Senior Project Proposal

Problem

Often when searching online for somewhere to eat or something to do, users are inundated with a large number of business & restaurant listings. Searching for a restaurant in a metropolitan area like San Francisco or New York on CitySearch will return so many entries that most users will not bother to navigate past the first page of results. With the enormous amount of entries returned to the user it can be a problem if none of the entries are relevant to that user's interests. There may be highly rated restaurants in the entries shown to the user but if the user doesn't like a certain type of food, those entries will be irrelevant to them.

Solution

In order to provide the user with results that are relevant to them, their personal tastes and tastes of people like them should be considered. Our solution will feature an algorithm that recommends restaurants and activities based on your personal preferences instead of the average of every users opinion. In order to retrieve initial information about these establishments, we will implement a data mining tool that visits sites which aggregate such information and automate the process to enter it into our database. We will track users visits to each page and business listing and by doing so we will be able to tell, by number of visits to a certain category, what types of food or activities that the user might be interested in.

Data Mining Tool

RSS feeds and XHTML & HTML pages from all around the internet will be parsed for relevant event/business details such as address, type of food/activity, cost, etcetera. The Data Miner will then enter this information in our database. After successfully parsing a business listing, it will be entered into the database allowing access to that data from dynamic pages on the website. These pages will contain the scraped information as well as be open to additional entries from the public.

Suggestion Algorithm

Our suggestion algorithm will take into account a users personal preferences as ascertained
from their ratings and present the user with activities and businesses that others with their tastes rate highly. Our implementation will use keywords, or tags, that a user can quickly click to apply to the business or activity. When many users agree on a tag, it becomes a feature of that activity or business and will be compared with others sharing the same or similar tags. Passive factors, such as how many times a page gets viewed may also account into suggestions. Care must be taken to avoid abuse and we will have to design in mechanisms to prevent this. Large amounts of testing will be needed to make sure we don't suggest irrelevant activities or businesses on the edge cases.

Schedule
Derek Johnson and myself will be implementing this project over Winter and Spring 2007-2008 quarters. We will start by researching data sources and suggestion algorithms then we will implement the data mining tool. Once we have a properly populated database, we will progress towards the final product of our web front end.

Meeting Minimum Criterion
This project meets the minimum criterion of independence, background research, and creativity.

Independence
Derek Johnson and myself will be solely responsible for the development of both stages of this project. I will be mostly focusing on the implementation of the data mining tool as well as the database design, while Derek Johnson will be focusing on the web front end. Although we have separated the project into back end and front end sections, during implementation we will both be helping and doing work on both sections.

Background Research
A large amount of research will be required to implement the suggestion algorithm. We will have to research current collaborative filtering algorithms and develop one to fit our needs. Beyond that, we will need to research the best methods to data mine for the information we need. The final step will involve researching into tools to develop our web front end as well as developing the server environment for it to run on.

Creative
The implementation of the web front end and suggestion algorithm will require creativity to pique the users interest as well as provide relevant results. Currently none of the top city listing sites have a suggestion mechanism which will make this project unique and novel.