Building an Efficient Web Portal for Students at Institutions of Higher Education Based on Web Crawlers

Haibo Yi[1, a] and Zhe Nie[1,b]

[1] School of Computer Engineering, Shenzhen Polytechnic. 518055 Shenzhen, China

[a]haiboyi@126.com, [b]niezhe@szpt.edu.cn

Abstract. Online applications of Internet are ubiquitous influences in the development of students at institutions of higher education. Web portal is one type of online application of Internet that has grown rapidly in prevalence and popularity in recent years. Web portal is most often one specially designed Internet web site that brings information together from diverse sources in a uniform way. However, few of them can offer efficient services of learning and entertainment for students at institutions of higher education. In this paper, we design an efficient web portal for students at institutions of higher education, which can advances student learning and entertainment. The web portal is designed based on a multi-process web crawler, which can bring information of learning and entertainment together from diverse sources on Internet in a uniform way. Our web portal provides a consistent look and feel with access control and procedures for multiple applications and databases of learning and entertainment, which otherwise would have been different web entities at various URLs. We divide the web portal into five functional modules, i.e. login module, administrator module, crawler module, learning module and entertainment module. By integrating the designs of five modules, we use Microsoft ASP.Net (version 4.0) and SQL Server (version 2008) to implement the web portal based on three-tier B/S architecture.

1. Introduction

Internet is a ubiquitous influence in the development of students at institutions of higher education, accounting for a large portion of their time, with 18-24 year old youth in China investing approximately 7.5 h per day with Internet. One type of online application of Internet that has grown rapidly in prevalence and popularity in recent years is web portal [1-5].

Web portal is most often one specially designed Internet web site that brings information together from diverse sources in a uniform way, such as government web portals, cultural portals, and search portals. Examples of public web portals are Google, Bing, MSN and Yahoo. However, few of them can offer efficient services of learning and entertainment for students at institutions of higher education.

In this paper, we design an efficient web portal for students at institutions of higher education, which can advances student learning and entertainment. The web portal is designed based on a multi-process web crawler, which can bring information of learning and entertainment together from diverse sources on Internet in a uniform way. Our web portal provides a consistent look and feel with access control and procedures for multiple applications and databases of learning and entertainment, which otherwise would have been different web entities at various URLs.

We divide the web portal into five functional modules, i.e. login module, administrator module, crawler module, learning module and entertainment module. By integrating the designs of five modules, we use Microsoft ASP.Net (version 4.0) and SQL Server (version 2008) to implement the web portal based on three-tier B/S architecture. Three-tier architecture is a software architecture pattern in which the user presentation, business logic and data access are developed and maintained as independent modules. It was developed by John J. Donovan in his tool company - Open Environment Corporation (OEC) in Cambridge, Massachusetts.

2. Web Portal Design

We adopt three-tier B/S architecture of Microsoft ASP.Net (Version 4.0) in our system design. Typically, three-tier B/S architecture includes user presentation tier, business logic tier and data access tier. The user presentation tier is the topmost level of the architecture, which displays information via web browsers (e.g. IE, Chrome). It receives requests of users (students, administrators and others) and returns processing results to them via web applications (e.g. ASP. net). The business logic tier is pulled out from the user presentation tier and, as its own layer, it controls a system functionality by performing detailed processing. The data access tier provides an Application Programming Interface (API) to the business logic tier that exposes methods of managing the stored data without exposing or creating dependencies on the data storage mechanisms (e.g. SQL Server).

We choose the three-tier system architecture and divide the web portal into five functional modules, i.e. login module, administrator module, crawler module, learning module and entertainment module, where the hierarchical modules of the web portal is shown in Fig. 1.
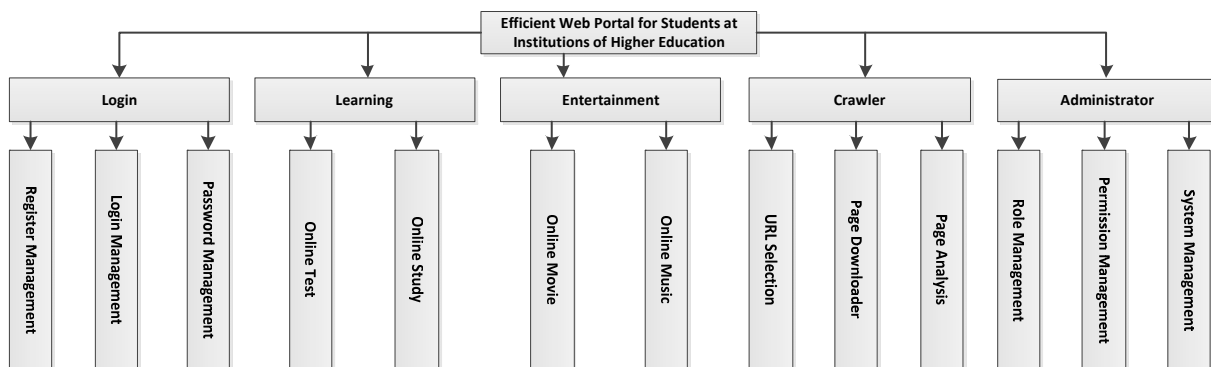

Fig. 1 Hierarchical Modules of the Web Portal

(1) Login module: login module is used to identify user's roles, which includes register management, login management and password management. New users can register the web portal with their usernames and passwords via the register page. Users use their usernames and passwords to login the web portal via the login page. They can change their passwords to protect their information via password page. Login module builds an access between the back-end system on the server and the system website on user's terminals.

(2) Administrator module: administrator module includes role management, permission management and system management, which is responsible for the users and group administration,

server configuration, and server maintenance. System administrators can manage information of users, learning and entertainment via administrator page.

(3) Crawler module: crawler module is used to bring information of learning and entertainment together from diverse sources on Internet in a uniform way and the information is displayed on websites. It includes URL selection, page downloader and page analysis.

(4) Learning module: learning module is used to offer efficient services of learning for users, which include online test and online study. Users can study and test online via test pages and study pages respectively.

(5) Entertainment module: entertainment module is used to offer efficient services of entertainment for users, which include online music and online movie. Users can listen to music and watch movie online via music pages and movie pages respectively.

We design a multi-process web crawler for updating the web content of web portal, which is depicted in Fig. 2. The web crawler starts with a list of URLs to visit. As the web crawler visits these URLs, it identifies all the inks in pages and adds them to the list of URLs, called the crawl frontier. URLs from the frontier are recursively visited according to a set of policies. If the web crawler is performing archiving of pages it copies and saves the information as it goes. The archives are stored in database.

We use Microsoft SQL Server 2008 to design entities for the web portal, which are shown in Fig. 3. Login information table stores usernames, passwords and types for users. User information table stores user's ID, name and college. Music information table stores music's ID, name, type and author. Movie information table stores movie's ID, name, type and author. Study information table stores lesson's ID, name, type and content. Test information table stores test's ID, name, type and content.
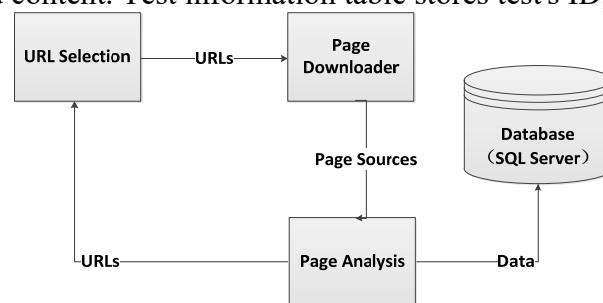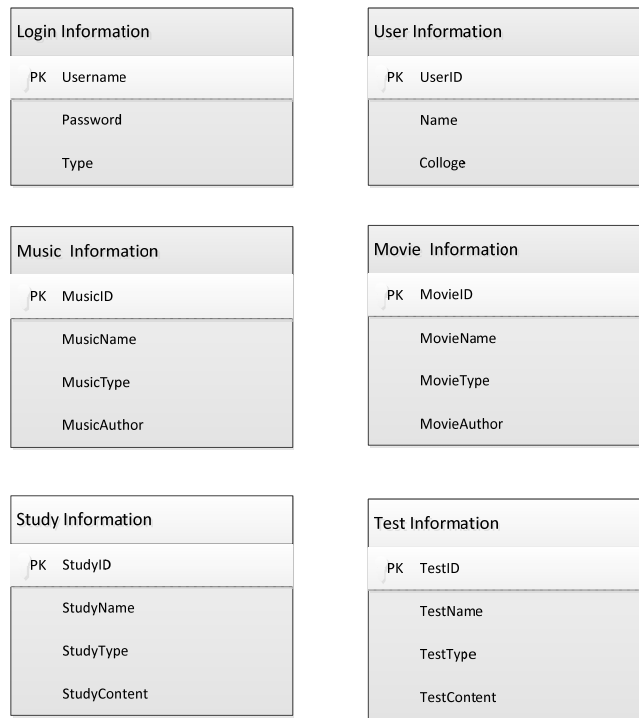

Fig. 2 Crawler Design

Fig. 3 Entities of the System

3. Implementation

According to our system design, we implement the efficient web portal for students at institutions of higher education based on web crawlers, which includes login module, user module, learning module, entertainment module and administrator module. We adopt Microsoft SQL Server 2008 for database design, ASP. net 4.0 for web design and Microsoft Visual Studio 2010 for development environment.

Learning module is used to offer efficient services of learning for users, which include online test and online study. Users can study and test online via test pages and study pages respectively. Fig. 4 shows the flowchart of the learning module. Entertainment module is used to offer efficient services of entertainment for users, which include online music and online movie. Users can listen to music and watch movie online via music pages and movie pages respectively. Fig. 5 shows the flowchart of the entertainment module.
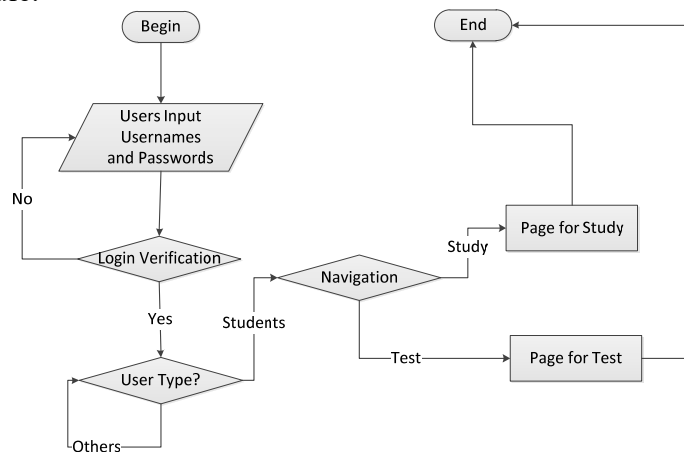

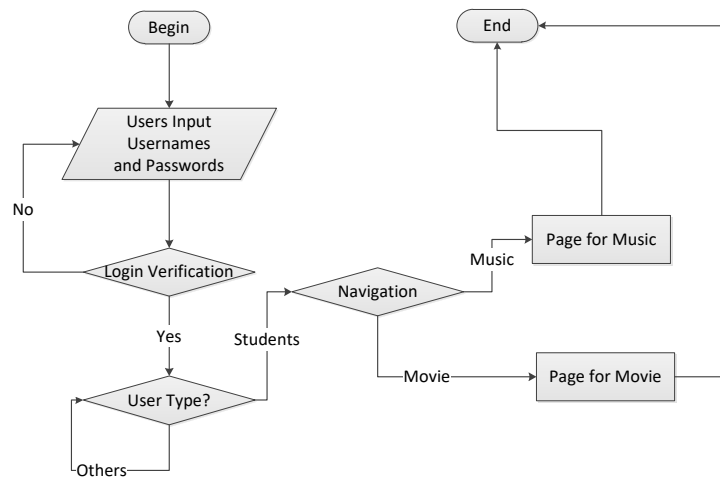Fig. 4 The Flowchart of Learning Module

Fig. 5 The Flowchart of Entertainment Module

4. Conclusions

In this paper, we design an efficient web portal for students at institutions of higher education, which can advances student learning and entertainment. The web portal is designed based on a multi-process web crawler, which can bring information of learning and entertainment together from diverse sources on Internet in a uniform way. Our web portal provides a consistent look and feel with access control and procedures for multiple applications and databases of learning and entertainment, which otherwise would have been different web entities at various URLs.

Acknowledgements

References

[1] Albert, R., Jeong, H. (1999). Internet: diameter of the world-wide web. Nature, 401(6), 130-131.
[2] Porter, M. E. (2001). Strategy and the Internet. Harvard Business Review, 79(3), 62-78.
[3] Turkle, S. (1995). Identity in the age of the Internet. Experimental Brain Research, 87(1), 173-80.
[4] Luzi, L., Hailemikael, S., Bindi, D., Pacor, F., Mele, F., Sabetta, F. (2008). Itaca (Italian Accelerometric Archive): a web portal for the dissemination of Italian strong-motion data. Seismological Research Letters, 79(5), 716-722.
[5] Gant J. P., Gant D. B. Web portal functionality and state government e-service[C]// Hawaii International Conference on System Sciences. IEEE, 2002:1627-1636.