2015

Vacation Package Recommender System

Anusha Allaparthi

Grand Valley State University

Vacation Package Recommender System

By
Anusha Allaparthi
April 2015

Vacation Package Recommender System

By
Anusha Allaparthi
April 2015

A project submitted in partial fulfillment of the requirements for the degree of

Master of Science in

Computer Information Systems

Grand Valley State University
April 2015

Yonglei Tao **Date**

Table of Contents

Abstract

Vacations play a vital role in our lives. Taking a good vacation can help our physical health; helps maintaining good family relations, improves mental health and reduces the chance of burnouts. However, most vacation recommender systems available are complicated and confusing and usually rely on explicit user ratings to recommend .travel packages. However, user ratings for travel data are sparse, therein reducing their effectiveness in recommending travel packages. I propose to develop a system aimed at exploiting a travel data set and creating travel package recommendations based on the user's interests and the spatial-temporal correlations that exist within sets of locations, seasons and attractions. Further, I will assess relationships between travel users so than common users can be arranged into travel groups or the people who wants to travel as a group with their family or friends can also be arranged into travel groups. This personalized vacation package recommendation based on the traditional models, which follow a recommendation strategy and has the ability to combine many possible constraints that exist in the real-world scenarios.

This data mining approach uses collaborative filtering method and performs much better than the traditional systems. It can be used both by the travel agencies and the travel groups at low maintenance and cost. The Graphical user interface is designed for both novice and expert users. This project has been developed using NetBeans with java and MySQL. I choose NetBeans because it is free, open-source, cross-platform IDE with built-in-support for Java programming language. This package system can be considered as an experimental prototype, we can see that the proposed recommendation approach works very well for predicting the user travel preferences by exploiting the unique characteristics of vacation package data.

Introduction

Recommender systems have been successfully applied to enhance the quality of service in a number of fields; it is natural choice to provide travel package recommendations. The first operative tourism recommender system was introduced by Delgado and Davidson. There are many technical and domain challenges inherent in designing and implementing an effective recommender system for personalized travel package recommendations.

First, travel data is much fewer and sparser than traditional items, such as movies for recommendation, because the costs for a travel are much more expensive than for watching a movie.

Second, every travel package consists of many landscapes (places of interest and attractions), and, thus, has intrinsic complex spatial temporal relationships. For example, a travel package could only include the landscapes which are geographically co-located together or could be developed for different travel seasons. Therefore, the landscapes in a travel package usually have spatial temporal autocorrelations.

Third, traditional recommender systems usually rely on user explicit ratings. However, for travel data, the user ratings are sparse, reducing the effectiveness of these systems.

Finally, traditional items for recommendation usually have a long period of stable value, while the values of travel packages can easily depreciate over time and a package usually only lasts for a certain period of time. Travel companies need to actively create new tour packages to replace the old ones based on the interests of the tourists.

To address these challenges, I will implement a different approach to personalized travel package recommendation. Specifically, I will first analyze the key characteristics of an existing travel data set. Along this line, travel time and travel destinations are divided into different seasons and areas and a tourist-area-season theme model is developed, which can represent travel packages and tourist's interests by different theme distributions. In this model, the extraction of themes is conditioned on both the tourists and intrinsic features such as locations and travel seasons of the landscapes. So, this model can well represent the content of the travel packages and the interests of the tourists. I will further enhance this model by considering some additional factors including the seasonal behaviors of tourists, the prices of travel packages, and the cold start problem of new packages leading to the tourist-relation-area-season theme model. , which helps understand the reasons why tourists form a travel group. This goes beyond personalized package recommendations and is helpful for capturing the latent relationships among the tourists in each travel group. Finally, I will showcase the model on a travel data set to show that this model can effectively capture the unique characteristics of travel and performs much better than traditional techniques.

A travel package is a general service package provided by a travel company for an individual or a group of tourists based on their travel preferences. A package consists of a combination of themes (area-season) and landscapes, which are the places of interest and attractions located in nearby areas the landscapes and some related information, such as the price, the travel period, and the transportation means. For this system, the users are the tourists and items are the existing packages, and we exploit a travel data set for building a recommender system.

Background and Approach

Recommender systems can be classified into two categories - Content-based filtering and Collaborative Filtering. Content-based filtering analyzes the association between user problems and the descriptions of items. To recommend new items to a user, the content-based filtering approach matches the new items descriptions to those items known to be of interest to the user. On the other hand, the collaborative filtering (CF) approach does not need content information to make recommendations. Collaborative Filtering has been developed and improved over the past decade to the point where a wide variety of algorithms exist for generating recommendations. Each algorithmic approach has adherents who claim it to be superior for some purpose. Clearly identifying the best algorithm for a given purpose has proven challenging, in part because researchers disagree on which attributes should be measured, and on which metrics should be used for each attribute. Researchers who survey the literature will find over a dozen quantitative metrics and additional qualitative evaluation techniques. I will make use of the Collaborative Filtering technique for creating recommendations for package definition wherein –

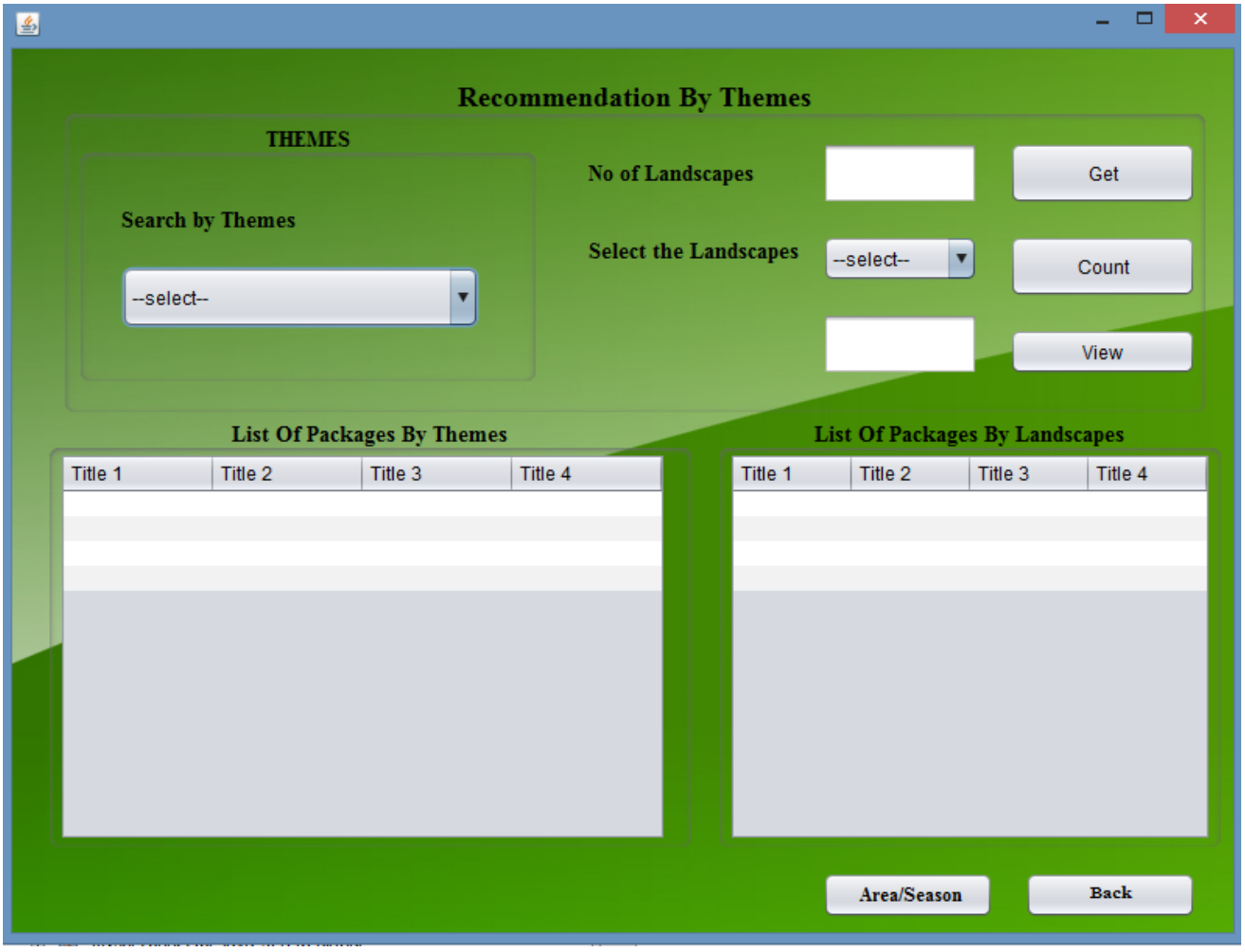
i. I capture the unique characteristics of travel data using the Tourist-Area-Season-Theme model and enhance it to create personalized travel packages based on additional factors including the seasonal behavior of the tourists
ii. I develop the personalized candidate package set for each tourist by the collaborative method.
iii. Extend the model to include the relationships between tourists which is helpful for capturing the latent relationships among the tourists in each travel group and to understand why tourists form groups.

The System

The travel package recommender system will allow the user or administrator to search for packages based on various criteria. The options that the user can search by are:

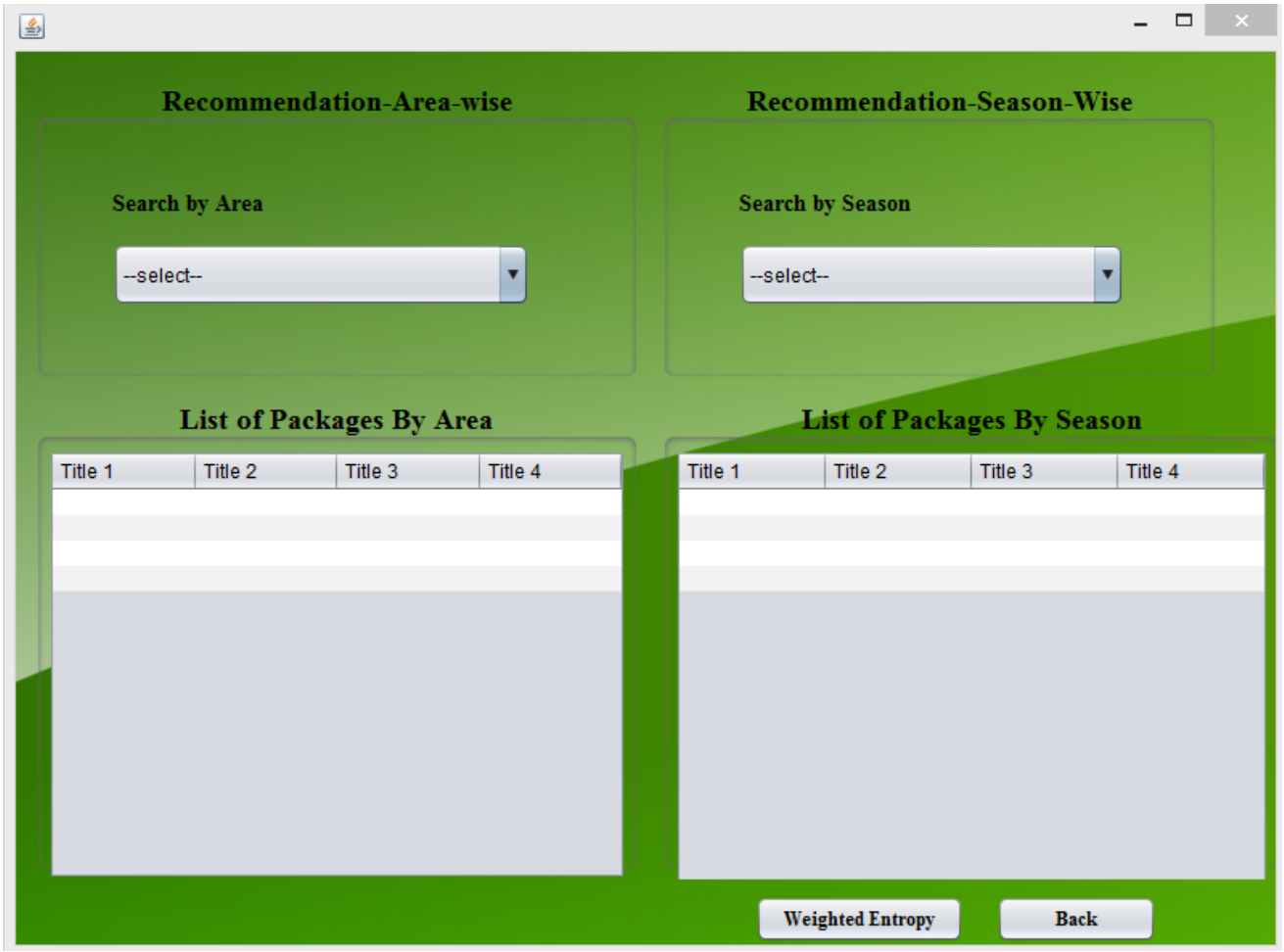
a. Recommendation by Themes

This option does not consider travel area and travel season factors and recommends packages based on the theme selected. Results are retrieved from a travel data set through a simple selection by Theme.



b. Recommendation by Area

This option provides package recommendations based on user-entered Area and also provides results for the Area by Season. Results are retrieved from a travel data set through a simple selection by Area.

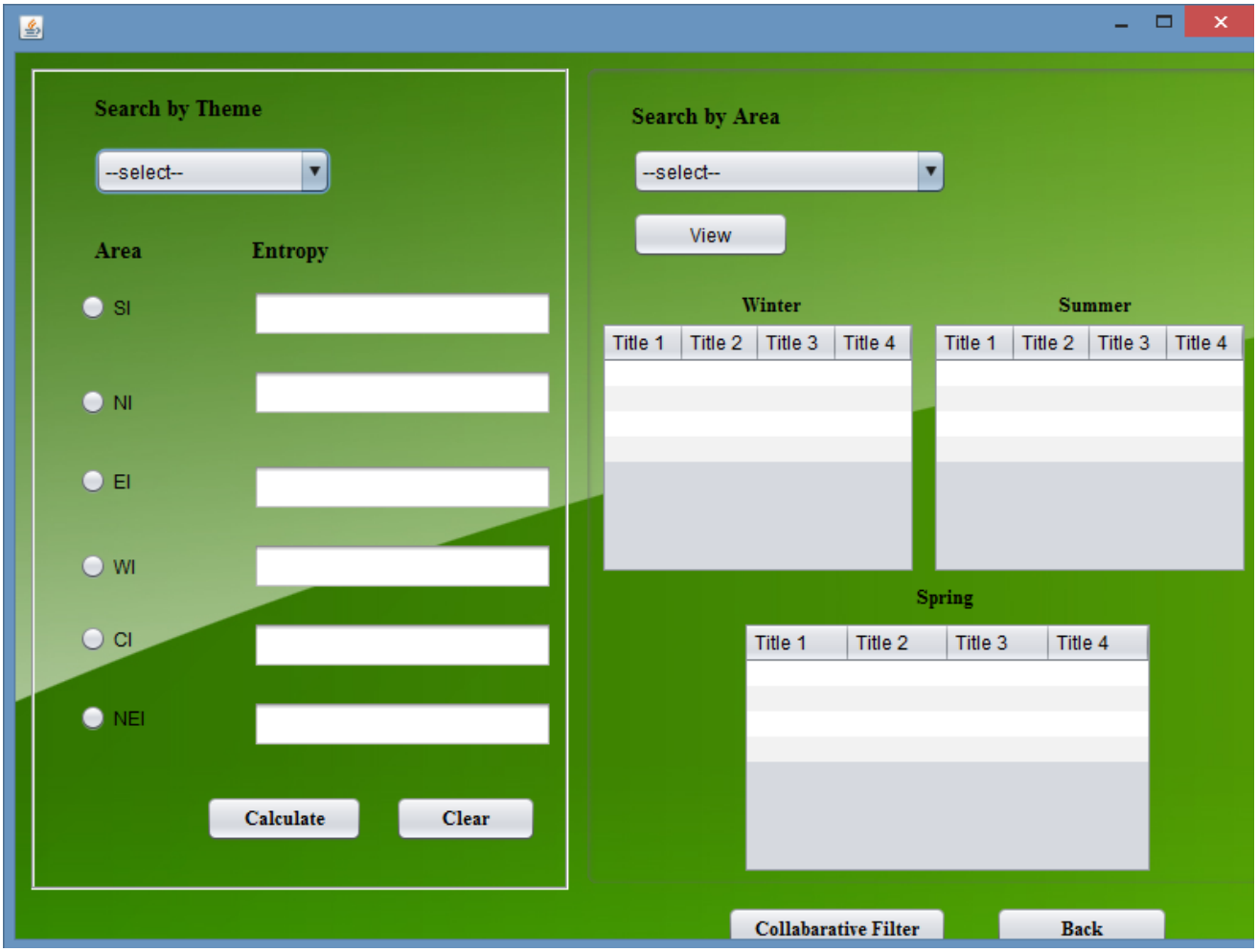


c. *Recommendation by Seasons*

This option calculates the Average Weighted Entropy for each Area for a user-entered Theme. The weighted Entropy is then used to recommend packages by Season for a selected Area.

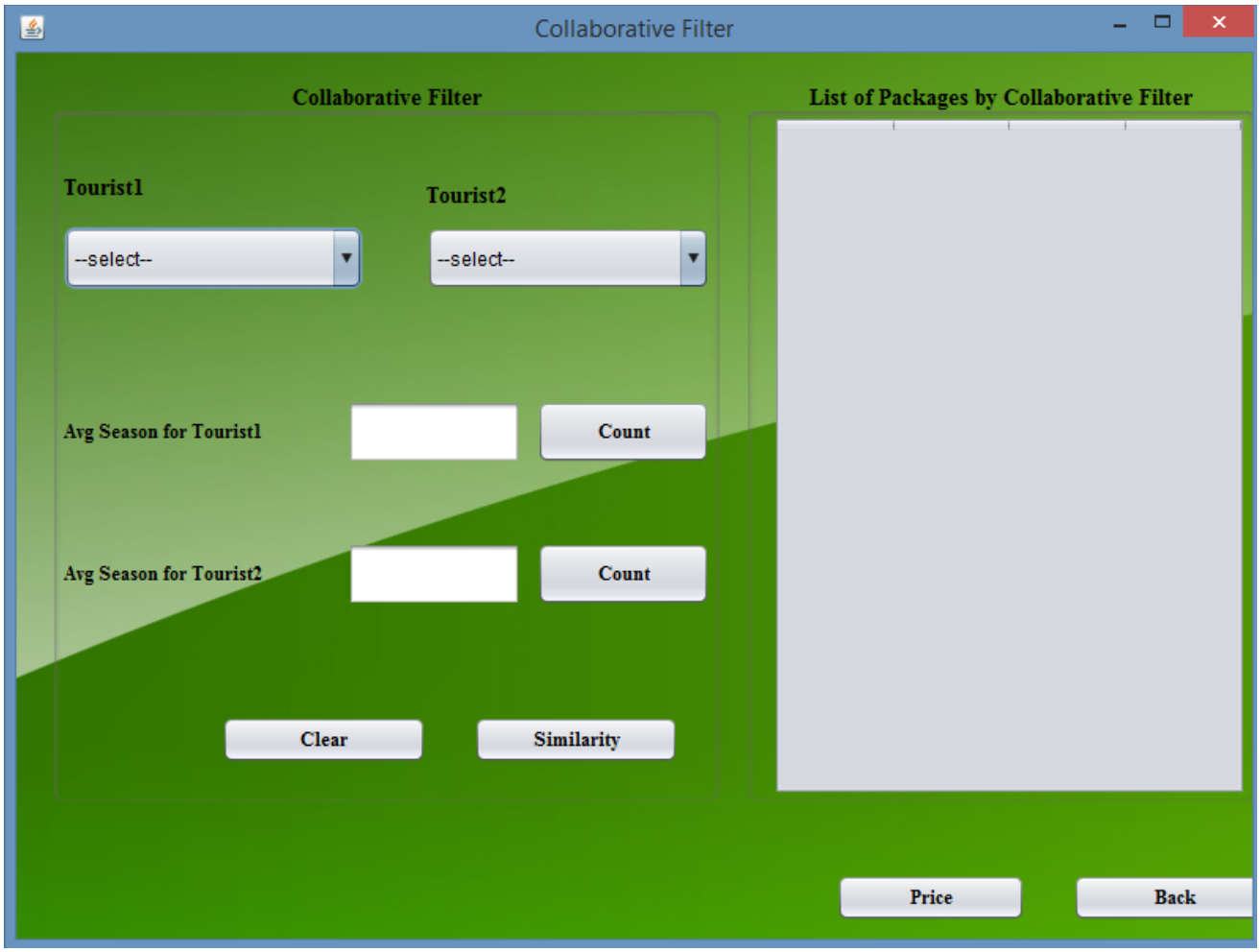
Entropy of the season S^P is $Ent(S^P) = \sum_{i=1}^{|Sp|} pilog(pi)$

Where $|S^p|$ is the number of different packages in S^P and p_i is the proportion of package P_i in this season.



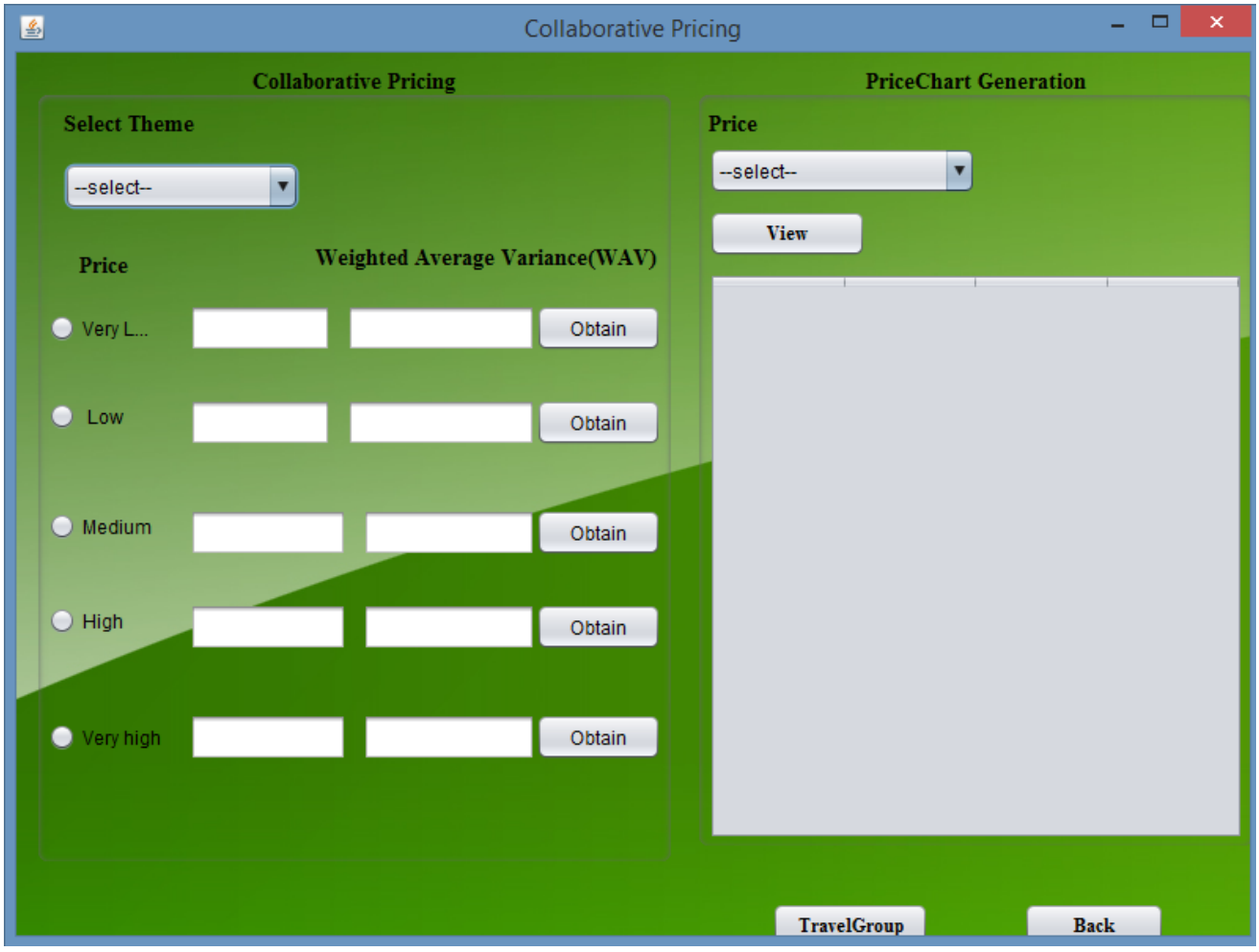
d. Recommendation by Collaborative Filtering

This option uses Collaborative filtering to determine similar packages for two tourists after they have selected a Season of travel.



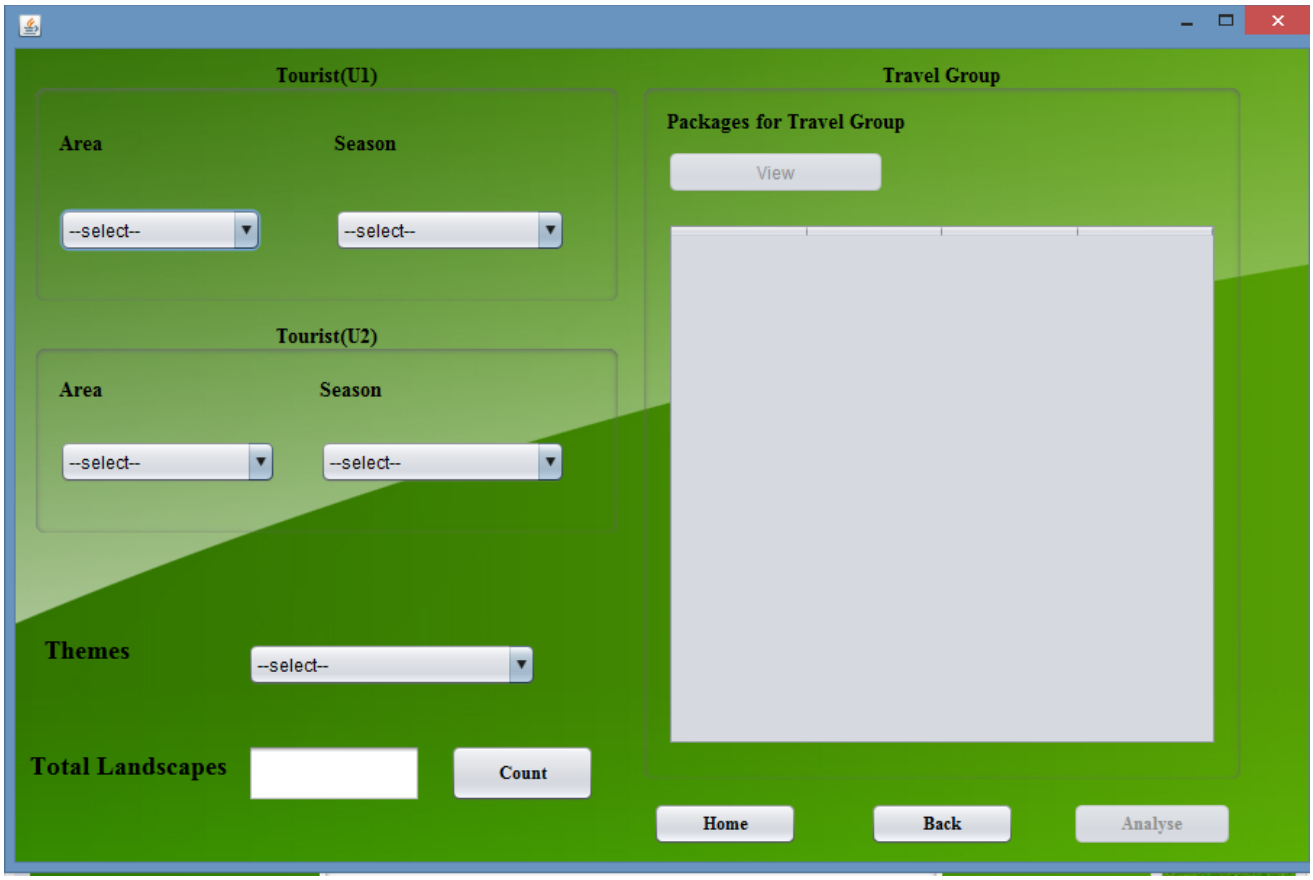
e. Recommendation by Collaborative Pricing

In this option, I calculate a Weighted Average Variance (WAV) for prices of all packages in a selected Theme. The best price range is defined as the one with the minimal Weighted Average Variance (WAV) value. These values are then used to recommend packages when a Price range is seleted.



f. Travel Group Recommendation

This option provides common packages in a particular Theme for two tourists when they select an Area and a Season.



Technologies Used

1. Net Beans:

   The Net Beans Platform is a reusable framework for simplifying the development of Java Swing desktop applications. The platform offers reusable services common to desktop applications, allowing developers to focus on the logic specific to their application. Among the features of the platform are:

   - User interface management (e.g. menus and toolbars)
   - Storage management (saving and loading any kind of data)
   - Window management
   - Wizard framework (supports step-by-step dialogs)
   - Net Beans Visual Library

2. WAMP Server:

   WAMPs are packages of independently-created programs installed on computers that use a Microsoft Windows operating system. WAMP is an acronym formed from the initials of the operating system Microsoft Windows and the principal components of the package: Apache, MySQL and one of PHP, Perl or Python. Apache is a web server. MySQL is an open-source database. PHP is a scripting language that can manipulate information held in a database and generate web pages dynamically each time content is requested by a browser.

3. MySQL

   MySQL is a freely available open source Relational Database Management System (RDBMS) that uses Structured Query Language (SQL). SQL is the most popular language for adding, accessing and managing content in a database. It is most noted for its quick processing, proven reliability, ease and flexibility of use.

References

[1] Q. Liu, Y. Ge, Z. Li, H. Xiong, E. Chen and X. Wu, "A Cocktail Approach for Travel Package Recommendation", *Proc. IEEE 13th Int',l Conf. Data Mining (ICDM ',13),* pp. 278-293, 2013.

[2] Y. Koren and R. Bell, "Advances in Collaborative Filtering," Recommender Systems Handbook, chapter 5, pp. 145-186, 2011.

[3] Y. Koren, "Factorization Meets the Neighborhood: A Multifaceted Collaborative Filtering Model," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '08), pp. 426-434, 2008.

[4] S. Lai et al., "Hybrid Recommendation Models for Binary User Preference Prediction Problem," Proc. KDD- Cup 2011 Competition, 2011.

[5] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," Proc. Berkeley Symp. Math. Statistics and Probability (BSMSP), vol. 1, pp. 281-297, 1967.

[6] A. Mccallum, X. Wang, and A. Corrada-Emmanuel, "Topic and Role Discovery in Social Networks with Experiments on Enron and Academic Email," J. Artificial Intelligence Research, vol. 30, pp. 249-272, 2007.

[7] R. Pan et al., "One-Class Collaborative Filtering," Proc. IEEE Eighth Int'l Conf. Data Mining (ICDM '08), pp. 502-511, 2008.