 2020 • págs. 96-107

ARTÍCULO ORIGINAL

Sistema de Recomendación en una Solución Transaccional y Analítica para la Promoción de la Salud

Recommender System in a Transactional Analytical Soltion for Health Care and Health Promotion

Claudia Quintana-Wong
claudia.quintana@matcom.uh.cu • https://orcid.org/0000-0003-0551-8231
Lucina García Hernández
lucina@matcom.uh.cu • https://orcid.org/0000-0003-4551-1747
Amelia Rabanillo Echaniz
amelia.rabanillo@gmail.com • https://orcid.org/0000-0002-2249-9135
Javier Guillot Jiménez
javier.guillot@gmail.com • https://orcid.org/0000-0003-3723-3353
Zoraida Amable Ambrós
amacor@sld.cu • https://orcid.org/0000-0002-9151-7544
UNIVERSIDAD DE LA HABANA, CUBA

Recibido: 2020-02-04 • Aceptado: 2020-03-06

RESUMEN

La Sanología, como saber nacido en la Universidad de La Habana, propone una nueva concepción de la salud de las personas y busca enriquecer el estilo de vida. El desarrollo de las ciencias de la información y la madurez que ha alcanzado el enfoque constituyen las bases para la creación de nuevas soluciones computacionales. En el presente trabajo, se concibe y diseña una solución matemático-computacional basada en bases de datos *NoSQL* para la aplicación de la estrategia sanológica. La solución se caracteriza por favorecer el acceso de los usuarios desde diferentes lugares a través de los teléfonos móviles, el diseño de una arquitectura distribuida y el uso de modelos de datos flexibles. A partir de la aplicación transaccional, se hace necesario el análisis automático de los datos sociales y médicos almacenados. Con este fin se propone un sistema de recomendación híbrido que combina técnicas basadas en filtrado de conocimiento, de contenido y colaborativo, así como aprendizaje automático y procesamiento del lenguaje natural con el fin de apoyar el proce-

so de toma de decisiones. La solución propuesta es capaz de ofrecer recomendaciones personalizadas teniendo en cuenta datos históricos de los usuarios, su comportamiento y la similitud con otros usuarios. La validez de la solución fue verificada a partir de la implementación de un prototipo funcional y un conjunto de experimentos.

Palabras clave: Big data; aprendizaje automático; sistemas de recomendación; computación móvil.

ABSTRACT

The Sanologia, born at Havana University, proposes a new conception of people's health and aims to enrich their lifestyle. The broad development of the information sciences and the maturity that has reached the proposed approach constitute the bases for the creation of new computational solutions. In the present paper, a mathematical-computational solution was conceived and designed for the application of the approach proposed by Sanologia based on NoSQL databases. The solution is characterized to favor the access of the users from diverse locations via mobile devices, from the instrumentation of a distributed architecture and the use of flexible data models. From the operational application, it is feasible to analyze the medical and social data collected in an automated way. Knowledge-based, content-based and collaborative filtering recommendation techniques are hybridized in a recommender system, which using machine learning and natural language processing is incorporated in order to support the decision-making process. The proposed solution is capable of offering personalized recommendations taking into account the users' historical data, their behavior and the similarity with other users. The validity of the solution was verified by the implementation of a functional prototype and a set of experiments.

Keywords: *Big data; machine learning; recommender systems; mobile computing.*

INTRODUCTION

Health is a topic that has worried man since its inception, which is why he has embarked on a long journey dedicated to his study. For several centuries, health has been valued through disease, making medicine a hegemonic discipline and healing its guiding objective. The prac-

tice has shown that this model does not satisfy current problems. In response to this new challenge, at the end of the 20th century, at the Center for Human Health and Wellbeing Studies (CESBH) of the University of Havana, a new thought called "Sanologia" arose, even though it is a nonexistent word in the Spanish language, is defined as "an emerging knowledge and still under construction, which considers health as the result of free, harmonious and responsible development of the capabilities and potential of the human being in all its dimensions in the itinerary of his life. It does not exclude the disease, but its focus is on health promotion as the guiding axis of actions in the construction of a healthy culture" (Amable Ambrós, 2012).

The Sanologia outlines a strategy to promote the new health model and create awareness in the human person. The strategy is conceived within a certain context, called the sanological scenario, in which the lifestyle of the person is studied and actions are negotiated to improve their physical and social well-being (Amable Ambrós, 2012). The process of applying the strategy begins with a dialogue between the Sanologia expert and the interlocutor in order to collect personal information. After the data has been collected, is important to analyze it to make inferences about the individuals and communities, at the same time, create initial knowledge for the development of a healthy living project. To achieve these results with greater speed, accuracy, and efficiency, it is necessary to have a computational mathematical solution that minimizes the work and increases the good results in the data analysis.

Computer applications have made use of the latest technologies both in conception and in instrumentation, showing the rapid evolution and growth of technology in recent years (Amable Ambrós, 2012). The proposals of previous computational solutions make almost exclusive use of relational databases, which is not the most convenient given the unstructured nature of the data. The most recent contribution to Sanologia in terms of technology has been a solution, which makes use of NoSQL technologies, developed and presented in 2014 at (Guillot Jiménez, 2014). The solution of Guillot Jiménez guarantees the storage of all the information generated from the practice of the sanological strategy and its contribution as a transactional application is unquestionable. However, the information stored in the sanological spaces is not sufficiently exploited in order to make decisions, that is the reason why emerges the idea of a computational solution for the incorporation of a recommender system, within the framework of the transactional application distributed and oriented to documents, for the promotion of human health and well-being (Quintana-Wong, 2017). Finally, (Rabanillo Echaniz, 2018) presents both solutions, the transactional and analytical, as a software product.

METHODOLOGY

Since the year 1970, Edgar F. Codd proposed the relational model, based on which relational databases have triumphed in different contexts. With Web arrival, the exploitation of semi-structured and unstructured data has become an essential factor to bring wealth to current solutions. This phenomenon led to the need for data management systems to respond to new demands such as scalability, low cost, flexibility, and availability. Achieving

some of these objectives with relational databases was possible, but very difficult and required potentially high costs. In order to overcome the limitations of the relational model, the NoSQL ("Not only SQL") databases arose which group a large number of DBMSs that are characterized by storing data in a nonrelational way (Sullivan, 2015). These systems are not intended to replace the traditional RDBMS because, although they offer improvements in terms of horizontal scalability and more efficient storage of semi-structured and unstructured data, they have limitations in terms of complexity, consistency, and reliability. However, the NoSQL documental approach allows data analysis enrichment and decision-making process improvement (Pokorný, 2013).

With the explosive growth of the number of information sources, it has become a necessity for users to automate the search and processing of information (Jure Leskovec, 2014). Recommender Systems (SR) help solve this problem by finding information to provide users with personalized services and content. According to (Francesco Ricci, 2015), "recommendation systems are information filtering systems that deal with the problem of information overload by filtering fragments of vital information with a large amount of dynamically generated information according to the user's preferences, interest or behavior observed on the item. The system has the ability to predict whether a particular user would prefer an item or not based on the user's profile". Previous contributions, (Aggarwal, 2016) and (Nikzad–Khasmakhi, 2019), distinguish three main kinds of recommendation techniques: collaborative filtering, content-based and knowledge-based. Each of these systems has its advantages and disadvantages. In order to supply the limitations of one approach with the advantages of another and create a more complete recommendation system, hybrid recommendation systems arose.

Although hybrid systems have great advantages, such as improving the recommendations accuracy, the handling of the cold-start problem and the dispersion of data, their presence on the web is low due to the complexities in the development. It is also not possible to undertake all the possible combinations of hybrid recommendation systems in the same computational solution (Morisio, 2019).

Context

The sanological approach draws a sanological strategy that has as basic principles the promotion of a healthy lifestyle, as well as the production of changes. The sanological strategy starts with the assistance of the person to a sanological space where he can openly discuss and debate with the specialist. Self-assessment allows estimating a person's lifestyle according to human dimensions such as family, physical activity, nutrition, etc. In the second and third moments, a detailed interview and physical examination are carried out and a health state is proposed. At the last moment, the health route, there is a specialist-interlocutor negotiation process to establish objectives to be achieved in a comprehensible time range. Each of the moments relies on the use of various instruments in the form of questionnaires to collect information (Amable Ambrós, 2012) (Aldereguía Henriques, 1993).

The application of the sanological strategy allowed us to determine that there was insufficient use of the potentialities provided by the heterogeneity of data and the ubiquity of applications by existing computer solutions. To respond to the needs for the sanological approach, a mechanism for representing, capturing and preserving the data obtained from the application of instruments must be offered. As well as allowing the automatic generation of questionnaires, offering the possibility of favoring the in-depth interviews to design the health routes together with the interlocutors and to be able to resort to them regularly to verify the degree of compliance with the goals outlined and evaluate the changes in behavior, as well as future actions.

Distributed transactional solution

It is common to think of the relational model as the first alternative due to its popularity in different contexts. However, when the data is complex and unstructured, its management and recovery from relational databases are very complex. The questionnaires are made up of a variable number of questions, which offer a set of response options. Therefore, for a user to answer what is asked in a questionnaire, it is necessary to recover the structure of the database -disseminated in several tables- for this it is necessary to gather data from multiple tables, a very expensive operation in relational data. After a thorough study of the NoSQL databases, it was determined that the use of document-oriented databases is the most appropriate option for representing the data in the current solution since they make it possible to efficiently model variable structures.

The system aims to guarantee access to the different functionalities of the system through two fundamental channels: web clients and mobile applications. The first variant allows interaction with a web application that would make available to specialists the creation of questionnaires with possible associated evaluation rules and ensures that the interlocutors complete questionnaires adequately and keep the respective results to be consulted later. The mobile application, which would communicate with the data servers without mediating in this process the web application, would be in function of a primary objective: to enhance the individual participation of the experts and the interlocutors in the sanological spaces, without forgetting the possibility of exchange between them. This variant should satisfy the need for highly available applications, and for this purpose, local copies of the necessary data should be available so that the applications can work independently of whether the user is connected or not to the network. Figure 1 shows the general architecture of the distributed transactional solution, where **SanoLive** and **SanoLite,** which reflect the two dimensions of conception are intentionally highlighted.

The process of application of the instruments will generate a huge amount of information, this information must be analyzed and processed by the sanologist. The analysis of this data volume manually can be very complex and even results in poor use of stored information, which is necessary to devise an analytical model that allows transforming the information into knowledge that can be used by the experts in the function of the decision making.

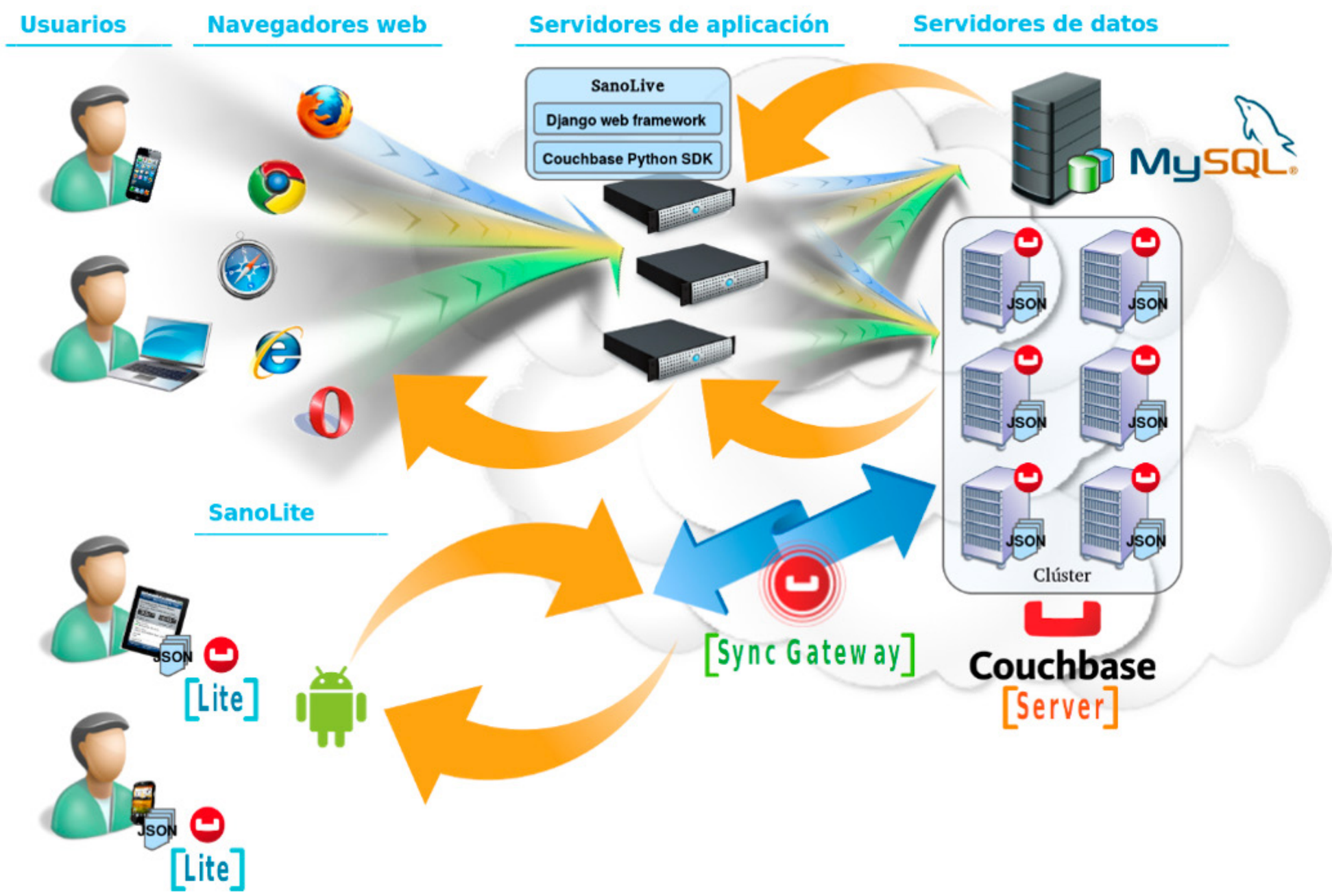

Figure 1. General architecture of the distributed transactional solution.

Hybrid Recommender System

This analytical solution must allow the analysis of the completed instruments and provide personalized recommended actions that contribute to shaping the health route in correspondence with the characteristics of each interlocutor. It was conceived the development of a hybrid recommender system where items are the actions and users, interlocutors. Figure 2 shows the general architecture that corresponds to the analytical solution referring to the recommendation system. For each action, the system must predict a rating that represents its importance for a particular user. The hybrid recommender system takes advantage of the completed instruments, the items description and the evaluation of the items by interlocutors and sanologists. Each of the inputs is used by a recommendation technique, which will be explained in the next paragraphs. In this regard, the hybrid recommendation system combined three filtering techniques: knowledge-based, content-based and collaborative filtering.

The knowledge-based filtering technique, which makes recommendations based on the needs and interests of the user, is based on the explicit features of actions. In this case, the direct participation of the sanologist as an expert is required. To build the user profile of an interlocutor, from the completed instruments the system learns the essential features of the interlocutors. The relationship between an interlocutor and an action is determined by the application of a similarity measure. Once the user profile and the action's profile are built, it is possible to determine the recommendations taking into consideration the similarity measure. The solution incorporates the knowledge-based technique, that way information can be

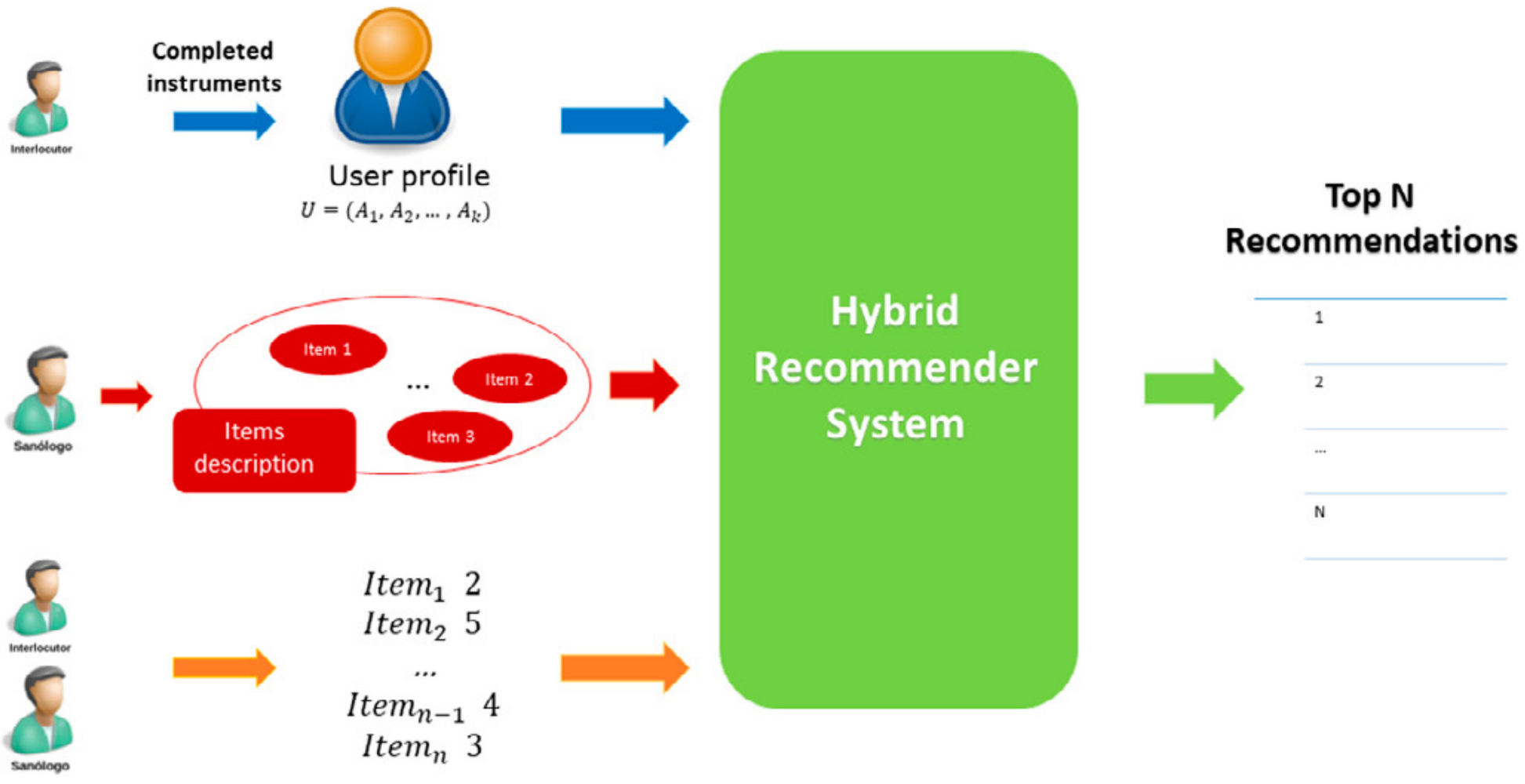

Figure 2. General architecture of the hybrid recommender system.

extracted directly from the answers of the interlocutors to the questionnaires. In addition, it solves the cold-start problem (Aggarwal, 2016) because once a speaker completes an instrument, a set of recommendations is already obtained.

The generation of recommendations based on the choices that the user made in the past is covered by the content-based techniques. This way the system exploits the evaluations issued by the sanologist and the interlocutor in a sanological space, which expresses the change in behavior and the improvement experienced or not by the interlocutor. Based on the set of evaluated actions, it is possible to predict the relevance of the actions not yet evaluated. This technique has been included in the conception of the analytical solution in order to guarantee the learning of the interests and preferences of a user for certain actions, a knowledge that cannot be inferred from the completion of the instruments. Evaluations are also exploited by the collaborative filtering technique, which makes recommendations to each interlocutor based on the information of users who most resemble him. The collaborative approach helps us to identify similar user groups that share lifestyle characteristics to suggest actions to an interlocutor, considering those that were effective for similar users. In this last technique, each user will be identified by the evaluations of the actions contemplated in their health route. Additionally, the inclusion of this type of technique allows discovering new preferences, which brings a certain degree of serendipity and novelty in the recommendations.

The system designed takes advantage of each of the techniques described and builds a set of hybrid recommendations by combining the particular results. For hybridization, the weighted method has been selected not only because it is easy to implement but also because, under certain criteria, it is possible to define specific weights that allow establishing priorities or relevance among the results. In this case, a relevance criterion is used because it can be foreseen that the precision of one technique will be greater than that of another. Next, each of the techniques employed is explained independently and, then, the hybridization strategy is discussed as the basis for the delivery of the recommendations to the users.

Knowledge-based filtering technique

The knowledge-based filtering technique learns from the user's interests and is based on the items (actions) features. This current technique uses the information given by the interlocutors in the filled instruments to build a user profile. The user profile is built based on the score reached in each of the healthy-action areas constituting a features vector. Actions are represented as a vector of attributes as well. To calculate the relationship between users and items, the attributes of the user profile are matched with the attributes of the items, from which it is possible to apply a similarity metric between an interlocutor and an action, in this case, the Euclidean distance. The user's profile depends on their answers to the instruments, which may be open or selection responses. In the case of selection responses, the evaluation rules are well defined and respond to a particular sanological area. In the case of open questions, an unsupervised learning algorithm is used to perform Natural Language Processing (NLP) based on word similarity.

Content-based filtering technique

The content-based filtering technique makes recommendations according to the user's history, which preserves their preferences and interests. The value of the actions that have not been evaluated explicitly is estimated, based on the actions that the interlocutor and the sanologist have evaluated in advance. The prediction can be achieved with the application of machine learning algorithms. In this case, it is possible to use a supervised algorithm using the rated actions as training data. Support Vector Machine (SVM[1]) with polynomial kernel was the chosen algorithm due to its achieved accuracy in classification problems in free-domain contexts. However, not having a large set of initial training has not been feasible to verify the SVM performance in the sanological context.

Collaborative filtering technique

The collaborative filtering technique includes recommendations taking into account actions that were effective for similar users considering the expert's evaluation. Similarity can be defined as the numerical representation of the similarity degree between two interlocutors, based on evaluations of common actions. Interlocutors clustering by affinity can be achieved through the use of unsupervised learning methods. The k- Nearest Neighbors (kNN) algorithm was used and the Pearson Correlation was selected as the similarity measure.

For the model to deliver hybrid recommendations, an integration strategy is applied, making use of the weighted method. Each of the techniques exposed predicts the relevance of action as a recommendation and the final rating is the result of a linear combination. Once the data is processed, the actions are shown to the user in decreasing order.

1 https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47

RESULTS

The service web client **SanoLive** and the mobile application **SanoLite** were implemented. Figure 3 shows the solution architecture through a deployment diagram in which the main technologies and components are identified for its implementation. The process of implementing the prototype has been oriented to the development of a web application that involves several functionalities, from the creation of the instruments and sanological spaces to the online interaction of the specialists and the interlocutors. ~~*Also*~~*In addition,* it has been introduced elemental statistical analysis and the hybrid recommender engine that uses the data collected from the sanological spaces. Furthermore, the implementation involved the development of a mobile application that favors the individual participation of the sanologists and the interlocutors at the time and place that is convenient for them, including the eventual exchange of opinions between them.

An assessment of several DBMS NoSQL was made and finally Couchbase Server[2] was chosen to store and process JSON[3] data. The selected web development platform for the web client **SanoLive** was Django. The mobile application -called **SanoLite**- is a native application for the Android operating system (Guillot Jiménez, 2014). As programming language for data analysis, Python was selected, since it is a general-purpose language and has an accessible learning curve. Python libraries were used for the analysis such as Scikit-learn[4], a Python open source library for machine learning and Natural Language ToolKit (NLTK[5]) for natural language processing. To complement this tool was used WordNet, for its wide use in linguistic computing and natural language processing.

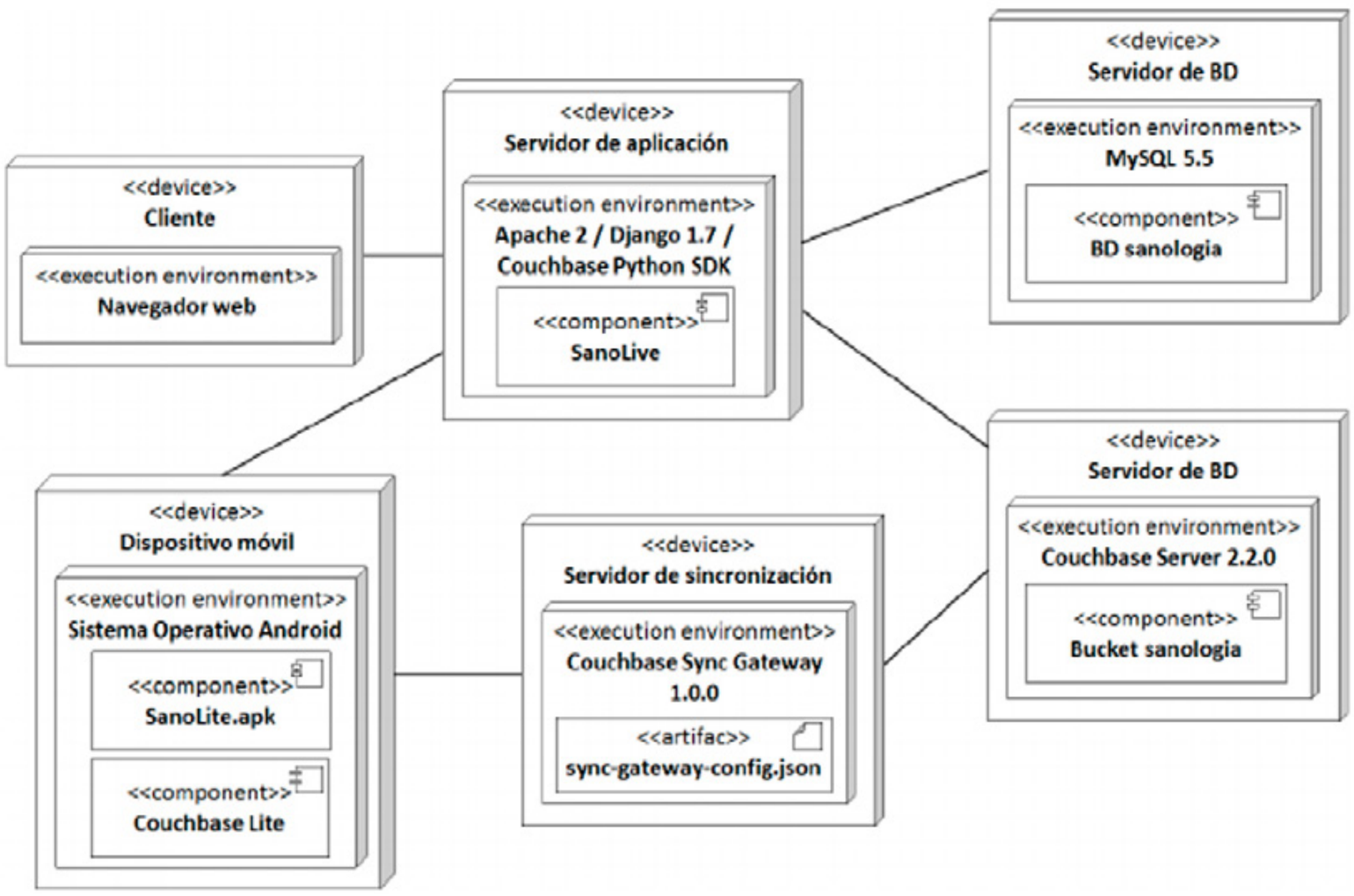

Figure 3. Deployment diagram of the computational solution.

[2] http://www.couchbase.com/

[3] https://www.json.org/

[4] http://www.scikit-learn.org

[5] http://www.nltk.org

To verify the validity of the proposed computational solution model, a prototype was implemented as a first approximation, on whose platform a set of experiments was conceived and executed. The presented experiment is oriented to validate the creation and/or modification of the documents in the local database, as well as the pertinent synchronization with the remote database and the recommended actions given by the recommender system. The processes of creating a health route, delivering and evaluating recommendations, adding actions to a health route and evaluating recommendations can be observed in Figure 4.

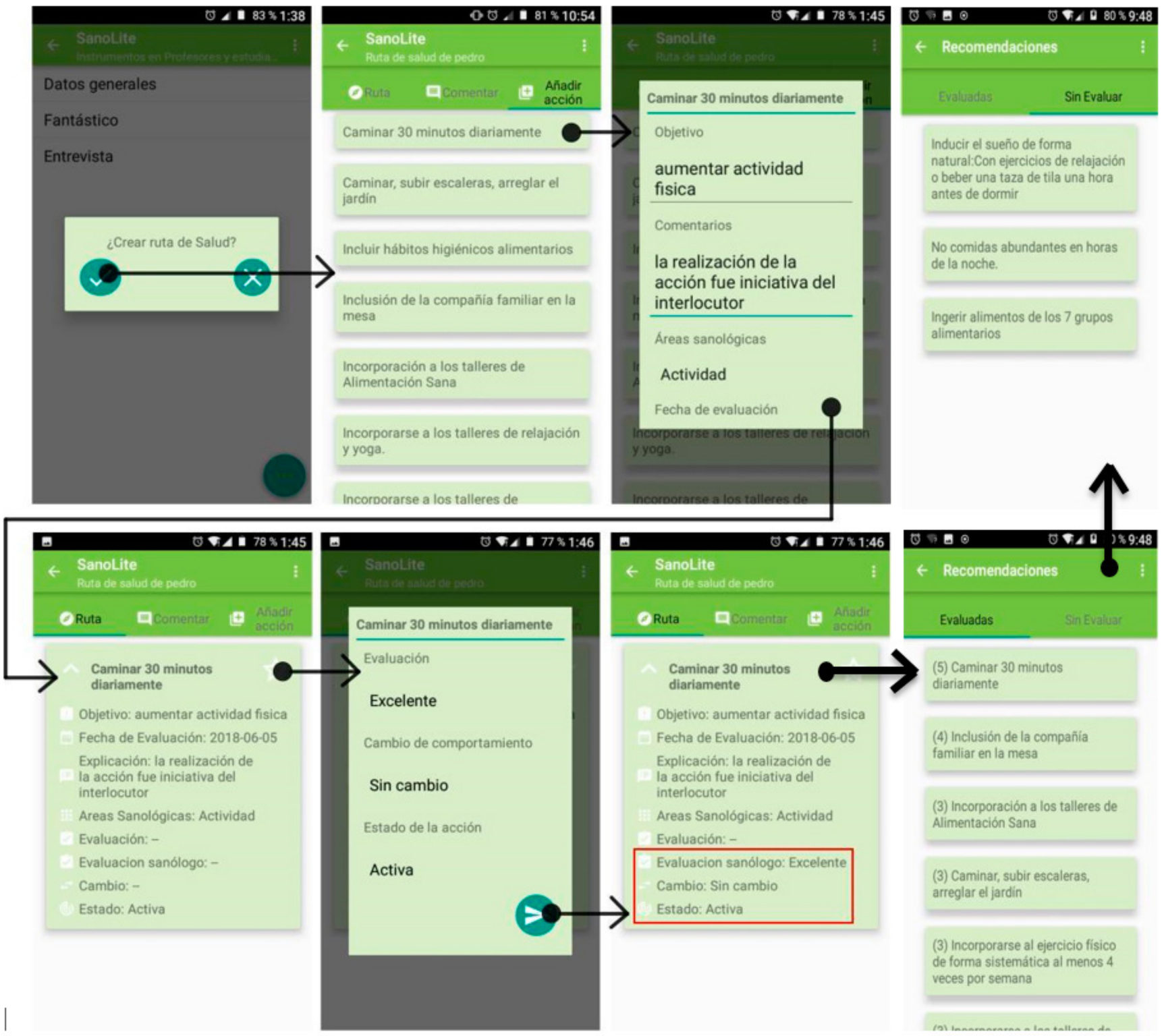

Figure 4. Creation and management of a health route

As a result of this paper, the creation of a mathematical-computational solution that responds to the analytical needs of the experts in Sanologia was achieved, in which statistical techniques and information recommendations are integrated. The solution is characterized by the enrichment of health routes through the suggestion of relevant actions. In the development of the recommender system, the transformation of the information contained in the sanological spaces into knowledge is evidenced for its exploitation based on ~~*decision making*~~*decision-making,* supporting expert decisions by analyzing the interlocutor development in the different sanological areas and delivering personalized recommendations according to his profile.

From the experiment's results, it is possible to affirm that the construction of a prototype was achieved that responds to the priority requirements of the sanological strategy and is in correspondence with the model of the proposed transactional and analytical solutions.

CONCLUSIONS

In the present paper, the creation of a mathematical-computational solution for the management and application of the sanological strategy was achieved. The solution is characterized by favoring the access of users from diverse locations and at any time from the use of a distributed architecture and flexible data models.

From the deepening in the areas of knowledge associated with contemporary information systems, the approach to the sanological context and the study of the distributed transactional solution, a model was designed for the generation of recommendations in sanological spaces. The general design of a computational system based on the abstract model was achieved and a prototype was designed and implemented on the basis of which a set of experiments was applied to establish the validity of the global conception.

The creation of a mathematical-computational solution for distributed storage, the use of mobile technology, the analysis of heterogeneous data, the hybridization of filtering techniques in the recommendation engine and the natural language processing constitute contributions regarding to the computational tools implemented previously in the sanological field.

ACKNOWLEDGMENTS

This research was supported by the University of Havana. We thank our colleagues from the Information Systems and Business Intelligence Group from the university (ISBIG-UH) who provided insight and expertise that greatly assisted the research

REFERENCES

Aggarwal, C. C. (2016). *Recommender Systems: The Textbook.* IBM T.J. Watson Research Center: Springer. DOI:10.1007/978-3-319-29659-3

Aldereguía Henriques, J. C. (1993). *Salud y Sanología en Médicas de Familia.* INTERCIENCIA.

Amable Ambrós, Z. (2012). *Sanología Nueva forma de pensar y actuar en salud.* Universidad Nacional Autónoma de México: ISBN: 978-807-02-3634-1.

Francesco Ricci, L. R. (2015). *Recommender Systems Handbook* (Second Edition ed.). New York: Springer Science+Business Media . doi:10.1007/978-1-4899-7637-6

Guillot Jiménez, J. (2014). *Solución computacional transaccional para la promoción de la salud y el bienestar humanos.* Universidad de La Habana: Tesis de Maestría en Ciencias de la Computación.

Jure Leskovec, A. R. (2014). *Mining of Massive Datasets.* Stanford University Course.

Morisio, E. C. (2019). Hybrid Recommender Systems: A Systematic Literature Review. *CoRR,* (p. abs/1901.03888).

Nikzad–Khasmakhi, N. a.–D. (2019). The state-of-the-art in expert recommendation systems. *Engineering Applications of Artificial Intelligence.* Elsevier Ltd.

Pokorný, J. (2013). *NoSQL databases: a step to database scalability in web environment.* International Journal of Web Information Systems.

Quintana-Wong, C. (2017). *Solución Analítica para la Promoción de la Salud y el Bienestar Humanos.* Universidad de La Habana: Tesis de Licenciatura en Ciencia de la Computación.

Rabanillo Echaniz, A. (2018). *Reingienería de las Soluciones Transaccional y Analítica para la Promoción de la Salud y el Bienestar Humanos.* Universidad de La Habana: Tesis de Licenciatura en Ciencia de la Computación.

Sullivan, D. (2015). *NoSQL for Mere Mortals.* Addison Wesley.