京都大学防災研究所年報 第 49 号 C 平成 18 年 4 月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 49 C, 2006

Digitization of Disaster Prevention Printed Matters and Video Information, and Construction of the Search Engine, which can Search these on the Internet Website at High Speed (3)

Hideki MATSUURA, Kenichi TATSUMI, Yoshinori YOSHIDA, Tsutomu MIURA
Tetsuro TAKAYAMA, Hiroo WADA and Norio HIRANO

Synopsis

A large amount of disaster prevention research base Printed Matters has stored since 1951 at the Disaster Prevention Research Institute Kyoto University (DPRI). The aim of assignment for research subject of Division of Technical Affairs is that these base materials are widely exhibited to society, and can be used. However, the materials are not the one accumulated on the assumption of being able to use it on Web in the future. Thereby, it is difficult for us to pick up only necessary information of the materials. Accordingly, it is necessary to offer the comprehensible materials on the Internet to advance the disaster prevention study research.

In this paper, we provide method of converting the materials of the DPRI into digitalized materials and construction of the high-speed system that can search the information of disaster prevention on the Web. In last year, we converted the all previously-published Annual papers and the 20-45 Bulletin papers to the Portable Document Format (PDF) files, which were exhibited on the DPRI Internet website. Moreover, we constructed the search engines to search data of the materials published by the DPRI for users efficiently and conveniently. For this year, we digitalized the disaster prevention materials other than document papers. Moreover we can quickly search the materials on the Internet website.

Keywords: DPRI Annuals; OCR; Digital image; PDF; a-high-speed scanner; search engine;

1. Introduction

Disaster Prevention Research Institute (DPRI) was established in 1951. The field of research in DPRI now covers a wide range of disaster-related topics. DPRI plays a key role in the research on "Investigation of disaster theory and construction related to disaster prevention study".

Recently, telecommunication technology and multimedia technology have developed rapidly. As a result, much information is exchanged all over the world. Therefore, to retrieve only the most necessary information and effectively use the information is important. In particular, we expect that the disaster prevention sources will be more useful for the society, if a lot of the disaster prevention materials of DPRI were exhibited comprehensible on the Internet homepage.

After the first issue of the Annuals was published in 1958, they keep being published until today. Similarly,

the first issue of the Bulletin was published by DPRI in 1951. However, most of the preserved literature in the DPRI is a paper medium. In the last three years, we completely digitized the previously-published by DPRI paper materials of the Annuals and Vol.20-45 Bulletin.

The above-mentioned materials are the disaster prevention materials which digitized last year. Besides, a lot of other disaster prevention material exists in DPRI. At current year, we presented that we digitized the disaster prevention materials other than Document that we digitized before.

2. Web deliveries of the disaster prevention materials other than document

The procedure that is digitizing the paper literature can divide into three parts. In the disaster prevention basic material that the DPRI own, there is the material that exists only in the paper medium other than Document. The earthquake wave form chart, shown in Figure 1 and 2, is the one example.

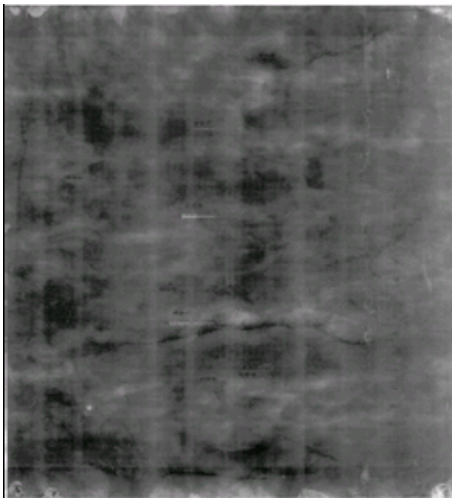

Fig. 1 The earthquake wave form record recorded in smoked paper

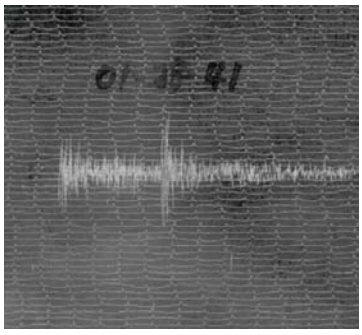

Fig. 2 enlarged view of Fig. 1

The maintenance of disaster prevention basic material other than the document that the disaster prevention laboratory has are also important.

However the basic information of the disaster prevention that is the paper medium, is difficult to keep completely preserving because of natural deterioration. Then, it was thought that making the electronic file was necessary to improve the state of preservation of the material in the disaster prevention and to make effective use of the information.

The DPRI has attached facilities of the earthquake observatory. The earthquake waveform chart recorded on the smoked paper in the facilities because the data processing technology did not develop enough.

At the beginning of digitalizing the document at current year, we digitalize file from the earthquake waveform chart on smoked paper. The digitalized object is the earthquake waveform chart that was observed, at Kamitakara Observatory in 1971 and at Sakurajima Volcano Research Center from 1967 to 1973. We digitalized the chart with the machine that can scan A0 size because of huge size as shown in Figure 3.

Fig. 3 Image scanner that can scan A0 size (CS2000 image scanner: GRAPHTEC.)

2.1 Digitalization for the paper medium of the disaster prevention basic material of large size

The earthquake waveform chart on smoked paper maintains record of even minute wave information. However, it is easy only to scan the A0 size paper medium disaster prevention basic material, on the other hand, it is really difficult to set the configuration of scan for maintenance of detailed information on the earthquake wave form. To make it to the electronic file with the information maintained, a setting parameter at

the scanning should be made appropriate. The setting parameter for appropriately converting it is shown in Table 1.

Table 1 CS2000 image scanner setting parameter

Color	grayscale（8bits）
Dpi	600
gammacorrection	1.9
Brightness	70
Contrast	80
White level	255
Black level	67

Thereby, each of the earthquake waveform charts saved as a digitalized image about 100-300MB in the hard disk.

2.2 Web delivery of digitalized still image file for big size

The size of the digitalized image file of the earthquake waveform charts is very big size. Therefore, it is necessary to reduce the size of the digitalized image file as much as possible to deliver on the Internet Web site. Because, the problem occur on the network delivered when the digital image over 100MB size is loaded.

Then, we came up with a way that the digital image file of the large size was reduced in using the computer software named ZOOMA. The ZOOMA is the computer software to which the high resolution image data can be delivered without putting a strain in the network on Web. (Fig. 4)

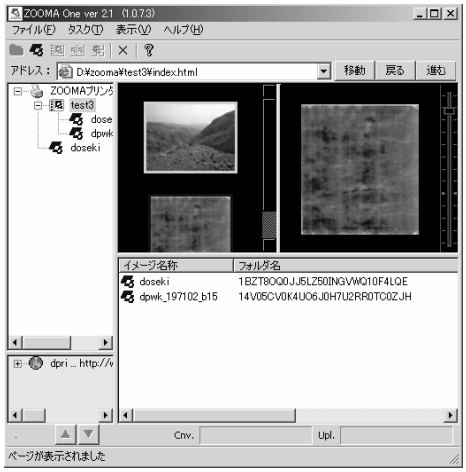

Fig. 4 Zooma software

And, using the Zooma software, we confirmed to be able to deliver the digitalized image file of the earthquake waveform charts on Web.

As a result, we thought that the result of the research preserved by a digital image of a big size of the file, for instance, the hazard map and the simulation result, can be delivered on the Web in the future.

3. The database that integrated disaster prevention information

Sharing a variety of disaster prevention information and data is indispensable for the development of the disaster prevention research. However, there is not a lot of cooperation in most of the fields that are the relations to disaster prevention, in the present condition that disaster prevention study is not matured as an interdisciplinary field.

Then, we draft the construction of the integrated database system as the place where various researchers exchanged information for disaster prevention by this research.

3.1 The database design concept for a researcher

When a researcher of majority obtains disaster prevention information from Web site, they retrieves from the search engine that the enterprise is offering and obtains the information. However, it is very difficult to acquire the right information that the researcher requests because there is all sorts of information.

Therefore, we thought that arranging information based on the standardization rule from which it is considered that the researcher and the society use is necessary to solve such a problem, and to develop the disaster prevention research in the future. Moreover, it is important to construct the rule that arrange information in the DPRI organization.

We aimed at constructing a data base that the disaster prevention researchers and also the different field researchers retrieve easily. It is necessary to construct such an inclusive data base to correspond with various data resources. We roughly divided the research data resource of disaster prevention into 12 categories. It is shown in Table 2. (document, study, model, voice,image,video, geospatial, internet, person, etc.)

Table 2 12 kinds of data resource types (XMDB conforming)

Japanese	English
文書	Document
静止画像	Image
音声	Audio
動画	Video
インターネット	Internet
モデル	Model
研究	Study
基礎データ	Data
イベント	Event
地理情報	Geospatial
人物	Person
組織	Organization

The search engine constructed in the previous year, which can search disaster prevention, can be retrieved based on three elements information (Title, Author, and Keyword). However, this search engine has a very easy retrieval technique.

Then, we made data base for a new search engine as follows first.

[1] The data base table of disaster which occurred from the 20th century

[2] The data base table of author

[3] The data base table of key word related to disaster prevention

In the above data bases, especially the disaster data divide into 9 categories as shown in Table 3.

Table 3 9 kinds of disaster type

Japanese	English
地震	Earthquake
台風	Typhone
集中豪雨	Torrential rain
津波	Seismic sea wave
噴火	Eruption
地すべり	Landslide
土石流	Debris flow
竜巻	Tornado
その他	Other disaster

We made the core data base table related to the data base of [1], [2] and [3] as shown in Table 4.

Table 4 The Core table of the database architecture (The Dublin Core metadata format conforming)

Field	Field Type
data_code	varchar(80)
data_title_j	text
data_title_e	text
data_subject_j	text
data_subject_e	text
data_publisher_j	Text
data_publisher_e	Text
data_creator_j	Text
data_creator_e	Text
data_description_j	text
data_description_e	text
data_format_id	int(11)
data_format	varchar(80)
data_type_id	int(11)
data_type	varchar(80)
data_saigai_id	int(11)
data_saigai_type_id	int(11)
data_right_id	int(11)
data_category_id	int(11)
data_source_filename	text
data_source_filename2	text
data_relation_link	text
data_language	varchar(80)
data_year	int(11)
data_maketime	datetime
data_starttime	datetime
data_endtime	datetime
data_check1	varchar(80)
data_check2	varchar(80)
data_check3	varchar(80)
data_ndc_9y1	int(11)
data_ndc_9s1	int(11)
data_ndc_9y2	int(11)
data_ndc_9s2	int(11)
data_ndc_9y3	int(11)
data_ndc_9s3	int(11)
data_ndc_k	int(11)
data_ndc_t1	int(11)
data_ndc_t2	int(11)
data_ndc_t3	int(11)

In addition, we made the new metadata to arrange these resources in accordance with the Dublin Core metadata format that was an international standard. The Dublin Core Metadata Element Set was standardized in February, 2003 as a network meta data resource, that is ISO15836. Moreover, we assigned the NDC (Nippon Decimal Classification) code number to all the registered disaster prevention data appropriately. NDC is one of the methods of classifying data by decimal notation. The NDC code becomes a standard when the book is classified.

As the result, user can obtain the retrieval result by the search engine based on the classified new data base, by inputting disaster information, disaster free word or a keyword in the NDC. The flow searching is shown in Figure 5.

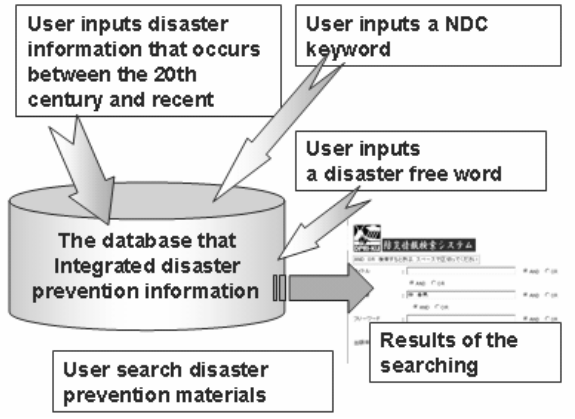

Fig. 5 Flow searching of the disaster prevention materials

3.2 The new search engine that can search the disaster prevention research materials of DPRI

The search engine that runs on DPRI website now can only search the DPRI documents. It is necessary to construct a new search engine to be able to search not only the DPRI documents but also the other disaster prevention research materials.

The last year, we constructed the prototype of a new search engine besides three existing search engines. At current year, we added the function to which narrowing was done by the data resource type to the retrieval of the search engine, as shown in Figure 6.

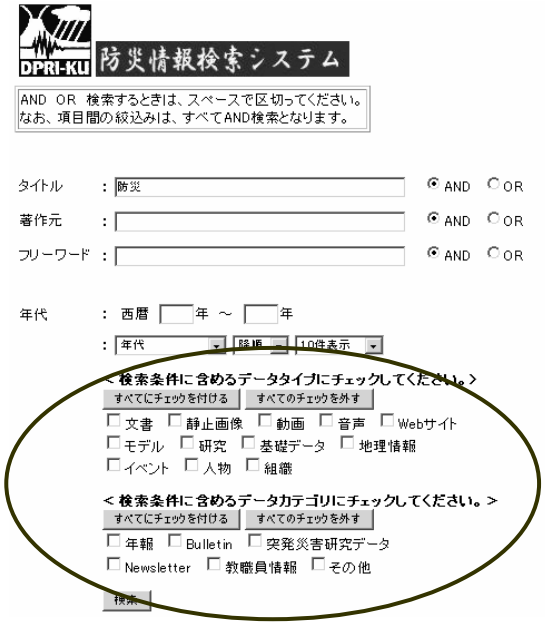

Fig. 6 Search engine that can retrieval to which narrowing is done by data resource type

In addition, we have improved previous search engine so that cooperation with other data bases, XMDB and SAIGAI, OPAC, may become possible as for the search engine (Fig.7).

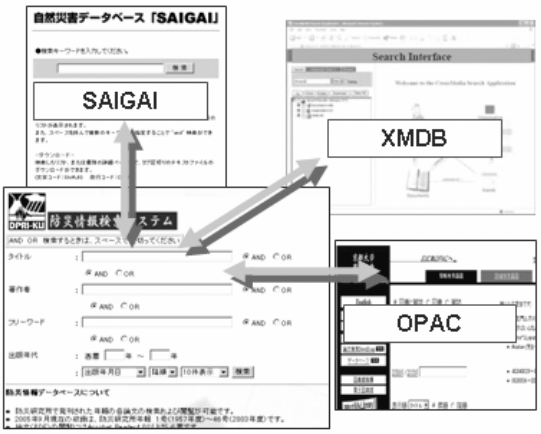

Fig. 7 Search engine that can cooperate with other data bases

4. Conclusions

In this year, we introduced a high-speed scanner and have improved the work system for the increase in the work efficiency. As a result, we completed digitizating the disaster prevention basic material made out of large size paper medium by the end of 2005 fiscal year. We make a database of the metadata of the paper medium disaster prevention basic material.

We show the result of the research preserved by a digitalized large size, for instance, the hazard map and the simulation result, can be delivered on the Web by the ZOOMA software in the future.

In addition, we constructed the retrieval to which narrowing was done by the data resource type to the function of the search engine. We have improved previous search engine so that would be able to cooperation with other data bases, XMDB and SAIGAI, OPAC, may become possible as for the search engine.

Acknowledgments

This research was supported by the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT) 21st Century COE Program for DPRI, Kyoto University (Program Leader: Prof. Yoshiaki KAWATA, Assignment research Leader: Prof. Haruo HAYASHI) as well as the Grant-in-Aid for Scientific Research (Principal Investigator: Norio HIRANO, Kyoto University).

The authors express their sincere gratitude to the colleagues, the adjuncts and professors that helped us on this project and the 21st Century COE Program committee for allowing us to research on this project.

References

Satoshi TATEOKA, (2003): A guide to a formal WeB Database system by MySQL+PHP -Point of Web application development Advanced server side programming -, Gijutsu-Hyohron co (in Japanese).

Hajime BABA, (2001): Construction and practical use of a system - Japanese full-text search thoroughness guide-, SOFTBANK co (in Japanese).

Haruo HAYAMI, (2002): Database IT Text, Ohmsha co (in Japanese).

Hideki MATSUURA, Kenichi TATSUMI, Hideo TAGAWA, Yoshinori YOSHIDA, Tsutomu MIURA, Tetsuro TAKAYAMA, Hiroo WADA and Norio, (2004): Digitization of disaster prevention Printed Matters and Video information, and construction of the search engine which can search these on the Internet website at high speed, Annuals of DPRI, Kyoto Univ., No. 47 C, pp. 117-126.

Tomohiro KUGAI, Yoshiaki KAWATA and Haruo HAYASHI, (2004): Development of Cross-Media Database for Sharing Disaster Information, Annuals of DPRI, Kyoto Univ., No. 47 C, pp. 331-336.

Go URAKAWA, Nozomu YOSHITOMI*, Tomohiro KUGAI, Hironori KAWAKATA, Kenneth C. Topping and Haruo. (2004): Development of Cross-Media Database for Sharing Disaster Information and A Case Study about Implementation Process, Annuals of DPRI, Kyoto Univ., No. 47 C, pp. 337-344 .

印刷物・映像情報の電子ファイル化と Ｗｅｂ上で高速検索可能なシステムの構築（３）

松浦秀起・辰己賢一・吉田義則・三浦勉・高山鐵朗・和田博夫・平野憲雄

要旨

本研究では、防災研究所に蓄積されてきた研究成果の印刷物・映像情報を電子ファイル化し、Ｗｅｂ上で高速検索可能なシステムの構築を目指している。平成17年度では、文献以外の紙媒体の防災資料の電子ファイル化に取り組み、大型サイズの静止画像のWeb配信手法を示した。また、昨年度以上に、研究者が求める文献をより効率よく検索できる防災情報検索システムの機能向上、及び、XMDB、OPAC、SAIGAI等の他のデータベースシステムとの連携できる新しいシステムの構築案を示した。

キーワード: 防災研究所; 研究資料; OCR; 電子ファイル; A0スキャナ; PDF; 高速検索システム