Abra iTour: A Semantic Web Recommender Using Hybrid Algorithm

Arpee M. Callejo, Amando P. Singun

ABSTRACT--- Information on the web is increasing at an exponential rate. This has resulted in a myriad of choices that are available for users on the web giving complex processes of the world’s largest database, the Internet. This has also given birth to the development of data filtering algorithms used in recommender technologies that help users find their best decision from the large unstructured database of the World Wide Web (WWW). Recommender systems have been widely used in e-commerce websites like Lazada, Amazon and other popular websites like YouTube, Spotify, LinkedIn, Facebook, and Instagram. Thus, the researchers came up with a study on the development of a recommender system.

This study is a research and development of a recommender system for the province of Abra, Philippines, titled “Abra iTour: A Semantic Web Recommender Using Hybrid Algorithm”. The system adopted a hybrid algorithm, a combination of Collaborative and Content-based filtering algorithms to extract data for the recommendation lists that are offered to the users of the system.

IT experts assessed the extent on the level of efficiency of the hybrid algorithm with a 4.59 rating described as Very Great Extent. Thus, the hybrid algorithm used by the system is proven. The recommender system ISO 25010 Software Quality Standards evaluation acquired an overall weighted mean of 4.52, Very Great Extent. This implies that the recommender system is ready for deployment and implementation.

Index Terms — hybrid filtering algorithm, machine-learning, recommender systems, web mining.

1. INTRODUCTION

The Internet is a dynamic technology that has changed the way we interact and exchange information on the web. Information on the web has been exponentially increasing because of the emergence of social media platforms and weblogs. Internet users adopted the use of the idea of folksonomy in which users of the Internet associate their posts through keywords and hashtags, one way to organize the data over the Internet. As the amount of information becomes greater than ever, experts and researchers are also finding ways to address the complexities of the unstructured database on the Internet for personalization and customization. Data extraction and filtering have been initially studied for e-commerce technologies.

Other than e-commerce, these technologies can also be applied to create a great impact on travel and tourism since most people are dependent on the use of the Internet and Google Maps on their searches of travel. As Google Trends reported that travel and tourism rose by 19%, millennials today often query “Where to go?” “How to travel alone?” and “What sites to see?” in their Google searches. Tourism is among those fastest growing industries of the market providing sources of employment and exports.

Statistica (2018) shows that the impact of the tourism industry on the global economy contributed to approximately $ 2.31 trillion dollars in 2016.

It was supported by the World Travel and Tourism Council (2011) that they expect tourism to contribute to the global economy yielding to 9% of the global overall market value of goods and services.

Murati (2013) stated that the ability of the national economy to benefit from tourism depends on the availability of investment to development the necessary infrastructure and to supply the needs of tourists.

Thus, tourism really is an important sector in the growth of the economic status of a country as mentioned from those works of literature presented.

In Asia, beautiful destinations are abundant with its many beaches, mountains and other natural resources. Business in Asia (2018) stated that major destinations in Asia and its percentage growth in 2014 are: (a) Myanmar posted 51% growth (b) Japan posted 29% growth (c) South Korea 17% growth (d) India posted 11% growth (e) Hong Kong posted 8% growth (f) Malaysia, Indonesia and Cambodia also posted solid growth, each posted at 7% growth figure (g) Vietnam saw a 4% growth.

According to the aforementioned destinations, it has been noticed that the Philippines is not listed among the major destinations in Asia. Thus, experts and governance in tourism should empower promotions of tourism in the country to improve the economic status, to augment exports, and to provide employment for the Filipinos of the country.

The Philippines is known for its natural beauty with its richness in resources composing of rivers, lakes, mountains and other natural habitats which make it considers as a tourist destination. The richness of the different cultures strengthens tourism in the country made it more attractive to tourists. The Philippines is composed of 7, 107 islands combined by geographical divisions.

Luzon, Visayas, and Mindanao are the main islands of the Philippines. Luzon is the largest island among the two (2) islands in the Philippines where one of the Seven (7) Wonder Cities in the world which is Vigan City is situated.

With the research locale, Abra is a province geographically located in the northern archipelago of Luzon particularly on the western side of the Cordillera, surrounded by Ilocos Norte and Apayao on the North,
Mountain Province and Ilocos Sur on the South, Kalinga on the East and Ilocos Sur on the West. Abra is often skipped by tourists because of its location which is very far from Manila and the province is also known as a political hotspot. But there is more to offer when you explore the province. There are many tourist attractions which Abrenians can be proud of that are still unknown.

According to the Department of Interior and Local Government (2018), Abra has fifty (52) tourist attractions in record from the Department of Trade and Industry, both explored and unexplored.

Hence, the local government of Abra, as well as the Department of Tourism - Provincial Office, shall seek for promotion for the tourism industry to enhance the tourism industry in the province as well as to increase travel satisfaction of tourists and travelers, appropriate technologies should be maximized for this purpose.

The development of a website, integrated with a recommendation system allows people to browse the beautiful and attractive tourist attractions, events and festivities that may result in the promotion of the tourism industry to leverage the economic status or the Gross Domestic Product (GDP), employment and exports of the province and of the country-at-large.

Recommender systems are widely used in the field of E-commerce. Netflix and other social media like Facebook and Twitter make use of these technologies for efficiency and effectiveness of the website.

According to Aggarwal (2017), recommender systems provide personalized recommendations of products or services among users based on their previous searches or purchases.

A study on recommender systems stated that recommender systems are practically a necessity for keeping a site's content current, useful, and interesting to visitors (Falk, 2018).

Likewise, Negre (2015) stated that a recommender system aims to guide users in their exploration of the large quantities of data available.

The above-stated definitions served as a guide in the development of the system. The recommender system is appropriate for the recommender system for the tourism industry in Abra because it can be a venue for giving suggestions and recommendations of tourist attractions as well as other establishments for the tourists and travelers in the province.

Thus, there is a need to develop an “Abra iTour: A Semantic Web Recommender Using Hybrid Algorithm” for the province to support the tourism in Abra to enable the government entity to put Abra on the tourist map of the best tourist destinations in the country in order to keep it prominent worldwide.

**Conceptual Framework**

The Input-Process-Output (IPO) Model was utilized to realize the needed data, tools and methodology to be used for process the data and most especially the output of the study.

Figure 1 shows the research paradigm of the study. The researchers adopted the Input-Process-Output (IPO) model to represent the framework of the study. The researchers used the Input-Process-Output which captured the logical view and concepts of the data used.

The concerns, issues, and problems in the province as regards to tourism aspects, data sets on tourist attractions, transportation services, trip schedules, hotel accommodations, and pasalubong centers were considered as inputs in the study.

On the other side, the process used in the study includes the Rapid Application Development (RAD). The Rapid Application Development methodology helped the researchers expedite analysis, design and implementation phases and these also helped the researchers improved the speed and quality of the system. Another is the integration of the hybrid algorithm which is the combination of the two (2) algorithms was integrated into the User Design and Construction phase of the Rapid Application Development.

It made use of the cosine similarity for Collaborative filtering algorithm and dice coefficient for the Content-based filtering algorithm. Cosine similarity was implemented in the system and computes the user-based similarity to get the rated and reviewed items by an active user to get the recommendations of attractions and establishments. Assessment using ISO 25010 Software Quality Standards was also used to test the quality of the developed system.

In the Output part of the conceptual framework is the study, the “Abra iTour: A Recommender System Using Hybrid Algorithm”.

The Input-Process-Output (IPO) Model was utilized in the realization of the needed data and tools and methodology to be used and most especially the output of the study which is the development of the Abra iTour, A Semantic Web Recommender.

**Statement of the Problem**

The study aimed to develop, design, and implement an Abra iTour: A Recommender System Using Hybrid Algorithm. Specifically, it sought to answer the following research questions:
1. What are the problems/issues and challenges encountered by tourists and travelers that are related to tourism in the province of Abra?
2. What algorithm can be proposed to address the identified problems found in the existing system of the tourism industry in Abra?
3. What proposed system can be developed to address the identified problems and issues encountered in the existing system of the tourism industry in Abra?
4. What is the level of efficiency of the hybrid algorithm in terms of the following criteria: a) Quality of Recommended Items, b) Interaction Adequacy, c) Interface Adequacy, d) Perceived Ease of Use, e) Perceived Usefulness, f) Control/Transparency, g) Attitudes and h) Behavioral Intention.
5. What is the extent of compliance of the developed application to ISO 25010 Software Quality Standards in terms of the a) Functional Sustainability, b) Performance Efficiency, c) Compatibility, d) Usability, e) Reliability, f) Security, g) Maintainability, and h) Portability.
6. Is there a significant difference in the extent of compliance of the developed application to ISO 25010 Software Quality standards as assessed by the IT Expert and Users in terms of a) Functional Sustainability, b) Performance Efficiency, c) Compatibility d) Usability, e) Reliability, f) Security, g) Maintainability, and h) Portability?
7. What enhancement can be done to improve the proposed system?

A. Research Design
The research design that was used in the study is a descriptive research design. The researchers also used the Rapid Application Development (RAD) methodology which is under the Agile Methodology. Rapid Application Development focuses on gathering customer requirements through workshops or focus groups, early testing of the prototypes by the customer using iterative concept, reuse of the existing prototypes (components), continuous integration and rapid delivery.

Software Development Methodology
Rapid Application Development (RAD) was used because it is the most appropriate one (Arruejo, 2013) and could develop faster through more expedient processes because of its iterative feature which enabled the researchers to develop the system while it is being evaluated until it becomes almost perfect. RAD has continuous interaction with the stakeholders of the system. It has four (4) phases, namely: Requirements Planning, User Design, Construction, and Implementation or the Cutover.

The researchers used front end and back end frameworks for the development of the system. Node.js Express Framework, Hypertext MarkUp Language (HTML), Cascading Style Sheet (CSS) and MySQL were used for the development. Other software utilized were Adobe Photoshop and Sublime Text for the Abra iTour Semantic Web Recommender System.

B. Algorithms of the Abra iTour Semantic Web Recommender

The researchers used a filtering and data extraction for the development of the system. The algorithms which were used are content-based and collaborative filtering algorithms or in other words, when combined, a hybrid algorithm.

![Fig. 2. The Hybrid Algorithm](image)

Collaborative Filtering Algorithm

Users of the system have the ability to input ratings on the tourist attractions and other establishments which then be used for recommendations to other users. Through this algorithm, it finds users from the database of the system who shared the same appreciation.

The implementation of the CF algorithm in the system is based on user-based recommender systems. This has required the following two (2) steps: 1) Calculating user similarities and 2) Predicting the targeted items for recommendations and ratings of a user to items.

The goal of the CF algorithm is to suggest new items to users. In order to do that, a utility matrix for items and user’s ratings is tabulated to determine the distance of similarity among users stored in the database.

The table below shows the utility matrix for the calculation of spots similarities with user ratings.

| Table 1. Utility Matrix of Spots and Users |

The researchers also used the cosine similarity for the computation of the distance between items to get the most similar items. The higher cosine similarity or the users having a greater distance among the items have the most similarity. Therefore, spots in the database with the greater distance from other spots which are determined through the ratings shall be recommended to the active user. Formally, cosine similarity is expressed through this formula:

\[
\text{cos}(A,B) = \frac{AB}{||A|| ||B||} = \frac{\Sigma (A_i B_i)}{\sqrt{\Sigma (A_i^2)} \sqrt{\Sigma (B_i^2)}}
\]

Cosine Similarity is equal to the dot product of A and B divided by the magnitude of A and B, and is equal to the summation of A and B ratings on an item i divided by the square root of the summation of A squared and the square root of B squared. Manually, to compute the similarity we have to see the most similar items in the utility matrix. In computing the cosine similarity, first, multiply both ratings of the active user to the most similar items, then add the quotient of ratings of the active user and most similar user. Next is to compute the square root of ratings of both active user and a similar user then add. Lastly, perform the sum of the first step divided into the product of the square root of
ratings of both active and most similar user. After computation, we have to see the greatest value but not equal to 0 and 1 values of the distance between users. In cosine rule having a greater angle has a smaller similarity.

**Implementation of the CF Algorithm**

The goal of the CF algorithm is basically to generate personalized recommendations for an active user from the most similar user. It was implemented on the database and the codes in JavaScript.

The database on ratings table is where we get the actual data for the computation of the similarities and prediction of recommended results. Another is the JavaScript codes integrated into the recommender system which creates a Utility Matrix for the recommender.

```javascript
let iUsers = [];
for (i = 0; i < rows.length; i++) {
    iUsers.push(rows[i].rating_value);
}
```

From the codes, the rating_value from the ratings table is transposed to the weighted mean of the ratings of the active user. Another code is as follows.

```javascript
let otherUsers = {};
for (i = 0; i < ratings.length; i++) {
    othersUsers[ratings[i].user_no] = [];
    othersUsers[ratings[i].user_no].push(ratings[i].rating_value)
}
```

The above codes are used to collate other user ratings on an array that are registered in the recommender system. The most important codes are presented below for the computation of the similarity.

```javascript
for (var key in otherUsers) {
    for (otherUsers.hasOwnProperty(key)) {
        if (key == ratings[i].user_no) {
            othersUsers[ratings[i].user_no] .push(ratings[i].rating_value)
        }
    }
}
```

The above codes perform multiplication then addition of each rating’s value of the active user to the other array of other user’s ratings. Another codes for the square root of the cosine similarity as coded from the formula is presented as follows.

```
var magA = Math.sqrt(dot(a, a))
var magB = Math.sqrt(dot(a, b))
if (magA && magB) return dot(a, b) / (magA *
else return false
```

Then the above codes perform the square root of both arrays. Lastly, (dot (a, b), which has a SUM value equates to the formula SUM divided to the product of (magA*magB) or the square roots of arrays of an active user and other users ratings.

This means that the active user whose User ID is 6 has the closest similarity to the greatest value not equal to 1 which is User 11 among the (kNN) k-Nearest Neighbors, which has a distance of .8111181311491903. Therefore, spots rated by User 11 are recommended to the active user who is User 6. In the database of the system User 6 with a user_id 6 has a user name of John Doe and User 11 with a user_id has a user name of Arpee Callejo.

**Table 2. Utility Matrix of Users 6 and User11**

From the Utility Matrix from the table ratings of the database, User 11 has rated spot3 with a name of Kapalkan Falls, spot15- Bandi River, spot20- Sapilang Twin Falls, spot28-Sinublan Falls, spot29- Tubong Falls, and spot51-Mount Bullagao. Therefore, these items are recommended to User 6 whose user name is John Doe.

From the table, User 6 who has the closest similarity to User 11 is recommended with items liked by User 11 which are spot3 with a name of Kapalkan Falls, spot15- Bandi River, spot20- Sapilang Twin Falls, spot28-Sinublan Falls, spot29- Tubong Falls, and spot51- Mount Bullagao.

Hence, the proposed system offers a personalized recommendation engine, because it doesn’t only look for the contents of the existing profile of each tourist attractions and establishments of the system, but also to the similarity of the users for it generates recommendation lists for the other users of the recommender system.

Another algorithm is the Content-Based Algorithm.

**Content-based Filtering Algorithm**

The researchers used another algorithm to meet the limitations of the other algorithm, a content-based algorithm. It works with existing profiles, descriptions, meta tags and keywords of the tourist attractions and establishments.


The researchers used the concept on the similarity of an unseen item with the user profile based on the keyword overlap particularly using the Dice coefficient. Formally, the formula is written below.

\[ R = \frac{2|keywords(Ai) \cap keywords(Aj)|}{|keywords(Ai)| + |keywords(Aj)|} \]  

(2)
From the formula, R is equal to the keywords of an item of the user searched and interacted describe the tourist attractions and establishments in the database with a set of keywords. For example, user 9 like spot_no 51 with a spot_name Mount Bullagao with a set of keywords of mountains, trekking, hiking, mountain climbing, and biking. This means that users who like Mt. Bullagao will likely rate or review items with keywords camping, forest, and biking. Thus, items with keywords camping, forest, and biking, will be recommended also to user 51.

**Implementation of the CBF Algorithm**

The database of the system included keywords which are used in the query of each item for the CBF algorithm.

The following Dice coefficient is converted into codes to query similar items with the same keywords from the database for the section “You Might Also Like” section of the recommender system for the content-based filtering algorithm.

```
Exports.userRecommendation = (spot_no) => {
  return 'SELECT spots.spot_no, spots.spot_name, k_keyword,
  spots_category.sc_name, spots_photo.imag_filename,
  round(SUM(rating_value)/COUNT(8),2) as RATES
  FROM ratings
  INNER JOIN spots ON ratings.spot_no = spots.spot_no
  INNER JOIN users ON ratings.user_no = users.user_no
  INNER JOIN spots_category ON spots_category.sc_no = spots.sc_no
  INNER JOIN spots_photo ON spots_photo.spot_no = spots.spot_no
  INNER JOIN keywords ON spots.spot_no = keywords.k_spot_no
  WHERE NOT ratings.spot_no = '${spot_no}' AND NOT
  AND spots_photo.img_isprimary = 1
  AND k_keyword IN
  (SELECT k_keyword FROM keywords
  WHERE keywords.k_spot_no = '${spot_no}')
  GROUP BY ratings.spot_no
  LIMIT 0,6';
```

**Fig. 3. Similar to this Item**

Figure 3 represents the CBF algorithm of the recommender system where the spot Lobot Eco Park has a similar attributes with Sinublan Falls and Piwek Rock Formations where keywords of both attractions are water adventure, water, summer and others from the spots table of the database of the recommender system. This feature of the system is essential, because it recommends other spots and establishments in Abra, which has the same attributes to other attractions in the province. This is considered as a personalization because it recommends similar items users prefer looking or searching from the developed recommender system.

**C. Study Participants**

The study participants were consist of the stakeholders of the Department of Tourism-Provincial Office who were the one-hundred thirty-five (135) tourism staff, hotel owners and the prospect tourists or travelers in the province. Also, fifteen (15) IT experts were considered to validate the quality of the developed recommender system.

**D. Sampling Technique**

Convenience sampling was used in the study to select the participants of the study and to further get the perfect results of the statistics of the study (Singun, 2018). The researchers have identified one hundred twenty-five (100) travelers/tourists, fifteen (15) tourism office staff, fifteen (15) hotel owners, and twenty (20) IT experts for the ISO 25010 Software Quality Standards Questionnaire and fifteen (15) IT experts for the Efficiency of the Hybrid Algorithm Evaluation. The number of IT experts is justified by the study of Singun (2018).

**E. Data Gathering Tools**

The researchers made use of interview, library research for book references and internet research as its main gathering tools.

The researchers gathered data from the different stakeholders who were the target users of the system. The researchers interviewed the stakeholders of Department of Tourism and tourists and stakeholders of the tourism in the Province of Abra.

The researchers also used the internet resources, books, journals, and previous studies from international to local prospective and other materials that is essential in the familiarization of the process, which served as the secondary sources of data to determine the issues encountered with the manual system and to know the latest technologies used.

The researchers also gathered information from books, graduate study researches and from the internet for the review of related literature to serve as a guide for the study.

**F. Data Gathering Procedures**

The procedures used in the study include questionnaire, and interview of the stakeholders of the tourism industry to collect data and other necessary information for the conduct of the study. Before gathering and collecting data, the researchers requested for the proper authorization and permission from the Tourism Office of Abra. A letter of request approved by the Provincial Office of Abra was prepared and sent to the Provincial Tourism Office to formally ask permission to conduct the study.

The system followed a guide on the ISO 25010 to satisfy the characteristics needed by the stakeholders for the system. In the questionnaire characteristics includes the functional suitability, performance efficiency, usability, reliability, security, maintainability and portability of the recommender system.
The researchers used a user-centric evaluation for recommender systems questionnaire to analyze the level of efficiency of the algorithm. From this study, the questionnaire has been validated purposely for recommender systems. It was adapted from Pu, P. et.al. (2011).

G. Data Analysis

The data collected were tabulated, analyzed and interpreted using both descriptive and inferential statistics. Descriptive statistics, like percentage, mean and independent TTest were used.

II. RESULTS

Problems, Issues and Challenges of Tourism in the Province of Abra

The researchers found four (4) significant problems, issues and challenges of tourism specifically to tourists and travelers, namely: (1) politics, (2) transportation (3) location (4) promotions among others. These findings imply that there is really a need to promote the tourism industry through a recommender system. Through the use of the system, tourists can have the opportunity to experience the province as a place of beautiful attractions which when tourists found peace, security and hospitality of the people in Abra on their visit, tourists can eventually spread to the world that the place is safe and can change the bad notion on political issues about Abra.

It is consistent to the recent studies of Stetic (2014) and Ali (2014) that the rapid development and omnipresence of ICT have a significant impact on modern development of tourism and tourism market; hence, ICT is a critical catalyst for tourism innovation.

Hybrid Algorithm for the Semantic Web Recommender

The hybrid algorithm that was used in the development utilized the cosine similarity to get the distance between users in the CF algorithm and Dice Coefficient for the CBF algorithm. Both were implemented through the database and JavaScript codes for the recommender system. The implementation of the hybrid algorithm is used in the “Recommended for You” section for CF algorithm and “You Might Also Like” section for CBF algorithm. These findings imply that a computation on the cosine similarity is needed to find the most similar users based on reviews and content-based algorithm needed a dice coefficient computation using keywords and tags to get the similar items recommended to unregistered users.

Fig. 4. Abra iTour Semantic Web Recommender

The development of a recommender system addresses the identified problems of the tourism industry in Abra. The system “Abra iTour: A Semantic Web Recommender” guides users to tourist attractions, helps them decide on where to go and recommends users with of their preferences. Features of the Abra iTour Semantic Web Recommender include primary features such as (1) Recommender Agent Using CF and CBF Algorithm, (2) Geo-References, (3) Travel Guide, (4) User Reviews (5) Search Feature through Filter by Popularity, Recently Added and Featured. Additional features also includes (1) Security Mechanism- Captcha and Email Verification, (2) Notifications, (3) Accounts Management (4) Record Management, (5) Guest and Registered Homepage, (6) Services, and (7) Tour for Itinerary. This finding implies that a feature on the recommender part essentially addresses the problem of tourism because unseen tourist attractions and other establishments were being recommended to users even if they have unrated the item before.

Level of Efficiency of the Hybrid Algorithm

IT experts evaluated the hybrid algorithm level of efficiency as 4.59 with a descriptive result of “Very Great Extent”. This finding implies that the efficiency of the algorithm is proven.

Extent of compliance of the developed application to ISO 25010 Software Quality Standards

The recommender system has an overall weighted mean of 4.52 which equates to a “Very Great Extent” rating on ISO 25010. This finding implies that the system is ready for utilization and deployment.

Significant Difference in the Extent of Compliance of the Abra iTour to ISO 25010 Software Quality standards as assessed by the IT Expert and Users

All factors on the evaluation between users and IT experts on the extent of compliance to ISO 25010 Software Quality Standards were resulted to no significant difference therefore both users and IT experts who evaluated the recommender system were almost the same therefore the assessment is reliable.
Enhancement can be done to improve the proposed system

Enhancements of the developed system were given by tourists and travelers, and IT experts who evaluated the system. IT experts also gave suggestions for the improvement of the system. One said that it may be included also with video clips to add more attractiveness. Another expert suggested including reservation on some establishments. These suggestions were taken for future enhancement of the developed system. These findings imply that the system can be improved and enhanced in the future.

III. CONCLUSION

The existing recommender system has an overall weighted mean of 4.52, with a descriptive result of “Very Great Extent” on ISO 25010: Software Quality Standards and 4.59 with a descriptive result of a “Very Great Extent” on Recommender Systems Questionnaire (ResQue) on the hybrid algorithm level of efficiency. These mean that it is recommended for use of the local government of the province particularly Department of Tourism-Provincial Office. The recommender system has its limitation and same issue on “cold start problem” as stated by authors of the same study. Thus, this requires for further utilization of users. In order to fully appreciate the recommender engine, the system shall store databases on user reviews on tourist attractions and establishments.

A recommender is suitable for the promotion of tourism because a technique of the CBF and CF algorithm of providing unseen spots and establishments are recommended to the users of the system. Thus, this really contributes to the tourism promotion of the province.

The recommender system is fully compliant to the ISO 25010 Software Quality Standards and Recommender System Evaluation. The recommender system can be enhanced through user’s collaboration because the database has time-variance attribute in both CF and CBF algorithms which were both implemented on the “Recommended for You” and “You Might Also Like” parts of the system’s recommendation lists.

REFERENCES