Construction on English Translation Corpus System for Chinese Cultural Classics

Fenghua Li

Teaching and Research Institute of Foreign Languages, Bohai University, Jinzhou, 121013, China

lzy8917891@163.com

Keywords: Chinese cultural classics; English translation; corpus; system construction; Oracle; database design

Abstract. Classic into English is an important means to promote the national culture, to promote the exchange and integration of Chinese and Western cultures, to carry forward the Chinese culture. Translation corpus research developed rapidly in recent years and provided a new approach and ideas for classic into English. For the difficult problems about corpus building, this paper is divided into four aspects: First, Oracle database system architecture and its advantages; Second, the conceptual structure design on database with E-R diagram; third, according to the logical structure design principles on database designed each table structure; fourth, created a corpus system database and completed the relevant settings. Contents of this paper filled the blank for translation corpus and have an important role in improving the English translation level of Chinese classics works and translation speed.

Introduction

Classics translation is the own development needs for Chinese classics. Although translations exist in a different space with the original work, the translation work was derived from the original; it is a continuation of the original life. Excellent cultural classic is one of particular product in the any times, if spread its deep thought and culture down from generation to generation, and enduring, needed the continuous efforts from scholars and translators. Classics translation not only makes China classical works shine again, but also spreads Chinese traditional excellent national culture to the world. Chinese cultural classics not only thinking deep, profound, but also contains a wealth of philosophy, as the modern Chinese people have the responsibility and obligation to translate these works fully and accurately to foreign countries, so that other nations of the world really understand the Chinese culture, and then to really understand China [1,2]. Chinese classics translation has been an important part of Western cultural exchange and is also an important way and means to spread the Chinese history and culture. Chinese classics not only have important ideological value, but also contain a wealth of cultural information, so that understanding and difficulty of translation increased accordingly. In the specific translation strategies, translators and scholars should take pluralism complementary translation strategies to pass out the cultural information in the Chinese classics works accurately and completely.

Corpus refers to the scientific sampling and processing large-scale electronic text library, is the basis of corpus linguistics, and is the main resource of empirical language research methods. Since the emergence of computerized corpus, corpus linguistics rapid development, comparative study of the use of the language corpus, language ontology research has achieved fruitful results [3,4]. Becker created the world's first translational English corpus, proposed parallel corpus, multilingual corpus and comparable corpus can be found and identified by conventional methods difficult semantic features found in study of the text style, language habits, such as redundancy of language, vocabulary co-occurrence, the extent of regulation, coherent form, syntactic patterns, even punctuation usage characteristics, and to help translate to choose corresponding translation strategies. Has translated works of Chinese classics is a valuable asset, an important reference and reference. Chinese cultural classics works translated into English Corpus system for storing the

essence, for the study of Chinese classics translation, dissemination of Chinese culture provides a new way.

Oracle Database System Architecture

This paper based on Oracle database system to construct corpus. Oracle database system is provided by the US Oracle distributed database as the core of a group of software products, is currently the most popular Client/Server or one database B/S architecture, and is currently the world's most extensive database management system. As a general-purpose database system, with a complete data management functions; as a relational database, is a relationship of complete product; as a distributed database, a distributed processing functions. Oracle database system architecture is shown in Fig. 1.
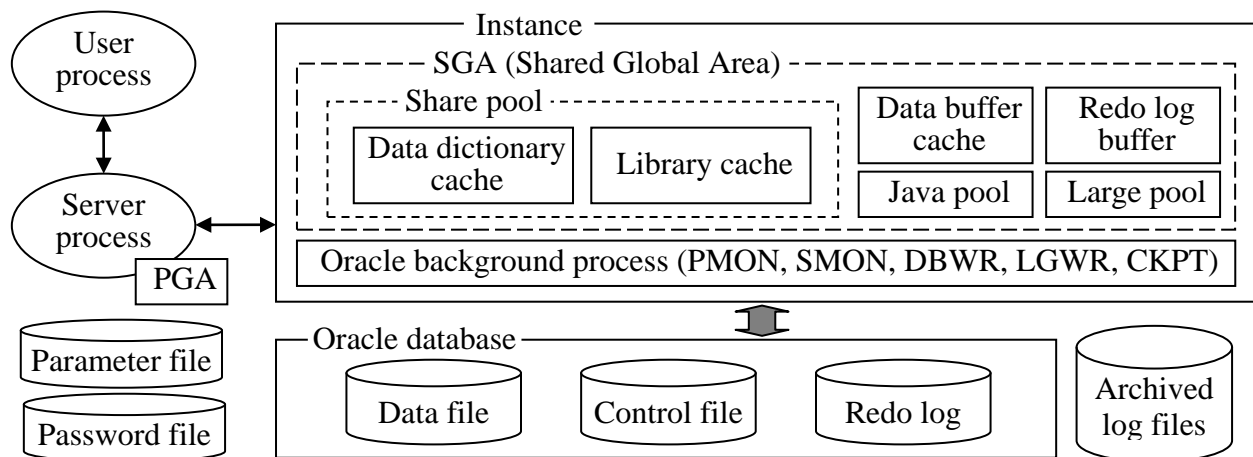

Fig. 1. Oracle database system architecture

The system architecture in Fig. 1 is divided into three parts [5,6], on the left, User Process, Server Process, PGA constitute Client, over Instance and Database, Parameter file, Password file and Archived log files consisting Oracle Server, entire diagram can be understood as a C/S architecture. Oracle Server consists of two entities, namely Instance and Database. These two entities are separate, but connected together. In the database creation process, instances are created firstly, and then create the database. In a typical single-instance environment, the relationship is one to one with the database instance, and an instance of a database connection, the instance and the database can also be many to one relationship, that multiple instances on different computers on a shared disk system open a common database. This n to 1 relationship is called the RAC (Real Application Clusters), greatly improved performance, fault tolerance and scalability of the database, and is an essential part of the Oracle grid.

Oracle Database has the following advantages: (1) the introduction of the shared SQL and more clues to the server architecture, reducing resource consumption, so that in the low-end hardware and software platforms with fewer resources can support more users; (2) provides role-based division of security and confidentiality management, database management, integrity checking, security, consistency have good performance; (3) supports a large number of multimedia data, such as binary graphics, sound, animation and multi-dimensional data structure ; (4) provides a high-level interface software and third-generation language, PRO*series, can be embedded SQL statements and procedural statements in C and C ++ and other main languages in the database for data manipulation. It can generate a rapid development platform for PC-based client applications, and has good portability; (5) provides a new distributed database capabilities. The distal end can be more easily read and write data in the database through the network, and there is a symmetric replication technology.

Conceptual Structure Design on Database

The conceptual structure design is abstract the result of needs analysis into a data model that does not rely on any particular machine, namely the conceptual data model. Conceptual data model enables the designer's attention freed from the complex implementation details, only focus on the most important organizational structure and processing mode information. Conceptual model to describe the relationship between the various entities in the system, and related entities, is the system features and static description. E-R diagram also called Entity Relationship Diagram provides methods represented entity types, attributes and contact, used to describe the conceptual model of the real world [7,8]. Entity-shaped frame with rounded corners, said contact with the diamond, said property represented by an oval frame. System includes three Chinese entities and three English entities. Three Chinese entities are "Classics Information (Chinese), Clause Information (Chinese) and Word Information (Chinese)"; three English entities are "Classics Information (English), Clause Information (English) and Word Information (English)." Each Chinese entity corresponds to an English entity; inter-entity is "1 to 1" link, namely a Chinese entity only has one Chinese entity corresponding to it, the name of contraction with the "Exist" representation. There are two "1 to n" among three Chinese entities, the name of contraction with the "Include" representation, for example, "Classics Information (Chinese)" entity and "Clause Information (Chinese)" are 1 to n , a "Classics Information (Chinese)" has many "Clause Information (Chinese)", a "Clause Information (Chinese)" has only one" Classics Information (Chinese)". The contraction among 3 English entities is same with Chinese entities; will not be described here. E-R model of the system database conceptual design is shown in Fig. 2.
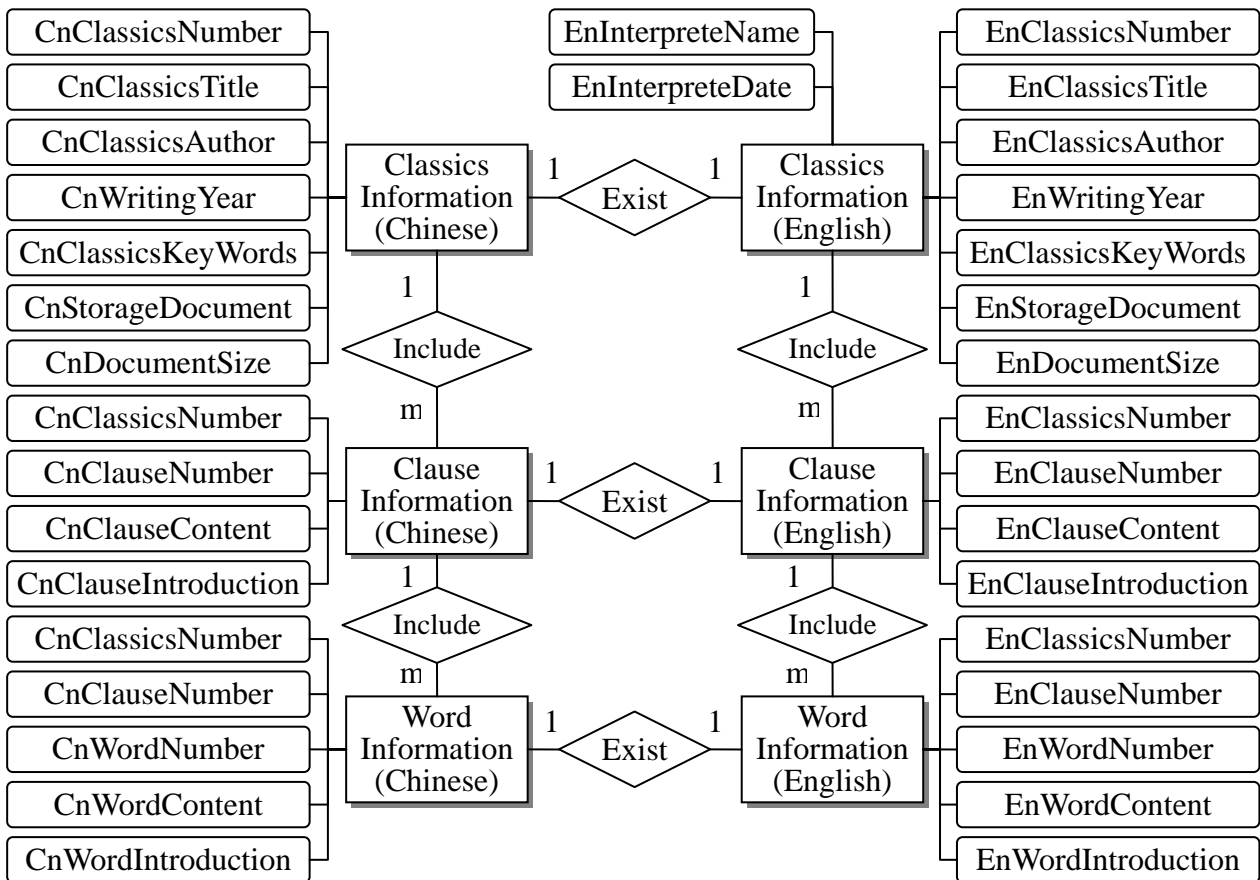

Fig. 2.   E-R diagram of system

Logical Structure Design on Database

The task of logical structure design on database is to convert the conceptual structure models into specific the data model supported by database management system. The conversion rule for the conceptual model to logic model is that an entity type converts into a relational schema. Properties for entities are the relationship properties, the code for entity is the relationship code; a 1: 1 contact can be converted to an independent relationship mode, you can also either end of the corresponding relational schema merge. If converted to an independent relationship model code of each entity is connected to the contact and contact property itself converted to property relations, the code of each entity are the relationship between the candidate codes. If one end of the corresponding entity relationship model merge, you need to add another relational schema code and links itself in the properties of the relational model; a 1: n link can be converted into a separate relational schema, but also with n end corresponding relational schema merge. If converted to an independent relationship model code of each entity is connected to the contact and contact property itself converted to property relations, and the relationship of the code is the code n-side entity.

Oracle offers a variety of data types, the system uses "VARCHAR2, BLOB, NUMBER, DATE" four types. Wherein, VARCHAR2, storing character data type, the advantage is to save storage space; BLOB, store unstructured binary file, supports transaction processing; NUMBER, for storing zero, positive, negative as well as fixed length floating point; DATE, for storing date and time data, using the 7-byte fixed-length. Each field logical structure design with [field name, type, width] of the format, the type of each field is one of "VARCHAR2, BLOB, NUMBER and DATE". Logical structure design of the system is expressed as follows:

(1) ClassicsInformationChinese: [CnClassicsNumber, VARCHAR2, 10], [CnClassicsTitle, VARCHAR2, 100], [CnClassicsAuthor, VARCHAR2, 50], [CnWritingYear, DATE, 7], [CnClassicsKeyWords, VARCHAR2, 100], [CnStorageDocument, BLOB], [CnDocumentSize, NUMBER, (10,4)].

(2) ClauseInformationChinese: [CnClassicsNumber, VARCHAR2, 10], [CnClauseNumber, VARCHAR2, 15], [CnClauseContent, VARCHAR2, 100], [CnClauseIntroduction, VARCHAR2, 200].

(3) WordInformationChinese: [CnClassicsNumber, VARCHAR2, 10], [CnClauseNumber, VARCHAR2, 15], [CnWordNumber, VARCHAR2, 20], [CnWordContent, VARCHAR2, 50], [CnWordIntroduction, VARCHAR2, 100].

(4) ClassicsInformationEnglish: [EnClassicsNumber, VARCHAR2, 10], [EnClassicsTitle, VARCHAR2, 200], [EnClassicsAuthor, VARCHAR2, 100], [EnWritingYear, DATE, 7], [EnClassicsKeyWords, VARCHAR2, 200], [EnStorageDocument, BLOB], [EnDocumentSize, NUMBER, (10,4)], [EnInterpreteName, VARCHAR2, 100], [EnInterpreteDate, DATE, 7] .

(5) ClauseInformationEnglish: [EnClassicsNumber, VARCHAR2, 10], [EnClauseNumber, VARCHAR2, 15], [EnClauseContent, VARCHAR2, 500], [EnClauseIntroduction, VARCHAR2, 1000].

(6) WordInformationEnglish: [EnClassicsNumber, VARCHAR2, 10], [EnClauseNumber, VARCHAR2, 15], [EnWordNumber, VARCHAR2, 50], [EnWordContent, VARCHAR2, 200], [EnWordIntroduction, VARCHAR2, 500].

Establishment on Corpus System Database

The process of creating the corpus in an Oracle database version "Oracle Database 11g Enterprise Edition Release 11.1" is following [9-11]:

(1) Creating the necessary relevant directory. You can create in the Windows graphical interface; and also create using the mkdir operating system command. Created the relevant directory under the "C: /oracle/product/11.1/admin/" include book, bdump, udump, cdump, pfile, create and so on.

(2) Creating an initialization parameter file. The sample initialization parameter file is located in "$ORACLE_HOME\admin\sample\pfile" directory initsmpl.ora copied to the relevant directory,

renamed initbook.ora, use Notepad to open after modification related projects. Initialization parameter files, including shared server configuration, database instance name, security and auditing, global data names, control file and so on.

(3) Creating an instance. Select "Programs -> Run", type "cmd" command, open the DOS interface: the environment variable ORACLE_SID is set to "ChineseClassicsEnglishCorpus"; create "ChineseClassicsEnglishCorpus" database use ORADIM command. ORADMIN is for instance management tools provided by Oracle, you can create, modify, deletes, start-up and shut down an instance.

(4) Creating a password file. Password file is used to store the sys user password, because the sys user is responsible for building a database, start the database, close the database and other special tasks, the sys user's password stored in the password file separately, so that the database is not open when you can be authenticated. orapwd is to create the password file name of the program tools, including file, password, entries, force parameters and other parameters.

(5) Connecting and starting the instance. Enter sqlplus environment, connect to a database (SQL>connect sys/mynewdb as sysdba) with the connect command; start the database (SQL> startup nomount) with the startup command.

(6) Using create database statement to create database. While executing create database statement automatically performs $ ORACLE_HOME \ rdbms \ admin \ sql.bsq script to create the SYSTEM tablespace and SYSTEM rollback segments, establish SYS and SYSTEM account, set up a data dictionary base tables, set up a data dictionary index, create Predefined roles.

(7) Using create tablespac command to create additional table space. The database create database statement creates only SYSTEM, UNDOTBS and TEMPTBS three table spaces, but also need to create some other additional table space.

(8) Creating the database objects. Database objects include tables, views, indexes, sequences, stored procedures, and other hair solution, respectively, using create table, create view, create index, create sequence, create procedure, create trigger commands [11].

Conclusion

Chinese classics are the crystallization of China's ancient civilization, are an important part of world civilization, and should let the world's people share in a fortune. However, due to language barriers and ideological differences, the West know very little about the Chinese ancient culture from long time, the plan to put Chinese books translated into English is a basic task, requiring many people to pay unremitting efforts and to conduct long accumulation of culture. Translation corpus research has developed rapidly in recent years; many translation corpus, parallel corpus and comparable corpus have been built and put into use. Contents of this paper filled the blank for translation corpus and have an important role in improving the English translation level of Chinese classics works and translation speed.

Acknowledgement

References

[1] H. Y. Wang, "Problems of Translating Chinese Classics and Solutions," Chinese Culture Research, vol. 23, no. 2, pp. 59-68, 2015.

[2] Y. Y. Fu, "Conveying the Meaning in its Full Flavor - Towards the Construction of the Theoretical System of Translating Chinese Classics into EnglishDetailed on Oracle database architecture," Doctor's degree of Soochow University, 2011.

[3] Velislava Stoykova, "Teaching Corpus Linguistics," Procedia - Social and Behavioral Sciences. 2014, 143(8): 437-441.

[4] Blanka Frydrychova Klimova, "Using Corpus Linguistics in the Development of Writing," Procedia - Social and Behavioral Sciences. 2014, 141(1): 124-128.

[5] TATAIT, "Detailed on Oracle database architecture," http://www.tatait.com/html/2013-10-29-20131029065838500001276.html, 2015-11-18.

[6] Mengxiangfeiyang's column, "Detailed diagram on Oracle database architecture," http://blog.csdn.net/mengxiangfeiyang/article/details/7841653, 2015-11-18.

[7] Sudha Ram, Vijay Khatri, "A comprehensive framework for modeling set-based business rules during conceptual database design," Information Systems, 2005, 30(2): 89-118.

[8] Y. Zhang, "Database design on enrolling management system of college sports class majors," Information Technology, vol. 38, no. 6, pp. 42-45, 2014.

[9] J. Xiong, Y. G. Liu, F. Gao, J. P. Feng, "Research of Oracle Bone Inscriptions Ontology Construction Based on Relational Database," Procedia Environmental Sciences, 2011, 11(A): 447-451.

[10] Seyed Reza Shahamiri, Wan Mohd Nasir Wan Kadir, "An automated framework for software test oracle," Information and Software Technology, 2011, 53(3): 774-788.

[11] Tom Laszewski, Prakash Nauduri, "Oracle Forms to Oracle Application Development Framework 11g," Elsevier, 2012.