

2011 3rd International Conference on Environmental
Science and Information Application Technology (ESIAT 2011)

Research on Field Characteristics of Shared Earth System Science Data Using Keyword Analysis and Visualization

Runda Liu^a, Haiqing Lin^b, Hui Zhao^a ^{*}

^a *Institute of Scientific & Technical Information of China, Beijing, China*

^b *School of Management and Economics, Beijing Institute of Technology, Beijing, China*

Abstract

This paper conducts analysis on the metadata information of shared Earth System Science data of China Scientific Data Sharing Project (SDSP) published online by Portal of Chinese Science and Technology Resource. High-frequency keywords are extracted from the metadata, keyword clusters are formed by means of co-occurrence matrix and K-core analysis, and then network visualization is made by Pajek software. This research reveals the field characteristics of the data through analyzing the keywords network structure, make conclusion on the hot acquisition spots of data in Earth System Science field shared on SDSP.

© 2011 Published by Elsevier Ltd. Open access under [CC BY-NC-ND license](#).

Selection and/or peer-review under responsibility of Conference ESIAT2011 Organization Committee.

Keywords: Scientific Data; Earth System Science; Co-occurrence Analysis; Social Network Analysis;

1. Introduction

Scientific Data are data resources relating scientific research, specifically refers to basic sci-tech data, materials generated by social sci-tech activities and data products and related information generated for different requirements ^[1]. Scientific data sharing means sharing scientific data among scientific research groups and individuals by certain strategies and techniques, it has become an important guarantee for the smooth evolution of scientific research, and it effectively promotes the sustainable development and innovation of contemporary science and technology.

^{*} Corresponding author. Tel.: +86-10-58882532; fax: +86-10-58882532.

E-mail address: liurd@istic.ac.cn

Worldwide, series of studies and practices on scientific data sharing methods based on computer network were conducted in developed countries and related international organizations [2]. In China, Scientific Data Sharing Project (SDSP) under National Science & Technology Infrastructure program (NSTI) is the main project of scientific data sharing activities implemented in sci-tech community [3]. In September 2009, Portal of Chinese Science and Technology Resource which is a portal of NSTI was launched and started providing service to the whole society [4], it gathers superior resources and achievements of NSTI so that general public can download various kinds of scientific data from it. The data cover the fields of medical hygiene, advanced manufacturing, earth system, geology & minerals, forestry, meteorology, communication, marine, agriculture and earthquake.

At present, metadata information of scientific data under the Portal of Chinese Science and Technology Resource is open for share. Data characteristics can be rediscovered by the analysis of the metadata information. However, there are few analysis cases on these data resource characteristics based on data attributes, especially based on keywords in the metadata. This article adopts keyword co-occurrence network method and uses visualization technology to analyze the field characteristics and hot acquisition spots of the earth system science field data in SDSP, and then propose analysis conclusions and suggestions to provide references for researchers and scientific data sharing activities.

2. Data sources and analysis methods

2.1. Data sources

International organization for standardization (ISO) regards metadata as the description of data content, quality, conditions and other characteristics [5]. The data used in this research are metadata of the scientific data from Earth System Science field data resources of China Scientific Data Sharing Project published online by Portal of Chinese Science and Technology Resource. Data acquisition time was in October 2010. 736 metadata were obtained. These metadata are originally created by Earth System Science Data Sharing Network [6] which is a project under SDSP. Earth System Science Data Sharing Network started collecting and integrating data resource in 2003, the amount of data accumulation rise consequently in the following years.

A new keyword field used for analysis was set by extracting and perfecting terms from the original keywords and abstract fields, the process includes the merger of synonyms or similar words and the removal of common words and noise words etc. Then new keywords field for analysis were revised, audited and verified by field experts to ensure data validity. The resulting new keywords field for each metadata item contains 2~5 new keywords separated by comma.

After statistics, there are 1046 different keywords in the new keywords field. High-frequency keywords are selected according to their cumulative frequencies. To make a focus for this study, we select top 51 keywords with at least 10 times of appearance shown in table 1.

Table 1. High-frequency keywords of Earth System Science field data

No.	Keywords	Frequency	No.	Keywords	Frequency	No.	Keywords	Frequency
1	China	160	18	Temperature	17	35	mud-rock flow	12
2	south pole	146	19	Russia	16	36	ACT	11
3	Zhongshan station	66	20	black soil	16	37	earth tide	11
4	Qinghai-Tibet plateau	64	21	Chengdu	15	38	Hailun	11

Download English Version:

<https://daneshyari.com/en/article/4403790>

Download Persian Version:

<https://daneshyari.com/article/4403790>

[Daneshyari.com](https://daneshyari.com)