A Design of the Extraction System for an Internet Business Region Variation based on OLAP

Sung Hun Kim[1], Nam Hun Park[2], Kil Hong Joo[3*], Yang Sun Lee[4]

[1,3] Dept. of Computer Education, Gyeongin National University of Education, Korea
[2] Dept. of Computer Science, Anyang University, Korea
[4] Dept. of Computer Engineering, Seokyeong University, Korea
[1] tomorrow@imtsoft.net, [2] nmhnpark@anyang.ac.kr, [3] khjoo@ginue.ac.kr,
[4] yslee@skuniv.ac.kr

Abstract. With the popularization of an internet, the quantity of web information and the online business have been increased. In the online business, it is important to predict to variation of the Internet business region variation. User can need more efficiently target marketing than general marketing. For this purpose, this paper proposes a design of OLAP using Discovery-driven exploration of domain data cube. This design suggests to analyze the Internet business region variation based on domain, and prediction of the change of Internet business region as well. In addition, the appropriation of the result from OLAP design applied by discovery-driven exploration is presented by the comparison between the statistical analysis of domain registration and the statistical analysis of the Internet Business. The performance of the proposed algorithm is analyzed by a series of experiments to identify its various characteristic.

Keywords: Data mining, OLAP, Datacube

1   Introduction

Domain registration organizations have long recognized the importance of their customers. However, since the product, domain relies on the purpose of use of a company a registrant belongs to in its nature, other than individual purpose of use, domain marketing hardly targets individuals. Domain registrants are not any specific group of people and their companies also provide far diversified services. In this situation, it is extremely difficult for domain registration service providers to match the domain registration with registrants' propensity and company service requirements in performing CRM. Moreover, the information possessed by domain registration organizations is limited only to domain registration-related aspects without those on how the registered domain is used in what kind of services online or how to classify information for management, etc. It is true that such organizations, even though they manage key information – address sources – for online business,

---

* Corresponding Author

they have not utilized it more effectively. To overcome the situation, they would need key information that encompasses aspects about a registrant, enterprise and online business altogether. By analyzing such information, domain registration organizations can anticipate online business trend earlier than any other and perform CRM (customer relationship management) for customers in the business areas expected to change.

2   Related works

In the past, CRM studies have focused mainly on the expenses to attract customers or to prevent losing customers. And today, as a further advancement, studies focus on e-CRM where customer relationships via the media of internet are researched and applied [1,2,3]. The term, OLAP was first used by E.F. Codd in 1993 to use in opposition to the idea of OLTP (On-Line Transaction Processing)[4]. Codd used a static model and dynamic model as en enterprise model and defined OLAP as a dynamic enterprise analysis necessary to generate, handle, animate and aggregate information from the dynamic model [4]. Data cube is a multidimensional database known as a main OLAP data model. Data cube performs multidimensional data modeling to make data be displayed. It is generated from table star schema [5,6].

3   Database design and modeling of internet business region

To identify and analyze changes in the domain-based internet business area, the relation OLAP was employed in this research. ROLAP system needs data based on domain data warehouse which have been accumulated for years. And to connect the changes in internet business areas in its analysis, the system needs separated external data accumulation. The external data, here, should have domain keys and attribute data on the internet business use area. Such data should be stored in an automated system. The accumulated internet business attribute data are linked to the basic domain data warehouse to provide multidimensional data structures for ROLAP. The basic domain data and internet business area information are linked by the domain values of each information set as a connection key. Corresponding data are collected to find out an exceptional or abnormal pattern by analyzing the relationship among such information. In this thesis, data cube was used for multidimensional data modeling and recognition. Figure 1 represents 4D data cube consisted of product, region, internet business and point of registration.
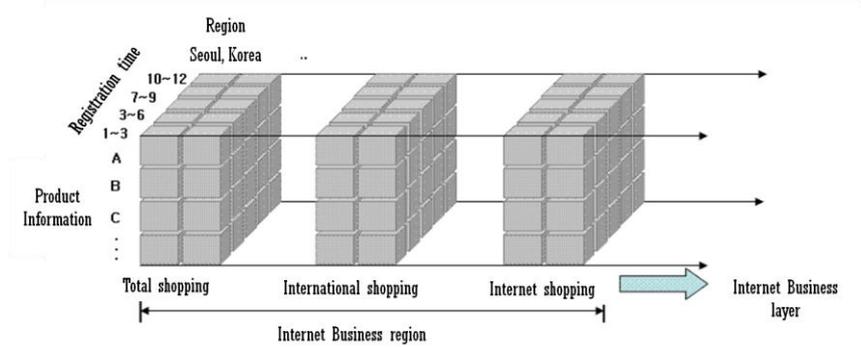
Fig. 1. Data cube of internet business region based on domain

To examine internet business area changes, a separate database needs to be built based solely on better organized and more necessary data among the initial data. Hence, it is fundamental to better organize product data on newly-generated domains having a registration point and region data possessed by the domain registering customers. It is also vital to form accurate connection with internet business area data generated separately from outside. Therefore, to meet such conditions, this paper defined the mutual relationship between product data and data on registration point, region and internet business area in a star schema structure as shown in Figure 4. Here, each data represents one single dimension in OLAP. Actually when analyzing changes in domain-based internet business areas, such data help provide the results from diverse changes in internet business areas according to registration point, region and product type. For multidimensional analysis of internet business area changes, each dimensional tables are based to form fact tables of domain-based internet business areas. And based on the fact tables, integrated tables are formed for subsequent OLAP analysis.
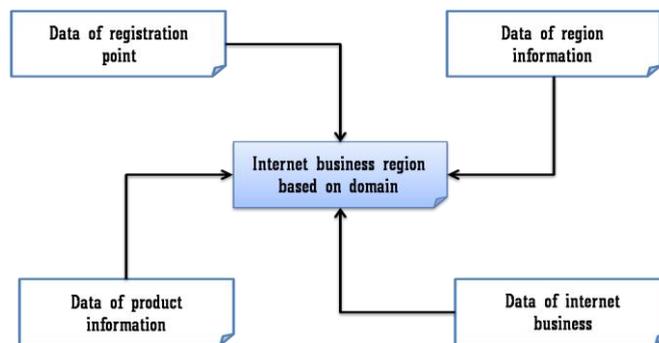

Fig. 2. Star schema for OLAP

4   Conclusion

This paper proposed an OLAP system design applying data cube measure and discovery-driven search method to identify changes and expectations of domain-based internet business areas. Data cube measure and indication values were presented herein as those that could resolve the problems of statistical method. Their appropriateness was presented after comparing with the general statistical analysis. Such OLAP is expected to be applicable to other similar services to the domain registration such as hosting and homepage production, etc.

References

1. Chen, Qimei, Chen, Hong-Mei : Exploring the success factors of eCRM strategies in practice. The Journal of Database Marketing & Customer Strategy Management, vol. 11, no. 4, pp. 333-343 (2004)
2. Balaji Padmanabhan, Alexander Tuzhilin : On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities, vol. 49 no. 10, pp. 1327-1343 (2003)
3. Firoozeh Fouladivanda, Mahsa Mohseni, Mani Shehni Karam Zadeh, Aidin Barbat : A Study on the Relation between Electronic Customer Relationship Management (ECRM) and Customers Loyalty in the International Market, Life science journal, vol. 10, no. 12, pp. 353-359 (2013)
4. Nigel Pense : THE OLAP REPORT : What is OLAP ?, Olap Report Corporation Whaite Paper, http://www.olapreport.com (2002)
5. Gaurav Gupta, Himanshu Aggarwal : Overview of CRM using Data Mining, International Journal for Multi Disciplinary Engineering and Business Management, vol. 1, no. 1, pp. 15-20 (2013)
6. J.Gray, S.Chaudhuri, A.Bosworth, A.Layman, D.Reichart, M.Venkatrao : Data Cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, and Sub-Totals, Data Mining and Knowledge Discovery vol. 1, pp. 29-53 (1997)