

Personalized Search for TV Programs Based on Software Man

Hai-yan Huang¹²

*Department of Computer Science, Zhengzhou College of Science & Technology
Zhengzhou, China 450064
E-mail: 492590002@qq.com*

Bao-long Zhang³

*Department of Computer Science, Zhengzhou College of Science & Technology
Zhengzhou, China 450064
E-mail: 115441690@qq.com*

Watching TV programs is one of the most popular ways of spending our free time. With the development of TV programs, it's a problem of how to search their favorite TV programs efficiently. In this paper, it proposes the personalized search for TV programs based on Software Man. According to user's interests, the personalized search can quickly sort the search results, and then meet the demand of the user's personalized search; besides, this software man can adapt to the environment and make autonomous decision. Finally, this software man can track users and collect the user's browse of TV programs regularly in order to build the users' interests model.

*CENet2015
12-13 September 2015
Shanghai, China*

¹Speaker

²Corresponding Author

³2013-2015, key project of science and technology research in Henan, Department of Education,(14B520005).

1. Introduction

There are too many duplicate messages and it costs lots of time to get useful information, which becomes the biggest problem of searching TV programs. No matter the user is teenager, adult or the elderly, the search process is the same. The current information retrieval technology meet the needs of people to some extent, but as a result of its general properties, it still cannot meet the query request from the different background of the different purposes in different periods. Personalized search is aimed at this problem, which can provide different services for different users to meet different needs. Personalized search will learn the user's interest and behavior by collecting and analyzing the customer's information in order to realize the active recommendation. Personalized search technology can change the search results of "one-size-fits-all" completely and improve the relevance of search results, so as to improve the user's usage experience.

2. Personalized Search Framework

The personalized search system of TV shows is divided into seven subsystems, as shown in Fig. 1.

Natural Language processing subsystem is the foundation of the whole system and also the most basic subsystem. Because the processing object is Chinese TV description text, it doesn't have obvious space like English word. As to Chinese word, there is no space between the word and the word, thus the word segmentation is required for further processing. Obviously, the whole performance of the system is directly influenced by the segmentation accuracy.

Network web spider's subsystem: which is used to grab the TV text description HTML files from TV web, convert TXT file, and be deposited in the TV text corpus.

Storage subsystem: the function of the subsystem television is to store files and compress index. In general, as the index will take up a lot of storage space, high efficiency of index compression can save a lot of disk space.

Indexing subsystem: which is responsible for the TV program content of the document analysis and establishment of the inverted index structure for merging and updating.

Universal search subsystem, which can realize the function to explain the query expression and present the retrieval result through the calculate similarity between inverted index database query and document.

Personalized search subsystem: which has realized the user's personalized search function. This function can realize process mainly through the query expansion, the user interest model and user feedback, etc. while the query expansion needs to use the interfaces provided by the semantic repository.

User query processing subsystem: which has realized the user's query preprocessing, including the word segmentation and parsing while displaying the final query results.

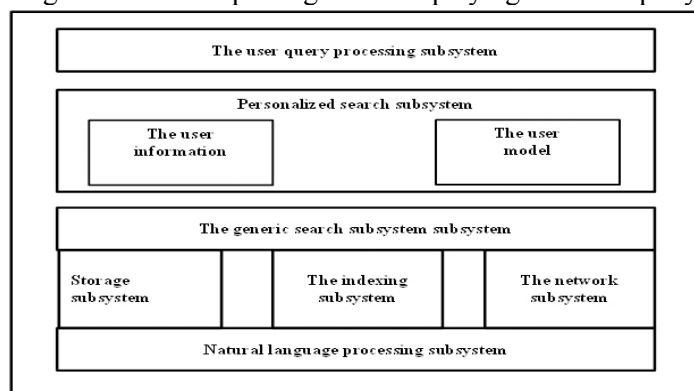


Figure 1: Personalized Search System of TV Show

3. Software Man

"Software man" is an artificial life from the point of view of generalized artificial life so as to extend and expand people's life. It survives in the software environment and activities in the internet world, just like people who have thinking, awareness, behavior characteristics and information acquisition, processing and utilization ability[1]. "Software man" features the functions of personification attributes, personification, personification, behavior and structure of personification.

3.1 Software Man Model

The software model is composed of six layers: core layer, behavior layer, functional layer, interface layer and data layer interaction [2], as shown in Fig.2.

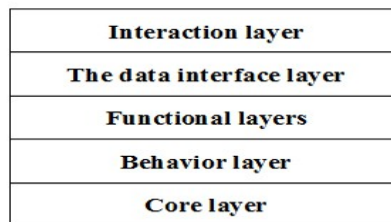


Figure 2: Software Hierarchical Model

The core layer is a set of basic properties of software man [2]. In this layer, it contains the initialization module, function of the registry, behavior and characteristics of the registry, the self-organizing assembly module and evolutionary clone module.

The behavior layer is a set of the software man's natural behavior, spontaneous behavior, self-consciousness and software group collection of environmental administrative tasks. This layer contains the environmental adaptive module, the self extinguishing and clearance module, the function loading module, the migration module, the object recognition module and the basic interaction primitive module.

The function layer is the set of all functions of software designed by the designer. This layer contains the spam cleanup module, the information push and pull module, the information security module and other planning and definition of the service module.

The interactive layer is the software man and other software man, environment and human interaction function set. The interactive behavior includes the transmission of information, looking for similar, work together, communication and exchange information.

The data interface layer is a set of software man and other software man, software man and the environment, software man and human interaction with data exchange and information storage stack [3].

4. Implementation of Software Man in Personalized Search of TV Program

4.1 Grabbing TV Program Information

These six properties can best represent a TV program including the program name, the film types, the actor, the release date, projection television and the audience rating. Seeing each attribute of a film as one dimensional vector, we can get a six - dimensional vector $m\{x_1, x_2, x_3, x_4, x_5, x_6\}$. Given the weight of W_i ($i=1, 2, 3 \dots 6$) of the six vector $(x_1, x_2, x_3, x_4, x_5, x_6)$, we can generate the interest samples on the basis of parsing these six properties. The weight must meet the following two formulas:

$$W_1 \sim W_6 > W_2 \sim W_3 > W_4 > W_5 \quad (4.1)$$

$$W_1 + W_2 + W_3 + W_4 + W_5 + W_6 = 1 \quad (4.2)$$

Then, we cluster all TV programs by using undirected graph according to the weight. The algorithm steps are shown as follows:

- (1) Select k films from n films at random as the k cluster centers;

- (2) Calculate each distance between the remaining n-k film and the k cluster centers, choose the shortest distance and add this film to the clustering center;
- (3) Recalculate the new clustering center of the cluster;
- (4) If there is any change of the previous cluster center, jump to tStep 2 and reassign films; otherwise, exit from the iterative process and the algorithm ends [4].

4.2 Semantic Extension of TV Program

With the mechanical word segmentation method, it divides the name of TV programs into words. This rule-based method will match the Chinese character string to the sufficiently large mechanized dictionary according to certain strategy. If the match is successful, it returns the match results as a participle; besides, it can be divided into the minimum and maximum matching algorithm according to different scanning directions, and can be divided into the simple word segmentation method and the method in combination of segmentation and labeling according to the principle whether the word labeling is combined.

4.3 Tree Structure of Searching TV Program

A tree structure can be used to store a lot of strings. With the help of public prefix of strings, it can save a lot of storage space. Then, it can store the information of the mapping table between search strings and pinyin initials into this tree structure (as shown in Fig.3).

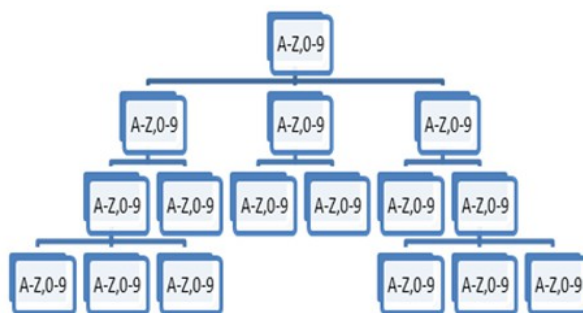


Figure 3: Searching Tree

The principle of this searching tree is shown as follows:

- (1) The first search starts from the root node;
- (2) Find the first letter of string, choose corresponding sub-tree according to the letter and jump to this sub-tree to continue this search;
- (3) In this sub-tree, find the second letter and select corresponding sub-tree to search;
- (4) By that analogy, find the rest letters of string and select the corresponding sub-tree to search;
- (5) At a node, if all of the letters have been taken out, read the information of this node and finish this search; finally, it will return the name of this TV program.

4.4 Detailed Design of Program Search

The search system of television program mainly includes the function modules as follows.

Collection of Program data. The module is mainly to collect data for structural analysis through the metadata of the program source, establish the corresponding program fetching templates, and then snatch in accordance with the fetching templates to carry out data acquirement, then finally analyze the acquirement of data and input metadata database for network modeling module using[5].

The network modeling module: which is mainly to get programs from Collect of Program data in order to the data mining, build method of social network based on the film properties and get social network program.

The program searching: which is mainly to search optimization program and improve the efficiency of search accuracy.

The process is as shown in Fig. 4~Fig. 6.

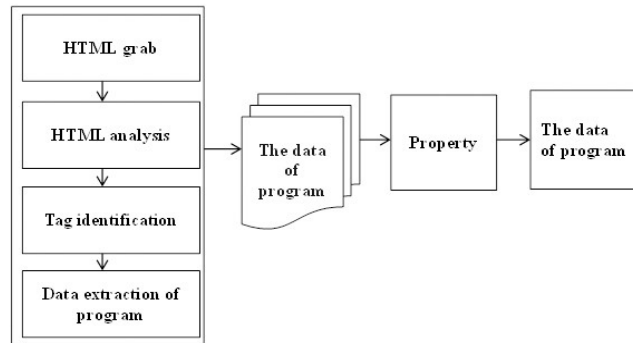


Figure 4: Data Collection of Program

- (1) Obtain program data sets and grab the alias and program evaluation attributes, establish data sets and finally integrate data sets.
- (2) Repeat (1) in the regular time and update the data

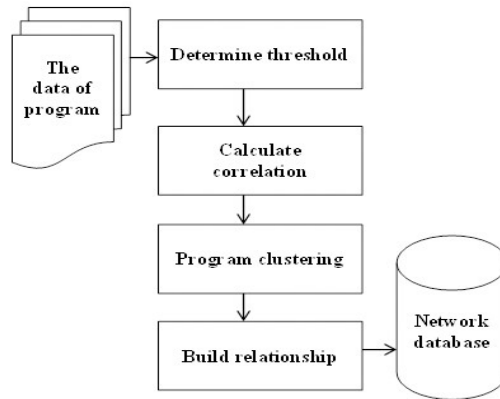


Figure 5: Network Database of Program

- (1) Get attributes of the threshold value through the off-line training;
- (2) Use the utility function to calculate the program correlation between each other.
- (3) Gain of correlation from (2) and cluster program from repository program by k-means algorithm [6].
- (4) When the network information database is established, send TV program information to the database.

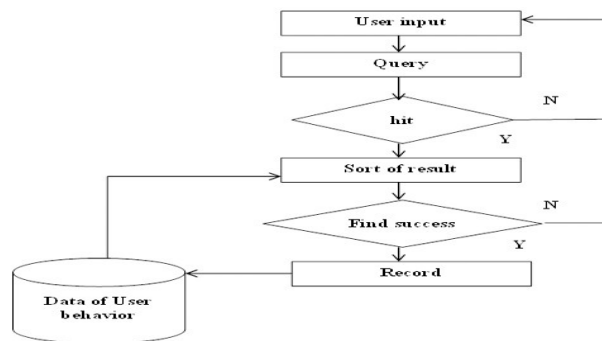


Figure 6: Program Searching

- (1) Interview the network module of information database and establish the programs - initials mapping table
- (2) Wait for the user input
- (3) When a user inputs a button, form a search string in front of the input as inquired.
- (4) Search the query through programs - initials mapping table. If the result of sequence is achieved successfully, return null and jump to (2).
- (5) Make the intelligent sort by result sequence and return the final query results
- (6) If user's choice is correct, find the finish; otherwise, jump to (2)
- (7) If the user selects a film, represent a successful query, record the user behavior, jump to the video module; otherwise, find the failure and jump to (2)

4.5 Search Implementation

It uses C/S structure design between client and server. As to the client, the program, which can be run on Windows and Linux platform, is developed by QT. As to the server side, the program, which can be run on the Linux platform and used sqlite and xml to store data, is developed by C++. The socket connection is established on the basis of TCP between client and server, and the message, consisting of message header and message body, is defined by TLV. Then, the length is used to define the length of message content and the value is used to store the variable message content. The client sends commands (such as init, search, pre, next, recommend) to the server through the corresponding port and the server responds to these commands.

The searching page is shown in Fig.7. After each user has input a letter, the system will automatically ash button which doesn't combine with later letter. It may help the user to save the number of keystrokes and improve the search efficiency. When the user presses the confirm button, the system from the search results page jumps to the target page. If you do not use smart sort of search results, the user presses R. Sequences generates the search results.



Figure 7: Operation Interface

5. Conclusion

According to the serious of problems that users encounter when search for TV information such as information repeating ,time consuming, and the same searching process and so on, combining the information properties of TV programs, this paper combines personalized search and software model by the program for clustering weights iteration, tree structure being adopted to improve the intelligent search, and through the history of the machine user data mining draws preference information, then realizing personalized search programs on television so as to meet the personalized needs of users.

References

- [1] Min Zuo. *A system based on parallel evolution about Robot and Software Man and its application* [M]. Beijing: National Defence Industry Press, 2~4(2012).(In Chinese)
- [2] Guangping Zeng, Xuyan Tu. *Software man's concept model and the tectonic characteristics* [J]. Journal of computer science, 2005(5): 135~ 136. (In Chinese)
- [3] Yi Song. *The research of personalized search for TV programs* [D]. Shanghai: Shanghai Jiaotong University, 2008. (In Chinese)
- [4] Guohui Liu. *The research and implementation of interactive television, film and television program search system* [D]. Shanghai: Fudan University, 2011. (In Chinese)
- [5] Heng Zhuang, Longbing Qi, *The experience design of digital TV man-machine interface, Modern TV technology* [J], 2010 (4): 144~147. (In Chinese)
- [6] Qingzhen Wang, Lili Qiu, *the design of children's TV show search and monitoring Softrobot soft robot , Intelligent computer and application* [J], 2014(2):40-44. (In Chinese)